THE MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY

JOURNAL

Physics and Mathematics

VOLUME 16, No. 4, 2023

Peter the Great St. Petersburg
Polytechnic University
2023

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL.
PHYSICS AND MATHEMATICS

The journal is included in the List of leading peer-reviewed scientific journals and other editions to publish major findings of theses for the research degrees of Doctor of Sciences and Candidate of Sciences.

The publications are presented in the VINITI RAS Abstract Journal and Ulrich's Periodical Directory International Database.

The journal is published since 2008 as part of the periodical edition 'Nauchno-tekhnicheskie vedomosti SPb-GPU'.

The journal is registered with the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR). Certificate ПИ № ФС77-52144 issued December 11, 2012.

The journal is distributed through the CIS countries catalogue, the «Press of Russia» joint catalogue and the «Press by subscription» Internet catalogue. The subscription index is **71823.**

The journal is in the **Web of Science** (Emerging Sources Citation Index), **Scopus**, the **Russian Science Citation Index** (RSCI) and the **Directory of Open Access Journals** (DOAJ) databases.

© Scientific Electronic Library (http://www.elibrary.ru).

Address: 195251 Politekhnicheskaya St. 29, St. Petersburg, Russia.

Phone: (812) 294-22-85.
http://ntv.spbstu.ru/physics

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ

САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Физико-математические науки

ТОМ 16, № 4

2023

Санкт-Петербургский политехнический
университет Петра Великого
2022

НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА. ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ

Журнал с 2002 г. входит в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученых степеней доктора и кандидата наук.

Сведения о публикациях представлены в Реферативном журнале ВИНИТИ РАН, в международной справочной системе «Ulrich's Periodical Directory».

С 2008 года выпускается в составе сериального периодического издания «Научно-технические ведомости СПбГПУ».

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-52144 от 11 декабря 2012 г.

Распространяется по Каталогу стран СНГ, Объединенному каталогу «Пресса России» и по Интернет-каталогу «Пресса по подписке». Подписной индекс **71823**.

Журнал индексируется в базе данных **Web of Science** (Emerging Sources Citation Index), **Scopus**, а также включен в базы данных **«Российский индекс научного цитирования»** (РИНЦ), размещенную на платформе Научной электронной библиотеки на сайте http://www.elibrary.ru, и **«Directory of Open Access Journals»** (DOAJ)

Адрес редакции и издательства:
Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29.
Тел. редакции (812) 294-22-85.
http://ntv.spbstu.ru/physics

Contents

Condensed matter physics

Simulation of physical processes

Mathematical physics

Experimental technique and devices

Physical electronics

Physical materials technology

Biophysics and medical physics

Nuclear physics

Radiophysics

Mathematics

Содержание

Физика конденсированного состояния

Математическое моделирование физических процессов

Математическая физика

Приборы и техника физического эксперимента

Физическая электроника

Физическое материаловедение

Биофизика и медицинская физика

Ядерная физика

Радиофизика

Математика

CONDENSED MATTER PHYSICS

Научная статья
Original article
DOI: https://doi.org/10.18721/JPM.16401

CURRENT–VOLTAGE CHARACTERISTICS OF $MAPBI_3$ PEROVSKITE FILMS FORMED BY THE SINGLE-STAGE SPIN-COAT METHOD

M. K. Ovezov[1]✉, *A. A. Ryabko*[1], *A. N. Aleshin*[1], *V. A. Moshnikov*[2], *V. M. Kondratyev*[3,4], *A. I. Maximov*[2]

[1] Ioffe Institute of RAS, St. Petersburg, Russia;

[2] St. Petersburg Electrotechnical University "LETI", St. Petersburg, Russia;

[3] Alferov University of RAS, St. Petersburg, Russia;

[4] Moscow Institute of Physics and Technology (National Research University), Moscow, Russia

✉strontiumx94@gmail.com

Abstract. In the paper, the properties of $MaPbI_3$ films made with or without a precipitant have been investigated. The samples had a planar geometry based on ceramic substrates with interdigitated gold electrodes and also based on glass substrates. The samples were irradiated with green light from an LED source, and a special setup was used to measure current–voltage (I–V) characteristics. The polycrystalline films exhibited high sensitivity (an increase in current by about 2 orders upon irradiation). The width of their optical band gap was the same regardless of the use of the precipitant but the maximum trap-filling voltages turned out to be very sensitive to such use. According to optical microscopy, the film microstructure was characterized by a growth of large long dendritic structures, i.e., the nucleation occurred in the solution mass during the films' making. This growth mechanism may be convenient for the use of $MaPbI_3$ films in photodetectors.

Keywords: organometallic perovskite, semiconducting polymer, transport mechanism, solar cell

Funding: The reported study was funded by Russian Science Foundation (Grant No. 23-42-10029; https://rscf.ru/en/project/23-42-10029/).

Citation: Ovezov M. K., Ryabko A. A., Aleshin A. N., Moshnikov V. A., Kondratyev V. M., Maximov A. I., Current – voltage characteristics of $MaPbI_3$ perovskite films formed by the single-stage spin-coat method, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 9–19. DOI: https://doi.org/10.18721/JPM.16401

УДК 538.91
DOI: https://doi.org/10.18721/JPM.16401

ВОЛЬТАМПЕРНЫЕ ХАРАКТЕРИСТИКИ ПЕРОВСКИТНЫХ ПЛЕНОК $MAPbI_3$, СФОРМИРОВАННЫХ ОДНОСТАДИЙНЫМ МЕТОДОМ ЦЕНТРИФУГИРОВАНИЯ

М. К. Овезов[1]✉, А. А. Рябко[1], А. Н. Алешин[1]
В. А. Мошников[2], В. М. Кондратьев[3,4], А. И. Максимов[2]

[1] Физико-технический институт им. А. Ф. Иоффе РАН, Санкт-Петербург, Россия;

[2] Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» имени В. И. Ульянова (Ленина), Санкт-Петербург, Россия;

[3] Академический университет имени Ж. И. Алфёрова РАН, Санкт-Петербург, Россия;

[4] Московский физико-технический институт (Национальный исследовательский университет), Москва, Россия

✉strontiumx94@gmail.com

Аннотация. В работе исследованы свойства пленок $MaPbI_3$., изготовленных с применением осадителя либо без него. Образцы обладали планарной геометрией на основе керамических подложек со встречно-штыревыми золотыми электродами, а также на основе стеклянных подложек. Образцы облучали зеленым светом от светодиодного источника, а для измерения вольтамперных характеристик использовали специальную установку. Поликристаллические пленки продемонстрировали высокую фоточувствительность (увеличение тока примерно на 2 порядка при облучении). Ширина их оптической запрещенной зоны была одинаковой вне зависимости от использования осадителя, однако предельные напряжения заполнения ловушек оказались весьма чувствительными к такому использованию. По данным оптической микроскопии, для микроструктуры пленки характерно образование крупных дендритных структур, т.е. при ее изготовлении происходило зародышеобразование в толще раствора. Этот механизм может быть удобным для использования пленок $MaPbI_3$ в фотодетекторах.

Ключевые слова: металлорганический перовскит, полупроводящий полимер, механизм транспорта, солнечный элемент

Финансирование: Исследование проведено при финансовой поддержке Российского научного фонда (РНФ), грант № 23-42-10029; https://rscf.ru/en/project/23-42-10029/.

Ссылка для цитирования: Овезов М. К., Рябко А. А., Алешин А. Н., Мошников В. А., Кондратьев В. М., Максимов А. И. Вольтамперные характеристики перовскитных пленок $MaPbI_3$, сформированных одностадийным методом центрифугирования // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. 9–19. DOI: https://doi.org/10.18721/JPM.16401

Introduction

Organometallic perovskites such as $(Fa)(Ma)MX_3$, where Fa is formamidinium $CH(NH_2)_2$, Ma is methylammonium CH_3NH_3, metal M = Pb, halogen X = Br or I, have been attracting increasing attention recently for applications in solar cells (SC) and photodetectors, X-ray detectors as well as memristor structures [1–3]. SC based on organometallic perovskites exhibit efficiency comparable to silicon solar cells (25.8%) [4]. In addition to organometallic perovskite films, films of inorganic perovskites $CsPbX_3$ are used for SE [5]. In this case, $CsPbX_3$ perovskite films are formed not only from a solution, but also from suspensions of

colloidal quantum dots, allowing to control the absorption spectrum of the SC by adjusting the size of the quantum dots [6–8]. Interest is also growing towards lead-free perovskite materials, primarily due to their low toxicity [9]. Even though SC based on lead-free perovskites exhibit lower efficiencies compared to lead-based ones (by about 6%), their stabilities are close to 2,000 hours of operation in a nitrogen atmosphere [10]. Finally, a promising direction is fabrication of tandem structures based on various SC combined with perovskite photovoltaic elements [11].

One of the best materials with a perovskite structure for creating SC is organometallic perovskite $CH_3NH_3PbI_3$ (referred to as $MaPbI_3$ from now on), also widely used to create photodetectors and X-ray detectors. Single-crtstal $MaPbI_3$ perovskite has been used for these detectors, demonstrating trap densities of the order of 10^{10} cm^{-3} and a carrier diffusion length exceeding 175 μm [12, 13]. Although single-crystal $MaPbI_3$ perovskite is also applicable for photovoltaic structures, it is economically impractical to use it for SC, as it is too difficult to produce a single crystal with a large diameter. Polycrystalline films are used to create SC, deposited from a solution by such methods as spray-coating, inkjet and screen printing, spin-coating and others [14].

As a rule, polycrystalline layers of $MaPbI_3$ are deposited by spin-coating in two steps if PbI_2 and CH_3NH_3I solutions are deposited separately, or in one step if $CH_3NH_3PbI_3$ solution is deposited. Studies producing the materials discovered that the two-step method allows to obtain more homogeneous films, but it is more technologically complex, while films made by two different methods yield comparable characteristics, some (for example, grain size) even better for films prepared by the one-step method [15, 16].

Analyzing the available data, we opted for the one-step method. A precipitator is often used for depositing the film, allowing to considerably improve the quality of the film [17–19]. Many reagents can be used as precipitators, but ethyl acetate (EA) stands out in that it allows to achieve a sufficiently high efficiency of the photosensitive layer (up to 19.53%). In addition, this layer can preserve up to 84.8% of the initial efficiency in an SC operating in an open air environment for a long time (over 1900 hours) [20]. Additionally, ethyl acetate has low cost and low toxicity, offering greater prospects for commercial applications.

In this paper, we considered the influence of the technology for manufacturing the MaPbI perovskite film on its current–voltage characteristics in planar structures was investigated; ethyl acetate was deposited in a one-step procedure in this case.

Materials and methods

Powders of $MaPbI_3$ organometallic perovskites purchased from Xi'an Polymer Light Technology Corp. (China) were used to prepare the samples. The films were applied by spin-coating from a solution of dimethylformamide and dimethyl sulfoxide (4:1 volume ratio) with a $MaPbI_3$ mass concentration of 300 mg/ml with subsequent annealing at a temperature of 110 °C for 10 minutes. The spin-coating rate was 3000 rpm (for 30 s) with pre-centrifugation at 1000 rpm (10 s). The samples were obtained by two techniques: using ethyl acetate as a precipitator and without it. The precipitator was introduced at the acceleration stage of spin coating.

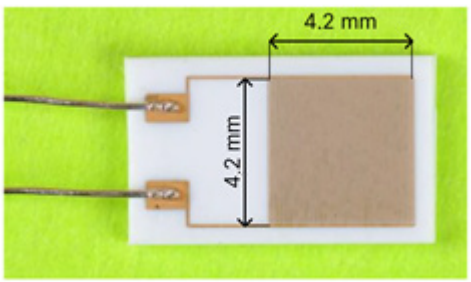

Fig. 1. Photo of ceramic substrate with interdigitated NiCr/Ni/Au electrodes (sensor platform, Tesla Blatná)
The size of the region with these electrodes is 4.2 × 4.2 mm

The current–voltage characteristics were measured by depositing the perovskite films on ceramic substrates with interdigitated gold electrodes (sensor platform from Tesla Blatná, Czech Republic). The thickness of the electrodes and the distance between them was 25 μm (Fig. 1).

The sample morphology was studied by optical and scanning electron microscopy, also depositing layers of $MaPbI_3$ perovskite film on silicon substrates. To determine the band gap width by processing the optical absorption spectra, such layers were deposited on glass substrates with ITO coating. The latter is a solution consisting of indium, oxygen and tin, i.e., indium oxide In_2O_3 and tin oxide Sn_2O_3.

The current–voltage (I–V) characteristics of the samples were measured with a Keithley 6487 picoammeter (USA) in the dark and under illumination with an LED source at a wavelength of 535 nm. The absorption spectra were measured with a PE-5400UF spectrophotometer (Russia). The microstructure of the objects was studied with a POLAM-312 polarizing microscope (Russia). The morphology of the coatings was also studied with a Zeiss Supra 25 scanning electron microscope (Germany).

Results and discussion

The samples of thin polycrystalline $MaPbI_3$ films obtained in this study exhibit a characteristic optical absorption spectrum (curve *1* in Fig. 2,*a*). The optical band gap of the samples was determined as a Tauc plot, $(\alpha h\nu)^{1/r}$ versus $h\nu$, by extrapolating the linear section to the abscissa axis, where the value of r expresses the type of dependence of the semiconductor's absorption coefficient on the irradiation wavelength (greater than the semiconductor's absorption edge). Since $MaPbI_3$ is a direct band gap semiconductor, the absorption coefficient is described by the root dependence on the wavelength ($r = 1/2$).

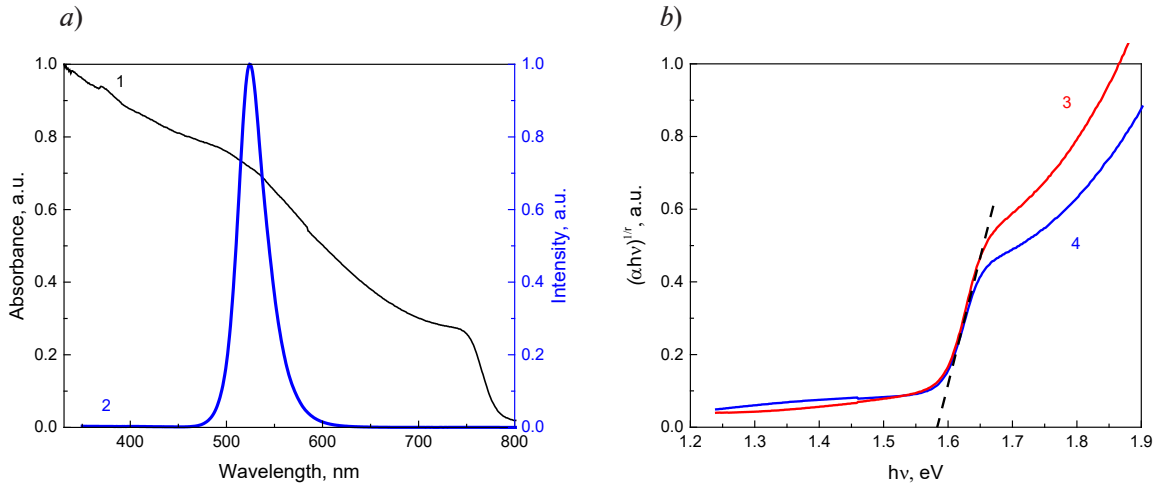

Fig. 2. Analysis of $MaPbI_3$ layers (curve *1* in Fig. 2,*a*, curves *3*, *4* in Fig. 2,*b*), and LED source (curve *2* in Fig. 2,*a*):
typical optical absorption spectrum *1*; optical absorption spectra *3*, *4* as Tauc plots for determining the optical band gap; dependences are shown for samples prepared with a precipitator (*3*) and without it (*4*); *2* is the electroluminescence spectrum of a green LED used to record the photo response of samples

According to the results obtained, the optical band gap E_g of the manufactured films was $E_g \approx 1.58$ eV, and this value did not depend on whether a precipitator was used; it is characteristic for polycrystalline $MaPbI_3$ films.

The results of I–V measurements in the dark and under illumination are shown in Figs. 3 and 4. The samples exposed to irradiation with green light leads exhibit a change in the current by about two orders of magnitude, while using a precipitator does not have a noticeable effect. The current–voltage characteristics in the dark exhibit hysteresis, which is associated with migration of ions (primarily I^-), as well as the space-charge limited current (SCLC) [20]. The influence of the technology by which the samples were prepared is also observed: the maximum trap-filling voltages V_{TFL} differ for samples obtained with and without the precipitator.

We also found a slight increase in V_{TFL} after the samples were irradiated with light (a green LED was used), pointing to an increase in trap concentration during exposure.

It was found in [20] that such irradiation can significantly accelerate or induce ion migration, and such migration is observed over a wide temperature range. A five-fold decrease in the activation energy of ion migration is also reported (from 0.82 to 0.15 eV with an increase in the irradiation intensity from 0 to 20 MW/cm^2) [21].

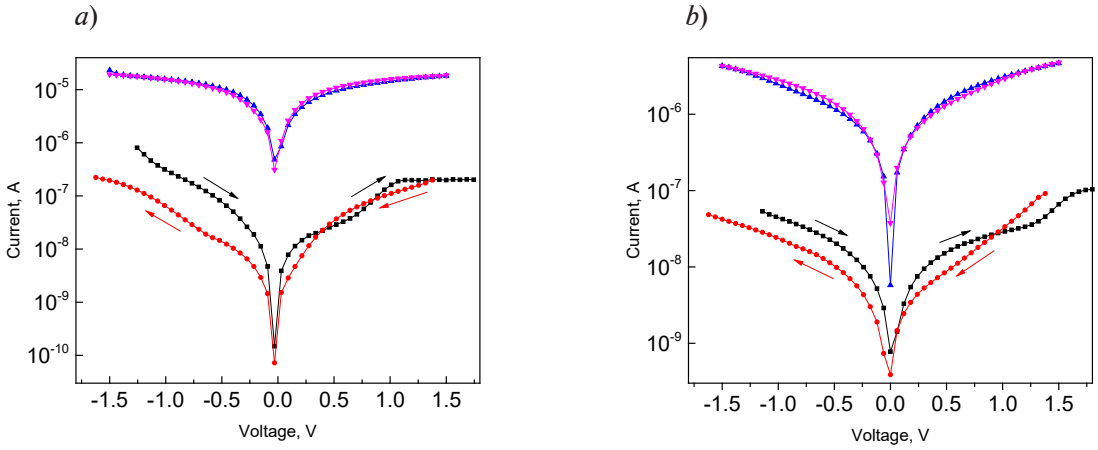

Fig. 3. $I-V$ curves of the samples prepared without precipitator (*a*) and with it (*b*) in the dark (lower curves) and under irradiation with green light (upper curves)
The arrows indicate the variation modes of the voltages applied to the samples

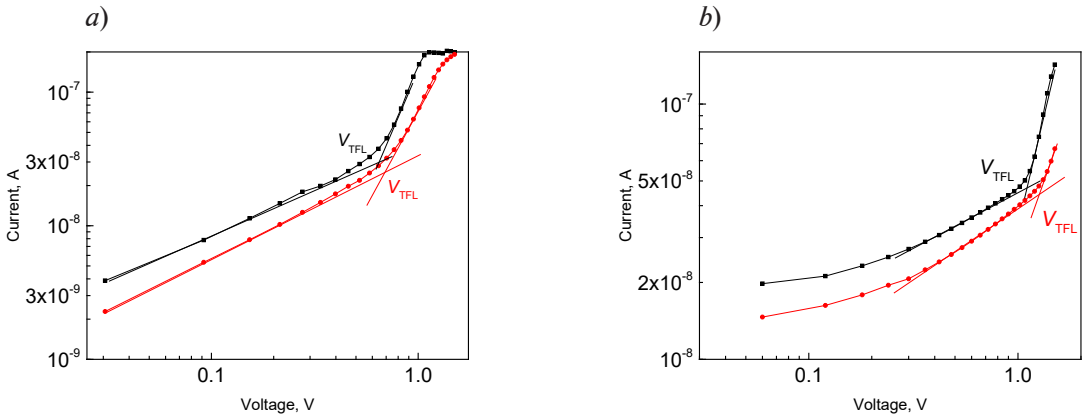

Fig. 4. Fragments of $I-V$ curves of the samples prepared without precipitator (*a*) and with it (*b*) in the dark, obtained before (black curves) and after (red curves) irradiation with light.
A difference in V_{TFL} values is observed.

The maximum trap-filling voltage V_{TFL} is directly proportional to their concentration:

$$V_{TFL} = \frac{eN_t L^2}{2\varepsilon\varepsilon_0}, \tag{1}$$

where e is the electron charge, N_t is the trap concentration, L is the distance between the electrodes (25 μm for our case), ε is the permittivity of the material (ε = 32 for $MaPbI_3$), ε_0 is the dielectric constant [18].

The concentration of traps can be calculated from this using the formula

$$N_t = \frac{2\varepsilon\varepsilon_0 V_{TFL}}{eL^2}. \tag{2}$$

N_t values equal to 3.68, 3.81, 6.09 and 7.01 (10^{14} cm^{-3}), respectively, were obtained for a voltage V_{TFL} of approximately 0.65–0.70 V for a sample without precipitator and about 1.13 – 1.30 V for a sample with precipitator.

The values of V_{TFL} and, consequently, the trap concentrations turned out to be lower for the sample obtained without precipitator, which is inconsistent with the literature data [16, 22].

The results obtained by scanning electron microscopy (Fig. 5, *c*, *d*) indicate that using the precipitator (ethyl acetate) under these conditions produces an increase in the size of film crystallites, compared with the film prepared without the precipitator. There are also pores in the films due to the conditions of solvent evaporation. The optical microscopy data (Fig. 5,*a*) for the sample obtained with a precipitator indicate that some substrate regions are not covered by the perovskite film, however, the regions covered by the $MaPbI_3$ film are homogeneous. On the other hand, not only heterogeneous nucleation is observed for the sample obtained without precipitator at the interface with the substrate, but also growth of large elongated (~50 μm) dendritic structures (Fig. 5,*d*).

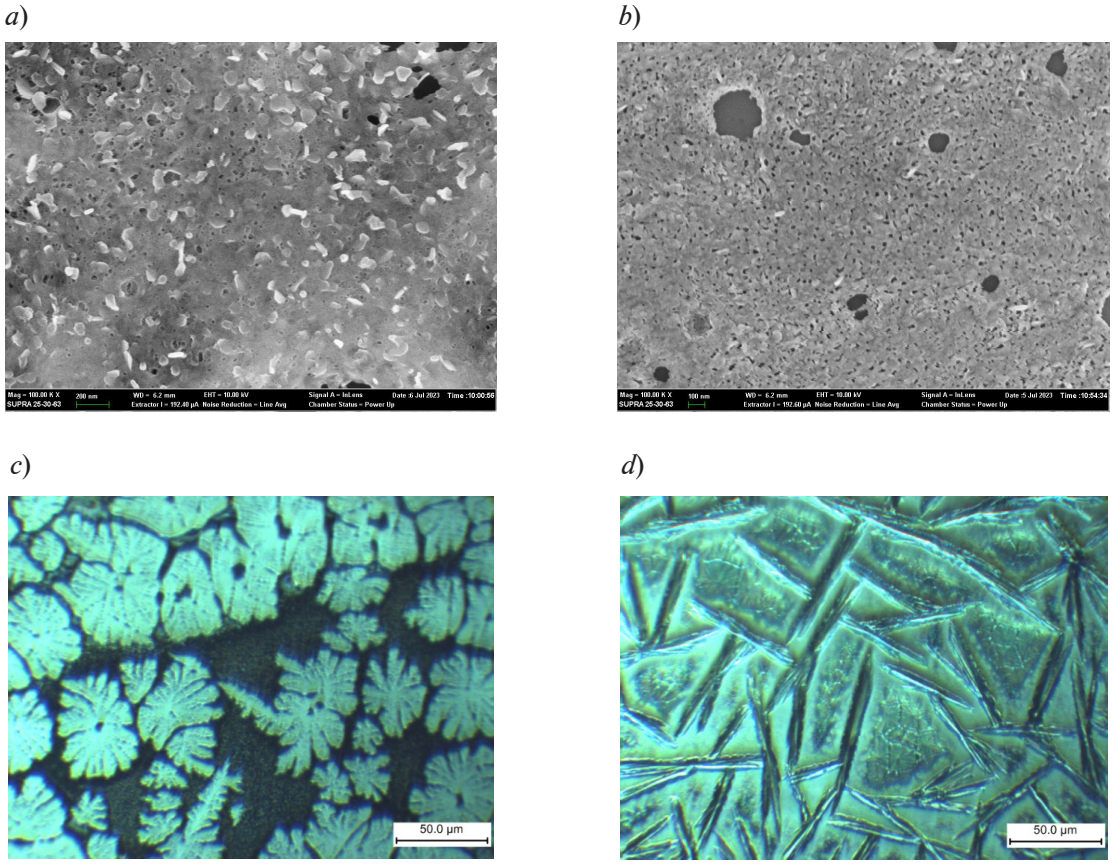

Fig. 5. Micrographs of $MaPbI_3$ layers for samples prepared with precipitator (*a*,*c*) and without it (*b*,*d*) obtained by optical (*a*, *b*) and scanning electron (*c*, *d*) microscopy

Evidently, as the solvent evaporates from the $MaPbI_3$ solution under heating, this can lead to an increase in the solute concentration near the surface. Moreover, the temperature gradient from the substrate surface to the film surface leads to a decrease in the solubility of the $MaPbI_3$ film in the near-surface region.

Thus, the conditions for nucleation are fulfilled near the surface of the solution film, with subsequent growth of dendritic structures. Nucleation at the phase boundary is energetically favorable, inducing growth of polycrystalline film directly on the surface of the substrate (Fig. 5,*b*). Consequently, the precipitator leads to an increase in the size of the film crystallites formed at the phase boundary, however, the conditions of film growth without precipitator lead to growth of elongated $MaPbI_3$ structures of a larger scale.

Although the conditions for film growth are different if polished surfaces of silicon substrates and ceramic substrates are used, due to the different concentrations of nucleation sites on the substrate surface and hydrophobicity, we assume that the character of film formation

should be similar. Thus, the film consisting of elongated structures has a lower concentration of grain boundaries, which leads to a decrease in the trap concentration and the voltage V_{TFL}. Although this growth mechanism hinder the application of polycrystalline films in photovoltaic structures, it may be convenient for using $MaPbI_3$ films in photodetectors. On the other hand, using a precipitator is preferable from the standpoint of morphology of such a film for creating photovoltaic structures.

Conclusions

Films of organometallic perovskite $MaPbI_3$ produced for the study exhibit a high response to visible-light irradiation (with a green LED) as well as to the characteristic mode of space charge limited current. The dark current–voltage (I–V) characteristics exhibit hysteresis due to ion migration.

We found that $MaPbI_3$ films prepared by the described procedure without the precipitator contain large dendritic structures providing a decrease in the maximum trap-filling voltage We believe that films with such morphology can be successfully used as photodetectors.

REFERENCES

1. **Sakhatskyi K., Türedi B., Matt G. J., et al.,** Stable perovskite single-crystal X-ray imaging detectors with single-photon sensitivity, Nat. Photon. 17 (June) (2023) 510–517.

2. **Wang H., Sun Y., Chen J., et al.,** Review of perovskite-based photodetectors and their applications, Nanomater. 12 (24) (2022) 4390.

3. **Nenashev G. V., Aleshin A. N., Shcherbakov I. P., Petrov V. N.,** Effect of temperature variations on the behavior of a two-terminal organic-inorganic halide perovskite rewritable memristor for neuromorphic operations, Solid State Commun. 348–349 (1 June) (2022) 114768.

4. **Roy P., Ghosh A., Barclay F., et al.,** Perovskite solar cells: A review of the recent advances, Coat. 12 (8) (2022) 1089.

5. **Wang J., Zhang J., Zhou Y., et al.,** Highly efficient all-inorganic perovskite solar cells with suppressed non-radiative recombination by a Lewis base, Nat. Commun. 11 (Jan) (2020) 177.

6. **Sadhu A. S., Huang Y.-M., Chen L.-Y., et al.,** Recent advances in colloidal quantum dots or perovskite quantum dots as a luminescent downshifting layer embedded on solar cells, Nanomater. 12 (6) (2022) 985.

7. **Srivastava A. K., Zhang W., Schneider J., et al.,** Luminescent down-conversion semiconductor quantum dots and aligned quantum rods for liquid crystal displays, Adv. Sci. 6 (22) (2019) 1901345.

8. **Aleksandrova O. A., Galieva D. M., Drobinceva A. O., et al.,** Nanochastitsy, nanosistemy i ikh primeneniye. Ch.1. Kolloidnyye kvantovyye tochki [Nanoparticles, nano systems and their application. Part 1. Colloidal quantum dots], Aeterna Publishing, Ufa, 2015 (in Russian).

9. **Shi Z. J., Guo J., Chen Y. H., et al.,** Lead-free organic-inorganic hybrid perovskites for photovoltaic applications: Recent advances and perspectives, Adv. Mater. 29 (16, 3 Febr.) (2017) 1605005.

10. **Gao W., Ran C., Li J., et al.,** Robust stability of efficient lead-free formamidinium tin iodide perovskite solar cells realized by structural regulation, J. Phys. Chem. Lett. 9 (24) (2018) 6999–7006.

11. **Li H., Zhang W.,** Perovskite tandem solar cells: From fundamentals to commercial deployment, Chem. Rev. 120 (18) (2020) 9835–9950.

12. **Rong Sh.-Sh., Faheem M. B., Li Y.-B.,** Perovskite single crystals: Synthesis, properties, and applications, J. Electron. Sci. Technol. 19 (2) (2021) 100081.

13. **Lian Zh., Yan Q., Gao T., et al.,** Perovskite $CH_3NH_3PbI_3(Cl)$ single crystals: Rapid solution growth, unparalleled crystalline quality, and low trap density toward 10^8 cm^{-3}, J. Am. Chem. Soc. 138 (30) (2016) 9409–9412.

14. **Zhang Y., Zhang H., Zhang X., et al.,** Major impediment to highly efficient, stable and low-cost perovskite solar cells, Metals. 8 (11) (2018) 964.

15. **Xiao M., Huang F., Huang W., et al.,** A fast deposition-crystallization procedure for highly efficient lead iodide perovskite thin-film solar cells, Angew. Chem. Int. Edit. 53 (37) (2014) 9898–9903.

16. **Ahmed D. S., Mohammed B. K., Mohammed M. K. A.,** Long-term stable and hysteresis-free planar perovskite solar cells using green antisolvent strategy, J. Mater. Sci. 56 (27) (2021) 15205–15214.

17. **Shaikh J. S., Shaikh N. S., Sheikh A. D., et al.,** Perovskite solar cells: In pursuit of efficiency and stability, Mater. Design. 136 (15 Dec) (2017) 54–80.

18. **Li D., Shi J., Xu Y., et al.,** Inorganic-organic halide perovskites for new photovoltaic technology, Natl. Sci. Rev. 5 (4) (2018) 559–576.

19. **Zhang W., Li Y., Liu X., et al.,** Ethyl acetate green antisolvent process for high-performance planar low-temperature SnO_2-based perovskite solar cells made in ambient air, Chem. Eng. J. 379 (1 Jan) (2020) 122298.

20. **Zhang P., Gu N., Song L., et al.,** The disappearing additive: introducing volatile ethyl acetate into a perovskite precursor for fabricating high efficiency stable devices in open air, Nanoscale. 14 (13) (2020) 5204–5213.

21. **Khan R., Ighodalo K.O., Z. Xiao Z.,** Ion migration in metal halide perovskites solar cells, soft-matter thin film solar cells: Physical processes and device simulation, AIP Publishing LLC, Melville, New York, 2020, p. 1-283.

22. **Zhao Y.-C., Zhou W.-K., Zhou X., et al.,** Quantification of light-enhanced ionic transport in lead iodide perovskite thin films and its solar cell applications, Light Sci. Appl. 6 (5, 5 May) (2017) e16243.

СПИСОК ЛИТЕРАТУРЫ

1. **Sakhatskyi K., Türedi B., Matt G. J., et al.** Stable perovskite single-crystal X-ray imaging detectors with single-photon sensitivity // Natural Photonics. 2023. Vol. 17. June. Pp. 510–517.

2. **Wang H., Sun Y., Chen J., Wang F., Han R., Zhang C., Kong J., Li L., Yang J. A.** Review of perovskite-based photodetectors and their applications // Nanomaterials. 2022. Vol. 12. No. 24. P. 4390.

3. **Nenashev G. V., Aleshin A. N., Shcherbakov I. P., Petrov V. N.** Effect of temperature variations on the behavior of a two-terminal organic-inorganic halide perovskite rewritable memristor for neuromorphic operations // Solid State Communications. 2022. Vol. 348–349. 1 June. P. 114768.

4. **Roy P., Ghosh A., Barclay F., Khare A., Cuce E.** Perovskite solar cells: A review of the recent advances // Coatings. 2022. Vol. 12. No. 8. P. 1089.

5. **Wang J., Zhang J., Zhou Y., Liu H., Xue Q., Li X., Jen A. K.** Highly efficient all-inorganic perovskite solar cells with suppressed non-radiative recombination by a Lewis base // Nature Communications. 2020. Vol. 11. January. P. 177.

6. **Sadhu A. S., Huang Y.-M., Chen L.-Y., Kuo H.-C., Lin C.-C.** Recent advances in colloidal quantum dots or perovskite quantum dots as a luminescent downshifting layer embedded on solar cells // Nanomaterials. 2022. Vol. 12. No. 6. P. 985.

7. **Srivastava A. K., Zhang W., Schneider J., Halpert J. E., Rogach A. L.** Luminescent down-conversion semiconductor quantum dots and aligned quantum rods for liquid crystal displays // Advanced Science. 2019. Vol. 6. No. 22. P. 1901345.

8. **Александрова О. А., Галиева Д. М., Дробинцева А. О. и др.** Наночастицы, наносистемы и их применение. Ч. 1. Коллоидные квантовые точки. Под ред. В. А. Мошникова, О. А. Александровой. Уфа: Аэтерна, 2015. 236 с.

9. **Shi Z. J., Guo J., Chen Y. H., Li Q., Pan Y. F., Zhang H. J., Xia Y. D., Huang W.** Lead-free organic-inorganic hybrid perovskites for photovoltaic applications: Recent advances and perspectives // Advanced Materials. 2017. Vol. 29. No. 16. 3 February. P. 1605005.

10. **Gao W., Ran C., Li J., Dong H., Jiao B., Zhang L., Lan X., Hou X., Wu Z.** Robust stability of efficient lead-free formamidinium tin iodide perovskite solar cells realized by structural regulation // The Journal of Physical Chemistry Letters. 2018. Vol. 9. No. 24. Pp. 6999–7006.

11. **Li H., Zhang W.** Perovskite tandem solar cells: From fundamentals to commercial deployment // Chemical Reviews. 2020. Vol. 120. No. 18. Pp. 9835–9950.

12. **Rong Sh.-Sh., Faheem M. B., Li Y.-B.** Perovskite single crystals: Synthesis, properties, and applications. // Journal of Electronic Science and Technology. 2021. Vol. 19. No. 2. P. 100081.

13. **Lian Zh., Yan Q., Gao T., Ding J., Lv Q., Ning Ch., Li Q., Sun J.-L.** Perovskite $CH_3NH_3PbI_3(Cl)$ single crystals: Rapid solution growth, unparalleled crystalline quality, and low trap density toward 10^8 cm^{-3}// Journal of the American Chemical Society. 2016. Vol. 138. No. 30. Pp. 9409–9412.

14. **Zhang Y., Zhang H., Zhang X., Wei L., Zhang B., Sun Y., Hai G., Li Y.** Major impediment to highly efficient, stable and low-cost perovskite solar cells // Metals. 2018. Vol. 8. No. 11. P. 964.

15. **Xiao M., Huang F., Huang W., Dkhissi Y., Zhu Y., Etheridge J., Gray-Weale A., Bach U., Cheng Y., Spiccia L.** A fast deposition-crystallization procedure for highly efficient lead iodide perovskite thin-film solar cells // Angewandte Chemie International Edition. 2014. Vol. 53. No. 37. Pp. 9898–9903.

16. **Ahmed D. S., Mohammed B. K., Mohammed M. K. A.** Long-term stable and hysteresis-free planar perovskite solar cells using green antisolvent strategy // Journal of Materials Science. 2021. Vol. 56. No. 27. Pp. 15205–15214.

17. **Shaikh J. S., Shaikh N. S., Sheikh A. D., Mali S. S., Kale A. J., Kanjanaboos P., Hong C. K., Kim J. H., Patil P. S.** Perovskite solar cells: In pursuit of efficiency and stability // Materials & Design. 2017. Vol. 136. 15 December. Pp. 54–80.

18. **Li D., Shi J., Xu Y., Luo Y., Wu H., Meng Q.** Inorganic-organic halide perovskites for new photovoltaic technology // National Science Review. 2018. Vol. 5. No. 4. Pp. 559–576.

19. **Zhang W., Li Y., Liu X., Tang D., Li X., Yuan X.** Ethyl acetate green antisolvent process for high-performance planar low-temperature SnO_2-based perovskite solar cells made in ambient air // Chemical Engineering Journal. 2020. Vol. 379. 1 January. P. 122298.

20. **Zhang P., Gu N., Song L., Chen X., Du P., Zha L., Chen W.-Hs., Xiong J.** The disappearing additive: introducing volatile ethyl acetate into a perovskite precursor for fabricating high efficiency stable devices in open air // Nanoscale. 2022. Vol. 14. No. 13. Pp. 5204–5213.

21. **Khan R., Ighodalo K.O., Z. Xiao Z.** Ion migration in metal halide perovskites solar cells, soft-matter thin film solar cells: Physical processes and device simulation. Melville, New York: AIP Publishing LLC, 2020. p. 1-283

22. **Zhao Y.-C., Zhou W.-K., Zhou X., Liu K.-H., Yu D.-P., Zhao K.** Quantification of light-enhanced ionic transport in lead iodide perovskite thin films and its solar cell applications // Light: Science & Applications. 2017. Vol. 6. No. 5. 5 May. P. e16243.

THE AUTHORS

OVEZOV Maksat K.
Ioffe Institute, RAS
26 Polytekhnicheskaya St., St. Petersburg, 194021, Russia
strontiumx94@gmail.com
ORCID: 0009-0009-6273-1478

RYABKO Andrey A.
Ioffe Institute, RAS
26 Polytekhnicheskaya St., St. Petersburg, 194021, Russia
a.a.ryabko93@yandex.ru
ORCID: 0000-0001-9626-7612

ALESHIN Andrey N.
Ioffe Institute, RAS
26 Polytekhnicheskaya St., St. Petersburg, 194021, Russia
aleshin.transport@mail.ioffe.ru
ORCID: 0000-0001-5449-4446

MOSHNIKOV Vyacheslav A.
St. Petersburg Electrotechnical University "LETI"
5 Professor Popov St., St. Petersburg, 197376, Russia
vamoshnirov@mail.ru
ORCID: 0000-0001-6500-5492

KONDRATYEV Valeriy M.
Alferov University, RAS
Moscow Institute of Physics and Technology (National Research University)
8/3 Khlopin St., St. Petersburg, 194021, Russia
kvm_96@mail.ru
ORCID: 0000-0002-3469-5897

MAXIMOV Alexander I.
St. Petersburg Electrotechnical University "LETI"
5 Professor Popov St., St. Petersburg, 197376, Russia
aimaximov@mail.ru
ORCID: 0000-0003-0195-8870

СВЕДЕНИЯ ОБ АВТОРАХ

ОВЕЗОВ Максат Кемалович – *младший научный сотрудник лаборатории неравновесных процессов в полупроводниках Физико-технического института имени А. Ф. Иоффе Российской академии наук.*
194021, *Россия*, г. Санкт-Петербург, Политехническая ул., 26
strontiumx94@gmail.com
ORCID: 0009-0009-6273-1478

РЯБКО Андрей Андреевич – *кандидат технических наук, младший научный сотрудник лаборатории неравновесных процессов в полупроводниках Физико-технического института имени А. Ф. Иоффе Российской академии наук.*
194021, *Россия*, г. Санкт-Петербург, Политехническая ул., 26
a.a.ryabko93@yandex.ru
ORCID: 0000-0001-9626-7612

АЛЕШИН Андрей Николаевич – *доктор физико-математических. наук, заведующий лабораторией неравновесных процессов в полупроводниках Физико-технического института имени А. Ф. Иоффе Российской академии наук.*
194021, *Россия*, *г.* Санкт-Петербург, Политехническая ул., 26
aleshin.transport@mail.ioffe.ru
ORCID: 0000-0001-5449-4446

МОШНИКОВ Вячеслав Алексеевич – *доктор физико-математических наук, профессор кафедры микро- и наноэлектроники Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» имени В. И. Ульянова (Ленина).*
197376, *Россия*, г. Санкт Петербург, ул. Проф. Попова, 5
vamoshnirov@mail.ru
ORCID: 0000-0001-6500-5492

КОНДРАТЬЕВ Валерий Михайлович – *аспирант Академического университета имени Ж. И. Алфёрова Российской академии наук, г. Санкт-Петербург; младший научный сотрудник Московского физико-технического института (Национального исследовательского университета).*
194021, *Россия*, г. Санкт-Петербург, ул. Хлопина, 8/3
kvm_96@mail.ru
ORCID: 0000-0002-3469-5897

МАКСИМОВ Александр Иванович – *кандидат физико-математических наук, доцент кафедры микро- и наноэлектроники Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» имени В. И. Ульянова (Ленина).*
197376, *Россия*, г. Санкт Петербург, ул. Проф. Попова, 5
aimaximov@mail.ru
ORCID: 0000-0003-0195-8870

Received 14.09.2023. Approved after reviewing 20.10.2023. Accepted 20.10.2023.
Статья поступила в редакцию 14.09.2023. Одобрена после рецензирования 20.10.2023. Принята 20.10.2023.

Original article
DOI: https://doi.org/10.18721/JPM.16402

THE POLARIZABILITY ANISOTROPY OF OXYGEN ATOMS AS A POSSIBLE REASON FOR THE FORMATION OF ANTIFERROELECTRICITY IN PEROVSKITES

A. E. Ganzha✉, *M. A. Kniazeva*, *A. V. Filimonov*, *R. G. Burkovsky*

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ alexander.ganzha@gmail.com

Abstract. The paper analyzes the structural features of perovskite-like ABO_3 type materials responsible for the formation of antiferroelectric phases. For this purpose, the descriptions of some single crystals have been compared using three models: the adapted Slater dipole model (I), the Cowley shell model (II) and the Born – Karman model supplemented with consideration dipole-dipole forces and parameterized basing on *ab initio* calculations by Ghosez (III). The parameters of model I were found at which qualitative agreement with the data on inelastic X-ray scattering obtained by experiments with lead hafnate was observed. An analysis of all the results led to the conclusion that model I and the Ghosez parameterization confirmed the hypothesis about the key role of the lateral component of the oxygen atoms' polarizability over its axial component for the antiferroelectricity formation. However, model II data did not support this.

Keywords: structural phase transition, antiferroelectric, shell model, dipole-dipole interaction

Funding: The reported study was carried out within the framework of the State Assignment for Fundamental Research (Subject Code FSEG-2023-0016).

Citation: Ganzha A. E., Kniazeva M. A., Filimonov A. V., Burkovsky R. G., The polarizability anisotropy of oxygen atoms as a possible reason for the formation of antiferroelectricity in perovskites, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 20–29. DOI: https://doi.org/10.18721/JPM.16402

Научная статья
УДК 538.913
DOI: https://doi.org/10.18721/JPM.16402

АНИЗОТРОПИЯ ПОЛЯРИЗУЕМОСТИ АТОМОВ КИСЛОРОДА КАК ВОЗМОЖНАЯ ПРИЧИНА ФОРМИРОВАНИЯ АНТИСЕГНЕТОЭЛЕКТРИЧЕСТВА В ПЕРОВСКИТАХ

А. Е. Ганжа✉, *М. А. Князева, А. В. Филимонов, Р. Г. Бурковский*

Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Россия

✉ alexander.ganzha@gmail.com

Аннотация. В работе проанализированы особенности структуры перовскитоподобных материалов вида ABO_3, ответственные за формирование антисегнетоэлектрических фаз. С этой целью сопоставлено описание ряда монокристаллов с помощью трех моделей: адаптированной дипольной Слэтера (I), оболочечной Каули (II) и модели Борна – Кармана, дополненной учетом диполь-дипольных сил и параметризованной на основе первопринципных расчетов Госеза (III). Определены параметры модели I, при которых наблюдается качественное согласие с данными по неупругому рассеянию рентгеновского излучения из экспериментов с гафнатом свинца. Анализ всех результатов привел к заключению, что модель I и параметризация Госеза подтверждают гипотезу о ключевой роли латеральной компоненты поляризуемости атомов кислорода над ее аксиальной компонентой для формирования антисегнетоэлектричества. Однако результаты использования модели II этого не подтверждают.

Ключевые слова: структурный фазовый переход, антисегнетоэлектрик, оболочечная модель, диполь-дипольное взаимодействие

Финансирование: Работа выполнена в рамках Государственного задания на проведение фундаментальных исследований (код темы FSEG-2023-0016).

Ссылка для цитирования: Ганжа А. Е., Князева М. А., Филимонов А. В., Бурковский Р. Г. Анизотропия поляризуемости атомов кислорода как возможная причина формирования антисегнетоэлектричества в перовскитах // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 4 № .16. С. 20–29. DOI: https://doi.org/10.18721/JPM.14602

Introduction

Structural instabilities in perovskites are the focus of much attention, since perovskite crystals and thin films based on them find numerous technical applications in electromechanical sensors and drives, pyroelectric sensors, electrocaloric coolers, energy storage and memory devices [1–5].

When the crystal becomes unstable with respect to one of the phonon modes, a structural phase transition occurs. This process can be described as follows. A crystal consisting of N atoms has $3N$ degrees of freedom associated with displacements of these atoms from fixed equilibrium positions (three degrees of freedom for each atom).

The internal energy can be expressed in quadratic form in the harmonic approximation:

$$\frac{1}{2}\mathbf{u}^T \cdot \mathbf{U} \cdot \mathbf{u}, \tag{1}$$

where $\mathbf{u}$ is the displacement vector with the length $3N$, $\mathbf{U}$ is the matrix of force constants for the vector $\mathbf{u}$.

Some quadratic eigenvalues are necessarily zero, for example, the eigenvalues corresponding to homogeneous displacement of the crystal as a whole. However, the remaining eigenvalues must be positive so that the crystal tends to return to the state of equilibrium under an external perturbation. This theory is outlined in a number of papers and textbooks on lattice dynamics [6, 7]. When a structural phase transition from a cubic phase to a low-symmetry phase (characterized by some kind of distortion of the crystal lattice) occurs in perovskite-like crystals, the eigenvalue corresponding to this distortion becomes zero or negative. As for ferroelectric phase transitions, then this eigenvalue should correspond to the zero wave vector, i.e., the Γ point of the cubic Brillouin zone. The eigenvalues for other transitions, for example, antiferroelectric, correspond to another point in the Brillouin zone.

Thus, the question of why Perovskites undergo phase transitions and why all these transitions are so different can be broken into more specific questions, namely, why some of the eigenvalues could turn out to be negative and why this negative number corresponds exactly to this, and not to any other point in the Brillouin zone. For example, an approach outlined in John Slater's study [9] for barium titanate $BaTiO_3$ proposes to divide all forces acting on ions into two groups.

The *first group* includes short-range repulsive forces arising (in the simplest interpretation) because electron clouds of the ion crystal tend to reduce the degree of mutual overlap. These repulsive forces tend to stabilize the high-symmetry cubic phase, in other words, preserving all eigenvalues of the quadratic energy form positive.

The *second group* of forces are dipole–dipole forces resulting from the Coulomb interaction. Unlike the first group, dipole–dipole forces are long-range. If the influence of such dipoles on each other is taken into account, it turns out that such dipoles tend to line up along the O–B bonds in the ABO_3 structure in perovskite-like crystals. Such a tendency towards forming bonds by dipole–dipole forces can lead to a negative value of the eigenvalue corresponding to this distortion. The above-mentioned distortion induces ferroelectricity, since the displacements are arranged identically in each of the ABO_3 cells. This structural rearrangement is formed due to predominant influence of dipole–dipole forces specifically on the center of the Brillouin zone [8].

Thus, Slater's theory answers both questions posed for a ferroelectric such as $BaTiO_3$: why an eigenvalue becomes negative, and exactly which one it is.

However, John Slater's views have been regularly re-examined critically since the 1990s. For example, it was discovered through advances made in quantum mechanical calculations that these crystals are far from absolutely ionic, i.e., the overlap of electron clouds leads not only to repulsion but also to attraction due to orbital hybridization and formation of partially covalent bonds. In other words, destabilization of the cubic phase occurs not only due to the action of dipole–dipole forces, but also due to the influence of partial covalence of the bond. The role of dipole–dipole forces is preserved in this case, since covalent forces are short-range. Otherwise, it is difficult to explain exactly how long-range ordering of the structure can be achieved and establish the specfic point of the Brillouin zone that the resulting distortions correspond to.

A recent paper by Burkovsky [10] uses the Slater model [9] to describe mismatched antiferroelectric phases: it is found that dipole–dipole forces can enhance not only homogeneous polarization (center of the Brillouin zone) but also incommensurate distortions; however, this requires more complex characteristics of short-range forces that could cancel the tendency of dipole-dipole forces towards enhancing polarization along O–B bonds. This is possible if the crystal structure is such that it is extremely energetically unfavorable to form a dipole along such a bond even in the presence of dipole–dipole forces. For example, considerable energy costs are required if the ions are packed very closely and cannot shift much along these O–B bonds, and their electron clouds also cannot stretch in this direction. The tolerance factor acts as a measure of atomic packing for perovskite-like crystals: when its values are less than unity, the atoms are tightly packed along the O–B chains [11]. The tolerance factor for a $PbZrO_3$ crystal of the lead zirconate family is equal to 0.756, corresponding to the following case: the ionic radius of the B-ion (Zr^{4+}) is large, and the oxygen anions O^{2-} are 'sandwiched' between zirconium cations. If the ability to form O–B dipole bonds is suppressed, other pathways emerge for dipole ordering. For example, if strong polarizability of oxygen atoms in the O–A planes (perpendicular to the O–B bonds) is combined with strong polarizability of ion A, such loss of cubic symmetry for which the dipoles form an incommensurate modulation wave propagating in the O–A plane may be energetically favorable.

Slater's model is by no means the only one capable of explaining why incommensurability and antiferroelectricity arise in perovskites. We should consider the shell model describing the total energy of the lattice rather than that of its dipole subsystem as well as the Born–Karman model taking into account dipole–dipole forces, also used earlier to parameterize the lattice energy of perovskite-like crystals. These two models have not been compared before, and we propose to take the first steps in this direction.

Objectives and experimental procedure

To find out how significantly the combination of the polarizability of oxygen atoms in the O–A planes with the strong polarizability of ion A affects the formation of experimentally observed disproportionate and antiferroelectric phases in crystals of the lead zirconate family, we compare the values of the polarizability of various groups of atoms in the cubic phase of lead hafnate ($PbHfO_3$). To this end, we analyzed experimental data on inelastic X-ray scattering on single crystals of lead hafnate antiferroelectric in the framework of two models: dipole [10] and shell [12, 13].

In our opinion, the mathematical framework based on the shell and dipole models, used to describe the phenomena under consideration, is too extensive to be considered in its entirety in this paper, especially since an explicit mathematical representation of these models was given by Cowley [13] and Burkovsky [10]. We compare the existing results with the polarizabilities obtained from the first-principle (*ab initio*) quantum mechanical calculations by Ghosez et al. [14].

The shell model was first proposed by Dick and Overhauser [15] and adapted for the Perovskites by Cowley [13]. Polarizability consists of the sum of ionic and electronic polarizabilities within the framework of the model. The first one is associated with displacement of the entire ion from the equilibrium site. The displacement of an ion under the action of external forces is prevented by short-range forces between the ion shell and the shells of its neighbors. In addition to displacement of the entire ion, the charged core is displaced inside the electron shell. This process is equated to the second, electronic, polarizability.

The dipole model is considered a simplification of the shell model. The simplification is that it is sufficient to account for the long-range forces and only parts of short-range ones describing the polarizability of atoms in the crystal lattice to describe the internal energy in the crystal [10].

The study by Ghosez et al., with whose results we compare the polarizabilities, performed decomposition of interatomic interactions into short- and long-range components for several perovskite-like crystals: antiferroelectric lead zirconate ($PbZrO_3$) and ferroelectrics barium titanate and lead titanate ($BaTiO_3$ and $PbTiO_3$). We recalculated the polarizabilities from the Ghosez parameters as Z^2/k, where Z are the Born effective charges and k are the corresponding force constants.

The main results of the analysis are presented below.

Experimental phonon dispersion curves and their description based on the shell model are shown in Fig. 1,*a* and *b* [12]. Apparently, the shell model adequately reproduces the anisotropy of phonon dispersion curves for transverse phonons.

The acoustic phonon branches were converted into dipole stiffness α using the dipole model to describe the experimental data (see Fig. 1, *a, b*) by the formula

$$\alpha = CE^2,$$

where E, J, is the phonon energy; C, $C^2/J^2 \cdot m^3$), is the dimensional coefficient.

We consider this conversion to be fairly reasonable, since the frequency of transverse acoustic phonons and dipole stiffness are symmetric quantities: the higher the value of dipole stiffness, the greater energy the phonons creating such dipole waves should have. Fig. 1,*c* shows a three-dimensional graph of dipole stiffness in the $hk0$ plane of the Brillouin zone: qualitative agreement with experimental data on inelastic scattering by single crystals of lead hafnate was achieved at $C = 0.12$ $C^2/(J^2 \cdot m^3)$. To represent stiffness within the framework of the dipole model in this manner, the polarizability of the A-ion α_A should be higher than the polarizability of the B-ion α_B; it is obtained that the anisotropy parameter of oxygen atoms $\delta = \alpha_{O\text{-}A}/\alpha_{O\text{-}B}$ is equal to 1.40 in this case.

Analyzing the results in Fig. 1, we can conclude that both models adequately reproduce the anisotropy of phonon dispersion curves along different directions, which means that a certain coincidence can be expected when dipole and shell models are used.

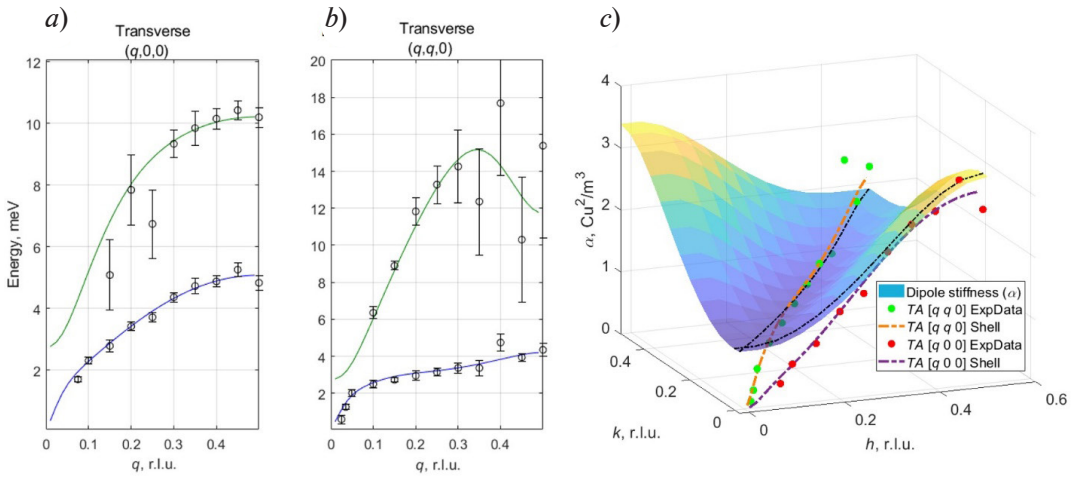

Fig. 1. Experimental phonon dispersion curves of lead hafnate in cubic phase (symbols) and their approximations by the shell model (solid lines) for transverse phonons propagating in the directions [100] (*a*) and [110] (*b*); description of acoustic phonon branch (TA) by the dipole model (*c*)

The table shows the polarizabilities of atoms obtained by the three models, as well as similar data for barium titanate $BaTiO_3$ obtained by Turik and Khasabov [16] using the shell model and the Slater model [9].

Table

Comparison of polarizabilities obtained by different calculation models for the oxygen anion in two crystal sites and for various cations

ABO_3	Model	Polarizability, Å^3			
		α_A	α_B	$\alpha_{O\text{-}A}$	$\alpha_{O\text{-}B}$
$PbHfO_3$	Shell [12, 13]	4.90	0.37	2.74	3.63
	Dipole [10]			3.90	2.78
$PbZrO_3$	*Ab initio* calculations by Ghosez [14]	177.43	22.33	55.80	12.48
$PbTiO_3$		89.86	52.73	21.70	32.39
$BaTiO_3$		10.07	52.17	9.97	38.86
	Shell [16]	1.95	0.18	0.64	3.80
	Slater [9]	1.94	0.19	2.38	

Notations: α_A, α_B are the polarizabilities of the A^{2+} and B^{4+} cations; $\alpha_{O\text{-}A}$, $\alpha_{O\text{-}B}$ are the polarizabilities of the O^{2-} anion in O–A planes and along O–B bonds. Note. The values of electronic polarizability are given for the shell model, and those of total polarizability for the other models.

According to our estimates, the polarizabilities α_A and α_B of A- and B-cations in lead hafnate turned out to be very similar for the dipole and shell models (the case in the table when they are strictly equal is accidental), but the polarizabilities of oxygen anions are different. Moreover, different predictions were obtained for the ratio of polarizabilities of oxygen atoms along different directions within the framework of the two models: the anisotropy parameter $\delta = \alpha_{O\text{-}A}/\alpha_{O\text{-}B} = 0.75$ for the shell model, and $\delta = 1.40$ for the dipole model.

This is an unexpected result, which can be explained by the simplification of the shell model, namely, that the dipole model does not account for all the particulars of the effect of short-range forces on atoms. However, the problem requires extensive research, and it is impossible to make an unambiguous choice in favor of any model at this stage.

According to the *ab initio* calculations by Ghosez, the value of the anisotropy parameter δ of oxygen atoms exceeds unity only for antiferroelectric $PbZrO_3$, while for ferroelectrics $BaTiO_3$ and $PbTiO_3$ this parameter is less than unity. This suggests that the anisotropy of oxygen polarizability, when dipole formation along the O–B bonds is suppressed and enhanced in the O–A planes, is a major factor in the origin of antiferroelectricity.

Notably, there is a large numerical difference between the polarizabilities obtained by Ghosez by calculations from first principles (*ab initio*) and those based on model predictions. In our opinion, the reason for the difference may lie in the instability of the crystal assumed by the Ghosez model (in contrast to the assumptions of the dipole and shell models), implying that the polarizabilities obtained by this model may be higher. In addition, we used Born charges to calculate polarizabilities using the Ghosez model; however, constraints are sometimes imposed on such charges in calculations of local crystal properties, which include the polarizability of individual atoms.

Discussion

In this paper, we analyzed the particulars of the structural configuration of perovskite-like materials to identify its specific properties inducing antisegnetoelectric states during structural phase transitions.

To this end, we described the experimentally observed phonon spectrum of lead hafnate ($PbHfO_3$) using two models: shell [12] and dipole [10]. These models are essentially similar: the internal energy of the crystal structure consists of two forces of a different nature that affect ions, the short-range and long-range ones. The former arise due to overlap of electron clouds of neighboring ions and include both repulsive forces aimed at maintaining the highly symmetric phase in the crystal and attractive forces caused by hybridization of electronic orbitals and the formation of partially covalent bonds. The latter, i.e., long-range forces are Coulomb in nature, determining the internal organization of the crystal at long distances, due to the ordering of dipoles in the crystal. The dipole model is a certain rationalization of the shell model and is convenient for representing incommensurate phases in antiferroelectrics. The simplification is that the description of internal energy in the crystal is limited, only accounting for a part of the short-range forces affecting the polarizability of the atoms of the crystal lattice (accounting for the long-range forces remains).

In general, both models adequately reproduce the experimental data, the descriptions are in agreement with respect to the polarizabilitiesof A- and B-cations, however, they yield different estimates for the polarizability of oxygen anions: according to the shell model, the polarizability of oxygen atoms $\alpha_{O\text{-}A}$ along the O–A planes is less than that for oxygen atoms along the O–B bonds; the results turn out to be exactly the opposite within the framework of the dipole model.

Comparison of the obtained parameters with the results of *ab initio* calculations by Ghosez et al. [14] for ferroelectric and antiferroelectric perovskites confirms the hypothesis of the key role that the special nature of polarizability anisotropy of oxygen atoms in the crystal plays in the structural phase transition into antiferroelectric or incommensurate phase. According to the results obtained by *ab initio* calculations, as well as by the dipole model, the value of $\alpha_{O\text{-}A}$ must exceed the value of $\alpha_{O\text{-}B}$ to achieve the transition of the crystal to the antiferroelectric phase. At the same time, there is no quantitative agreement between the parameters: the parameter values obtained by Ghosez are much higher than those for the dipole–dipole model. We attribute this discrepancy to the fact that we used Born charges to calculate local polarizabilities by the Ghosez model, which, strictly speaking, are determined for displacements of entire sublattices, and not individual ions, i.e., the polarizability calculated using Born charges may be significantly overestimated.

Conclusion

Each of the model descriptions obtained has some advantages over the others. The benefit of the dipole model is that the role of each of the parameters governing the experimental three-dimensional distribution of dipole stiffness is clear. The benefit of the shell model is that the calculated results can be directly compared with the experimental data on inelastic X-ray scattering. Meanwhile, the description by the Born–Karman model accounting for dipole–dipole forces allows to correlate the experimental data with the results obtained by calculating the electronic structure of many-particle systems (based on density functional theory (DFT)), i.e., DFT calculations.

The reasons for incommensurability and antiferroelectricity in perovskites are yet to be fully understood; attempts at interpretation include macroscopic models and atomistic models with varying degre of detail [10, 17–21]. In this paper, we limited ourselves to three models for which we were able to establish the role that the ratio of polarizabilities of various atoms plays in the given phenomenon. The predictions of two models are similar in this aspect, but contradict the predictions of the third one. So far, it seems premature to make an unambiguous choice in favor of any of the models, since each has its advantages with sound arguments to support them.

REFERENCES

1. **Haertling G. H.,** Ferroelectric ceramics: history and technology, J. Am. Ceram. Soc. 82 (4) (1999) 797–818.

2. **Scott J. F.,** Applications of modern ferroelectrics, Science. 315 (5814) (2007) 954–959.

3. **Wei X.-K., Tagantsev A. K., Kvasov A., et al.,** Ferroelectric translational antiphase boundaries in nonpolar materials, Nat. Commun. 5 (08 Jan) (2014) 3031.

4. **Geng W., Liu Y., Meng X., et al.,** Giant negative electrocaloric effect in antiferroelectric La-doped $Pb(ZrTi)O_3$ thin films near room temperature, Adv. Mater. 27 (20) (2015) 3165–3169.

5. **Xu B., Hciguez J., Bellaiche L.,** Designing lead-free antiferroelectrics for energy storage, Nat. Commun. 8 (30 May) (2017) 15682.

6. **Born M., Huang Kun,** Dynamical theory of crystal lattices, Clarendon Press, Oxford, UK, 1988.

7. **Maradudin A. A., Montroll E. W., Weiss G. H.,** Theory of lattice dynamics in the harmonic approximation, Academic Press, New York, London, 1963.

8. **Anderson P. W.,** Qualitative considerations on the statistics of the phase transition of $BaTiO_3$ type ferroelectrics, In book: A Career in Theoretical Physics, Second Ed., Pp. 61– 69 ("World Scientific Series in 20th Century Physics", Vol. 35), World Scientific Publishing Co. Pte. Ltd, USA, 2005.

9. **Slater J. C.,** The Lorentz correction in barium titanate, Phys. Rev. 78 (6) (1950) 748–760.

10. **Burkovsky R.,** Dipole-dipole interactions and incommensurate order in perovskite structures, Phys. Rev. B. 97 (18) (2018) 184109.

11. **Sato T., Takagi S., Deledda S., et al.,** Extending the applicability of the Goldschmidt tolerance factor to arbitrary ionic compounds, Sci. Rep. 6 (1) (2016) 23592.

12. **Burkovsky R., Andronikova D., Bronwald Yu., et al.,** An analysis of the phonon dispersion curves of lead hafnate in the cubic phase using lattice-dynamical models, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. (3 (248)) (2016) 9–14.

13. **Cowley R. A.,** Lattice dynamics and phase transitions of strontium titanate, Phys. Rev. 134 (4A) (1964) A981–A997.

14. **Ghosez Ph., Cockayne E., Waghmare U. V., Rabe K. M.,** Lattice dynamics of $BaTiO_3$, $PbTiO_3$, and $PbZrO_3$: A comparative first-principles study, Phys. Rev. B. 60 (2) (1999) 836–843.

15. **Dick Jr. B. G., Overhauser A. W.,** Theory of the dielectric constants of alkali halide crystals, Phys. Rev. 112 (1) (1958) 90–102.

16. **Turik A. V., Khasabov A. G.,** Shell model and single-particle ion potentials in $BaTiO_3$ crystal, Ferroelectr. 83 (1) (1988) 165–169.

17. **Hlinka J., Márton P.,** Phenomenological model of a 90° domain wall in $BaTiO_3$-type ferroelectrics, Phys. Rev. B. 74 (10) (2006) 104104.

18. **Tagantsev A. K., Vaideeswaran K., Vakhrushev S. B., et al.,** The origin of antiferroelectricity in $PbZrO_3$, Nat. Commun. 4 (29 July) (2013) 2229.

19. **Patel K., Prosandeev S., Yang Y., et al.,** Atomistic mechanism leading to complex antiferroelectric and incommensurate perovskites. Phys. Rev. B. 94 (5) (2016) 054107.

20. **Xu B., Hellman O., Bellaiche L.,** Order-disorder transition in the prototypical antiferroelectric $PbZrO_3$, Phys. Rev. B. 100 (2) (2019) 020102.

21.**Bussmann A., Bilz H., Roenspiess R., Schwarz K.,** Oxygen polarizability in ferroelectric phase transitions, Ferroelectr. 25 (1) (1980) 343 –346.

СПИСОК ЛИТЕРАТУРЫ

1. **Haertling G. H.** Ferroelectric ceramics: history and technology // Journal of the American Ceramic Society. 1999. Vol. 82. No. 4. Pp. 797–818.

2. **Scott J. F.** Applications of modern ferroelectrics // Science. 2007. Vol. 315. No. 5814. Pp. 954–959.

3. **Wei X.-K., Tagantsev A. K., Kvasov A., Roleder K., Jia C.-L., Setter N.** Ferroelectric translational antiphase boundaries in nonpolar materials // Nature Communications. 2014. Vol. 5. 08 January. P. 3031.

4. **Geng W., Liu Y., Meng X., Bellaiche L., Scott J. F., Dkhil B., Jiang A.** Giant negative electrocaloric effect in antiferroelectric La-doped $Pb(ZrTi)O_3$ thin films near room temperature // Advanced Materials. 2015. Vol. 27. No. 20. Pp. 3165–3169.

5. **Xu B., Hciguez J., Bellaiche L.** Designing lead-free antiferroelectrics for energy storage // Nature Communications. 2017. Vol. 8. 30 May. P. 15682.

6. **Борн М., Кунь Х.** Динамическая теория кристаллических решеток. Пер. с англ. М.: Изд-во иностранной литературы, 1958. 487 с.

7. **Марадудин А., Монтролл Э., Вейсс Дж.** Динамическая теория кристаллической решетки в гармоническом приближении. Пер. с англ. М.: Мир, 1965. 384 с.

8. **Андерсон П. В.** Качественное рассмотрение статистических фазовых переходов в сегнетоэлектриках типа $BaTiO_3$ // Физика диэлектриков (Сборник статей) Ред. Г. И. Сканави. М.: Изд. АН СССР, 1960. С. 290–301.

9. **Slater J. C.** The Lorentz correction in barium titanate // Physical Review. 1950. Vol. 78. No. 6. Pp. 748–760.

10. **Burkovsky R.** Dipole-dipole interactions and incommensurate order in perovskite structures // Physical Review B. 2018. Vol. 97. No. 18. P. 184109.

11. **Sato T., Takagi S., Deledda S., Hauback B. C., Orimo S. I.** Extending the applicability of the Goldschmidt tolerance factor to arbitrary ionic compounds // Scientific Reports. 2016. Vol. 6. No. 1. P. 23592.

12. **Burkovsky R., Andronikova D., Bronwald Yu., Filimonov A., Vakhrushev S.** An analysis of the phonon dispersion curves of lead hafnate in the cubic phase using lattice-dynamical models // St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 2016. No. 3 (248). Pp. 9–14.

13. **Cowley R. A.** Lattice dynamics and phase transitions of strontium titanate // Physical Review A. 1964. Vol. 134. No. 4A. Pp. A981–A997.

14. **Ghosez Ph., Cockayne E., Waghmare U. V., Rabe K. M.** Lattice dynamics of $BaTiO_3$, $PbTiO_3$, and $PbZrO_3$: A comparative first-principles study // Physical Review B. 1999. Vol. 60. No. 2. Pp. 836–843.

15. **Dick Jr. B. G., Overhauser A. W.** Theory of the dielectric constants of alkali halide crystals // Physical Review. 1958. Vol. 112. No. 1. Pp. 90–102.

16. **Turik A. V., Khasabov A. G.** Shell model and single-particle ion potentials in $BaTiO_3$ crystal // Ferroelectrics. 1988. Vol. 83. No. 1. Pp. 165–169.

17. **Hlinka J., Márton P.** Phenomenological model of a 90° domain wall in $BaTiO_3$-type ferroelectrics // Physical Review B. 2006. Vol. 74. No. 10. P. 104104.

18. **Tagantsev A. K., Vaideeswaran K., Vakhrushev S. B., et al.** The origin of antiferroelectricity in $PbZrO_3$ // Nature Communications. 2013. Vol. 4. 29 July. P. 2229.

19. **Patel K., Prosandeev S., Yang Y., Xu B., Hciguez J., Bellaiche L.** Atomistic mechanism leading to complex antiferroelectric and incommensurate perovskites // Physical Review B. 2016. Vol. 94. No. 5. P. 054107.

20. **Xu B., Hellman O., Bellaiche L.** Order-disorder transition in the prototypical antiferroelectric $PbZrO_3$ // Physical Review B. 2019. Vol. 100. No. 2. P. 020102.

21. **Bussmann A., Bilz H., Roenspiess R., Schwarz K.** Oxygen polarizability in ferroelectric phase transitions // Ferroelectrics. 1980. Vol. 25. No. 1. Pp. 343–346.

THE AUTHORS

GANZHA Alexander E.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
alexander.ganzha@gmail.com
ORCID: 0000-0002-9974-1073

KNIAZEVA Maria A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
kniazeva.maria225@yandex.ru
ORCID: 0000-0003-1120-2102

FILIMONOV Alexey V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
filimonov@rphf.spbstu.ru
ORCID: 0000-0002-2793-5717

BURKOVSKY Roman G.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
roman.burkovsky@gmail.com
ORCID: 0000-0003-0474-3242

СВЕДЕНИЯ ОБ АВТОРАХ

ГАНЖА Александр Евгеньевич – *аспирант, младший научный сотрудник Научно-образовательного центра «Физика нанокомпозитных материалов электронной техники» Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
alexander.ganzha@gmail.com
ORCID: 0000-0002-9974-1073

КНЯЗЕВА Мария Александровна – *инженер и младший научный сотрудник Научно-образовательного центра «Физика нанокомпозитных материалов электронной техники» Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
kniazeva.maria225@yandex.ru
ORCID: 0000-0003-1120-2102

ФИЛИМОНОВ Алексей Владимирович – *доктор физико-математических наук, профессор Высшей инженерно-физической школы, а также соруководитель Научно-образовательного центра «Физика нанокомпозитных материалов электронной техники» Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
filimonov@rphf.spbstu.ru
ORCID: 0000-0002-2793-5717

БУРКОВСКИЙ Роман Георгиевич – *кандидат физико-математических наук, доцент Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
roman.burkovsky@gmail.com
ORCID: 0000-0003-0474-3242

Received 07.09.2023. Approved after reviewing 08.11.2023. Accepted 12.11.2023.
Статья поступила в редакцию 07.09.2023. Одобрена после рецензирования 08.11.2023. Принята 12.11.2023.

Original article
DOI: https://doi.org/10.18721/JPM.16403

TEMPERATURE CHARACTERIZATION OF GaAs/AlGaAs CONNECTING TUNNEL DIODES

E. V. Kontrosh✉, *V. S. Kalinovskii, G. V. Klimko, B. Ya. Ber, K. K. Prudchenko, I. A. Tolkachev, D. Yu. Kazantsev*

Ioffe Institute of RAS, St. Petersburg, Russia

✉ kontrosh@mail.ioffe.ru

Abstract. The current-voltage characteristics of two types of GaAs-(δSi)/*i*-(GaAs/$Al_{0.2}Ga_{0.8}As$)/p^{++}-$Al_{0.2}Ga_{0.8}As$-(δBe) tunnel diode (TD) structures grown at different temperatures and epitaxial layer thicknesses have been investigated in the temperature range 100–400 K. Temperature dependences of the main TD parameters were determined: the peak value of the tunnel current density (J_p), the valley current density (J_v) and the differential resistance (R_d). TD samples of structure A grown at 500 °C exhibited the highest values of the peak current density ($J_p \leq 220$ A/cm^2) with temperature stability of 93 % over the whole temperature range. TD samples of structure B grown at 450 °C showed lower values of the peak tunneling current density ($J_p \leq 150$ A/cm^2), with significantly linear temperature dependence. Our findings can be used in the design and development of monolithic multijunction photoconverters of powerful laser radiation.

Keywords: current-voltage characteristics, tunnel diode, epitaxial layer, differential resistance, peak tunneling current

Citation: Kontrosh E. V., Kalinovskii V. S., Klimko G. V., Ber B. Ya., Prudchenko K. K., Tolkachev I. A., Kazantsev D. Yu., Temperature characterization of GaAs/AlGaAs connecting tunnel diodes, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 30–41. DOI: https://doi.org/10.18721/JPM.16403

Научная статья
УДК 621.315.592
DOI: https://doi.org/10.18721/JPM.16403

ТЕМПЕРАТУРНАЯ ХАРАКТЕРИЗАЦИЯ СОЕДИНИТЕЛЬНЫХ ТУННЕЛЬНЫХ ДИОДОВ GaAs/AlGaAs

Е. В. Контрош✉, *В. С. Калиновский, Г. В. Климко, Б. Я. Бер, К. К. Прудченко, И. А. Толкачев, Д. Ю. Казанцев*

Физико-технический институт им. А.Ф. Иоффе РАН, Санкт-Петербург, Россия

✉ kontrosh@mail.ioffe.ru

Аннотация. В температурном диапазоне 400 – 100 K исследованы вольтамперные характеристики двух типов структур соединительных туннельных диодов (ТД) n^{++}-GaAs-(δSi)/*i*-(GaAs/$Al_{0,2}Ga_{0,8}As$)/p^{++}-$Al_{0,2}Ga_{0,8}As$-(δBe), отличающихся температурой роста и толщинами эпитаксиальных слоев. Определены температурные зависимости основных параметров ТД: пикового значения плотности туннельного тока J_p, плотности тока долины J_v и дифференциального сопротивления R_d. Образцы ТД структуры А, выращенной при температуре 500 °C, обеспечивают в диапазоне 100 – 400 K наибольшие значения пикового тока $J_p \leq 220$ А/см^2 при температурной стабильности величины около 93 %. ТД структуры В, выращенные при температуре 450 °C, показали меньшие значения плотности пикового туннельного тока: $J_p \leq 150$ А/см^2, с существенной линейной температурной зависимостью. Полученные результаты могут быть использованы при разработке и создании монолитных многопереходных фотопреобразователей мощного лазерного излучения.

Ключевые слова: вольтамперная характеристика, туннельный диод, эпитаксиальный слой, дифференциальное сопротивление, пиковый туннельный ток

Ссылка для цитирования: Контрош Е. В., Калиновский В. С., Климко Д. В., Бер Б. Я., Прудченко К. К., Толкачев И. А., Казанцев Д. Ю. Температурная характеризация соединительных туннельных диодов GaAs/AlGaAs // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. 30–41. DOI: https://doi.org/10.18721/JPM.16403

Introduction

High-power monolithic multijunction photovoltaic (PV) cells converting monochromatic light have potential for applications in optoelectronics systems operating both on Earth and in space. Such systems include a radiophoton phased array antenna array [1], energy-independent transceiver stations for free-space optical communication [2], batteries for autonomous optoelectronic devices, etc. [3, 4]. Depending on the optical power and the field of application, multijunction PV cells can operate in a wide temperature range (100–400 K) [5, 6]. Monolithic multijunction PV cells include several series-connected photoactive *p*–*n* junctions, i.e., semiconductor-based subunits with the same band gap width but with different geometries and doping levels of the layers. The subunits in the PV cell are connected via backward tunnel diodes (TD). The efficiency and reliability of multijunction PV cells significantly depends on the temperature stability of the parameters of the connecting TD: peak tunneling current density J_p, differential resistance R_d of the tunnel branch and high optical transparency over a wide range of operating temperatures. A distinct characteristic of connecting TD is the high degree of degeneration of sub-nanosized layers, achieved by the delta-doping method. However, interdiffusion of donor and acceptor impurities occurs in degenerate layers of TD during epitaxial growth of the entire structure of the PV cell, leading to profile smearing and a decrease in free charge carrier concentration. These factors have a significant impact on the parameters of the TD and the behavior of their temperature dependence.

In this paper, we experimentally measured the current–voltage (I–V) characteristics of GaAs/AlGaAs connecting tunnel diodes in the temperature range from 100 to 400 K, establishing the temperature dependences of the parameters J_p and R_d and analyzing the results obtained.

Experimental

Connecting p^{++}–i–n^{++} TD were considered for two types of structures: A and B, grown by molecular beam epitaxy (MBE). The distributions of atomic concentrations in the given structures, determined by secondary ion mass spectrometry (SIMS) are shown in Fig. 1. 'Quasi-neutral' i-regions were formed between the degenerate regions of both TD structures, consisting of two layers of different thicknesses: i-GaAs and i-$Al_{0.2}Ga_{0.8}As$. Both TD structures were grown on GaAs (100) p-type substrates with a beryllium concentration $N_A = 1 \cdot 10^{19}$ cm^{-3}. After the buffer layers were grown, the epitaxial temperatures were decreased to 500 and 450 °C for structures A and B, respectively.

A significant diffusion of the beryllium dopant into the degenerate region of n^{++}-GaAs doped with silicon was observed in both TD structures (see Fig. 1).

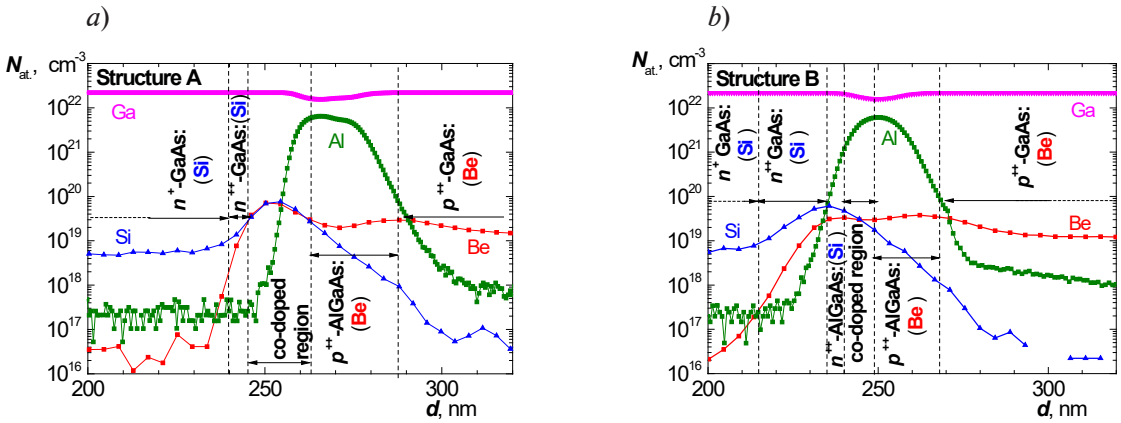

Fig. 1. Distribution of dopant concentration N_{at} over the thickness of the sample d in two types of TD structures: A (*a*) and B (*b*), determined by secondary ion mass spectrometry. Curves corresponding to different elements are shown in different colors

Analyzing the data in Fig. 1, we can conclude that diffusion of beryllium in structure A contributed to a decrease in the thickness of the n^{++}-GaAs and the concentration of free charge carriers in it, due to overcompensation of donor and acceptor impurities. The thickness of this layer, uncompensated by beryllium impurity, was approximately 5 nm with the concentration of silicon atoms varying from $1 \cdot 10^{19}$ to $3 \cdot 10^{19}$ cm^{-3} in structure A, and approximately 20 nm with the concentration varying from $9 \cdot 10^{18}$ to $6 \cdot 10^{19}$ cm^{-3} in structure B. Diffusion of beryllium atoms caused compensated quasi-neutral regions to appear between degenerate n^{++}- and p^{++}-layers. This is the co-doped region in Figs. 1,*a* and *b*. This region has a slightly larger thickness (about 25 nm) in structure A, consisting of two layers, GaAs:(Si, Be) and AlGaAs:(Si, Be) (see Fig. 1,*a*). This region consists of only one AlGaAs layer: (Si, Be) in structure B, with the thickness not exceeding 10 nm (see Fig. 1,*b*).

The measured peak concentrations of silicon and beryllium atoms approximately coincide in the overlap region of structure A, amounting to no more than $8 \cdot 10^{19}$ cm^{-3}. A different situation evolves in structure B, where the concentration of silicon atoms in the AlGaAs overlap region prevails over the concentration of beryllium atoms, with $N_D = 5 \cdot 10^{19}$ cm^{-3}, $N_A = 2 \cdot 10^{19}$ cm^{-3}.

Diode arrays with the mesa diameter of 225 μm, equipped with multilayer Ohmic contacts with n and p regions of AuGe-Ni-Au and AgMn-Ni-Au annealed in hydrogen at 500 °C, were formed on the TD structures by the post-growth technology.

I–V measurements of TD samples with structures A and B were performed at forward-bias voltages up to 1 V.

Results and discussion

Unlike structure B, structure A exhibited a spread in J_p values from 90 to 220 A/cm^2 at 300 K over an epitaxial wafer with a diameter of about 6 cm. The J_p values in the center of this wafer were close to the average of 116 A/cm^2, increasing to about 220 A/cm^2 at the periphery. A significantly smaller range of values, namely J_p = 125–150 A/cm^2, was obtained for samples in the center and at the periphery of the epitaxial wafer of structure B.

We selected TD samples from the central and peripheral regions of the epitaxial wafers of both structures to study the parameters J_p and R_d in the temperature range of 100–400 K. The I–V characteristics of the selected samples measured in the given range are shown in Fig. 2.

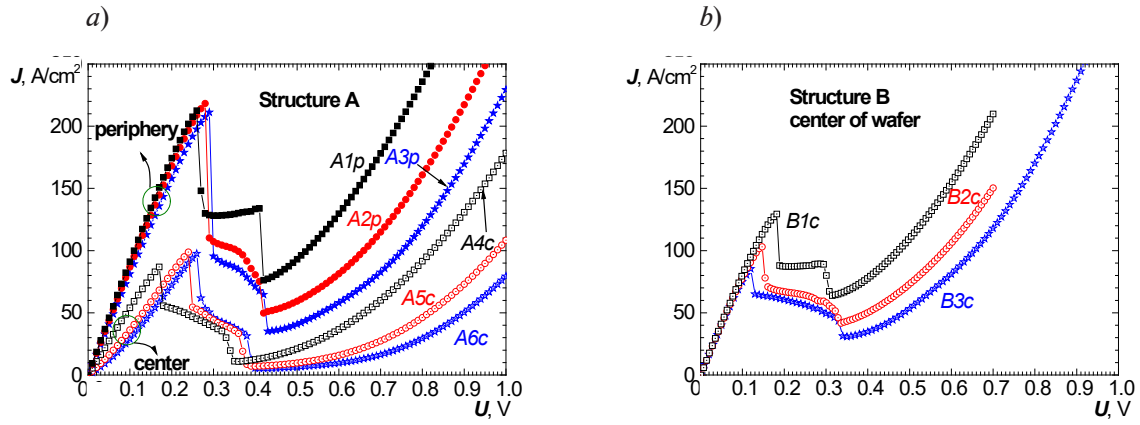

Fig. 2. Measurements of forward I–V characteristics in TD samples with two types of structures: A (*a*) and B (*b*) at different temperatures. Samples were taken at the periphery (curves *A1p*–*A3p*) and in the centers (*A4c*–*A6c*, *B1c*–*B3c*) of epitaxial wafers; *T*, K: 353 (*A1p*, *A4c*, *B1c*), 223 (*A2p*, *A5c*, *B2c*) and 123 (*A3p*, *A6c*, *B3c*)

Linear dependences are observed for the most part of the tunneling region in the experimental I–V curves obtained for any samples of structure B and peripheral structure A with a temperature decrease to 100 K (see Fig. 2). At the same time, an exponential dependence appears in the tunneling region of the I–V curves for samples from the center of structure A with a decrease in temperature. This behavior of dependences may be due to simultaneous influence of several factors.

As established in [7], the tunneling current in TD is determined by two transport mechanisms: interband quantum tunneling and trap-assisted tunneling. Both contribute to peak tunneling current density. The first mechanism is that electrons tunnel from occupied states of the conduction band to free states in the valence band through a potential barrier. The second, trap-assisted, transport mechanism, is due to the presence of localized impurity states (traps) in the band gap of the semiconductor. In this case, the tunneling electron is trapped, subsequently tunneling into the allowed states of the valence band. A decrease in temperature contributes to active freeze-out of charge carriers in traps; therefore, an increase in the electric field is required to overcome localized impurity states, leading to an exponential current-voltage dependence [8].

SIMS measurements for structure A (see Fig. 1,*a*) indicate that this transport mechanism may be predominant due to high degree of overlap of donor and acceptor impurities in the central region of the p^{++}–n^{++} tunnel junction.

We used the experimental data and expression (1) [9] to calculate the normalized temperature coefficient of the peak tunneling current density:

$$\Delta J_p = \frac{J_p^{T_j} - J_p^{T_{RT}}}{J_p^{T_{RT}}} \cdot 100\%, \qquad (1)$$

where ΔJ_p is the temperature coefficient; $J^{T_j}_p$, $J^{T_{RT}}_p$ are the peak tunneling current densities at fixed T_j and room temperature (T_{RT} = 300 K), respectively.

A positive value of ΔJ_p indicates an increase, and a negative value a decrease in the parameter J_p relative to its value at room temperature.

The dependences of J_p and ΔJ_p for structures A and B are shown in Fig. 3, *a*, *c*. TD samples from the central (*Ac* curves) and peripheral (*Ar* curves) parts of structure A exhibit higher thermal stability of peak tunneling current density, compared with samples from structure B (*Bc* curves). The variations in the maximum value of J_p amount to 17% for TD samples of structure A from the center and periphery of the wafer (*Ac* curve in Fig. 3,*a*) and 7% (*Ap* curve), respectively. The variation in the maximum value of J_p was 42% for TD samples of structure B (curve *Bc* in Fig. 3,*a*).

The parameter J_p decreases under heating from 300 to 400 K for TD samples taken from the center and periphery of the epitaxial wafer of structure A, with the value of ΔJ_p amounting to −9.5% in the center of the wafer and to −6.8% at the periphery(curves *Ac* and *Ap* in Fig. 3,*c*).

As the temperature decreased from 300 to 100 K, samples from the center of wafer A exhibited a nonlinear increase in the parameter J_p with the temperature coefficient $\Delta J_p = 7.5\%$, while the parameter J_p in peripheral samples decreased with the coefficient $\Delta J_p = -4.0\%$.

Notably, smooth maxima of peak tunneling currents are observed for the samples of structure A in Fig. 3,*a* in the temperature ranges of 150−250 K (*Ac* curve) and 200−300 K (*Ap* curve), respectively. The dependence of J_p for samples of structure B exhibits linear growth over the entire temperature range (curve *Bc* in Fig. 3,*a*). The value of ΔJ_p under heating from 300 to 400 K was 14%, and 29% under cooling to about 100 K.

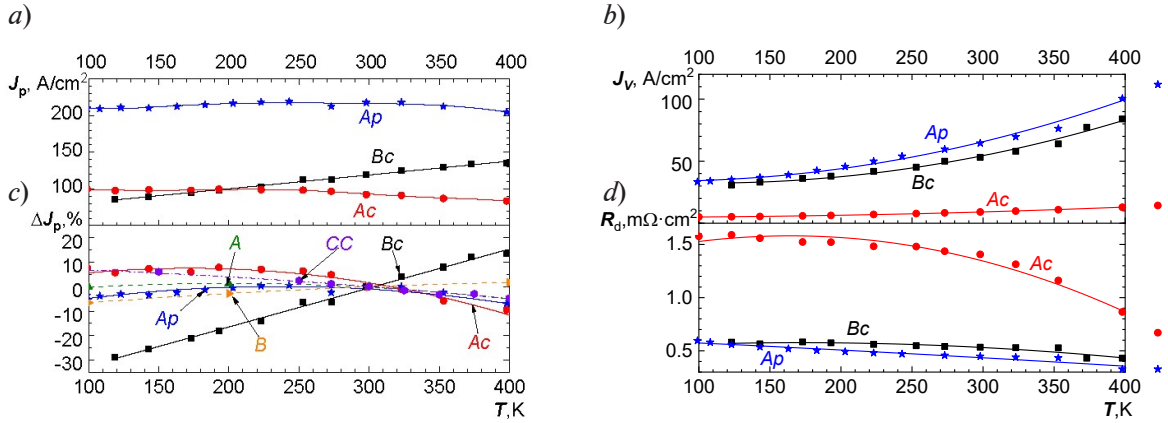

Fig. 3. Experimental (curves *Ac*, *Ap*, *Bc*) and calculated (*A*, *B*) temperature dependences for key parameters of TD samples of structures A and B: J_p (*a*), ΔJ_p (*c*), J_v (*b*), R_d (*d*). The samples were taken from the center (*Ac*, *Bc*) or from the periphery (*Ap*) of epitaxial wafers of structures A and B.
The calculated curve (*CC*) from [17] for a two-barrier resonant AlGaAs/GaAs tunnel diode is given for comparison

The behavior of the temperature dependence of J_p in TD samples with structures A and B depends on several factors with opposing influences on the value of J_p [6, 9]. Firstly, the band gap E_g of the semiconductor decreases with increasing temperature, leading to a decrease in the height of the potential barrier and an increase in the probability of quantum tunneling and the magnitude of J_p. Secondly, an increase in temperature reduces the degree of degeneracy of energy levels due to redistribution of electrons along them. The number of electrons in the conduction band at levels below the Fermi level E_F in the *n*-region E_c decreases, as some of the free electrons move to higher energy levels, and the Fermi level shifts downwards. As a result, the number of electrons capable of tunneling decreases, and the value of J_p decreases as well.

SIMS profiling (see Fig. 1) indicates that the doping level of the beryllium acceptor impurity in the degenerate p^{++}-AlGaAs region for both TD structures is approximately the same and is at least $2 \cdot 10^{19}$ cm^{-3}; in this case, the n^{++}-GaAs region has the predominant influence on the temperature characteristics of TD. As the concentration of free charge carriers decreases (below $1 \cdot 10^{19}$ cm^{-3}) in the n^{++}-GaAs region, the change in the position of the Fermi level starts to have the predominant

influence on the tunneling current J_p with an increase in temperature. The latter shifts closer to the bottom of the conduction band, and the value of J_p decreases. The change in the band gap has the predominant influence on the value of J_p in the range of 100–400 K at a higher concentration of free charge carriers (above $1 \cdot 10^{19}$ cm^{-3}), while the change in the position of the Fermi level has little effect; as a result, the value of J_p increases with increasing temperature.

We calculated the temperature dependences of ΔJ_p in TD similar to structures A and B (see Fig. 1) by the model described in [10]. The behavior of the calculated dependences of ΔJ_p (compare curves *A* and *B* in Fig. 3,*c*) is in qualitative agreement with the experimental ones (curves *Ac*, *Ar*, *Bc* in Fig. 3,*c*).

The presence of a negative temperature coefficient in the range of 300–400 K for the TD samples of structure A, in contrast to TD of structure B, indicates that the concentration level of free charge carriers of the n^{++}-GaAs region in the center and at the periphery of structure A is significantly lower than in structure B, due to the greater thickness of the overlap region and compensation of silicon and beryllium impurities (see Fig. 1,*a* and 2,*a*). Since the temperature coefficient of structure A from the center of the wafer is higher in absolute value for TD samples ($\Delta J_p = -9.5\%$) than for TD samples from the periphery ($\Delta J_p = -6.8\%$) at 400 K, the concentration of free charge carriers in the n^{++}-GaAs region of peripheral samples is slightly higher than in the center of the wafer with structure A but lower than in the center of the wafer with structure structure B. This is because the value of J_p increases for TD with structure B under heating with the temperature coefficient $\Delta J_p = 13.6\%$. The difference in parameters between the peripheral and central TD samples in the wafer of structure A may be due to a temperature gradient along the wafer even during the growth of the structure (the temperature at the periphery is several degrees lower than in the center). This leads to a decrease in the degree of overcompensation at the periphery.

The value of J_p of the TD considered is determined by the doping level and the degeneration degree of the n^{++}-GaAs TD layer. However, the J_p value in peripheral samples of TD with structure A is approximately 220 A/cm^2, while the corresponding maximum value for TD with structure B is about 150 A/cm^2. Our hypothesis is that this may be due to both the thickness of the 'quasi–neutral' overcompensated silicon-beryllium region [10] and the presence of a higher defect concentration in it due to overcompensation of impurity atoms [11–16]. This is also confirmed by the densities of valley currents in the dark *I–V* curves of the samples of the considered structures (see Figs. 2 and 3,*b*). As the forward bias voltage is increased (see Fig. 2), the current density first increases to the J_p value at a U_p voltage, and then decreases to the minimum value of the valley current density J_v at a U_v voltage due to a decrease in the degree of overlap of the conduction band with the valence band [9]. The valley current density is associated with the excess component of the current density of TD I–V characteristic. In turn, this excess component is determined by the concentration of deep levels inside the band gap of the semiconductor and the presence (or absence) of various types of structural defects. The defects contribute to dominance of an additional charge carrier transport mechanism associated with the resonant tunneling mechanism.

Since the overcompensation region is located between the degenerate n^{++} and p^{++} regions, it is depleted of the main charge carriers and is an 'effective' *i*-layer. We used numerical simulation [10] to establish that the dependence of J_p in tunnel diodes

$$n^{++}\text{-GaAs-}(\delta\text{Si})/i\text{-}(\text{GaAs/Al}_{0.2}\text{Ga}_{0.8}\text{As})/p^{++}\text{-Al}_{0.2}\text{Ga}_{0.8}\text{As-}(\delta\text{Be})\ (p\text{–}i\text{–}n)$$

on the thickness of the *i*-region is nonmonotonic. The peak current density first increases, reaching a maximum, and then decreases due to an increase in the thickness of the potential barrier through which the charge carriers tunnel.

Thus, the thickness of the 'effective' *i*-region caused by overcompensation of donor and acceptor impurities can affect the magnitude of J_p.

In addition, due to high concentration of silicon donors and beryllium acceptors, a higher concentration of defects and associated localized impurity states is present in the overcompensated 'effective' *i*-layer. Interband quantum tunneling acts as the main mechanism for transport of charge carriers in tunnel diodes with a sharp doping profile [11]. However, resonant tunneling (RT) begins to prevail in the presence of a sufficiently large region with overcompensation of donor and acceptor doping profiles and a high concentration of localized impurity states in the potential barrier [12]. In this case, resonance occurs if the energy of the states in the conduction

band coincides with the energies of the impurity states in the potential barrier and the allowed states in the valence band. The theory of RT through two localized impurity states is described in [13]. According to [14–16], localized states coinciding in energy and induced by a donor-acceptor pair contribute to a significant increase in J_p at low positive bias voltages in the TD. These assumptions are confirmed by the temperature dependence of J_p in resonant tunnel diodes.

For comparison, Fig. 3,*c* (*CC* curve) shows the temperature dependence of ΔJ_p for an AlGaAs/GaAs two-barrier resonant tunnel diode (RTD), simulated by the Hartree model, obtained in [17]. Analyzing this dependence, obtained under heating from 300 to 400 K, we can observe a decrease in the value of J_p at a negative temperature coefficient $\Delta J_p = -5\%$, the same as for structure A (see curves *Ac*, *Ap*). On the other hand, J_p increases smoothly under cooling from 300 to 100 K, giving the value of $\Delta J_p = 7\%$. The presence of a negative temperature coefficient in the I–V curves of RTD heated to 400 K is explained in [17] by scattering of charge carriers by phonons and electron-electron interaction.

We calculated the differential resistance responsible for parasitic losses at a voltage drop in TD in multijunction PV cells at different temperatures from the experimental I–V curves of the TD (Fig. 3,*d*). The connecting elements must provide resistances below 10 mOhm·cm^2 for effective operation of multijunction PV cells [6]. The dependences of differential resistance of the studied TD samples on temperature, accounting for resistance of the electrical circuit in the cryostat (about 0.7 mOhm·cm^2), are shown in Fig. 3,*d*. The best temperature stability R_d is observed for TD samples of structure B (see Fig. 3, *d*, curve *Bc*). The value of R_d varies from 0.58 to 0.42 mOhm·cm^2 over the entire temperature range for TD samples of structure B. The value of R_d varies from 1.59 to 0.67 mOhm·^{2} for TD samples from the central part and from 0.58 to 0.34 mOhm·cm^2 for samples from the peripheral part for structure A heated from 100 to 400 K.

The conducted temperature studies of forward I–V characteristics lead us to conclude that the developed connecting TD with structure A ensure the temperature stability of the parameters J_p, ΔJ_p at about 93% and the parameter R_d at about 59%, while structures B provide about 83% and 72%, respectively.

Conclusion

Forward current–voltage characteristics were studied for connecting tunnel diodes (TD) in the temperature range of 100–400 K, taking the form

$$n^{++}\text{-GaAs-}(\delta\text{Si})/i\text{-}(\text{GaAs}/\text{Al}_{0.2}\text{Ga}_{0.8}\text{As})/p^{++}\text{-Al}_{0.2}\text{Ga}_{0.8}\text{As-}(\delta\text{Be})$$

for two types of structures obtained by molecular beam epitaxy. The type A structure contained an overcompensated 'effective' *i*-layer with a thickness of ≤ 25 nm, formed by interdiffusion of silicon and beryllium at an epitaxial growth temperature of 500 °C. The type B structure had an 'effective' *i*-layer with a thickness of ≤ 10 nm, formed similarly at a growth temperature of 450 °C.

We observed a smooth decrease in the tunneling current density J_p for the TD with the structure A under heating to 400 K, while a linear growth of J_p was characteristic for the structure B. The change in J_p was 7% for TD samples with the structure A, and 42% for TD samples with the structure B. This J_p dependence on temperature for structure A is associated with the influence of temperature diffusion of the predominantly acceptor beryllium impurity from the layer

$$p^{++}\text{-Al}_{0.2}\text{Ga}_{0.8}\text{As-}(\delta\text{Be})$$

to the region

$$n^{++}\text{-GaAs-}(\delta\text{Si})/i\text{-}(\text{GaAs}/\text{Al}_{0.2}\text{Ga}_{0.8}\text{As}).$$

This diffusion contributed to a decrease in the degeneracy of the n^{++}-GaAs-(δSi) layer of the TD due to formation of an overcompensated silicon and beryllium region.

The linear increase in J_p with an increase in temperature for structure B is due to a lower thickness of the 'effective' *i*-layer, a smaller diffusion depth of beryllium impurity into the n^{++}-GaAs region and, accordingly, the higher doping and degeneracy levels of this region.

The maximum values of J_p obtained in the temperature range of 100–400 K were about 220 A/cm^2 for structure A and about 150 A/cm^2 for structure B. Such values of J_p can be due to both the thickness of the 'effective' *i*-layer and by the level of localized impurity states initiating resonant tunneling.

Analyzing the measured I–V characteristics, we found that the differential resistances for TD samples of structure A lie within R_d = 0.58–0.34 mΩ·cm^2 in the temperature range from 100 to 400 K, and within 0.58–0.42 mΩ·cm^2 for structure B.

The TD samples collected at the periphery of the epitaxial wafer of structure A grown at 500 °C exhibit higher temperature stability and maximum values of J_p at minimum values of R_d. A high degree of overcompensation of doping impurities in the active region of the TD located in the center of the wafer leads to a decrease in the peak tunneling current density. Samples of structure B grown at 450 °C provide better stability of the peak current density across the wafer, but with a lower maximum value of J_p and worse temperature stability. We assume that the optimal conditions for growth of the TD structure of type

$$n^{++}\text{-GaAs-}(\delta\text{Si})/i\text{-}(\text{GaAs/Al}_{0.2}\text{Ga}_{0.8}\text{As})/p^{++}\text{-Al}_{0.2}\text{Ga}_{0.8}\text{As-}(\delta\text{Be})$$

with the maximum values of J_p, minimum values of R_d and high temperature stability lie in the epitaxial growth temperature range of $450 < T < 500$ °C.

The temperature diffusion of impurities in highly doped layers of connecting TD must be taken into account in designs of multijunction photovoltaic cells converting high-power optical radiation Including an undoped *i*-layer up to 10 nm thick between the highly doped TD layers as well as optimizing the epitaxial growth temperature can prevent parasitic diffusion of the impurity. Furthermore, it is important to use a dopant with a lower diffusion coefficient, for example carbon as an acceptor impurity.

We should note that taking into account the contribution of resonant tunneling in connecting TD with high optical transparency has potential for constructing highly efficient multijunction photovoltaic converters of high-power laser radiation.

SIMS studies were performed using the equipment of the Center for Collective Use “Materials Science and Diagnostics in Advanced Technologies” (Ioffe Institute).

REFERENCES

1. **Zaitsev D. F., Andreev V. M., Bilenko I. A., et al.,** First radiophoton phased antenna array, Radiotekhnika. 85 (4) (2021) 153–164 (in Russian).

2. **Kalinovskii V. S., Terukov E. I., Kontrosh E. V., et al.,** Energy-informational hybrid photovoltaic converter of laser radiation, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (1.2) (2023) 47–51.

3. **Kalinovskiy V. S., Kontrosh E. V., Gusev G. A., et al.,** Study of PV characteristics of $Al_xGa_{1-x}As$/GaAs photodiodes, J. Phys. Conf. Ser. 993 (2018) 012029.

4. **Wang A.-Ch., Sun Y.-R., Yu Sh.-Zh., et al.,** Characteristics of 1520 nm InGaAs multijunction laser power converters, Appl. Phys. Lett. 119 (24) (2021) 243902.

5. **Hoheisel R., Bett A. W., Warner J. H., et al.,** Low temperature low intensity effects in III-V photovoltaic devices for deep space missions, Proc. IEEE 7th World Conf. Photovoltaic Energy Conversion (WCPEC) (A Joint Conf. 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC). June 10–15 (2018) 3763–3767.

6. **Lumb M. P., González M., Yakes M. K., et al.,** High temperature current–voltage characteristics of InP-based tunnel junctions, Prog. Photovoltaic. 23 (6) (2015) 773–782.

7. **Baudrit M., Algora C.,** Tunnel diode modeling, including nonlocal trap-assisted tunneling: A focus on III–V multijunction solar cell simulation, IEEE Trans. Electron Devices. 57 (10) (2010) 2564–2571.

8. **Tabe M., Tan H. N., Mizuno T., et al.,** Atomistic nature in band-to-band tunneling in two-dimensional silicon *pn* tunnel diods, Appl. Phys. Lett. 108 (9) (2016) 093502.

9. **Sze S. M.,** Physics of semiconductor devices, 2nd Ed., A Wiley-Interscience Publ., J. Wiley & Sons, New York, Chichester, Brisbar, Toronto, Singapore, 1981.

10. **Kalinovskii V. S., Kontrosh E. V., Klimko G. V., et al.,** Development and Study of the *p–i–n* GaAs/AlGaAs tunnel diodes for multijunction converters of high-power laser radiation, Semiconductors. 54 (3) (2020) 355–361.

11. **Esaki L.,** New phenomenon in narrow germanium *p–n* junctions, Phys. Rev. 109 (2) (1958). Pp. 603–604.

12. **Savchenko A. K., Kuznetsov V. V., Woolfe A., et al.,** Resonant tunneling through two impurities in disordered barriers, Phys. Rev. B. 52 (24) (1995) R17021–R17024.

13. **Larkin A. I., Matveev K. A.** Current-voltage characteristics of mesoscopic semiconductor contacts, Soviet Physics. JETP. 66 (3) (1987) 580–584.

14. **Jandieri K., Baranovskii S. D., Rubel O., et al.,** Resonant electron tunneling through defects in GaAs tunnel diodes, Appl. Phys. 104 (9) (2008) 094506.

15. **Prabhudesai G., Muruganathan M., Anh L.T., et al.,** Single-charge band-to-band tunneling via multiple-dopant clusters in nanoscale Si Esaki diodes, Appl. Phys. Lett. 114 (24) (2019) 243502.

16. **Prabhudesai G., Yamaguchi K., Tabe M., Moraru D.,** Coulomb-blockade charge-transport mechanism in band-to-band tunneling in heavily-doped low-dimensional silicon Esaki diodes, Proc. IEEE Silicon Nanoelectronics Workshop (SNW), Honolulu, Hawaii, USA, June 13–14 (2020) 109–110.

17. **Saha S., Biswas K., Hasan M.,** Temperature comparison of GaAs/AlGaAs based double barrier resonant tunneling diode considering NEGF, Proc. 4th Int. Conf. on Advances in Electrical Engineering (ICAEE). Bangladesh, Sept. 28–30 (2017) 44–47.

СПИСОК ЛИТЕРАТУРЫ

1. **Зайцев Д. Ф., Андреев В. М., Биленко И. А. и др.** Первая радиофотонная фазированная антенная решетка // Радиотехника. 2021. № 4. С. 153–164.

2. **Kalinovskii V. S., Terukov E. I., Kontrosh E. V., et al.** Energy-informational hybrid photovoltaic converter of laser radiation // St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 2023. Vol. 16. No. 1.2. Pp. 47–51.

3. **Kalinovskiy V. S., Kontrosh E. V., Gusev G. A., Sumarokov A. N., Klimko G. V., Ivanov S. V., Yuferev V. S., Tabarov T. S., Andreev V. M.** Study of PV characteristics of $Al_xGa_{1-x}As$/GaAs photodiodes // Journal of Physics: Conference Series. 2018. Vol. 993. P. 012029.

4. **Wang A.-Ch., Sun Y.-R., Yu Sh.-Zh., Yin J.-J., Zhang W., Wang J.-Sh., Fu Q.-X., Han Y.-H., Qin J., Dong J.-R.** Characteristics of 1520 nm InGaAs multijunction laser power converters // Applied Physics Letters. 2021. Vol. 119. No. 24. P. 243902.

5. **Hoheisel R., Bett A. W., Warner J. H., Walters R. J., Jenkins P.** Low temperature low intensity effects in III-V photovoltaic devices for deep space missions // Proceedings of the IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC). June 10–15, 2018. Pp. 3763–3767.

6. **Lumb M. P., González M., Yakes M. K., Affouda C. A., Bailey C. G., Walters R. J.** High temperature current–voltage characteristics of InP-based tunnel junctions // Progress in Photovoltaics. 2015. Vol. 23. No. 6. Pp. 773–782.

7. **Baudrit M., Algora C.** Tunnel diode modeling, including nonlocal trap-assisted tunneling: A focus on III–V multijunction solar cell simulation // IEEE Transactions on Electron Devices. 2010. Vol. 57. No. 10. Pp. 2564–2571.

8. **Tabe M., Tan H. N., Mizuno T., Muruganathan M., Anh L. T., Mizuta H., Nuryadi R., Moraru D.** Atomistic nature in band-to-band tunneling in two-dimensional silicon *pn* tunnel diods // Applied Physics Letters. 2016. Vol. 108. No. 9. P. 093502.

9. **Зи С.** Физика полупроводниковых приборов. В 2-х кн. Пер. с англ. 2-е, перераб. и доп. изд. М.: Мир, 1984. Кн. 1 – 456 с., Кн. 2 – 456 с.

10. **Калиновский В. С., Контрош Е. В., Климко Г. В., Иванов С. В., Юферев В. С., Бер Б. Я., Казанцев Д. Ю., Андреев В. М.** Разработка и исследование туннельных *p–i–n*-диодов GaAs/AlGaAs для многопереходных преобразователей мощного лазерного излучения // Физика и техника полупроводников. 2020. Т. 54. № 3. С. 285–291.

11. **Esaki L.** New phenomenon in narrow germanium *p–n* junctions // Physical Review. 1958. Vol. 109. No. 2. Pp. 603–604.

12. **Savchenko A. K., Kuznetsov V. V., Woolfe A., Mace D. R., Pepper M., Ritchie D. A., Jones G. A. C.** Resonant tunneling through two impurities in disordered barriers // Physical Review. B. 1995. Vol. 52. No. 24. Pp. R17021–R17024.

13. **Ларкин А. И., Матвеев К. А.** Вольт-амперная характеристика мезоскопических полупроводниковых контактов // Журнал экспериментальной и теоретической физики. 1987. Т. 93. № 3 (9). С. 1030–1038.

14. **Jandieri K., Baranovskii S. D., Rubel O., Stolz W., Gebhard F., Guter W., Hermle M., Bett A. W. J.** Resonant electron tunneling through defects in GaAs tunnel diodes // Applied Physics. 2008. Vol. 104. No. 9. P. 094506.

15. **Prabhudesai G., Muruganathan M., Anh L.T., Mizuta H., Hori M., Ono Y., Tabe M., Moraru D.** Single-charge band-to-band tunneling via multiple-dopant clusters in nanoscale Si Esaki diodes // Applied Physics Letters. 2019. Vol. 114. No. 24. P. 243502.

16. **Prabhudesai G., Yamaguchi K., Tabe M., Moraru D.** Coulomb-blockade charge-transport mechanism in band-to-band tunneling in heavily-doped low-dimensional silicon Esaki diodes // Proceedings of the IEEE Silicon Nanoelectronics Workshop (SNW). Honolulu, Hawaii, USA, June 13–14, 2020. Pp. 109–110.

17. **Saha S., Biswas K., Hasan M.** Temperature comparison of GaAs/AlGaAs based double barrier resonant tunneling diode considering NEGF // Proceedings of the 4th International Conference on Advances in Electrical Engineering (ICAEE). Bangladesh, September 28–30, 2017. Pp. 44–47.

THE AUTHORS

KONTROSH Evgeniy V.
Ioffe Institute of RAS
26, Polytekhnicheskaya St., St. Petersburg, 194021, Russia
kontrosh@mail.ioffe.ru
ORCID: 0000-0003-1812-3714

KALINOVSKII Vitaliy S.
Ioffe Institute of RAS
26, Polytekhnicheskaya St., St. Petersburg, 194021, Russia
vitak.sopt@mail.ioffe.ru
ORCID: 0000-0003-4858-7544

KLIMKO Grigory V.
Ioffe Institute of RAS
26, Polytekhnicheskaya St., St. Petersburg, 194021, Russia
gklimko@mail.ru
ORCID: 0000-0001-8893-7751

BER Boris Ya.
Ioffe Institute of RAS
26, Polytekhnicheskaya St., St. Petersburg, 194021, Russia
boris.ber@mail.ioffe.ru
ORCID: 0000-0003-2934-4176

PRUDCHENKO Kseniia K.
Ioffe Institute of RAS
26, Polytekhnicheskaya St., St. Petersburg, 194021, Russia
prudchenkokk@mail.ioffe.ru
ORCID: 0000-0003-4437-2984

TOLKACHEV Ivan A.
Ioffe Institute of RAS
26, Polytekhnicheskaya St., St. Petersburg, 194021, Russia
TolkachevIA@mail.ioffe.ru
ORCID: 0000-0001-8202-7087

KAZANTSEV Dmitry Yu.
Ioffe Institute of RAS
26, Polytekhnicheskaya St., St. Petersburg, 194021, Russia
Dukazantsev@mail.ioffe.ru
ORCID: 0000-0003-2173-1278

СВЕДЕНИЯ ОБ АВТОРАХ

КОНТРОШ Евгений Владимирович – *научный сотрудник Физико-технического института имени А. Ф. Иоффе Российской академии наук.*
194021, Россия, г. Санкт-Петербург, Политехническая ул., 26
kontrosh@mail.ioffe.ru
ORCID: 0000-0003-1812-3714

КАЛИНОВСКИЙ Виталий Станиславович – *старший научный сотрудник Физико-технического института имени А. Ф. Иоффе Российской академии наук.*
194021, Россия, г. Санкт-Петербург, Политехническая ул., 26
vitak.sopt@mail.ioffe.ru
ORCID: 0000-0003-4858-7544

КЛИМКО Григорий Викторович – *младший научный сотрудник Физико-технического института имени А. Ф. Иоффе Российской академии наук.*
194021, Россия, г. Санкт-Петербург, Политехническая ул., 26
gklimko@mail.ru
ORCID: 0000-0001-8893-7751

БЕР Борис Яковлевич – *старший научный сотрудник Физико-технического института имени А. Ф. Иоффе Российской академии наук.*
194021, Россия, г. Санкт-Петербург, Политехническая ул., 26
boris.ber@mail.ioffe.ru
ORCID: 0000-0003-2934-4176

ПРУДЧЕНКО Ксения Константиновна – *младший научный сотрудник Физико-технического института имени А. Ф. Иоффе Российской академии наук.*
194021, Россия, г. Санкт-Петербург, Политехническая ул., 26
prudchenkokk@mail.ioffe.ru
ORCID: 0000-0003-4437-2984

ТОЛКАЧЕВ Иван Андреевич – *младший научный сотрудник Физико-технического института имени А. Ф. Иоффе Российской академии наук.*
194021, Россия, г. Санкт-Петербург, Политехническая ул., 26
TolkachevIA@mail.ioffe.ru
ORCID: 0000-0001-8202-7087

КАЗАНЦЕВ Дмитрий Юрьевич – *старший научный сотрудник Физико-технического института имени А. Ф. Иоффе Российской академии наук.*
194021, Россия, г. Санкт-Петербург, Политехническая ул., 26
Dukazantsev@mail.ioffe.ru
ORCID: 0000-0003-2173-1278

Received 14.07.2023. Approved after reviewing 03.08.2023. Accepted 03.08.2023.
Статья поступила в редакцию 14.07.2023. Одобрена после рецензирования 03.08.2023. Принята 03.08.2023.

Original article
DOI: https://doi.org/10.18721/JPM.16404

PECULIARITIES OF STRUCTURE DAMAGE ACCUMULATION UNDER THE IMPLANTATION OF IONS OF DIFFERENT MASSES INTO ALPHA-GALLIUM OXIDE AT LOW DAMAGE LEVELS

A. I. Klevtsov[1,2]✉, *P. A. Karaseov*[1], *K.V. Karabeshkin*[1], *A. I. Titov*[1]

[1] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;

[2]Joint-Stock Company ""Research and Production Enterprise "ELAR"", St. Petersburg, Russia

✉ klevtsov_ai@spbstu.ru

Abstract. In the paper, the distributions of structure damage created in alpha-phase of gallium oxide by keV fluorine, phosphorus and xenon ion irradiation, have been obtained at room temperature. A noticeable effect of the average individual collision cascade density on the stable damage production efficiency at the surface was established. In contrast to many other semiconductors, an intermediate damage peak appeared in the alpha-Ga_2O_3 between the surface and bulk maxima. This intermediate peak visible in the RBS/C spectra at low damage levels was discovered for the first time. Characteristic peculiarities of the discovered maximum were investigated.

Keywords: gallium oxide, ion implantation, radiation defect, collision cascades, Rutherford backscattering

Funding: The reported study was funded by Russian Science Foundation (Grant No. 22-19-00166).

Citation: Klevtsov A. I., Karaseov P. A., Karabeshkin K.V., Titov A. I., Peculiarities of structure damage accumulation under the implantation of ions of different masses into alpha-gallium oxide at low damage levels, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 42–49. DOI: https://doi.org/10.18721/JPM.16404

Научная статья
УДК 539.1.043
DOI: https://doi.org/10.18721/JPM.16404

ОСОБЕННОСТИ НАКОПЛЕНИЯ СТРУКТУРНЫХ НАРУШЕНИЙ ПРИ ИМПЛАНТАЦИИ ИОНОВ РАЗНЫХ МАСС В АЛЬФА-ОКСИД ГАЛЛИЯ ПРИ МАЛЫХ УРОВНЯХ ПОВРЕЖДЕНИЯ

А. И. Клевцов[1,2]✉, *П. А. Карасев*[1], *К. В. Карабешкин*[1], *А. И. Титов*[1]

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия;

[2] Научно-производственное предприятие «ЭЛАР», Санкт-Петербург, Россия

✉ klevtsov_ai@spbstu.ru

Аннотация. В работе получены распределения структурных нарушений при облучении альфа-фазы оксида галлия ионами фтора, фосфора и ксенона с энергией, измеряемой килоэлектронвольтами (температура комнатная). Установлено заметное влияние усредненной плотности индивидуальных каскадов столкновений на эффективность введения стабильных нарушений для поверхностного пика радиационных дефектов. В отличие от случаев ионной имплантации во многие другие полупроводники, впервые обнаружено, что в альфа-Ga_2O_3 между поверхностным и объемным максимумами структурных нарушений возникает дополнительный пик. Этот промежуточный максимум ясно виден на спектрах резерфордовского обратного рассеяния при малых уровнях повреждения. Изучены характерные особенности впервые обнаруженного максимума.

Ключевые слова: оксид галлия, ионная имплантация, радиационный дефект, каскады столкновений, резерфордовское обратное рассеяние

Финансирование: Исследование выполнено при финансовой поддержке Российского научного фонда (грант № 22-19-00166).

Ссылка для цитирования: Клевцов А. И., Карасев П. А., Карабешкин К. В., Титов А. И. Особенности накопления структурных нарушений при имплантации ионов разных масс в альфа-оксид галлия при малых уровнях повреждения // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. 42–49. DOI: https://doi.org/10.18721/JPM.16404

Introduction

There are diverse modern technologies and tools for studying the properties of various materials; a crucial method is ion irradiation, allowing to modify the structure of matter. Implantation of ions into semiconductors is always accompanied by stable structural damage. Analysis of ion-stimulated processes, in particular the occurrence of structural damage in materials is mainly required for two applied problems. Firstly, radiation damage is the main limitation to ion beam machining technologies for manufacturing electronic devices. Secondly, it is often necessary to determine the durability of electronic devices operating under high radiation loads, finding ways to improve it. Studies into these problems have long been underway, however, radiation defects in binary or more sophisticated materials have a complex nature and remain poorly understood.

Damage accumulation in the crystal structure under ion bombardment is often studied by Rutherford backscattering by MeV helium ions in combination mode (RBS/C). The method was used to establish that the depth distributions of the accumulated damages have a bimodal nature for many semiconductors irradiated by light ions at least [2–5]. A bulk defect peak (BDP)

appears, typically located at the depth of the maximum elastic energy loss of stopping ions [3, 4], i.e., in the region where the majority of the primary point defects are generated. In addition, the crystal structure becomes disordered directly at the surface of the bombarded target. This surface defect peak (SDP) occurs due to diffusion of primary defects to the surface of the semiconductor and their subsequent coagulation on this surface (see, for example, [6]). In addition, sometimes another maximum is found in the depth distribution of defects, located between SDP and BDP; an example is the result obtained by irradiation of zinc oxide with heavy ions. Such a defect peak is commonly defined as intermediate (IDP) [7–9]. In the case of irradiation of zinc oxide with 500 keV xenon ions, IDP appears do to a highly Zn-rich layer [9].

One of the most promising semiconductor materials for high-power electronics and optoelectronics of a new generation is gallium oxide Ga_2O_3, offering such advantages as a wide band gap (4.5–5.3 eV for different phases), high breakdown voltage (about 8 MV/cm), etc. [10]. Some of the earliest data were obtained for the accumulation of structural damage in αGa_2O_3 [11, 12] and βGa_2O_3 [11, 13] for bombardment with accelerated atomic ions. The detected distribution of stable structural defects for both the stable βphase and the metastable αpolytype has a bimodal character. The ion dose required to achieve approximately the same level of disordering for αGa_2O_3 is about 10 times higher than for the stable βphase [11]. Further investigations established conditions for IDP to appear are observed under ion irradiation of alpha gallium oxide.

The goal of this study was to describe the detection of IDP in an αGa_2O_3semiconductor material under ion irradiation and to find out the conditions under which this peak appears.

Experimental procedure

We considered epitaxial layers of alpha gallium oxide (α-Ga_2O_3) with a corundum structure approximately 2 μm thick with the orientation (0001) grown on the *c*-plane of a sapphire substrate by hydride vapor-phase epitaxy (HVPE).

The samples were irradiated with fluorine, phosphorus and xenon ions at room temperature on a 500 kV HVEE implanter (Netherlands). The irradiation was carried out at an angle of 7° from the direction [0001] to minimize the channeling effects. The irradiation parameters were selected in such a way that the generation of primary defects by inhibited ions was approximately the same over the depth of the target in all cases. For this purpose, the ion energies and fluxes were chosen such that the depth distribution profiles of displaced atom concentrations were similar and differed only in the height of the peak under irradiation with different ions.

Displacement generation profiles were calculated in the binary collision approximation (BCA) [14]. Ion doses were expressed as displacements per atom (dpa), calculated at the maximum depth of the generation function. The dpa magnitude was calculated by the TRIM code (version SRIM-2013, http://www.srim.org) [14]. The ion flux during irradiation with various ions was maintained the same in dpa/s units. The doses were selected such that the damage levels near the SDP were not too high and close to each other so that we could conveniently compare the distribution profiles of structural defects during irradiation with different types of ions. In addition, irradiation with different ions (phosphorus) and with a higher energy (65 keV) was carried out at the same doses and fluxes (in dpa and dpa/s units) that were used earlier for more detailed comparative analysis of the detected effect.

The disordering degree of the crystal structure after irradiation was measured by RBS/C. A 0.7 MeV He^{++} beam probe in the direction [0001] was used for the measurement. The scattered particle detector was positioned at an angle of 103° relative to the direction of the incident beam. The obtained RBS/C spectra were processed using a standard algorithm to construct the distributions of relative depth disordering [15].

Experimental results and discussion

Fig. 1,*a* shows the distributions of structural defects along the depth of the α-Ga_2O_3 target obtained after implantation of ions with different masses. As already noted, the radiation doses were chosen in such a way that the resulting damage levels near the SDP were not too high and sufficiently close to each other. Clearly, the doses that meet these requirements are the lower the greater the ion mass. Indeed, a dose of 0.30 dpa is required to achieve a disordering of the order of 0.15 in SDP by irradiation with Xe ions, a dose of 0.44 dpa is required in the case of irradiation with P ions, while in the case of irradiation with F ions it is as high as 1.50 dpa. Recall that the the

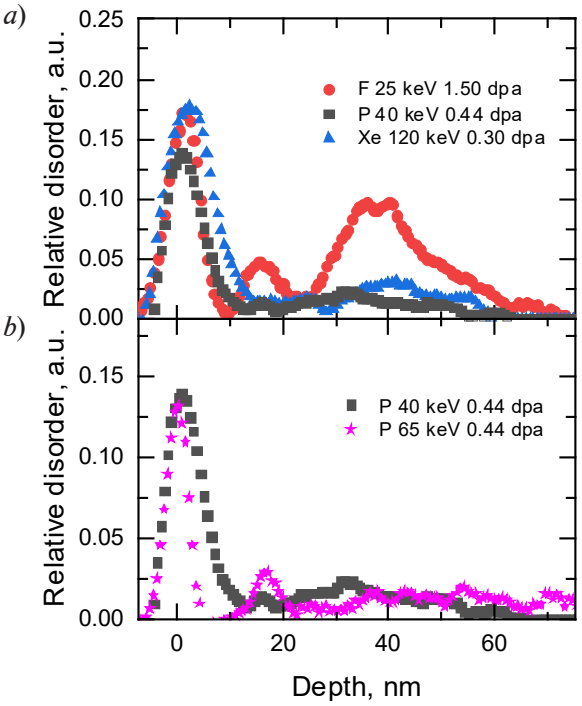

Fig. 1. Distributions of relative disordering concentration over the depth of the α-Ga_2O_3 target after irradiation with ions of different masses with different energies and doses (given in the caption):
fluorine, phosphorus and xenon ions (*a*); only phosphorus ions with two different energies and the same dose (*b*) Ion fluxes (10^{-3}dpa/s): 2.41 (for fluorine ions) and 0.08 (for other ions)

conditions for ion bombardment were chosen such that the generation rates of primary defects in the BCA [14] and the depth distributions of defects coincide for ions of different types.

It follows from these results that the level of damage to the α-Ga_2O_3 crystal structure can be greatly affected by another parameter that varies from ion to ion in such an experimental setting; this parameter is the average density of individual displacement cascades. Previous experimental studies found that irradiation with molecular ions is more efficient for producing the SDP; this was indicated by the results obtained for irradiation of gallium oxide with higher ion doses [12]. We earlier proposed to calculate the value of this parameter based on the binary collision approximation [16].

Fig. 2 shows the calculated density dependences of the cascades generated by ions along the depth of the alpha gallium oxide target. Evidently, the density of cascades generated by fluorine ions is less than that generated by phosphorus and xenon ions. Thus, an increase in the density of displacement cascades leads to an increase in the SDP production efficiency even in this case, at rather small ion doses and fluxes.

Notably, another pronounced peak appears between SDP and BDP on the curve shown in Fig. 1,*a* (distribution of defect concentration during bombardment with 25 keV F ions). This intermediate peak is located at a depth about 17 nm in the α-Ga_2O_3 target. In our opinion, a similar peak is present in the distribution obtained

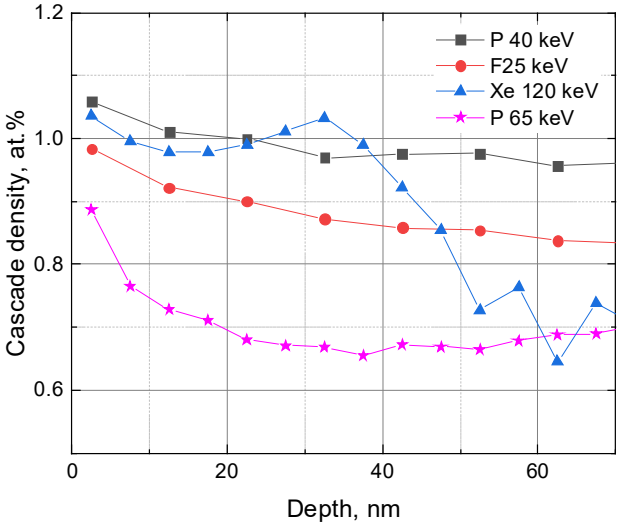

Fig. 2. Density distributions of individual displacement cascades over the depth of the α-Ga_2O_3 target after irradiation with ions of different masses with different energies (given in the caption).
The calculations were performed based on the binary collision approximation [14] by the technique proposed in [16]

by irradiation with 40 keV phosphorus ions. This peak is not detected in the case of bombardment of gallium alpha oxide with heavier xenon ions. We should note that the appearance of the IDP is a new phenomenon that we have not observed for implantation of ions to higher doses.

To further investigate this phenomenon, irradiation of phosphorus ions with higher energy (65 keV) and the same dose (0.44 dpa) was carried out. The resulting distribution of structural defects is shown in Fig. 1,*b*. Evidently, the width of the surface peak becomes slightly smaller with an increase in the energy of phosphorus ions. Furthermore, a rather pronounced intermediate peak is observed in gallium alpha oxide in this case, approximately at the same depth as the one previously observed for fluorine ions (see Fig. 1, *a* and *b*). Thus, the IDP is produced not only by irradiation with fluorine ions, but also with phosphorus ions. The IDP becomes more pronounced with an increase in the energy of phosphorus ions.

As noted above, the IDP was previously detected during implantation of ions into zinc oxide. However, the mechanism by which this peak is produced in this case (α-Ga_2O_3) is likely different than for ZnO. Indeed, the IDP in zinc oxide appears under irradiation with heavy ions and is noticeable in a wide range of doses. The magnitude of the IDP in zinc oxide did not depend on the type of the ion. In the case of α-Ga_2O_3 oxide, we observe the IDP only under irradiation with light ions and at low doses. The magnitude of IDP in gallium oxide under the same conditions of ion bombardment is different for phosphorus and fluorine ions, increasing with increasing energy for phosphorus ions.

Notable differences are observed in the behavior of density distribution curves of individual displacement cascades for all experimental cases (see Fig. 2). It can be seen that the highest value of the parameter is detected for stopping of heavy xenon ions near the surface. Phosphorus ions with an energy of 40 keV produce cascades with a lower density, and the densities are even lower for fluorine ions. The density of the displacement cascades decreases with an increase in the energy of phosphorus ions from 40 to 65 keV. Thus, the IDP in the α-Ga_2O_3 target appears at a low cascade density rather than at a high one as is the case with ZnO.

The reasons why the IDP appears in the spectrum are not yet clear. Further in-depth studies covering all aspects of the phenomenon are necessary for understanding the behavior and mechanism of IDP production in α-Ga_2O_3.

Conclusion

We experimentally obtained the depth distributions of structural damage under implantation of small doses of fluorine, phosphorus and xenon ions into the α-Ga_2O_3 semiconductor material in the keV energy range. The average density of individual displacement cascades was calculated; the calculation results, along with experimental data, indicate an increase in the efficiency of radiation damage to gallium oxide with an increase in such density.

We found that ion bombardment of a gallium oxide target (unlike other target materials) causes an additional structural defect peak to appear for the ions with the average masses and selected technological doses; this peak is located between the surface and bulk peaks on the corresponding distribution curves.

REFERENCES

1. **Feldman L. C., Mayer J. W.,** Fundamentals of surface and thin films analysis, Pearson College Div., London, 1986.

2. **Gerasimov A. I., Zorin E. I., Pavlov P. V., Tetelbaum D. I.,** On the peculiarities of silicon amorphization at ion bombardment, Phys. Status Solidi. 12 (2) (1972) 679–685.

3. **Kucheyev S. O., Williams J. S., Pearton S. J.,** Ion implantation into GaN, Mat. Sci. Eng. R Rep. 33 (2—3) (2001) 51—107.

4. **Azarov A. Yu., Titov A. I., Karaseov P. A., Hallén A.,** Effect of collision cascade density on radiation damage in SiC, Nucl. Instrum. Methods Phys. Res. B. 267 (8—9) (2009) 1247—1250.

5. **Kucheyev S. O., Williams J. S., Jagadish C., et al.** Ion-beam-produced structural defects in ZnO, Phys. Rev. B. 67 (9) (2003) 094115.

6. **Titov A. I., Belyakov V. S., Azarov A. Yu.,** Formation of surface amorphous layers in semiconductors under low-energy light-ion irradiation: Experiment and theory, Nucl. Instrum.Meth. B. 212 (Dec) (2003) 169–178.

7. **Azarov A. Yu., Kucheyev S. O., Titov A. I., Karaseov P. A.,** Effect of density of collision cascades on ion implantation damage in ZnO, J. Appl. Phys. 102 (8) (2007) 083547.

8. **Azarov A. Yu., Titov A. I., Karaseov P. A., et al.,** Structural damage in ZnO bombarded by heavy ions, Vacuum. 84 (8) (2010) 1058–1061.

9. **Myers M. T., Charnvanichborikarn S., Wei C. C., et al.,** Defect microstructure in heavy-ion-bombarded (0001) ZnO, Acta Mater. 60 (17) (2012) 6086–6090.

10. **Pearton S. J., Ren F., Mastro M.** (Eds.), Gallium oxide. Technology, devices and applications, Elsevier Inc., Amsterdam, 2019.

11. **Titov A. I., Karabeshkin K. V., Struchkov A. I. et al.,** Comparative study of radiation tolerance of GaN and Ga_2O_3 polymorphs, Vacuum. 200 (June) (2022) 111005.

12. Karaseov P. A., Karabeshkin K.V., Struchkov A. I. et al., Radiation damage accumulation in α-Ga_2O_3 under P and PF_4 ion bombardment, Semiconductors. 56 (9) (2022) 664–669.

13. **Azarov A., Venkatachalapathy V., Karaseov P., et al.,** Interplay of the disorder and strain in gallium oxide, Sci. Rep. 12 (1; 13 Sept.) (2022) 15366.

14. **Ziegler J. F., Biersack J. P., Littmark U.,** The stopping and range of ions in solids, Pergamon Press, NY, 1985.

15. **Schmid K.,** Some new aspects for the evaluation of disorder profiles in silicon by backscattering, Radiat. Eff. 17 (3–4) (1973) 201–207.

16. **Kucheyev S. O., Azarov A. Yu., Titov A. I., et al.,** Energy spike effects in ion-bombarded GaN, J. Phys. D: Appl. Phys. 42 (8) (2009) 085309.

СПИСОК ЛИТЕРАТУРЫ

1. **Фелдман Л., Майер Д.** Основы анализа поверхности и тонких пленок. М.: Мир, 1989. 344 с.

2. **Gerasimov A. I., Zorin E. I., Pavlov P. V., Tetelbaum D. I.** On the peculiarities of silicon amorphization at ion bombardment // Physica Status Solidi. 1972. Vol. 12. No. 2. Pp. 679–685.

3. **Kucheyev S. O., Williams J. S., Pearton S. J.** Ion implantation into GaN // Materials Science and Engineering: R: Reports. 2001. Vol. 33. No. 2–3. Pp. 51–107.

4. **Azarov A. Yu., Titov A. I., Karaseov P. A., Hallén A.** Effect of collision cascade density on radiation damage in SiC // Nuclear Instruments and Methods in Physics Research B. 2009. Vol. 267. No. 8–9. Pp. 1247–1250.

5. **Kucheyev S. O., Williams J. S., Jagadish C., Zou J., Evans C., Nelson A. J., Hamza A. V.** Ion-beam-produced structural defects in ZnO // Physical Review B. 2003. Vol. 67. No. 9. P. 094115.

6. **Titov A. I., Belyakov V. S., Azarov A. Yu.** Formation of surface amorphous layers in semiconductors under low-energy light-ion irradiation: Experiment and theory // Nuclear Instruments and Methods B. 2003. Vol. 212. December. Pp. 169–178.

7. **Azarov A. Yu., Kucheyev S. O., Titov A. I., Karaseov P. A.** Effect of density of collision cascades on ion implantation damage in ZnO // Journal of Applied Physics. 2007. Vol. 102. No. 8. P. 083547.

8. **Azarov A. Yu., Titov A. I., Karaseov P. A., Kucheyev S. O., Hallén A., Kuznetsov A. Yu., Svensson B. G., Pathak A. P.** Structural damage in ZnO bombarded by heavy ions // Vacuum. 2010. Vol. 84. No. 8. Pp. 1058–1061.

9. **Myers M. T., Charnvanichborikarn S., Wei C. C., Luo Z. P., Aitkaliyeva A., Shao L., Kucheyev S. O.** Defect microstructure in heavy-ion-bombarded (0001) ZnO // Acta Materialia. 2012. Vol. 60. No. 17. Pp. 6086–6090.

10. **Pearton S. J., Ren F., Mastro M.** (Eds.). Gallium oxide. Technology, devices and applications. Amsterdam: Elsevier Inc., 2019. 479 p.

11. **Titov A. I., Karabeshkin K. V., Struchkov A. I. Nikolaev V. I., Azarov A. Yu., Gogova D. S., Karaseov P. A.** Comparative study of radiation tolerance of GaN and Ga_2O_3 polymorphs // Vacuum. 2022. Vol. 200. June. P. 111005.

12. **Карасев П. А., Карабешкин К. В., Стручков А. И., Печников А. И., Николаев В. И., Андреева В. Д., Титов А. И.** Накопление структурных нарушений при облучении α-Ga_2O_3 ионами P и PF_4 // Физика и техника полупроводников. 2022. Т. 56. № 9. С. 882–887.

13. **Azarov A., Venkatachalapathy V., Karaseov P., Titov A., Karabeshkin K., Struchkov A., Kuznetsov A.** Interplay of the disorder and strain in gallium oxide // Scientific Reports. 2022. Vol. 12. No. 1. 13 September. P. 15366.

14. **Ziegler J. F., Biersack J. P., Littmark U.** The stopping and range of ions in solids. New York: Pergamon Press, 1985. 321 p.

15. **Schmid K.** Some new aspects for the evaluation of disorder profiles in silicon by backscattering // Radiation Effects. 1973. Vol. 17. No. 3–4. Pp. 201–207.

16. **Kucheyev S. O., Azarov A. Yu., Titov A. I., Karaseov P. A., Kuchumova T. M.** Energy spike effects in ion-bombarded GaN // Journal of Physics D: Applied Physics. 2009. Vol. 42. No. 8. P. 085309.

THE AUTHORS

KLEVTSOV Anton I.
Peter the Great St. Petersburg Polytechnic University,
Joint-Stock Company "Research and Production Enterprise «ELAR»"
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
klevtsov_ai@spbstu.ru
ORCID: 0009-0004-6988-9685

KARASEOV Platon A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
platon.karaseov@spbstu.ru
ORCID: 0000-0003-2511-0188

KARABESHKIN Konstantin V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
yanikolaus@yandex.ru
ORCID: 0000-0003-1770-1877

TITOV Andrei I.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
andrei.titov@rphf.spbstu.ru
ORCID: 0000-0003-4933-9534

СВЕДЕНИЯ ОБ АВТОРАХ

КЛЕВЦОВ Антон Игоревич – *аспирант Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого, инженер Акционерного общества* "Научно-производственное предприятие «ЭЛАР»".
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
klevtsov_ai@spbstu.ru
ORCID: 0009-0004-6988-9685

КАРАСЕВ Платон Александрович – *доктор физико-математических наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
platon.karaseov@spbstu.ru
ORCID: 0000-0003-2511-0188

КАРАБЕШКИН Константин Валерьевич – *кандидат физико-математических наук, старший научный сотрудник Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
yanikolaus@yandex.ru
ORCID: 0000-0003-1770-1877

ТИТОВ Андрей Иванович – *доктор физико-математических наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
andrei.titov@rphf.spbstu.ru
ORCID: 0000-0003-4933-9534

Received 27.10.2023. Approved after reviewing 12.11.2023. Accepted 12.11.2023.
Статья поступила в редакцию 27.10.2023. Одобрена после рецензирования 12.11.2023. Принята 12.11.2023.

SIMULATION OF PHYSICAL PROCESSES

Original article
DOI: https://doi.org/10.18721/JPM.16405

A COMPARISON OF TWO APPROACHES TO THE GLOBAL STABILITY ANALYSIS USING THE EXAMPLE OF THE CYLINDER FLOW PROBLEM

V. D. Golubkov✉, *A. V. Garbaruk*

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ golubkovvd@gmail.com

Abstract. In the paper, the two main approaches to calculating the Jacobian of the Navier–Stokes equations, namely, the continuum (CA) and discrete (DA) approaches, have been directly compared for the first time. The DA to calculating this Jacobian was implemented based on in-house finite-volume code for hydrodynamics simulation (in addition to the already existing CA). The DA was successfully verified by comparison between the obtained numerical result and that of solving the transient Navier–Stokes equations. The comparison of these approaches was carried out using the example of a laminar flow past a cylinder by a perfect gas at the near-critical Reynolds numbers (Re = 50 and 60). It was established that the CA predicted the growth rate of perturbations more accurately, while the DA did their frequency and amplitude in toto. The results obtained allow to assert that both CA and DA are equivalent in terms of accuracy, and the choice of a particular approach for analyzing the stability may determine by other criteria, e. g., ease of implementation, computational work and so on.

Keywords: global stability analysis, Navier–Stokes equations, Jacobian, automatic differentiation

Funding: The reported study was funded by Russian Science Foundation (Grant No. 22-11-00041).

Citation: Golubkov V. D., Garbaruk A. V., A comparison of two approaches to the global stability analysis using the example of the cylinder flow problem, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 50–62. DOI: https://doi.org/10.18721/JPM.16405

Научная статья
УДК 532.5.013.4
DOI: https://doi.org/10.18721/JPM.16405

СРАВНЕНИЕ ДВУХ ПОДХОДОВ К ГЛОБАЛЬНОМУ АНАЛИЗУ ГИДРОДИНАМИЧЕСКОЙ УСТОЙЧИВОСТИ НА ПРИМЕРЕ ЗАДАЧИ ОБТЕКАНИЯ ЦИЛИНДРА

В. Д. Голубков✉, *А. В. Гарбарук*

Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Россия

✉ golubkovvd@gmail.com

Аннотация. В работе впервые проведено прямое сравнение двух основных подходов к вычислению якобиана уравнений Навье – Стокса: континуального (КП) и дискретного (ДП). На базе собственного конечно-объемного кода для моделирования течений реализован ДП к вычислению якобиана (в дополнение к уже существующему КП). ДП был успешно верифицирован путем сравнения полученного численного результата с решением нестационарных уравнений Навье – Стокса. Сравнение двух подходов проведено на примере ламинарного обтекания цилиндра идеальным газом при околокритических числах Рейнольдса (Re = 50 и 60). Установлено, что КП точнее предсказывает показатель роста возмущений, а ДП – их частоту и амплитуду в целом. Полученные результаты позволяют утверждать, что КП и ДП равнозначны по порядку точности и выбор конкретного подхода для проведения анализа устойчивости может определяться другими критериями (например, простота реализации, вычислительные затраты и др.).

Ключевые слова: глобальный анализ устойчивости, якобиан уравнений Навье – Стокса, автоматическое дифференцирование

Финансирование: Работа выполнена при финансовой поддержке Российского научного фонда (грант № 00041-11-22).

Ссылка для цитирования: Голубков В. Д., Гарбарук А. В. Сравнение двух подходов к глобальному анализу гидродинамической устойчивости на примере задачи обтекания цилиндра // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. 50–62. DOI: https://doi.org/10.18721/JPM.16405

Introduction

One of the most powerful and advanced tools for studying the stability of viscous fluid flows is the linear theory of stability, which considers the development of small perturbations that do not interact with each other. Most 20th century studies were based on the linear theory of hydrodynamic stability within the framework of the locally parallel approach (the Orr–Sommerfeld equation) or two-dimensional parabolized equations (see books [1, 2] and a review [3]). By the end of the 20th century, the advances in computer technologies made it possible to conduct linear stability analysis of two-dimensional and even three-dimensional solutions of the Navier–Stokes equations; this approach came to be known as global stability analysis (GSA) in the literature [4].

The dynamics of the evolution of small perturbations within the GSA is determined by the matrix of derivatives of the governing equations with respect to all variables, i.e. the Jacobian of the stationary Navier–Stokes equations (more precisely, its discrete form). Currently, two different approaches are used to calculate this Jacobian. For example, [5–9] covering a wide range of problems of GSA for two-, three- and quasi-three-dimensional flows used the approach called continuum in [10]. It consists in the initial linearization of the Navier–Stokes equations, which leads to an analytical expression for their Jacobian, for which a discrete approximation is

then formed using one or another finite-difference scheme. In contrast to this method, [11–18] used an approach called discrete, in which the governing equations are initially discretized and then linearized.

The Jacobian matrices obtained using these approaches differ, since in general the linearization and discretization operations are noncommutative [10]. However, as the mesh is refined, the difference between the results of these approaches should decrease. Different aspects of the continuum and discrete approaches have been studied in the context of solving conjugate equations for optimization problems [19, 20]. However, these approaches were not compared within the framework of GSA in the literature and the choice of a specific approach in [5– 18] was not substantiated.

The goal of this study consists in comparing the results of GSA using various methods for calculating the Jacobian matrix using the example of laminar flow around a cylinder with perfect gas at near-critical Reynolds numbers.

Global stability analysis of steady laminar flows

The procedure for studying the global stability of laminar flows contains two main stages.

The first one is finding a numerical solution of a generalized system of steady Navier–Stokes equations, including equations of continuity, conservation of motion and energy, which can be written in operator form:

$$R(q) = 0, \tag{1}$$

where $q = \{\rho, \rho u, \rho v, \rho E\}^T$ is the vector of conservative variables; R is the nonlinear differential operator of steady Navier–Stokes equations.

The solution of the steady Navier–Stokes equations satisfying Eq. (1) and obtained by analyzing the stability of the flow is often called the basic one. The stability of this solution, denoted as $\bar{q}$, is in fact the subject of our analysis.

At the second stage, the evolution of perturbations of the basic solution over time is considered. The equation for perturbations can be obtained from the transient Navier–Stokes equations; they are written in the following operator form:

$$\frac{\partial q}{\partial t} = -R(q). \tag{2}$$

GSA uses the traditional approach for linear stability analysis, which is based on representation of the solution of the system of equations (2) as the sum of its steady solution $\bar{q}$ and small perturbations q':

$$q = \bar{q} + q'. \tag{3}$$

To obtain equations that are linear with respect to q', linearization of the operator $R(q)$ is carried out in the vicinity of the basic solution for these perturbations:

$$R(\bar{q} + q') = R(\bar{q}) + \frac{\partial R}{\partial q}(\bar{q})q', \tag{4}$$

where $\frac{\partial R}{\partial q}(\bar{q}) \equiv J(\bar{q})$ is the Jacobian of the Navier–Stokes equations (a differential operator depending on the basic solution).

The equation of relatively small perturbations is obtained by substituting expansion (3) into Eq. (2), taking into account Eqs. (1) and (4):

$$\frac{\partial q'}{\partial t} + J(\bar{q})q' = 0. \tag{5}$$

Due to linearity of the system of differential equations (5), its general solution is represented as a sum of terms (modes of perturbations), each of which is also a solution of system (5). Each mode can be represented as

$$q'(x,y,t) = \hat{q}(x,y)\exp(\omega t), \tag{6}$$

where $\hat{q}$ is the complex vector of the perturbation amplitudes; ω is the complex number $\omega_r + i\omega_i$, whose real part ω_r is the rate of growth/attenuation of the perturbation, and the imaginary part ω_i is its frequency (only the real part of relation (6) has a physical meaning).

Substituting equality (6) into system (5) leads to the eigenvalue problem for the Jacobian of the governing equations:

$$J\hat{q} = \omega\hat{q}. \tag{7}$$

The numerical solution of this problem is carried out on a finite difference mesh, so all continuous vectors and operators are replaced by their discrete approximations. Discretization of derivatives at each point of the computational mesh in accordance with an existing stencil of a numerical scheme determines the dependence of these derivatives on the values of variables at adjacent points. Thus, problem (7) is reduced to the eigenvalue problem of the discrete approximation of the Jacobian J, that is, the matrix M_{kl}:

$$M_{kl}\hat{\alpha}_l = \omega\hat{\alpha}_k. \tag{8}$$

Here, the vector $\hat{\alpha}_l$ is the discretized field of the amplitude of perturbations $\hat{q}$, and the discretized Jacobian M_{kl} is the matrix of derivative equations with respect to all variables at all points of the computational mesh, therefore, the indices k and l in Eq. (8) take values from 1 to $N_p \times N_v$, where N_p is the number of nodes of the computational mesh, N_v is the number of variables.

It should be noted that instead of linearization of expression (7) at the boundary points of the computational domain, linearization of the corresponding boundary conditions is used, therefore, the following equation is used for these points, instead of expression (8):

$$M_{kl}\hat{\alpha}_l = 0. \tag{9}$$

Eqs. (8), (9) can be combined if we formulate a generalized eigenvalue problem:

$$M_{kl}\hat{\alpha}_l = \omega T_{km}\hat{\alpha}_m, \tag{10}$$

where T_{km} is a diagonal matrix with $T_{ii} = 0$ at the boundary points and $T_{ii} = 1$ at the inner points.

Thus, the determination of the stability of the flow within the framework of GSA is reduced to solving the generalized eigenvalue problem (10). The eigenvalues of the matrix M_{kl} correspond to different modes of perturbation, and the real part of the eigenvalues is equal to the rate of growth of perturbations, and the imaginary part is the frequency of their vibrations.

The eigenvectors correspond to the spatial distributions of the mode amplitudes. The flow is unsteady if at least one eigenvalue has a positive real part (i.e., there is a growing perturbation mode), and stable otherwise.

As already noted in the introduction, two different approaches are currently used to determine the elements of the M_{kl} matrix at the inner points of the computational domain. According to the method for calculating this matrix, the GSA is called continuum or discrete, respectively.

Within the framework of the first of approach (see, for example, [5]), called continuum in [10], an analytical expression is derived for the Jacobian J, and then its discretization is carried out using some numerical scheme, which, generally speaking, may differ from that used to solve system of equations (1) obtaining the basic solution.

In contrast, within the second approach (see, for example, [11, 12]), called discrete in [10], the calculation of the Jacobian in problem (7) is carried out not at the differential, but at the discrete level, i.e., it is not the operator R itself that is differentiated, but its discrete form, used to obtain

a basic solution, called the right-hand side of system (2) (traditionally denoted as RHS_k); index k takes values from 1 to $N_p \times N_v$, as in Eq. (10).

The discrete form of the Jacobian in this case is the matrix of partial derivatives RHS_k with respect to the variables α_l (the discrete form of the vector of the principal variables q) at each inner point of the computational mesh:

$$M_{kl} = \frac{\partial \mathrm{RHS}_k}{\partial \alpha_l}. \tag{11}$$

There are two approaches to differentiating Eq. (11). Within the framework of the first, the explicit dependence $\mathrm{RHS}_k(\alpha_l)$ is formulated for the numerical scheme used, and then differentiated analytically. Even though this problem is very time-consuming, especially for modern schemes with high-order accuracy, it was solved in [21], and the developed approach was successfully applied in [11, 13, 16, 18].

This paper uses an alternative approach based on the technology of automatic differentiation (AD). Even though AD as a concept has appeared quite long ago [22], interest in it arose only in the last two decades, with efforts to solve related problems on optimizing the shapes of airfoils [23].

The basis for the AD is that the algorithm for calculating any complex function (including RHS) consists of sequential application of elementary operations φ_1 (addition, multiplication, exponentiation, etc.):

$$\mathrm{RHS} = \varphi_1 \circ \varphi_2 \circ \ldots \circ \varphi_n. \tag{12}$$

The values of the derivative of the elementary function at each step are known analytically, so the Jacobian of the RHS function can be calculated by the rule of differentiation of a complex function:

$$J = \varphi_1' \circ \varphi_2' \circ \ldots \circ \varphi_n'. \tag{13}$$

Libraries implementing AD (see, for example, [24, 25]) accumulate the results of this differentiation during the calculation of the initial function and calculate the discretized Jacobian. Notably, the AD method is not automatic in the full sense of the word and requires editing the source code of the program.

In the absence of gas-dynamic discontinuities, theoretically (i.e., with computational meshes that provide grid-independent solutions for emerging perturbations), the continuum and discrete approaches should provide the same result. However, in practice, the results obtained using different approaches on finite grids may differ dramatically.

It should be noted that the evolution of perturbations can be considered not only within the GSA, but also within direct numerical simulation of transient Navier–Stokes equations (2). In this case, the solution of the steady Navier–Stokes equations (1) is used as an initial approximation. The initial perturbations are determined by the error of the numerical solution of transient equations. If the flow is unsteady, then an increase in the amplitude of perturbations is observed as a result of the calculation. At the linear stage, when the exponential nature of the growth of perturbations is observed, their evolution should be consistent with the results of the GSA in the discrete calculation of the Jacobian.

In this paper, we verified our implementation of the discrete approach to calculating the Jacobian based on such a comparison.

Statement of the problem on the stability of steady flow around a cylinder and its computational aspects

The results of two stability analysis methods were compared using the example of the problem on laminar flow around a cylinder with perfect gas, using meshes that are sequentially refined in both directions. The problem was considered in a compressible statement with the Mach number $M = 0.2$ and two values of the Reynolds number, $Re = 50$ and 60, slightly exceeding the Reynolds

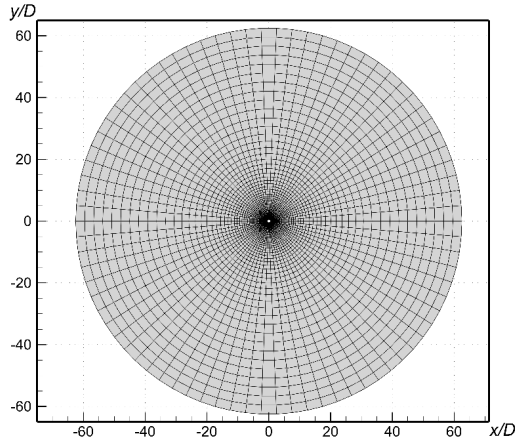

Fig. 1. Example of *O*-type computational mesh (*L*1 mesh)

number for the stability threshold, Re ≈ 47 (see, for example, [26]), when the Reynolds number is constructed from the velocity of the incident flow U_0 and the diameter of the cylinder D.

The size of the computational domain was $120D$. This size was sufficient to eliminate the influence of boundary conditions on the basic solution and the results of the GSA. A series of *O*-type computational meshes was constructed in this region (an example of such a mesh is shown in Fig. 1) with a uniform distribution of nodes along the angular coordinate and clustering towards the wall along the radial coordinate (the parameters of the constructed meshes are shown in Table 1).

In this paper, the finite volume Numerical Turbulence Simulation (NTS) CFD code was used for calculations [27]. In this code, the finite-difference relaxation method is used to find steady solutions to the governing equations. A hybrid scheme is used to approximate inviscid flows in calculations of compressible flows:

$$\Delta_H = \alpha_U \Delta_{Roe} + (1 - \alpha_U)\Delta_{4C}, \tag{14}$$

where α_U is the weight of upwind approximation; Δ_{Roe}, Δ_{4C} are the finite difference operators of the third-order upwind-biased Roe scheme and the fourth-order central difference scheme, respectively.

Table 1

Parameters of O-type computational meshes used and their values

Mesh	N_φ	N_r	$\Delta h_1/D$	$\Delta h_{i+1}/\Delta h_i$	$\Delta h_{\max}/D$
*L*1	80		$1.0\cdot10^{-2}$	1.098	2
*L*2	160		$5.0\cdot10^{-3}$	1.040	2
*L*3	240		$2.5\cdot10^{-3}$	1.028	2
*L*4	320		$1.0\cdot10^{-3}$	1.023	2
*L*5	800		$1.0\cdot10^{-4}$	1.011	1

Notations: N_φ, N_r is the number of nodes in the circumferential and radial directions, respectively, Δh_i is the grid pitch, $\Delta h_{\max}$ is its maximum value, D is the diameter of the cylinder.

The viscous components of the flows are approximated using a second-order central difference scheme.

To calculate the evolution of small perturbations by solving transient Navier–Stokes equations, numerical time integration was carried out using an implicit second-order Euler scheme with a time step $\Delta t = 0.3\cdot D/U_0$, which provided values of the Courant number less than unity in almost the entire computational domain and approximately 1,000 steps per Kármán vortex street for all meshes.

The indicators of the growth or attenuation of perturbations and their frequency were determined by processing the dependences of the transverse velocity on time obtained by unsteady calculations at several points in space. A linear stage of perturbation evolution was identified, when their amplitude increases exponentially.

Solving the spectral problem, the calculation of the discrete form of the Jacobian was carried out by both methods (discrete and continuum). Within the framework of the continuum method

implemented earlier in the NTS code, a finite-difference scheme was used to discretize the Jacobian J (it is described in more detail in [5]), which is a combinations of a third-order upwind scheme and a fourth-order central difference scheme:

$$\Delta_H = \alpha_U \Delta_{3U} + (1 - \alpha_U)\Delta_{4C}, \tag{15}$$

where α_U is the weight of the upwind approximation; Δ_{3U}, Δ_{4C} are the finite difference operators of the upwind scheme and the central difference scheme, respectively.

To use the discrete approach, we implemented in this paper, we applied the automatic differentiation method (using the ADF95 library [25]). For the numerical solution of the eigenvalue problem, the Krylov–Schur method was used, which is implemented using the open library SPEPc/PETSc [28]. This method is designed to solve eigenvalue problems with sparse non-Hermitian matrices of large size (this is the type of matrix considered). It is a modification of an implicitly restarted version of the Arnoldi method, which belongs to the class of Rayleigh–Ritz methods based on projection onto the Krylov subspace (see, for example, monograph [29]). The Krylov–Schur method allows to obtain the requested number of the eigenvalues the largest in absolute value and their corresponding eigenvectors. Therefore, to use it, the initial matrix is pre-transformed in such a way that the most important eigenvalues in terms of stability with the largest real part become the largest in absolute value. This transformation is a combination of shifting and inverting the matrix (this approach is called the "Shift-Invert Approach" in the literature [30]).

Verification of the GSA results obtained with the discrete approach to calculating the Jacobian

Fig. 2 shows the spatial distributions of perturbations of the longitudinal velocity U' at the Reynolds number Re = 60 on the *L4* mesh, obtained by discrete GSA and direct numerical solution of the transient Navier–Stokes equations. For the latter, the local amplitudes of perturbations are obtained as a result of subtracting the fields of the instantaneous and basic solution with normalization to the maximum value $|U'_{\max}|$.

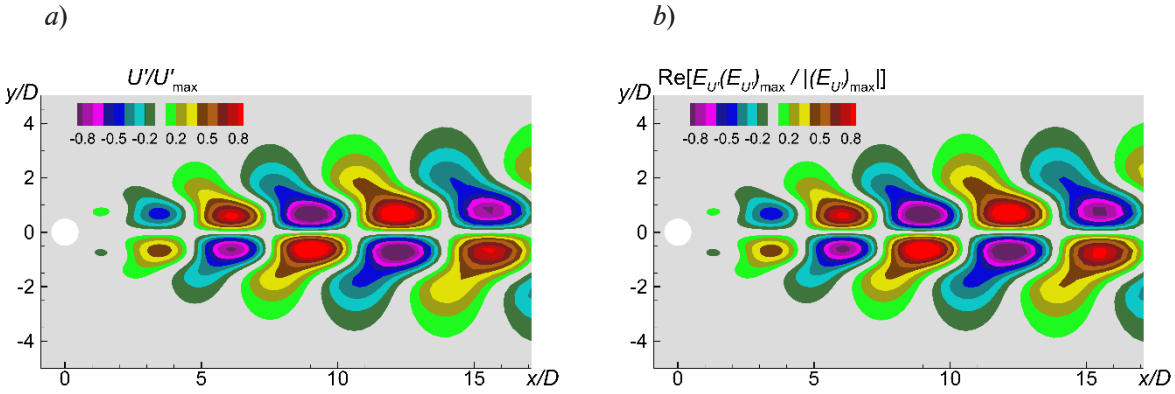

Fig. 2. Spatial distributions of longitudinal velocity perturbations obtained on the *L4* mesh by direct numerical solution of the transient Navier–Stokes equations (*a*) and using discrete GSA (*b*)
Reynolds number Re = 60, Mach number M = 0.2

In the framework of a discrete GSA, the spatial distribution of perturbations is determined by the real component of the eigenvector corresponding to the most unstable eigenvalue. For comparison, the complex components of the E_U vectors corresponding to the longitudinal velocity perturbations were reduced in phase and amplitude to the value at the point where the amplitude of the perturbations $|U'_{\max}|$ is maximum. The analysis of the data in Fig. 2 allows us to conclude the discrete GSA not only correctly predicts the shape of perturbations developing due to instability on the *L4* mesh but also provides good quantitative agreement.The growth rate and the frequency of development of the most unstable perturbations at Re = 60 on a series of meshes

$L1-L5$ are shown in Table 2. The growth rate and frequency obtained by the discrete approach coincide with high accuracy (on all meshes for the flow, the error does not exceed 0.4%) with the solution of the transient Navier–Stokes equations, which indicates that the approach was implemented correctly.

Table 2

Comparison of computational parameters of the most unstable perturbations obtained by two methods on a series of meshes

Mesh	Computational value of parameter			
	Growth rate ω_r		Frequency ω_i	
	I	II	I	II
$L1$	0.0132		0.754	0.753
$L2$	0.0389		0.740	0.741
$L3$	0.0420	0.0421	0.738	
$L4$	0.0430	0.0431	0.737	
$L5$	0.0437		0.736	

Notations: I corresponds to direct numerical solution of transient Navier–Stokes equations; II to GSA, discrete approach. Note. Reynolds number Re = 60, Mach number M = 0.2.

Comparison of the results of two methods of global stability analysis

The direct comparison of the discrete and continuum approaches implemented in the NTS code is complicated by differences both in the methods for calculating the Jacobian and in the numerical schemes used to calculate the inviscid part of the flows.

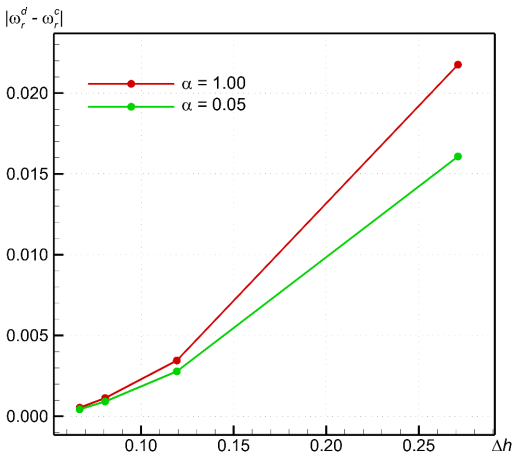

Fig. 3. Effect of the grid step on the difference in growth rates ω_r calculated using discrete (d) and continuum (c) GSA methods
Hybrid schemes with two weights of the upwind term α were used

The discrete GSA uses the same computational scheme as for calculating the basic flow, i.e., a hybrid upwind Roe scheme. This correction is significantly nonlinear, which does not allow it to be used within the framework of continuum GSA, therefore it uses a simplified linear upwind term. It would be possible to avoid differences between the schemes by using identical central difference schemes, but in practice this is impossible due to loss of stability when obtaining a basic solution. Nevertheless, if we reduce the weight of the upwind term, this can drastically reduce the difference in the schemes used.

This possibility is illustrated in Fig. 3, which shows the dependence on the grid step of the modulo difference in growth indicators $|\omega_r^d - \omega_r^c|$ in the vicinity of the cylinder obtained from the results of discrete and continuum GSA. If the weight of the upwind term is reduced, the difference decreases. The following are the results obtained using hybrid schemes with the weight of the upwind term $\alpha = 0.05$.

The growth rate and frequency obtained on the smallest $L5$ grid using the continuum and discrete approaches (Table 3) practically coincide. The same table shows a comparison with the results from [26, 31], confirming that the GSA results are representative.

Table 3

Comptutational parameters of unsteady perturbation mode obtained by two methods on the *L*5 mesh with varying Reynolds numbers, as well as comparison with the literature data

Computational approach	Computational value of parameter			
	ω_r		ω_i	
	Re=50	Re=60	Re=50	Re=60
GSA, discrete approach	–0.01099	–0.04368	0.72965	0.73637
continuum	–0.01093	–0.04372	0.72955	0.73633
[26], GSA, discrete approach	–0.013	–0.047	0.745	0.754
[31], direct numerical solution of Navier–Stokes equations	–0.012	–0.050	0.750	0.757

The arithmetic mean of the eigenvalues obtained using discrete and continuum approaches was used as a "reference" value of $\omega^{ref} = \left(\omega_r^{ref}, \omega_i^{ref}\right)$ to estimate the error in calculating the growth rate and the frequency of unsteady perturbation mode on on the coarser mesh *L*5.

Dependences of the error of the GSA results

$$\Delta\omega = \frac{(\omega - \omega^{ref})D}{U_0}$$

on the characteristic step of the mesh Δh, defined as the average step along the angular coordinate at a distance of 4*D* from the surface of the cylinder, are shown in Fig. 4 and allow us to draw the following conclusions.The calculation error is almost the same for both considered Reynolds numbers. The real order of accuracy of the GSA, which was determined by power-law approximations of the dependence of the error on the grid step, turned out to be approximately the same for both approaches: its value is approximately 3.1 for the growth rate, and 1.8 (discrete GSA) and 2.0 (continuum GSA) for frequency. These values are consistent with the formal order of the

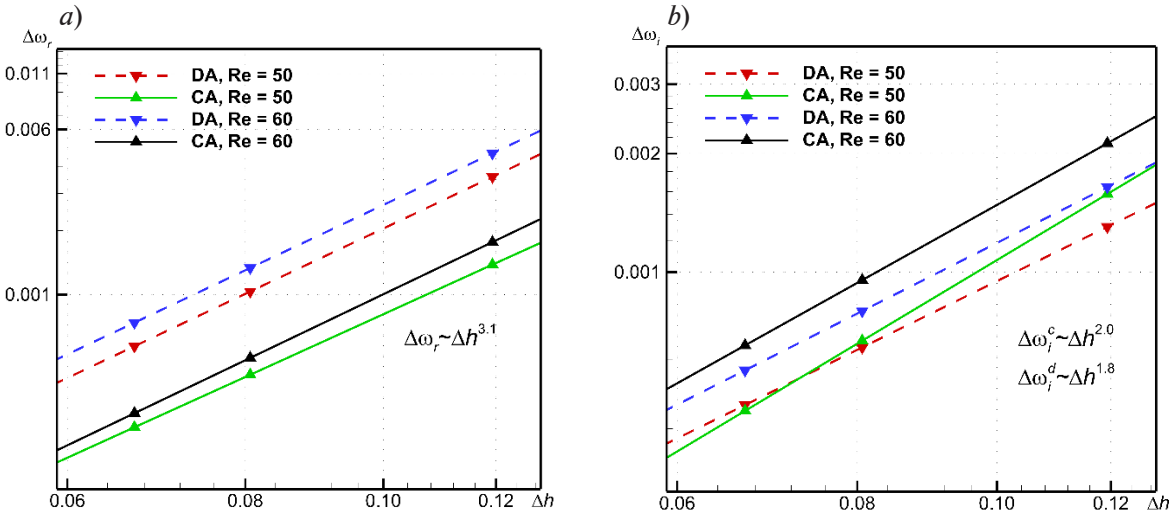

Fig. 4. Stepwise dependences of errors in calculating the growth rate (*a*) and frequency (*b*). The dependences were obtained by discrete (DA) and continuum (CA) approaches on meshes *L*1 – *L*4, with varying Reynolds numbers (dependences are given by symbols), and their approximation by exponential functions (straight lines on a logarithmic scale)

schemes used, in which convective terms are approximated by the third order, and viscous ones by the second. In addition, it should be borne in mind that the actual order of the schemes may decrease on non-uniform meshes (that is, the meshes are used in this work). The analysis of the data in Fig. 4 also allows us to conclude that the error in predicting the growth rate was about three times less when using the continuum approach, and the error in predicting the frequency of perturbations was less when using the discrete approach.

Conclusion

Two approaches to global stability analysis (GSA) were compared using the example of the problem on laminar flow around a cylinder at Reynolds numbers close to critical, differing in the methods for calculating the Jacobian of the Navier–Stokes equations: discrete (linearization of these discretized equations) and continuum (discretization of these linearized equations).

The discrete GSA approach we implemented was verified by comparison with the results of direct numerical simulation of unsteady laminar flow around the cylinder at Reynolds number $\mathrm{Re} = 60$. The results of the comparison showed that the growth rate and the vibration frequency of the most unsteady mode coincided with high accuracy on all the considered meshes.

The order of accuracy of the GSA turned out to be the same for continuum and discrete methods for calculating the Jacobian, and corresponded to the formal order of accuracy of spatial discretization by the numerical schemes used to obtain the solution whose stability was analyzed The error in predicting the growth rate of perturbations is less when using the continuum approach, and the error in predicting the vibration frequency of perturbations is less when using the discrete approach.

Thus, it can be argued that the continuum and discrete approaches are equivalent in order of accuracy and the choice of a specific approach for conducting stability analysis can be determined by other criteria (ease of implementation, computational costs, etc.).

This study was supported by the Russian Science Foundation (grant 21-72-20029). The simulations were run on the Polytechnic RSC Tornado cluster of the Polytechnic Supercomputer Center (http://www.scc.spbstu.ru).

REFERENCES

1. **Boiko A. V., Dovgal A. V., Grek G. R., Kozlov V. V.,** Physics of transitional shear flows: Instability and laminar-turbulent transition in incompressible near-wall shear layers, Book Ser. “Fluid Mechanics and its Applications”, Springer, 2012.

2. **Schmid P. J., Henningson D. S.,** Stability and transition in shear flows, Book Series “Applied Mathematical Sciences”, Vol. 142. Springer, New York, 2001.

3. **Theofilis V.,** Advances in global linear instability analysis of nonparallel and three-dimensional flows, Prog. Aerosp. Sci. 39 (4) (2003) 249–315.

4. **Theofilis V.,** Global linear instability, Annu. Rev. Fluid Mech. 43 (2011) 319–352.

5. **Crouch J. D., Garbaruk A., Magidov D.,** Predicting the onset of flow unsteadiness based on global instability, J. Comput. Phys. 224 (2) (2007) 924–940.

6. **Crouch J. D., Garbaruk A., Magidov D., Travin A.,** Origin of transonic buffet on aerofoils, J. Fluid Mech. 628 (10 June) (2009) 357–369.

7. **Garbaruk A., Crouch J. D.,** Quasi-three-dimensional analysis of global instabilities: Onset of vortex shedding behind a wavy cylinder, J. Fluid Mech. 677 (25 June) (2011) 572–588.

8. **Crouch J. D., Garbaruk A., Strelets M.,** Global instability in the onset of transonic-wing buffet, J. Fluid Mech. 881 (25 Dec.) (2019) 3–22.

9. **Garbaruk A., Strelets M., Crouch J. D.,** Effects of extended laminar flow on wing buffet-onset characteristics, AIAA J. 59 (8) (2021) 2848–2854.

10. **De Pando M. F., Sipp D., Schmid P. J.,** Efficient evaluation of the direct and adjoint linearized dynamics from compressible flow solvers, J. Comput. Phys. 231 (23) (2012) 7739–7755.

11. **Thormann R., Widhalm M.,** Linear-frequency-domain predictions of dynamic-response data for viscous transonic flows, AIAA J. 51 (11) (2013) 2540–2557.

12. **Mettot C., Renac F., Sipp D.,** Computation of eigenvalue sensitivity to base flow modifications in a discrete framework: Application to open-loop control, J. Comput. Phys. 269 (15 July) (2014) 234–258.

13. **Xu S., Timme S., Badcock K. J.,** Krylov subspace recycling for linearized aerodynamics analysis using DLR-TAU, Proc. Int. Forum Aeroelasticity Struct. Dyn. (IFASD 2015). 28 June–2 July, 2015. St. Petersburg, Russia. In 3 Vols. 2 (2016) 1462–1479.

14. **Sartor F., Metot C., Bur R., Sipp D.,** Unsteadiness in transonic shock-wave/boundary-layer interactions: experimental investigation n and global stability analysis, J. Fluid Mech. 781 (25 Oct.) (2015) 550–577.

15. **Busquet D., Marquet O., Richez F., et al.,** Global stability analysis of turbulent flows around an airfoil near stall, Proc. Eurogen 2017 Conf., Sept. 13–15, 2017, Madrid, Spain. Pp. 1–7.

16. **Timme S.,** Global instability of wing shock-buffet onset, J. Fluid Mech. 885 (25 Febr.) (2020) P. A37.

17. **Plante F., Dandois J., Beneddine S., et al.,** Link between subsonic stall and transonic buffet on swept and unswept wings: From global stability analysis to nonlinear dynamics, J. Fluid Mech. 908 (10 Febr.) (2021) A16.

18. **He W., Timme S.,** Triglobal infinite-wing shock-buffet study, J. Fluid Mech. 925 (25 Oct.) (2021) A27.

19. **Giles M. B., Pierce N. A.,** An introduction to the adjoint approach to design, Flow, Turbul. Combust. 65 (3–4) (2000) 393–415.

20. **Peter J. E. V., Dwight R. P.,** Numerical sensitivity analysis for aerodynamic optimization: A survey of approaches, Comput. Fluids. 39 (3) (2010) 373–391.

21. **Dwight R. P.,** Efficiency improvements of RANS-based analysis and optimization using implicit and adjoint methods on unstructured grids, Dtsch. Zent. fur Luft-und Raumfahrt–Forschungsberichte. (11) (2006) 1–162.

22. **Wengert R. E.,** A simple automatic derivative evaluation program, Commun. ACM. 7 (8) (1964) 463–464.

23. **Lyu Z., Kenway G. K. W.,** Automatic differentiation adjoint of the Reynolds-averaged Navier–Stokes equations with a turbulence model, Proc. 21st AIAA Comput. Fluid Dyn. Conf., June 24–27, 2013. San Diego (USA) (2013) 1–24.

24. **Hascолт L., Pascual V.,** TAPENADE 2.1 user's guide, INRIA Tech. Rep. No. 0300. Sept. 2004. 78 p. http://www.inria.fr/rrrt/rt-0300.html.

25. **Straka C. W.** ADF95: Tool for automatic differentiation of a FORTRAN code designed for large numbers of independent variables, Comput. Phys. Commun. 168 (2) (2005) 123–139.

26. **Gianetti F., Luchini P.,** Structural sensitivity of the first instability of the cylinder wake, J. Fluid Mech. 581 (25 June) (2007) 167–197.

27. **Shur M. L., Strelets M. K., Travin A. K.,** High-order implicit multi-block Navier – Stokes code: Ten-year experience of application to RANS / DES / LES / DNS of turbulence, Proc. 7th Symp. Overset Grids Solut. Technol. Oct. 5–7, 2004. Huntington Beach, CA, USA (2004) 1–52.

28. **Hernandez V., Roman J. E., Vidal V.,** SLEPC: A scalable and flexible toolkit for the solution of eigenvalue problems, ACM Trans. Math. Softw. 31 (3) (2005) 351–362.

29. **Golub G. H., van Loan C. F.,** Matrix computations, 3-rd Ed., John Hopkins University Press, Baltimore, Maryland, USA,1996.

30. **Mack C. J., Schmid P. J.,** A preconditioned Krylov technique for global hydrodynamic stability analysis of large-scale compressible flows, J. Comput. Phys. 229 (3) (2010) 541–560.

31. **Canuto D., Taira K.,** Two-dimensional compressible viscous flow around a circular cylinder, J. Fluid Mech. 785 (25 Dec.) (2015) 349–371.

СПИСОК ЛИТЕРАТУРЫ

1. **Бойко А. В., Грек Г. Р., Довгаль А. В., Козлов В. В.** Физические механизмы перехода к турбулентности в открытых течениях. Москва-Ижевск: НИЦ «Регулярная и хаотическая динамика», Институт компьютерных исследований, 2006. 304 с.

2. **Schmid P. J., Henningson D. S.** Stability and transition in shear flows. Book Series "Applied mathematical Sciences", Vol. 142. New York: Springer, 2001. 556 p.

3. **Theofilis V.** Advances in global linear instability analysis of nonparallel and three-dimensional flows // Progress in Aerospace Sciences. 2003. Vol. 39. No. 4. Pp. 249–315.

4. **Theofilis V.** Global linear instability // Annual Review of Fluid Mechanics. 2011. Vol. 43. Pp. 319–352.

5. **Crouch J. D., Garbaruk A., Magidov D.** Predicting the onset of flow unsteadiness based on global instability // Journal of Computational Physics. 2007. Vol. 224. No. 2. Pp. 924–940.

6. **Crouch J. D., Garbaruk A., Magidov D., Travin A.** Origin of transonic buffet on aerofoils // Journal of Fluid Mechanics. 2009. Vol. 628. 10 June. Pp. 357–369.

7. **Garbaruk A., Crouch J. D.** Quasi-three-dimensional analysis of global instabilities: Onset of vortex shedding behind a wavy cylinder // Journal of Fluid Mechanics. 2011. Vol. 677. 25 June. Pp. 572–588.

8. **Crouch J. D., Garbaruk A., Strelets M.** Global instability in the onset of transonic-wing buffet // Journal of Fluid Mechanics. 2019. Vol. 881. 25 December. Pp. 3–22.

9. **Garbaruk A., Strelets M., Crouch J. D.** Effects of extended laminar flow on wing buffet-onset characteristics // AIAA Journal. 2021. Vol. 59. No. 8. Pp. 2848–2854.

10. **De Pando M. F., Sipp D., Schmid P. J.** Efficient evaluation of the direct and adjoint linearized dynamics from compressible flow solvers // Journal of Computational Physics. 2012. Vol. 231. No. 23. Pp. 7739–7755.

11. **Thormann R., Widhalm M.** Linear-frequency-domain predictions of dynamic-response data for viscous transonic flows // AIAA Journal. 2013. Vol. 51. No. 11. Pp. 2540–2557.

12. **Mettot C., Renac F., Sipp D.** Computation of eigenvalue sensitivity to base flow modifications in a discrete framework: Application to open-loop control // Journal of Computational Physics. 2014. Vol. 269. 15 July. Pp. 234–258.

13. **Xu S., Timme S., Badcock K. J.** Krylov subspace recycling for linearized aerodynamics analysis using DLR-TAU // Proceedings of the International Forum on Aeroelasticity and Structural Dynamics (IFASD 2015). 28 June–2 July, 2015. St. Petersburg, Russia. In 3 volumes. Vol. 2. Pp. 1462–1479.

14. **Sartor F., Metot C., Bur R., Sipp D.** Unsteadiness in transonic shock-wave/boundary-layer interactions: experimental investigation n and global stability analysis // Journal of Fluid Mechanics. 2015. Vol. 781. 25 October. Pp. 550–577.

15. **Busquet D., Marquet O., Richez F., Juniper M., Sipp D.** Global stability analysis of turbulent flows around an airfoil near stall // Proceedings of the Eurogen 2017 Conference. September 13–15, 2017. Madrid, Spain. Pp. 1–7.

16. **Timme S.** Global instability of wing shock-buffet onset // Journal of Fluid Mechanics. 2020. Vol. 885. 25 February. P. A37.

17. **Plante F., Dandois J., Beneddine S., Laurendeau E., Sipp D.** Link between subsonic stall and transonic buffet on swept and unswept wings: From global stability analysis to nonlinear dynamics // Journal of Fluid Mechanics. 2021. Vol. 908. 10 February. P. A16.

18. **He W., Timme S.** Triglobal infinite-wing shock-buffet study // Journal of Fluid Mechanics. 2021. Vol. 925. 25 October. P. A27.

19. **Giles M. B., Pierce N. A.** An introduction to the adjoint approach to design // Flow, Turbulence and Combustion. 2000. Vol. 65. No. 3–4. Pp. 393–415.

20. **Peter J. E. V., Dwight R. P.** Numerical sensitivity analysis for aerodynamic optimization: A survey of approaches // Computers & Fluids. 2010. Vol. 39. No. 3. Pp. 373–391.

21. **Dwight R. P.** Efficiency improvements of RANS-based analysis and optimization using implicit and adjoint methods on unstructured grids // Deutsches Zentrum für Luft-und Raumfahrt – Forschungsberichte. 2006. No. 11. Pp. 1–162.

22. **Wengert R. E.** A simple automatic derivative evaluation program // Communications of the ACM. 1964. Vol. 7. No. 8. Pp. 463–464.

23. **Lyu Z., Kenway G. K. W.** Automatic differentiation adjoint of the Reynolds-averaged Navier–Stokes equations with a turbulence model // Proceedings of the 21st AIAA Computational Fluid Dynamics Conference. June 24–27, 2013. San Diego (USA), Pp. 1–24.

24. **Hascoлt L., Pascual V.** TAPENADE 2.1 user's guide. INRIA Technical Report No. 0300. September, 2004. 78 p. http://www.inria.fr/rrrt/rt-0300.html

25. **Straka C. W.** ADF95: Tool for automatic differentiation of a FORTRAN code designed for large numbers of independent variables // Computer Physics Communications. 2005. Vol. 168. No. 2. Pp. 123–139.

26. **Gianetti F., Luchini P.** Structural sensitivity of the first instability of the cylinder wake // Journal of Fluid Mechanics. 2007. Vol. 581. 25 June. Pp. 167–197.

27. **Shur M. L., Strelets M. K., Travin A. K.** High-order implicit multi-block Navier – Stokes code: Ten-year experience of application to RANS / DES / LES / DNS of turbulence // Proceedings of

the 7th Symposium on Overset Grids & Solution Technology. October 5–7, 2004. Huntington Beach, CA, USA. 2004. Pp. 1–52.

28. **Hernandez V., Roman J. E., Vidal V.** SLEPC: A scalable and flexible toolkit for the solution of eigenvalue problems // ACM Transactions of Mathematical Software. 2005. Vol. 31. No. 3. Pp. 351–362.

29. **Голуб Дж., Лоун Ван Ч.** Матричные вычисления. Пер. с англ. Под ред. Воеводина В. В. М.: Мир, 1999. 548 с.

30. **Mack C. J., Schmid P. J.** A preconditioned Krylov technique for global hydrodynamic stability analysis of large-scale compressible flows // Journal of Computational Physics. 2010. Vol. 229. No. 3. Pp. 541–560.

31. **Canuto D., Taira K.** Two-dimensional compressible viscous flow around a circular cylinder // Journal of Fluid Mechanics. 2015. Vol. 785. 25 December. Pp. 349–371.

THE AUTHORS

GOLUBKOV Valentin D.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
golubkovvd@gmail.com
ORCID: 0000-0001-9473-7430

GARBARUK Andrey V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
agarbaruk@mail.ru
ORCID: 0000-0002-2775-9864

СВЕДЕНИЯ ОБ АВТОРАХ

ГОЛУБКОВ Валентин Денисович – *инженер лаборатории «Вычислительная гидроакустика и турбулентность» Высшей школы прикладной математики и вычислительной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
golubkovvd@gmail.com
ORCID: 0000-0001-9473-7430

ГАРБАРУК Андрей Викторович – *доктор физико-математических наук, главный научный сотрудник лаборатории «Вычислительная гидроакустика и турбулентность» Высшей школы прикладной математики и вычислительной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
agarbaruk@mail.ru
ORCID: 0000-0002-2775-9864

Received 20.12.2022. Approved after reviewing 26.09.2023. Accepted 26.09.2023.
Статья поступила в редакцию 20.12.2022. Одобрена после рецензирования 26.09.2023. Принята 26.09.2023.

Original article
DOI: https://doi.org/10.18721/JPM.16406

SIMULATION OF A SUPERSONIC NOZZLE OF A MASSIVE HELIUM JET INJECTOR

V. M. Timokhin✉, *D. D. Korobko*, *L. I. Nurtdinova*,
V. G. Kapralov, *V. Yu. Sergeev*

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ v.timokhin@spbstu.ru

Abstract. The paper describes the design and operating principle of a valve designed for testing the prototypes of key elements of a massive gas injection system: a seat and a supersonic nozzle. The calculation results for supersonic nozzle parameters close to optimal for the selected valve design have been presented. The simulation of gas flows through the seat and nozzle was carried out, the parameters of the formed jet at the injection system outlet in the nozzle near field were calculated. The choice of the "stepped" nozzle profile for the first tests was justified. An assessment was made of the requirements for the accuracy of manufacturing prototype nozzles.

Keywords: numerical simulation, magnetic confinement, high-temperature plasma, gas jets, discharge shutdown

Funding: The research was supported by "ROSATOM" State Corporation and the Ministry of Science and Higher Education of Russian Federation within the framework of the Federal Project 3 (U3) "Development and Creation of Jet & Pellet Injection Systems with Increased Performance and Resource" (No. FSEG-2023-0018).

Citation: Timokhin V. M., Korobko D. D., Nurtdinova L. I., Kapralov V. G., Sergeev V. Yu., Simulation of a supersonic nozzle of a massive helium jet injector, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 63–75. DOI: https://doi.org/10.18721/JPM.16406

Научная статья
УДК 533.17
DOI: https://doi.org/10.18721/JPM.16406

МОДЕЛИРОВАНИЕ СВЕРХЗВУКОВОГО СОПЛА ИНЖЕКТОРА МАССИВНОЙ ГЕЛИЕВОЙ СТРУИ

В. М. Тимохин✉, *Д. Д. Коробко, Л. И. Нуртдинова,*
В. Г. Капралов, В. Ю. Сергеев

Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

✉ v.timokhin@spbstu.ru

Аннотация. В работе описана конструкция и принцип действия клапана, предназначенного для испытаний прототипов ключевых элементов системы массивной газовой инжекции: седла и сверхзвукового сопла. Приведены результаты расчетов параметров сверхзвукового сопла, близких к оптимальном для выбранной конструкции клапана. Выполнено моделирование газовых потоков через седло и сопло, рассчитаны параметры формируемой струи на выходе системы инжекции в ближнем поле сопла. Обоснован выбор «ступенчатой» формы профиля сопла для первых испытаний. Сделана оценка требований к точности изготовления прототипов сопел.

Ключевые слова: численное моделирование, магнитное удержание, высокотемпературная плазма, газовая струя, выключение разряда

Финансирование. Работа поддержана госкорпорацией «Росатом» и Министерством науки и высшего образования Российской Федерации в рамках Федерального проекта 3 (U3), № FSEG0018-2023- «Разработка и создание систем струйной и пеллет инжекции с повышенными производительностью и ресурсом».

Ссылка для цитирования: Тимохин В. М., Коробко Д. Д., Нуртдинова Л. И., Капралов В. Г., Сергеев В. Ю. Моделирование сверхзвукового сопла инжектора массивной гелиевой струи // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 4 № .16. С. 63–75. DOI: https://doi.org/10.18721/ JPM.16406

Introduction

One of the fundamental tasks in thermonuclear energy is to prevent or significantly reduce the probability of disruption of the tokamak discharge. Efforts are currently underway to solve this problem [1].

Let us briefly outline it. If disruption of the discharge current is unavoidable, rapid mitigation measures should be taken as a last resort to reduce the likelihood of significant damage to the vacuum chamber and other structural elements of the tokamak reactor. Since the thermal and magnetic energy of the plasma is effectively stored in the chamber during disruption, it can only be redistributed inside the chamber, but cannot be removed. This redistribution can be achieved, for example, by introducing a sufficiently large amount of impurities. A group of noble gases (helium, neon and argon) is commonly used as such impurities, which is designed to generate isotropic radiation of plasma energy on the first wall and is aimed at preventing concentrated loading.

The most common injection technology used for rapid delivery of impurities is massive gas injection (MGI). One of the prototypes of the MGI is described in the following section of this article.

An important element of the MGI system is the supersonic Laval nozzle, which ensures the formation of a jet of injected impurity into the tokamak vacuum chamber [2]. The nozzle parameters determine the velocity and distribution of the injected gas, which affects the penetration depth of the impurity and determines the distribution of the radiation source inside the

plasma filament. Gas flow in the subcritical (tapering) section of the nozzle occurs at subsonic speeds. The local gas velocity in the critical (narrowest) section of the nozzle reaches sonic levels, and gas flow moves at supersonic speeds in the supercritical (expanding) section. The internal energy of the gas is converted into the kinetic energy of its directional motion. In addition, the gas passing through the nozzle at a significant speed does not have time to transfer a noticeable amount of its thermal energy to its walls. This feature of the process allows to consider it adiabatic, which greatly simplifies its modeling.

The goals of this study included modeling the process of outflow of a massive gas jet from the MGI system for various types of nozzle and choosing its optimal type.

The criterion for the effectiveness of the nozzle is to achieve the maximum gas velocity with a minimum jet angle at the nozzle outlet (this allows increasing the penetration depth of the injected gas into the plasma).

Another important criterion for choosing a prototype nozzle is the simplicity of its manufacturing technology. The small size of the critical section and the complex shape of the nozzle profile require laser and electroplating technologies. The possibility of manufacturing a nozzle without complex metalworking procedures creates a significant advantage with comparable parameters of the jet formed by the nozzle.

The simulation was performed to support bench tests of a prototype valve for massive gas injection and did not involve changes in gas flow parameters and nozzle geometry over a wide range.

The main purpose of the calculations was to find the optimal distribution of gas flow parameters in the near field at the nozzle outlet. The optimal values of the set of parameters are necessary for designing the diagnostics of the MGI test system, as well as for modeling the interaction of gas jets with high-temperature plasma of magnetic confinement setups.

The obtained results are presented as follows. First, the operating principles and configuration of the gas valve designed to form a supersonic jet are described. Then the calculation results of the main parameters of the supersonic nozzle are presented. Next, we describe the algorithm for constructing the computational grid and the results obtained during the simulation, as well as analyze the gas flow through a supersonic nozzle. In conclusion, the main conclusions of the work are formulated.

Gas valve for testing seat and nozzle prototypes

The MGI method is relatively simple to implement, but tends to form a local source of impurity radiation in the peripheral regions of the plasma, which can lead to low efficiency of impurity injection into the central regions of the plasma cord [3]. According to the original design [4], the positioning of the MGI valve allows it to be brought closer to the plasma, which increases the efficiency of impurity injection. The main idea of the design is to ensure the mobility of the valve at a considerable distance from the actuator, which allows to shift the outlet section of the valve nozzle up to the last closed magnetic surface (LCMS). The valve system allows for it to move from a position outside the gate valve to the nozzle to a position next to the LCMS. The movement of the nozzle can reach a distance of more than a meter, which usually exceeds the size of the neutron protection of neutron setups, including the ITER reactor (International Thermonuclear Experimental Reactor). Such movement can be quite fast, which opens up prospects for using such a configuration of MGI in the radiation environment of a tokamak reactor. This eliminates the need for guide tubes to deliver the jet from the valve nozzle to the plasma, reduces the system response time, and reduces the angular expansion of the gas jet.

The following calculated characteristics of a massive helium gas jet injector are set:

gas flow is at least 10^{23} atm/s (maximum);

total number of injected particles is at least $5 \cdot 10^{23}$;

response time of the system (appearance of gas at the nozzle outlet after receiving the trigger signal from the system predicting the disruption of the discharge current) is no more than 3 ms;

gas delivery time from the nozzle outlet to the plasma is less than 1 ms (when the nozzle is located at a distance of less than 10 cm from the LCMS).

For the purpose of experimental studies of the formation of a gas jet, a special valve was developed that provides relatively easy access to the nozzle, as well as to the seat that is a structural element designed to securely attach the nozzle and block the gas flow.

Fig. 1 shows a general view of the valve for testing the seat and nozzle prototypes. Both a replaceable seat *4* and a replaceable nozzle *2* can be independently installed in the valve. The valve allows operation with gas pressures up to several tens of atmospheres. Outlet nozzle *1* provides connection to the vacuum circuit of the setup through a port with a poppet valve, used to install vacuum sensors of the PMT-6M type.

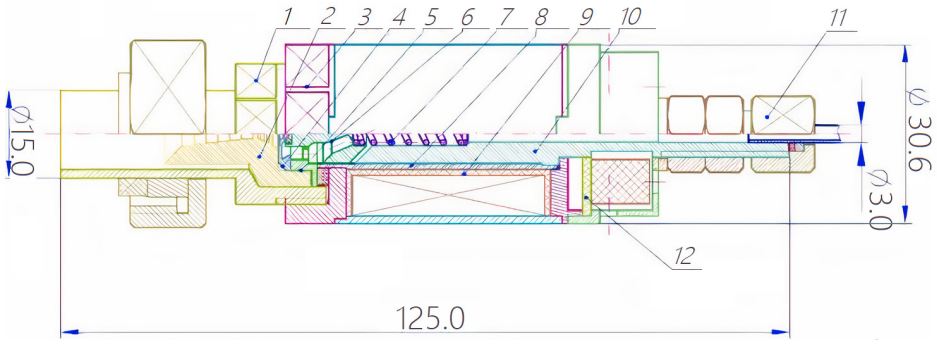

Fig. 1. Gas valve circuit for testing seat and nozzle prototypes:
outlet nozzle *1*, replaceable nozzle *2*, lower magnetic core *3*, replaceable seat *4*, separator *5*, armature with plate *6*, spring *7*, stainless separation tube *8*, solenoid coil *9*, stop *10*, gas inlet *11*, printed circuit board *12*

The valve design includes a solenoid, whose magnetic circuit consists of armature *6*, stop *10* and a set of magnetic cores. The working gap of the magnetic circuit is located between the armature and the stop. Solenoid *9* is mounted on stainless tube *8*, into which the stop is soldered. The coil is connected to an external switching power supply via terminals mounted on the printed circuit board *12*. External gas line *11* is connected to the inlet fitting of the stop to supply the working gas to the valve. Spring *7* is located inside the stop and armature, which ensures that the valve closes at the end of the electrical pulse.

The operating principle of the gas valve is as follows. The working gas flows from the inlet line into the inner channel of the stop, then through the channels in the armature, enters the volume surrounding the armature inside the stainless tube and is fed through separator *5* to replaceable seat *4*, whose channel is closed by the plate of armature *6*. When the solenoid is triggered, the valve plate moves away from the seat and the gas passes through the seat into replaceable nozzle *2*, in which a supersonic gas jet is formed, flowing through nozzle *1* into the vacuum volume on which the valve is installed. To replace the seat and/or nozzle, it is enough to turn off the gas line and fill the vacuum volume with atmosphere, then unscrewing the outlet nozzle from lower magnetic core *3*. After that, new seat and nozzle can be inserted, replacing the seals (if necessary).

The operating cycle of the valve includes the following sequence of actions.

At the stage of preparing the valve for operation, it is necessary to connect it to a switching power source and check the operation of the solenoid.

The next step is to connect inlet gas line *11* and pump it out and the internal volume of the valve, consisting of channels in stop *10*, anchor *6* and separator *5*, to remove the working gas and air that filled the main and the internal volume of the valve before operation starts.

After pumping, the inlet gas line and the inlet internal volume of the valve should be filled with new working gas.

Then, in compliance with safety procedures, it is necessary to check the operation of the valve with the working gas and its exhaust into the atmosphere.

Next, the valve is installed on a vacuum volume, into which the working gas is injected, and it is necessary to pump out the vacuum volume.

During operation, the switching power supply discharges the capacitor bank to solenoid *9*. As a result, magnetic flux is formed through a chain of magnetic conductors, stop *10* and armature *6*. The armature is attracted to the stop and detaches the plate from seat *4*. At the outlet of the seat, the leading edge of the gas flow is formed, which passes through replaceable nozzle *2*, where a supersonic gas jet is formed.

The calculation of parameters and optimization of gas flows formed by these structural elements is the subject of this study. As the source battery is discharged, the magnetic field in the solenoid weakens and the valve closes with spring *7*, which forms the trailing edge of the gas pulse and completes the operating cycle of the valve.

Calculation of supersonic nozzle parameters for MGI

The Laval nozzle parameters were calculated using methods and formulas for one-dimensional steady flows of ideal gas in channels [5, 6]. A variant of the Laval nozzle with a conical profile was calculated, which was taken as a basis for the manufacture of prototypes of the nozzle and seat. Well-known gas-dynamic relations and conservation laws were used:

$$\frac{T}{T_0} = 1 - \frac{\gamma - 1}{\gamma + 1}\lambda^2, \tag{1}$$

$$\frac{\rho}{\rho_0} = \left(1 - \frac{\gamma - 1}{\gamma + 1}\lambda^2\right)^{\frac{1}{\gamma - 1}}, \tag{2}$$

$$\frac{p}{p_0} = \left(1 - \frac{\gamma - 1}{\gamma + 1}\lambda^2\right)^{\frac{\gamma}{\gamma - 1}}, \tag{3}$$

where p_0, ρ_0, T_0 are the pressure, density and temperature of the resting gas; γ is the adiabatic index; λ is the reduced velocity,

$$\lambda = \frac{v}{c_{S*}} = \frac{v}{\sqrt{\frac{2\gamma R}{(\gamma + 1)M}T_0}}$$

(c_{S*} is the speed of sound at critical values of the parameters ρ_*, T_*, p_*).

The parameters ρ_*, T_*, p_* follow the expressions:

$$T_* = \frac{2}{\gamma + 1}T_0,\ \rho_* = \left(\frac{2}{\gamma + 1}\right)^{\frac{1}{\gamma - 1}}\rho_0,\ p_* = \left(\frac{2}{\gamma + 1}\right)^{\frac{\gamma}{\gamma - 1}}p_0.$$

The gas parameters should take their extreme values in the narrowest section of the Laval nozzle, called the "critical section". The designed nozzle should provide a mass flow of helium $G = 3{,}8 \cdot 10^{-4}$ kg/s. In accordance with the law of conservation of flow and the set values of pressure, temperature and flow at the nozzle inlet, the diameter of the critical section is determined:

$$d_{crit} = 2\sqrt{\frac{G\sqrt{T_0}}{B\pi p_0}}, \tag{4}$$

where $B = \left(\frac{2}{\gamma + 1}\right)^{\frac{\gamma + 1}{2(\gamma - 1)}}\sqrt{\gamma\frac{m_{He}}{R}} = 0.016\,\frac{\mathrm{K}^{0.5}\mathrm{s}}{\mathrm{m}}$.

We obtain the value d_{crit} = 0.73 mm in this simulation for helium.

When designing a conical nozzle, it is also necessary to choose the correct opening angle of the subsonic and supersonic sections of the nozzle [7–9]. The recommended values of the opening angle β of the supersonic section of the nozzle in conical nozzles should not exceed 15° [9].

For the calculations presented below, the angles of the subsonic and supersonic sections of the nozzle were selected (α and β, respectively: α = 26°, β = 12°), which determined the lengths of these sections as 10 and 20 mm, respectively.

When calculating the diameter of the outlet section, the gas dynamic function of the reduced flow rate is commonly used:

$$q(\lambda)=\frac{\rho v}{\rho_* c_{S*}}=\left(\frac{\gamma+1}{2}\right)^{\frac{1}{\gamma-1}}\lambda\frac{\rho}{\rho_0}=\left(\frac{\gamma+1}{2}\right)^{\frac{1}{\gamma-1}}\lambda\left(1-\frac{\gamma-1}{\gamma+1}\lambda^2\right)^{\frac{1}{\gamma-1}}. \quad (5)$$

It follows from the condition of conservation at steady flow that

$$\rho_2 v_2 S_2=\rho_* c_{S*} S_{crit}\rightarrow S_2=\frac{\rho_* c_{S*}}{\rho_2 v_2}S_{crit}=\frac{S_{crit}}{q(\lambda_2)}, \quad (6)$$

and the cross section at the nozzle outlet is expressed as

$$d_2=d_{crit}\sqrt{\frac{1}{q(\lambda_2)}}. \quad (7)$$

In the case under consideration, the diameter of the outlet section d_2 = 4 mm was set by the design parameters of the valve (see Fig. 1). Therefore, the value of the gas-dynamic function of the reduced flow rate, found from equations (5), allows to obtain the value of the reduced velocity λ_2 and the velocity v_2 at the nozzle outlet:

$$q(\lambda_2)=\left(\frac{d_{crit}}{d_2}\right)^2=0.014\rightarrow\lambda_2=1.97\rightarrow v_2=c_{S*}\lambda_2=1.74\ \text{km/s}. \quad (8)$$

Simulation results

The conical profile of the Laval nozzle calculated in the above manner had the following parameters:

diameter of inlet section d_1 = 4.0 mm;
diameter of critical section d_{crit} = 0.7 mm;
diameter of outlet section d_2 = 4.0 mm;
length of supersonic section of the nozzle l_2 = 20 mm;
total opening angle of the walls β = 12°;
length of subsonic section l_1 = 10 mm;
total opening angle of the walls α = 26°.

We chose the geometry of the conical nozzle with these parameters for the first experiments with the prototype of the MGI valve, simplified for manufacturing. A nozzle with stepped subsonic and supersonic sections, close to a conical profile, was designed, which can be achieved technologically by sequential drilling of the workpiece (the term «stepped nozzle» is used from now on for brevity). Seven steps were set for both sections of the nozzle. The step diameter was chosen to be 0.5 mm, the step length was 1.4 mm for the subsonic section and 2.8 mm for the supersonic section. A cylindrical seat was installed in front of the inlet section of the nozzle, with the length equal to the diameter of the inlet section, 4 mm, necessary to lock the gas flow and anchor this element in the valve structure.

A sketch of the nozzle with the seat and an example of a constructed computational grid for simulation is shown in Fig. 2. The computational domain was divided into two zones. A more refined grid was built near the nozzle and seat, where the gas flow parameters change faster, with the element size of 0.05 mm; the size of the grid cell was reduced to 0.01 mm for 10 cells adjacent to the wall to improve the resolution of the boundary layer in the main volume in the near-wall region. The size of the grid element was 0.15 mm in the area of the diagnostic chamber, where the jet is injected.

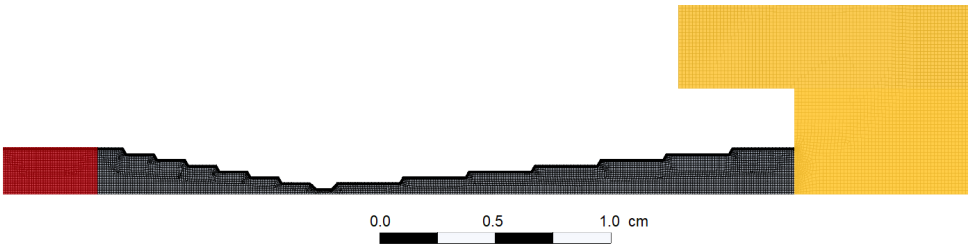

Fig. 2. Example of computational grid for stepped Laval nozzle with seat (highlighted in red). The region with the diagnostic chamber (highlighted in yellow) is shown

This section presents the results of numerical simulation of helium flow through such a nozzle. The computational domain was divided into similar zones, with the same characteristic dimensions of the grid element for calculations with other nozzle profiles (their main results are given in the next section).

The simulation was performed using the ANSYS Fluent CFD package, where a stationary system of continuity and momentum balance equations was numerically solved. An axisymmetric problem was solved using an implicit integration scheme, with second order approximation. The k-ε Realizable turbulence model was included in the calculations [10]. The following boundary conditions were imposed for solving the problem:

nozzle inlet pressure.....10 atm (1 MPa);
nozzle outlet pressure.....1 Pa;
temperature at the inlet and outlet of the nozzle.....300 K;
no-slip condition for the walls assumed to be smooth.

The choice of the calculation method was based on the following considerations. The estimation of the Knudsen number for the given problem parameters gives values not exceeding $2 \cdot 10^{-3}$ inside the nozzle and $3 \cdot 10^{-2}$at a distance of 5 cm from the nozzle, which allows using a continuum approach for this problem with reasonable accuracy at sufficiently small distances from the nozzle exit. The valve design assumes it is placed in close proximity to the plasma, at a distance of no more than 10 cm from the LCMS of the setup, which allows to expect a high degree of reliability of the calculated results obtained.

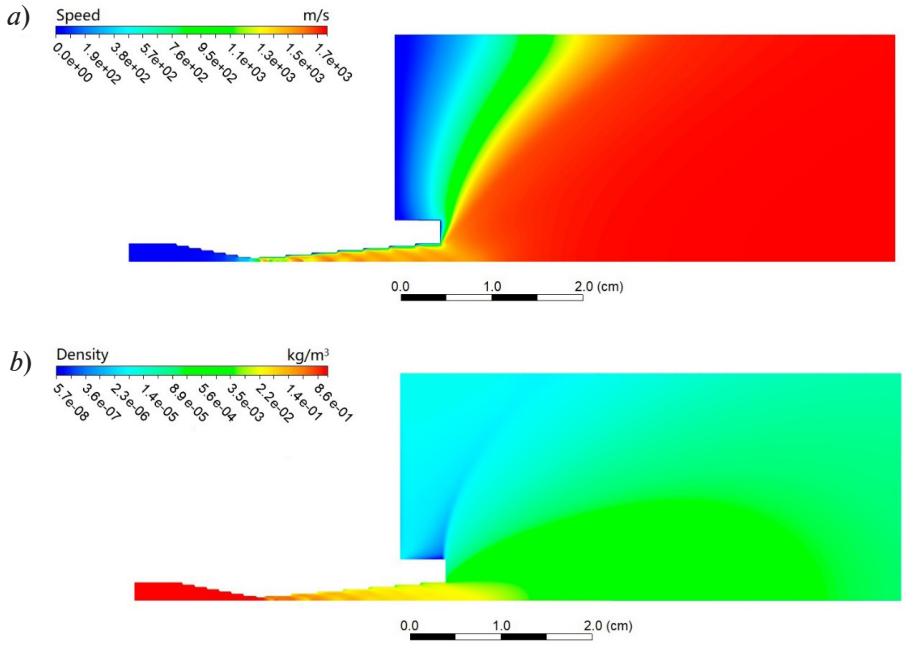

Fig. 3. Calculated distributions of velocity (*a*) and density (*b*) of helium jet at steady flow from stepped nozzle with seat

Fig. 3 shows the calculated velocity and density distributions for the stepped prototype of the Laval nozzle after steady flow is established. Inside the supersonic section of the nozzle, periodic jumps in gas parameters can be observed, resulting from the reflection of shock waves from the steps inside the nozzle. This behavior of gas outflow can lead to a turbulent gas flow regime inside the nozzle and affect its performance.

It is logical to assume that an increase in the number of steps in the nozzle profile will cause a decrease in the amplitude of these jumps and, accordingly, a more uniform behavior of the gas flow inside the nozzle. In this case, the risk of transition to a turbulent gas flow regime will be reduced. It should be noted that a jet with uniform outflow profiles of gas velocity and density is obtained at the outlet of the nozzle with such a profile with 7 steps, and the divergence of the jet is not large.

The velocity and density distributions along the jet injection axis are shown in Fig. 4. The figure clearly shows the nature of the flow parameter jumps along the injection axis inside the nozzle. The amplitude of fluctuations in density does not exceed 80% of the maximum value, and in terms of flow velocity it is about 40%. The maximum oscillations are concentrated near the critical region of the supersonic section of the nozzle. The oscillation amplitude decreases approaching the nozzle outlet.

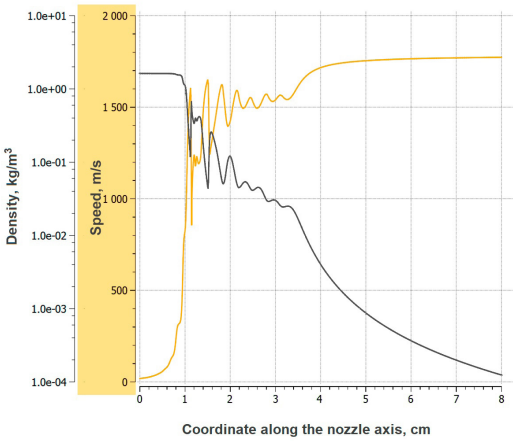

Fig. 4. Distributions of velocity (yellow curve) and density (black curve) of the jet along the nozzle axis for stepped nozzle with seat

Despite the peculiarities of the distribution of gas flow parameters inside the stepped nozzle, the general pattern of the distribution of jet parameters behind the nozzle remains the same as that of a conventional conical nozzle, except for a slightly larger gas expansion angle. The velocity directly at the nozzle outlet decreases to about 1.5 km/s, but the jet in the diagnostic chamber accelerates to the velocities $v_{max} = 1.7$ km/s, which is close to the theoretical maximum (see equations (8)). The pressure at the nozzle exit at the problem parameters set is about 800 Pa, which leads to a significantly underexpanded jet flow mode [11]. The characteristic dimensions of the hanging bow shock, obtained from empirical formulas [12], are about 10 cm and are comparable to the dimensions of the diagnostic chamber, which fully corresponds to the calculated flow pattern. Studies of the large-scale structure of the jet are beyond the scope of this work. The calculation results for nozzles with other profiles, including smooth conical ones, are given in the next section of the study.

Optimization of the nozzle profile

The conical shape of the nozzle is the limit for a nozzle with a stepped profile, with an increase in the number of steps to infinity. Theoretically, a parabolic nozzle should give optimal values of the velocity and gas expansion angle at the outlet [7]. The computational grid for the parabolic nozzle was constructed as follows: the dimensions of the inlet, outlet and critical sections were assumed to be the same as when designing a conical nozzle (to preserve mass flow), and the shapes of the profiles of the subsonic and supersonic sections were parabolical along the axis

of the nozzle. The shape of the profiles of the subsonic and supersonic parts was «merged» near the critical section, using the method with the first derivative of the shape of the nozzle profile tending to zero along its axis. Otherwise, a kink appear in the nozzle profile, which can lead to shock waves in this region.

To compare the angle of expansion for three types of nozzle, the half-width of the jet was estimated. The distribution of the jet density was calculated in the plane transverse to the axis of the jet injection, at a distance of 3 cm from the outlet section of the nozzle. Next, the distance from the injection axis was determined, at which the gas density drops by half relative to its maximum value, which determines the half-width of the jet, characterizing the scale of gas expansion at the nozzle outlet.

Density distributions along the cross section of the jet for the three types of nozzle for which calculations were performed are shown in Fig. 5, and the main characteristics of the jet in the diagnostic chamber are shown in the table. The results obtained indicate the smallest expansion of the helium jet and its highest exit velocity for a smooth conical nozzle. The parabolic shape of the nozzle section gives intermediate values of speed and width, which is apparently due to the suboptimal choice of the profile of this type of nozzle at given cross-section values.

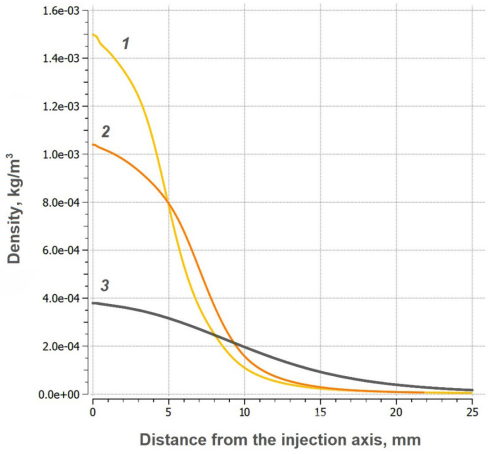

Fig. 5. Density distribution of helium jet along its cross-section for three types of nozzles at a distance of 3 cm from the exit from the outlet section.
The calculations were carried out for conical (curve *1*), parabolic (*2*) and stepped (*3*) nozzle shapes

The selection of the optimal nozzle profile with a parabolic cross section is a complex multi-parameter problem, whose solution is beyond the scope of this study. The stepped nozzle provides a less than 10% decrease in the jet velocity at the outlet and leads to an increase in the jet angle of expansion by about half, compared with the conical nozzle profile. Such a deterioration in the parameters of the jet is acceptable and is fully compensated by the technological simplicity of its manufacturing.

Analysis of Fig. 5 allows us to conclude that the density distributions of the jet in the cross section are monotonous, with maxima on the injection axis for all types of nozzles. The absolute density value on the injection axis for a stepped nozzle drops by 4 times relative to a conical one. There is also a significant increase in the angle of expansion of the jet for the case of a stepped nozzle.

During the simulation, the diameter of the critical section for the stepped nozzle was also varied to assess the effect of the nozzle manufacturing accuracy on the parameters of the formed jet. The computational grid corresponded to the one shown in Fig. 2, with accuracy up to the diameter of the critical section, which was set to ±0.2 mm from the calculated optimal value of 0.7 mm (obtained by formula (4). As a result, the velocity and density distributions of the gas at the outlet of the nozzle were calculated. The main parameters of the jet are summarized in the same table.

Table

Calculated key parameters of the nozzle

Nozzle type	Jet outlet velocity, m/s	Jet half-width, mm	Mass flow rate, g/s
Conical	1715	5.0	0.32
Parabolic	1666	7.0	0.32
Stepped .....standard	1551	10.3	0.32
.....with seat	1549	11.0	0.32
...with critical section increased to 0.9 mm	1551	10.6	0.55
...with critical sectiondecreased to 0.5 mm	1531	10.9	0.17

Thus, an increase in the mass flow rate of gas by 1.7 times was obtained with an increase in the diameter of the critical section to 0.9 mm. If this diameter was decreased to 0.5 mm, the mass flow rate decreased by 1.9 times. Calculations also show that the jet velocity in the diagnostic chamber does not significantly depend on the type of nozzle, while the smallest jet angle is obtained for a standard stepped nozzle. The drop in velocity at the nozzle outlet and the broadening of the jet profile with varying critical cross-sectional area do not exceed several percent. However, a significant change in the mass flow rate of gas with varying critical cross-sectional area indicates that high manufacturing accuracy is required for the dimensions of the stepped nozzle profile (diameter tolerance should be at least 0.01 mm).

Conclusion

The optimal parameters of the supersonic nozzle were calculated for a given mass flow rate of helium of $3.8 \cdot 10^{-4}$ kg/s. The diameters of the inlet and outlet sections were 4 mm, the diameter of the critical section was 0.7 mm, the lengths of the subsonic and supersonic sections of the nozzle were 10 and 20 mm, respectively; the total opening angle of the walls for the subsonic section was 26°, and 12° for the supersonic section. The outflow of gas through nozzles of various profiles into the diagnostic chamber was simulated using the ANSYS Fluent CFD package. The parameters of the gas flow through a stepped nozzle with a seat in the near field of the jet were calculated in detail. It was found that such a nozzle allows to obtain a gas flow rate of about 1550 m/s on the injection axis, with a half-width of the jet of 11 mm at a distance of 3 cm from the nozzle exit.

The simulation confirmed that a stepped nozzle with a seat gives an insignificant difference in the values of the velocity and half-width of the jet, compared with a nozzle without a seat. The maximum outlet velocity and the minimum jet angle can be achieved at the given cross-sectional values with a nozzle with a conical profile shape, while a stepped nozzle with a seat shows a decrease in quality by no more than 10% in terms of the jet velocity parameter, and a decrease by about 3 times in terms of the half-width of the jet, which is quite acceptable if the technological simplicity of manufacturing is taken into account.

Thus, the best option is the design and manufacturing of a stepped nozzle with a seat to be used in the first experiments with the prototype of the MGI valve in setups with magnetic plasma retention.

The critical cross-sectional area of a stepped nozzle with a seat was varied during the simulation. The calculation results showed a strong dependence of the mass flow rate on the critical section of the nozzle; the velocity and half-width of the jet at the outlet changed insignificantly.

In general, the conducted study confirms the stringent requirements for the manufacturing accuracy of nozzle profiles. The diameter tolerances of the nozzle profile must be no worse than the grade of 10.

REFERENCES

1. **Eidietis N. W.,** Prospects for disruption handling in a tokamak-based fusion reactor, Fusion Sci. Technol. 77 (7–8) (2021) 738–744.

2. **Landau L. D., Lifshitz E. M.,** Fluid mechanics (§ 97. Flow of gas through a nozzle), Second Ed. Vol. 6 of Course of Theoretical Physics. Pergamon Press, Oxford, New York, 1987.

Hollmann E. M., Aleynikov P. B., Fülöp T., et al., Status of research toward the ITER disruption mitigation system, Phys. Plasmas. 22 (2) (2015) 021802.

4. **Dremin M. M., Kapralov V. G., Kislov A. Ya., et al.,** Effect of noble gas injection on discharge disruption in T-10 tokamak, Probl. At. Sci. Technol. Ser. "Thermonucl. Fusion" (4) (2012) 58–70 (in Russian).

5. **Dulov V. G., Lukyanov G. A.,** Gazodinamika protsessov istecheniya [Gas dynamics of outflow processes], Nauka, Novosibirsk,1984 (in Russian).

6. **Izotov B. A.,** Raschet i postroyeniye profilya sopla Lavalya [Calculation and profiling of De Laval nozzle], Orenburg State University Publishing, Orenburg, 2009 (in Russian).

7. **Kirillov A. B., Rotinyan E. M.,** Mekhanika zhidkosti i gaza. Soplo Lavalya [Fluid-flow mechanics. De Laval nozzle], Polytechnical Institute Publishing, St. Petersburg, 2017 (in Russian).

8. **Pipko A. I., Pliskovskii V. Ya., Penchko E. A.,** Konstruirovaniye i raschet vakuumnykh system [Design and calculation of vacuum systems], Energiya, Moscow, 1979 (in Russian).

9. **Shishkov A. A., Panin S. D., Rumyantsev B. V.,** Rabochiye protsessy v raketnykh dvigatelyakh tverdogo topliva: Spravochnik [Operation processes in solid rocket motors: Handbook], Mashinostroyeniye, Moscow, 1988 (in Russian).

10. **Shih T. H., Liou W. W., Shabbir A., et al.,** A new k-ε eddy-viscosity model for high Reynolds number turbulent flows, Comput. Fluids. 24 (3) (1995) 227–238.

11. **Volkov K. N., Emelyanov V. N., Zazimko V. A.,** Turbulentnyye strui: staticheskiye modeli i modelirovaniye krupnykh vikhrey [Turbulent jet flows: Statistical models and simulation of large vortices], Fizmatlit, Moscow, 2014. (in Russian).

12. **Shelukhin N. N.,** Issledovaniye kharakteristik sverkhzvukovoy nedorasshirennoy strui [Study of characteristics of a supersonic underexpanded jet flow], TsaGI Sci. J. 26 (1–2) (1995) 78–87 (in Russian).

СПИСОК ЛИТЕРАТУРЫ

1. **Eidietis N. W.** Prospects for disruption handling in a tokamak-based fusion reactor // Fusion Science and Technology. 2021. Vol. 77. No. 7–8. Pp. 738–744.

2. **Ландау Л. Д., Лифшиц Е. М.** Теоретическая физика. В 10 тт. Т. 6. Гидродинамика (§ 97. Истечение газа через сопло). М.: Физматлит, 728 .2021 с.

3. **Hollmann E. M., Aleynikov P. B., Fülöp T., et al.** Status of research toward the ITER disruption mitigation system // Physics of Plasmas. 2015. Vol. 22. No. 2. P. 021802.

4. **Дремин М. М**., **Капралов В. Г., Кислов А. Я. и др**. Влияние напуска благородных газов на срыв разряда в токамаке Т-10 // Вопросы атомной науки и техники (ВАНТ). Серия. Термоядерный синтез. 2012. № 4. С. 58–70.

5. **Дулов В. Г., Лукьянов Г. А.** Газодинамика процессов истечения. Новосибирск: Наука, 1984. 226 с.

6. **Изотов Б. А.** Расчет и построение профиля сопла Лаваля. Оренбург: Изд-во Государственного Оренбургского университета, 20 .2009 с.

7. **Кириллов А. В., Ротинян Е. М.** Механика жидкости и газа. Сопло Лаваля. СПб.: Изд-во Политехнического университета, 2017. 22 с.

8. **Пипко А. И., Плисковский В. Я., Пенчко Е. А.** Конструирование и расчет вакуумных систем. -3е изд., перераб. и доп. М.: Энергия, 504 .1979 с.

9. **Шишков А. А., Панин С. Д., Румянцев Б. В.** Рабочие процессы в ракетных двигателях твердого топлива: Справочник. М.: Машиностроение, 240 .1988 с.

10. **Shih T. H., Liou W. W., Shabbir A., Yang Z. Zhu J.** A new k-ε eddy-viscosity model for high Reynolds number turbulent flows // Computers & Fluids. 1995. Vol. 24. No. 3. Pp. 227–238.

11. **Волков К. Н., Емельянов В. Н., Зазимко В. А.** Турбулентные струи: статические модели и моделирование крупных вихрей. М.: Физматлит, 360 .2014 с.

12. **Шелухин Н. Н.** Исследование характеристик сверхзвуковой недорасширенной струи // Ученые записки ЦАГИ (Центральный аэрогидродинамический институт). 1995. Т. 26. № 1–2. С. 78–87.

THE AUTHORS

TIMOKHIN Vladimir M.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
v.timokhin@spbstu.ru
ORCID: 0000-0002-4700-6122

KOROBKO Dmitriy D.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
korobko.dd@edu.spbstu.ru
ORCID: 0000-0002-8559-3209

NURTDINOVA Linara I.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
nurtdinova.li@edu.spbstu.ru
ORCID: 0000-0002-1029-2049

KAPRALOV Vladimir G.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
kapralov15@mail.ru
ORCID: 0000-0002-1091-6405

SERGEEV Vladimir Yu.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
V.Sergeev@spbstu.ru
ORCID: 0000-0002-4572-4120

СВЕДЕНИЯ ОБ АВТОРАХ

ТИМОХИН Владимир Михайлович – *кандидат физико-математических наук, старший научный сотрудник Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
v.timokhin@spbstu.ru
ORCID: 0000-0002-4700-6122

КОРОБКО Дмитрий Дмитриевич – *студент Института электроники и телекоммуникаций Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
korobko.dd@edu.spbstu.ru
ORCID: 0000-0002-8559-3209

НУРТДИНОВА Линара Ильфатовна – *студентка Института электроники и телекоммуникаций Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
nurtdinova.li@edu.spbstu.ru
ORCID: 0000-0002-1029-2049

КАПРАЛОВ Владимир Геннадиевич – *кандидат физико-математических наук, доцент Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
kapralov15@mail.ru
ORCID: 0000-0002-1091-6405

СЕРГЕЕВ Владимир Юрьевич – *доктор физико-математических. наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
V.Sergeev@spbstu.ru
ORCID: 0000-0002-4572-4120

Статья поступила в редакцию 05.09.2023. Одобрена после рецензирования 04.10.2023. Принята 04.10.2023.
Received 05.09.2023. Approved after reviewing 04.10.2023. Accepted 04.10.2023.

Original article
DOI: https://doi.org/10.18721/JPM.16407

SIZE EFFECTS IN MOLECULAR DYNAMICS SIMULATIONS OF A FULLERENE ION IMPACT ON THE SILICON SURFACE

K. P. Karasev[1,2]✉, *S. D. Strizhkin*[1], *P. A. Karaseov*[1], *A. I. Titov*[1]

[1] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;

[2] Alferov University of RAS, St. Petersburg, Russia

✉ kir.karasyov2017@yandex.ru

Abstract. In the paper, the interaction of an accelerated C_{60} fullerene ion with silicon monocrystal surface has been studied using molecular dynamics simulation. The dependence of a resulting crater size and sputtering yield on the initial size of the target was obtained. We proposed that computational artifacts revealed in simulations appeared due to two main reasons: shock waves raised by impinging the C_{60} ion, came back through the periodic boundary increasing the temperature around the impact point; dissipation of the energy, brought to the surface by the fullerene molecule, between small amount of atoms in the small cell might also affect the simulated results. It was established that 11 × 11 nm is the least size of lateral crystal dimensions required for the valid results of the simulation of the 8–14 keV C_{60} ion impact.

Keywords: molecular dynamics simulation, C_{60} fullerene ion, silicon, size effect, crystal, sputtering

Funding: The reported study was carried out within the framework of the State Assignment for Fundamental Research (Subject Code FSEG-2023-0016).

Citation: Karasev K. P., Strizhkin S. D., Karaseov P. A., Titov A. I., Size effects in molecular dynamics simulations of a fullerene ion impact on the silicon surface, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 76–85. DOI: https://doi.org/10.18721/JPM.16407

Научная статья
УДК 539.21
DOI: https://doi.org/10.18721/JPM.16407

РАЗМЕРНЫЕ ЭФФЕКТЫ ПРИ МОЛЕКУЛЯРНО-ДИНАМИЧЕСКОМ МОДЕЛИРОВАНИИ ПАДЕНИЯ ИОНА ФУЛЛЕРЕНА НА ПОВЕРХНОСТЬ КРЕМНИЯ

К. П. Карасев[1,2]✉, *Д. А. Стрижкин*[1], *П. А. Карасев*[1], *А. И. Титов*[1]

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия;

[2] Академический университет им. Ж. И. Алфёрова РАН, Санкт-Петербург, Россия

✉ kir.karasyov2017@yandex.ru

Аннотация. В работе выполнено моделирование взаимодействия ускоренных ионов фуллерена C_{60} с монокристаллом кремния. Исследовано влияние размеров модельного монокристалла на получаемые параметры кратера, образующегося в мишени при ударе, и распыление веществ мишени и фуллерена. Предлагаются причины возникновения вычислительных артефактов: это возврат энергии ударной волны через периодическую границу и не вполне корректное описание распределения принесенной энергии между атомами мишени. Установлено, что для получения достоверных (без размерных эффектов) результатов моделирования акта падения на монокристалл ионов C_{60}, обладающих энергиями 8 – 14 кэВ, необходимо использовать монокристаллы с размерами поверхности не менее 11 × 11 нм.

Ключевые слова: молекулярно-динамическое моделирование, ион фуллерена C_{60}, кремний, размерный эффект, кристалл, распыление

Финансирование: Работа выполнена в рамках Государственного задания на проведение фундаментальных исследований (код темы FSEG-2023-0016).

Ссылка для цитирования: Карасев К. П., Стрижкин Д. А., Карасев П. А., Титов А. И. Размерные эффекты при молекулярно-динамическом моделировании падения иона фуллерена на поверхность кремния // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 4 № .16. С. 76–85. DOI: https://doi.org/10.18721/JPM.16407

Introduction

Molecular dynamics (MD) simulation is widely used in modern science to study various phenomena at the micro level. It consists in simulating the time evolution of a system of objects (atoms), calculated by numerical integration of equations of motion. The motion of a particle ensemble can be uniquely given by a Hamiltonian in the classical-mechanics approximation, determined by a set of generalized coordinates and momenta. This Hamiltonian characterizes the total energy of the system and fully describes its dynamic nature. The MD method can be applied to analyze both simple crystal structures and complex biological molecules [1]. The MD method is also often used to study the effects occurring under irradiation of various targets with accelerated ions. Studies based on MD simulation established the mechanisms behind the evolution of structural defects [2] and the relief of the target surface [3] under ion bombardment. In particular, the dependences of the mass transfer function and the surface morphology on the initial beam incidence angle were determined [3]. Additionally, numerous papers focused on sputtering of ions incident on the surface: the number and composition of outgoing particles, their angular distribution over incident energies of bombarding ions [4].

Simulations are generally aimed at determining the forces with which particles interact with each other, since their values are necessary to calculate the atomic coordinates and momenta at subsequent points in time. The resulting force depends on the interaction potential and the spatial configuration of the particles and is calculated for each point in time.

Many functions have been developed to describe the interaction between particles, for example, the Lennard–Jones pairwise potential [5], allowing to calculate the force of interaction between two atoms depending on the distance between them. This is often used to simulate two-dimensional structures such as graphene or transition metal dichalcogenides. However, pairwise potentials do not take into account the dependence of the binding force on the directions and positions of neighboring particles in space, which narrows the scope of their applicability. For this reason, multiparticle potentials have been developed, in particular, the Tersoff potential [6] and Stillinger–Weber potential [7]. These potentials adequately describe the properties of silicon single crystals at certain parameter values, so they are often used in calculations.

An important aspect to be taken into account in MD simulations is the limited size of the system under consideration, which means it is nearly impossible to model the behavior of macroscopic objects by this approach. For example, beam diameter in experiments on bombardment of the surface with fullerene ions [8] ranged from 0.1 to 5 mm, while the sizes of the irradiated samples were even larger. The exposure times range from tens of seconds to several hours. However, simulating the motion of such a large number of atoms requires too much computing power, so only models with tens or hundreds of thousands of atoms have to be used. Periodic boundary conditions are imposed on the sides of the target, while thermostatic conditions are additionally imposed on the motion of external atomic layers in the crystal to match the model to the real sample, allowing to regulate heat flows, further reducing the computational time.

Earlier studies [9] found that a crater appears in the region around the cluster ion's impact point, with a small cluster of atoms forming above the surface along the crater's edge, known as its rim. In addition, was found that C_{60} ions impinging on the surface of a single silicon crystal completely destroy the cluster structure if their initial energy exceeds 1 keV [10]. Carbon atoms penetrate deep into the target and are distributed in a certain way, while some of the particles escape from the surface. Such particles are called sputtered and can consist of both single atoms and their agglomerates, i.e., clusters.

As an accelerated ion impinges on the surface, its energy is transferred to the target atoms, while part of it (especially in the case of molecules or clusters consisting of dozens or more atoms) can propagate deep into the crystal as a shock wave. As noted above, the dimensions of the cell used in calculations significantly affect the computing power and time. On the other hand, propagation of a shock wave in a small crystal model and the dissipation of the energy brought by an ion in this model may differ significantly from those in a real target. Accordingly, the simulation will give incorrect values for the effects obtained, such as sputtering, crater formation, etc.

This paper compares the single cases of C_{60} fullerene ions with energies of 8 and 14 keV impinging on the surface of silicon single crystals with different sizes, analyzing the influence of these sizes on the results.

Recommendations were formulated based on the data obtained for choosing the optimal dimensions of the computational cell for modeling the interaction of accelerated fullerene molecules with a silicon single crystal.

Description of the model

The open-source Lammps package was used to run the MD simulations [11]. Pairwise interaction of all types of atoms was described by the Tersoff potential [6], smoothly splined to the ZBL potential to describe the interaction of high-energy particles [12]. The initial system consisted of a silicon crystal with the (100) open surface, with a C_{60} fullerene molecule located at some height above it.

Periodic boundary conditions were imposed in the lateral directions; the three lower atomic layers were fixed. A Berendsen thermostat [13] consisting of a layer of silicon atoms with a thickness of one unit cell was used on the sides and bottom of the crystal.

The target temperature was approximately 0 K. The energy losses of fast particles due to interaction with target electrons (electron losses) were taken into account as a quasi-friction force applied to particles with energies over 10 eV. Several configurations of the lateral dimensions of the target (in nm) were considered:

$$5.4 \times 5.4;\ 8.0 \times 8.0;\ 11.0 \times 11.0;\ 22 \times 22;\ 33 \times 33;\ 44 \times 44,$$

which adequately corresponds to the values (in the unit cell lengths)

$$10 \times 10,\ 15 \times 15,\ 20 \times 20,\ 40 \times 40,\ 60 \times 60,\ 80 \times 80.$$

The target thicknesses were taken to be 11, 17 and 34 nm, i.e.,20, 31 and 63 in edge lengths of the unit cell; the edge length of the silicon unit cell is 0.543 nm.

All atoms of the C_{60} molecule were given the same velocity in the normal direction to the target surface at the initial time instant. The incident energy of fullerene amounted to 8 and 14 keV, and the initial temperature of the silicon crystal was 0 K. A variable time step was used to increase the computational accuracy at the initial stage as well as to speed up calculations after the particle energy decreased sufficiently.

The highest value of the time step was 1 fs. The total simulation time was selected depending on the initial fullerene energy: specifically, it was 5 and 10 ps for 8 and 14 keV, respectively. After the simulation was completed, the structures formed on the surface and the parameters of the sputtered particles were analyzed by the techniques described in [14]. Next, the system was restored to its initial state: the fullerene molecule was moved to a small random distance in the lateral directions (within 2 × 2 unit cells) and the calculation of a new trajectory began. For each combination of crystal size and C_{60} ion energy, 50 independent trajectories were calculated to reduce the statistical spread of the results.

Results and discussion

As mentioned in the introduction, MD simulations are performed for crystals with small sizes and periodic boundary conditions. Adding a thermostat along the boundary layers provides dissipation of excess energy brought to the target by an accelerated ion. At the same time, if the crystal size is too small, various computational artifacts may appear (i.e., results that do not reflect the real processes), therefore, it is necessary to avoid such phenomena as much as possible. We performed a series of calculations for molecules with energies of 8 and 14 keV impinging on silicon crystals of various sizes.

Fig. 1 shows how the appearance of a 1 nm thick cross section changes in the region where 8 keV fullerene impinges on it with an increase in the lateral size of the target. Here, the depth of the model crystal was 17 nm in all cases (31 is the length of the silicon unit cell). Evidently, the shape of the crater changes with increasing size in the lateral directions: it becomes wider and more spherical. Furthermore, it is highly likely that the volume of the amorphized target region changes to some extent.

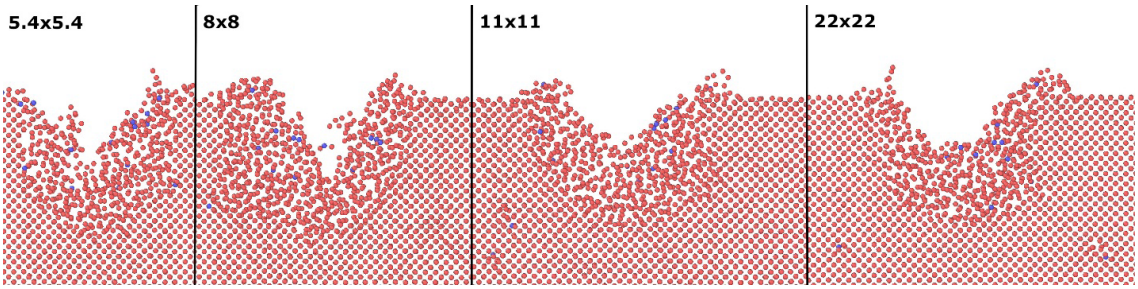

Fig. 1. Cross sections of a region of 10 Å thick silicon target after impact from 8 keV C_{60} fullerene ion with an energy (4 cases of lateral dimensions of the target (in nm))

Let us consider the changes in the shape of the crater in more detail, using the techniques for determining its volume, depth and opening area proposed in [14]. Fig. 2,*a* shows the dependence of the volume of the formed crater on the lateral dimensions of the model crystal at a constant thickness of 17 nm. It is clear that the volume of the crater increased by about 2 times at both energies, with an increase in the model size from 5.43 to 21.70 nm (from 10 to 40 unit cells). This effect turns out to be even more pronounced if the crater's opening area is increased, determined

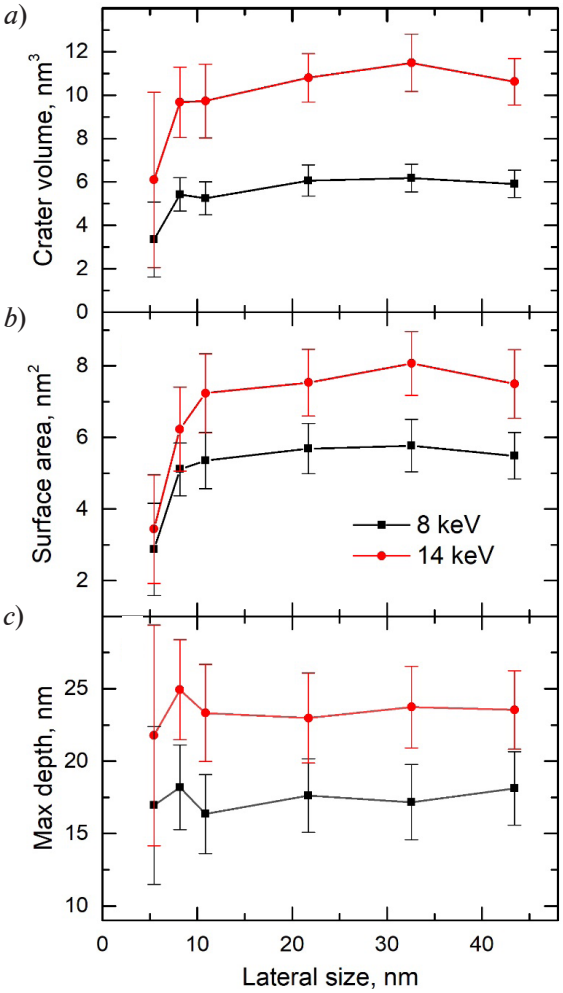

Fig. 2. Dependences of volume (*a*), surface area (*b*) and maximum depth (*c*) of the crater formed in a silicon target with a depth of 17 nm on its lateral dimensions
The results are given for the C_{60} ion energies of 8 and 14 keV

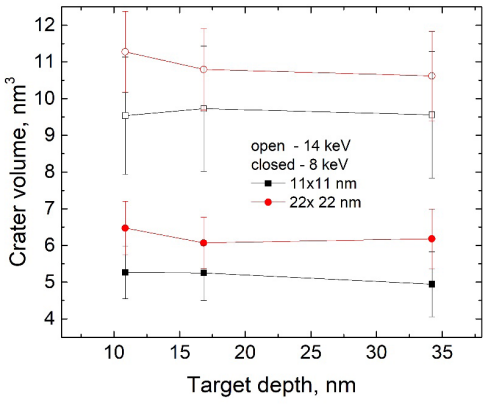

Fig. 3. Calculated dependences of mean volume of crater formed by impinging 8 and 14 keV fullerene ions on the depth of the model target with lateral dimensions of the crystal equal to 11 × 11 and 22 × 22 nm

at the level of the initial surface of the target [14]. The area varies from 3.0 to 5.5 nm^2 in the case of bombardment with fullerene at an energy of 8 keV, and from 3.5 to 8.0 nm^2 in the case of bombardment with fullerene at an energy of 14 keV. On the other hand, increasing the lateral dimensions of the model to 20 × 20 nm or more practically does not affect the formation of the crater (see Fig. 2). We should note (see Fig. 2,*c*) that the maximum depth of the crater practically does not depend on the lateral dimensions of the target used.

In addition to the lateral dimensions, the thickness of the target layer considered can also play a major role. To clarify this issue, we performed a series of calculations with different model depths given that the lateral dimensions of the crystal were equal to 11 × 11 and 22 × 22 nm. The obtained volumes of the crater are shown in Fig. 3. Apparently, a decrease in depth from 17 to 11 nm leads to some changes in the values of the obtained volumes, while its increase to 34 nm does not affect the results in any way. The opening area of the crater and its depth also practically do not depend on the thickness of the calculated model in the considered range. Thus, the lateral dimensions of the model used in the simulation of impinging fullerenes at energies of 8–14 keV play a more significant role in the formation of possible computational artifacts than its thickness, if the latter exceeds 10–15 nm.

As mentioned above, as a fullerene ion impinges on the target, some of the silicon atoms gain kinetic energy sufficient to overcome the forces of interatomic attraction and escape from the surface as sputtered particles. It is obvious that the dimensions of the calculated model can strongly influence the sputtering characteristics obtained as a result of the simulated effect.

Fig. 4,*a* shows the obtained dependences of the total number of sputtered atoms and the number of backscattered carbon atoms on the lateral dimensions of the target. Evidently, increasing them from 5 × 5 to 11 × 11 nm in the lateral dimensions leads to a decrease in the total number of escaping particles by about 2 times. The number of backscattered carbon atoms changes even more noticeably. Indeed, the calculation with the smallest cell yields 12–13 sputtered d carbon atoms and depends on the energy of incident fullerene. This number decreases to 6–7 for a cell with a size of 11 × 11 nm. A further increase in the lateral dimensions of the crystal does not affect the resulting sputtering.

Analyzing the angular distributions of the sputtered particles, we can observe a similar

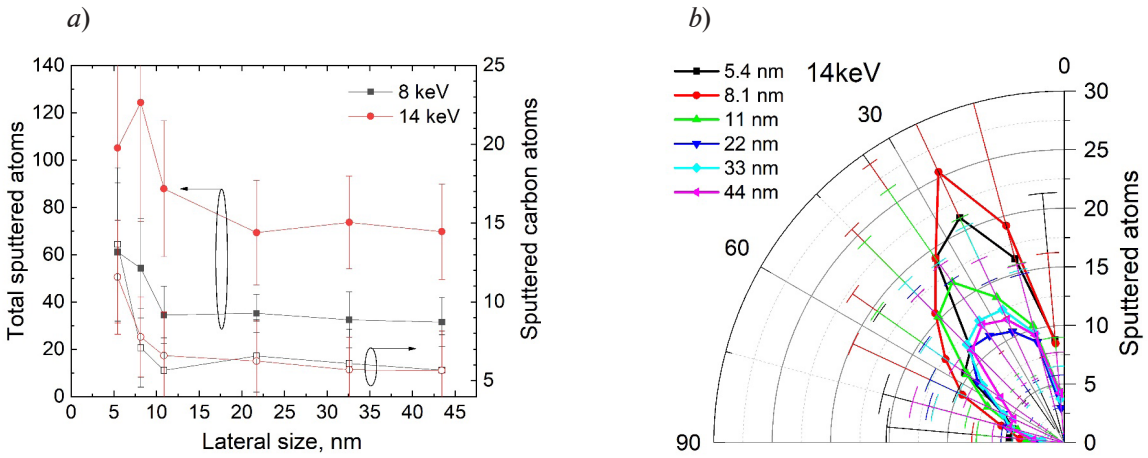

Fig. 4. Dependence of the total number of sputtered particles and backscattered carbon atoms on the lateral dimensions of the target (*a*); angular distributions of sputtered atoms for different lateral dimensions of the crystal with a depth of 17 nm

Cases of impact from C_{60} fullerene ions at energies of 8 and 14 keV (*a*) as well as 14 keV (*b*) are shown

trend (Fig. 4,*b*). The distribution has a pronounced maximum in the direction of 25° from the normal for a crystal with a small surface area. As the size increases, the distribution becomes more symmetrical, and the maximum shifts towards 35–40°. A further increase in the size of the cell in the range of 22–44 nm does not lead to a change in the resulting distribution.

Notably, the magnitude of the statistical spread in the obtained values of both total and differential sputtering yields is significantly reduced for targets of 22 × 22 nm and larger. The variation in the depth of the model crystal practically does not affect either the absolute number of sputtered particles or their distributions.

To understand the causes behind the discovered computational artifacts, we analyzed the propagation of shock waves occurring in the target. Fig. 5 shows the characteristic patterns obtained by MD simulation, which can occur during interaction in models with a depth of 17 nm at different lateral dimensions. The colors of the atoms (small circles) correspond to different kinetic energies.

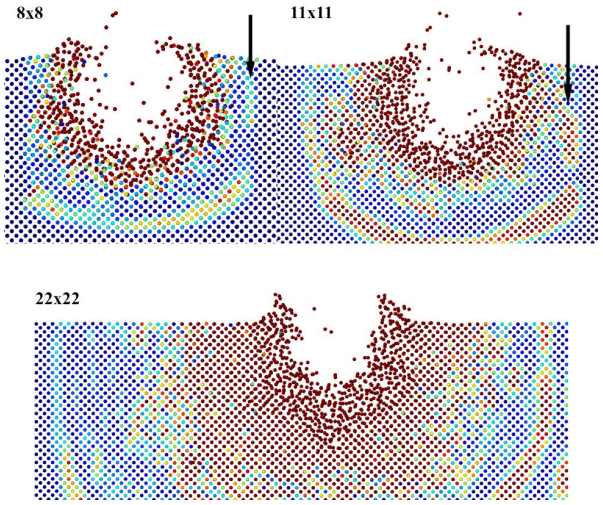

Fig. 5. Characteristic views of cross section (thickness 10 E) of model with the depth of 17 nm, evolving in the interaction of the fullerene ion with the target at three different lateral dimensions of the target (in nm)

The colors of the atoms (small circles) correspond to different kinetic energies (see Table); the arrows indicate the segments of waves passing through the periodic boundary and propagating to the region where the crater forms

Table

Parameters of target atoms (see Fig. 5)

Lateral size of target, nm	Kinetic energy of silicon atoms, eV	Time period, ps
8 × 8 × 17	0–0.20	0.4
11 × 11 × 17	0–0.05	0.5
22 × 22 × 17	0–0.01	1.1

Notes. 1. The time intervals between the instants when the C_{60} ion impinges on the surface and the patterns shown in Fig. 5 are recorded.
2. The C_{60} ion had an energy of 8 keV.

The arrows in the two upper patterns (see Fig. 5, models with sizes of 8 × 8 and 11 ×11 nm) show the segments of waves passing through the periodic boundary and propagating towards the region where the crater formed rather than away from it. Such a phenomenon could not be detected in the case of simulation with a 22 × 22 nm cell: all visible waves propagated from the impact point and attenuated near the thermostat.

The table shows the time instants at which the images were taken and the energy ranges used to conveniently describe the wave processes in different cases. We should note that the energies of silicon atoms reach 0.20 and 0.05 eV in cells with the dimensions of 8 × 8 and 11 × 11 nm, respectively, i.e., the calculated temperatures near the crater exceed 900 K. All atoms have an energy of less than 0.01 eV in a 22 × 22 nm cell. Thus, the energy transferred by the fullerene ion does not have time to dissipate into the target volume at small cell sizes and returns to the impact area through periodic boundaries; this leads to increased relaxation of produced defects and a decrease in the size of the crater by maintaining an elevated temperature for a longer time than in the case of larger model crystals. The effect of the return shock wave may also be associated with increased sputtering, including in directions closer to the normal. In the case of large cells, the scattering and dissipation of the energy transferred between atoms are more uniform and better correspond to the picture observed experimentally.

Conclusion

The paper presents a molecular dynamics simulation of the interaction between the C_{60} fullerene ions, with energies of 8 and 14 keV, and the surface of a silicon single crystal. In particular, we considered the influence of the size of this single crystal on the results obtained.

We found that various computational artifacts arise at an energy of 8 keV and lateral dimensions of less than 11 × 11 nm, both as a crater evolves in the target and during sputtering of particles caused by its bombardment. Varying the depth of the model crystal in the range from 11 to 33 nm had practically no effect on the results obtained. The cell dimensions of 11 × 11 nm turned out to be insufficient at the incident energy of 14 keV, since slightly overestimated sputtering yields are produced in this case, although no artifacts were observed during the formation of the crater.

It was established that the causes of the artifacts were, firstly, the return of energy transferred by fullerene through periodic boundaries to the region where the crater formed, and secondly, somewhat incorrectly calculated distribution of the energy transferred between the target atoms in the case of small models.

Analyzing our findings, we can conclude that a model with the dimensions of 11 × 11 nm is preferable for simulating the interaction of C_{60} ions at a kinetic energy of less than 8 keV with the target; it is recommended to use cells with larger lateral dimensions for the case of 14 keV.

REFERENCES

1. **Arnittali M., Rissanou A. N., Amprazi M., et al.,** Structure and thermal stability of wtRop and RM6 proteins through all-atom molecular dynamics simulations and experiments, Int. J. Mol. Sci. 22 (11) (2021) 5931.

2. **Ullah M. W., Kuronen A., Nordlund K., et al.,** Atomistic simulation of damage production by atomic and molecular ion irradiation in GaN, J. Appl. Phys. 112 (4) (2012) 043517.

3. **Maciążek D., Kański M., Postawa Z.,** Intuitive model of surface modification induced by cluster ion beams, Anal. Chem. 92 (10) (2020) 7349–7353.

4. **Aoki T., Matsuo J.,** Molecular dynamics study of surface modification with a glancing angle gas cluster ion beam, Nucl. Instrum. Methods Phys. Res. B. 255 (1) (2007) 265–268.

5. **Lennard-Jones J. E.,** On the determination of molecular fields. – II. From the equation of state of a gas, Proc. R. Soc. Lond. 106 (738) (1924) 463–477.

6. **Tersoff J.,** New empirical approach for the structure and energy of covalent systems, Phys. Rev. B. 37 (12) (1988) 6991–7000.

7. **Stillinger F. H., Weber T. A.,** Computer simulation of local order in condensed phases of silicon, Phys. Rev. B. 31 (8) (1985) 5262–5271.

8. **Khadem M., Pukha V., Penkov O., et al.,** Formation of wear-resistant graphite/diamond-like carbon nanocomposite coatings on Ti using accelerated C_{60}-ions, Surf. Coat. Technol. 424 (25 Oct) (2021) 127670.

9. **Aoki T., Seki T., Matsuo J.,** Molecular dynamics simulations for gas cluster ion beam processes, Vacuum. 84 (8) (2010) 994–998.

10. **Skripov I. N., Karasev K. P., Strizhkin D. A., Karasev P. A.,** Issledovaniye poverkhnostnykh yavleniy pri padenii uskorennogo iona C_{60} na monokristall kremniya [Study of surface phenomena in the impact of an accelerated C_{60} ion on a Si monocrystal], In book of abstracts: Proceedings of All-Russian Conference "Scientific week at Institute of Electronics and Telecommunications of St. Petersburg Polytechnical University", Nov. 15–19 2021, SPb. (2021) 55–57 (in Russian).

11. **Thompson A. P., Aktulga H. M., Berger R., et al.,** LAMMPS – a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, Comp. Phys. Commun. 271 (Febr) (2022) 108171.

12. **Ziegler J. F., Biersack J. P., Littmark U.,** The stopping and range of ions in matter (Book series: Stopping and range of ions in matter. Vol. 1), Pergamon, New York, 1985.

13. **Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., et al.,** Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (8) (1984) 3684–3690.

14. **Karasev K., Strizhkin D., Karaseov P.,** The way to analyze MD simulation results of cluster ion bombardment, IEEE Xplore Proc. 2023 Int. Conf. Electric. Eng. Photon. (EExPolytech-2023), Oct. 19–20 2023, St. Petersburg, SPbPU, (2023) 282–284.

СПИСОК ЛИТЕРАТУРЫ

1. **Arnittali M., Rissanou A. N., Amprazi M., Kokkinidis M., Harmandaris V.** Structure and thermal stability of wtRop and RM6 proteins through all-atom molecular dynamics simulations and experiments // International Journal of Molecular Sciences. 2021. Vol. 22. No. 11. P. 5931.

2. **Ullah M. W., Kuronen A., Nordlund K., Djurabekova F., Karaseov P. A., Titov A. I.** Atomistic simulation of damage production by atomic and molecular ion irradiation in GaN // Journal of Applied Physics. 2012. Vol. 112. No. 4. P. 043517.

3. **Maciążek D., Kański M., Postawa Z.** Intuitive model of surface modification induced by cluster ion beams // Analytical Chemistry. 2020. Vol. 92. No. 10. Pp. 7349–7353.

4. **Aoki T., Matsuo J.** Molecular dynamics study of surface modification with a glancing angle gas cluster ion beam // Nuclear Instruments and Methods in Physics Research. B. 2007. Vol. 255. No. 1. Pp. 265–268.

5. **Lennard-Jones J. E.** On the determination of molecular fields. II. From the equation of state of a gas // Proceedings of the Royal Society A, London. 1924. Vol. 106. No. 738. Pp. 463–477.

6. **Tersoff J.** New empirical approach for the structure and energy of covalent systems // Physical Review B. 1988. Vol. 37. No. 12. Pp. 6991–7000.

7. **Stillinger F. H., Weber T. A.** Computer simulation of local order in condensed phases of silicon // Physical Review B. 1985. Vol. 31. No. 8. Pp. 5262–5271.

8. **Khadem M., Pukha V., Penkov O., Khodos I., Belmesov A., Nechaev G., Kabachkov E., Karaseov P., Kim D.-E.** Formation of wear-resistant graphite/diamond-like carbon nanocomposite coatings on Ti using accelerated C_{60}-ions // Surface and Coatings Technology. 2021. Vol. 424. 25 October. P. 127670.

9. **Aoki T., Seki T., Matsuo J.** Molecular dynamics simulations for gas cluster ion beam processes // Vacuum. 2010. Vol. 84. No. 8. Pp. 994–998.

10. **Скрипов И. Н., Карасев К. П., Стрижкин Д. А., Карасев П. А.** Исследование поверхностных явлений при падении ускоренного иона C_{60} на монокристалл кремния // Неделя науки ИЭиТ (Институт электроники и телекоммуникаций). Материалы Всероссийской конференции. СПб., СПбПУ. 15–19 ноября 2021. С. 55–*57*.

11. **Thompson A. P., Aktulga H. M., Berger R., et al.** LAMMPS – a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales // Computer Physics Communications. 2022. Vol. 271. February. P. 108171.

12. **Ziegler J. F., Biersack J. P., Littmark U.** The stopping and range of ions in matter (Book series: Stopping and range of ions in matter. Vol. 1). New York: Pergamon, 1985. 321 p.

13. **Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., DiNola A., Haak J. R.** Molecular dynamics with coupling to an external bath // The Journal of Chemical Physics. 1984. Vol. 81. No. 8. Pp. 3684–3690.

14. **Karasev K., Strizhkin D., Karaseov P.** The way to analyze MD simulation results of cluster ion bombardment // IEEE Xplore Proceedings of the 2023 International Conference on Electrical Engineering and Photonics (EExPolytech-2023). St. Petersburg, SPbPU, October 19–20, 2023. Pp. 282–284.

THE AUTHORS

KARASEV Kirill P.
Peter the Great St. Petersburg Polytechnic University,
Alferov University of RAS
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
kir.karasyov2017@yandex.ru
ORCID: 0000-0002-0969-0162

STRIZHKIN Denis A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
strdenis02@gmail.com
ORCID: 0009-0003-1062-8360

KARASEOV Platon A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
platon.karaseov@spbstu.ru
ORCID: 0000-0003-2511-0188

TITOV Andrei I.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
andrei.titov@rphf.spbstu.ru
ORCID: 0000-0003-4933-9534

СВЕДЕНИЯ ОБ АВТОРАХ

КАРАСЕВ Кирилл Платонович – *инженер-исследователь Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого, студент Академического университета им. Ж. И. Алфёрова РАН.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
kir.karasyov2017@yandex.ru
ORCID: 0000-0002-0969-0162

СТРИЖКИН Денис Александрович – *лаборант-исследователь Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
strdenis02@gmail.com
ORCID: 0009-0003-1062-8360

КАРАСЕВ Платон Александрович – *доктор физико-математических наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
platon.karaseov@spbstu.ru
ORCID: 0000-0003-2511-0188

ТИТОВ Андрей Иванович – *доктор физико-математических наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
andrei.titov@rphf.spbstu.ru
ORCID: 0000-0003-4933-9534

Received 01.12.2023. Approved after reviewing 04.12.2023. Accepted 04.12.2023.
Статья поступила в редакцию 01.12.2023. Одобрена после рецензирования 04.12.2023. Принята 04.12.2023.

MATHEMATICAL PHYSICS

Original article
DOI: https://doi.org/10.18721/JPM.16408

THE FOURIER ANALYSIS IN INHOMOGENEOUS MEDIA

V. G. Gnevyshev[1], *T. V. Belonenko*[2]✉

[1] Shirshov Institute of Oceanology, RAS, Moscow, Russia;

[2] St. Petersburg State University, St. Petersburg, Russia

✉ btvlisab@yandex.ru

Abstract. In the paper, the definition and basic properties of the Fourier transform (FT) are discussed. It has been shown with specific examples that integral solutions of the model inhomogeneous equation, the nonstationary Cauchy problem on an inhomogeneous shear flow, and the boundary value problem on the transformation of internal waves in the vicinity of the focus in the inhomogeneous medium can be found by FT and using its properties. The constructed Fourier integrals refuted the widely held claim that the Fourier analysis is unusable for the study of inhomogeneous media.

Keywords: Fourier analysis, Laplace transform, Cauchy problem, waves, inhomogeneous medium

Funding: The reported study was carried out within the framework of the of the State Assignment No. 0128-2021-0003 of Shirshov Institute of Oceanology, RAS, and was funded by St. Petersburg State University (Grant No. 93016972).

Citation: Gnevyshev V. G., Belonenko T. V., The Fourier analysis in inhomogeneous media, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 86–100. DOI: https://doi.org/10.18721/JPM.16408

Научная статья
УДК 532.5: 532.591: 532.592.2
DOI: https://doi.org/10.18721/JPM.16408

ФУРЬЕ-АНАЛИЗ В НЕОДНОРОДНЫХ СРЕДАХ

В. Г. Гневышев[1], *Т. В. Белоненко*[2]✉

[1] Институт океанологии им. П. П. Ширшова, РАН, Москва, Россия;

[2] Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

✉ btvlisab@yandex.ru

Аннотация. В работе обсуждаются определение и основные свойства преобразования Фурье. На конкретных примерах показано, что с его помощью, а также через использование его свойств можно найти интегральные решения модельного неоднородного уравнения, нестационарной задачи Коши на неоднородном сдвиговом потоке и краевой задачи о трансформации внутренних волн в окрестности фокуса в неоднородной среде. Построенные интегралы Фурье опровергают широко распространенное утверждение, что Фурье-анализ непригоден для исследования неоднородных сред.

Ключевые слова: Фурье-анализ, преобразование Лапласа, задача Коши, волны, неоднородная среда

Финансирование: Исследование выполнено в рамках государственного задания Института океанологии РАН № 0003-2021-0128 и при финансовой поддержке гранта Санкт-Петербургского государственного университета № 94033410.

Ссылка для цитирования: Гневышев В. Г., Белоненко Т. В. Фурье-анализ в неоднородных средах // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 4 № .16. С. 86–100. DOI: https://doi.org/10.18721/ JPM.16408

Introduction

While the Fourier transformation is not the only method for solving differential equations, it is one of the most effective approaches to solving boundary-value problems, when variables are separated in a multidimensional problem. However, there is a widespread misconception about the inapplicability of Fourier analysis for inhomogeneous media. For example, the following is stated in the monograph by Whitham [1, p. 365]:

"For an inhomogeneous medium, or for nonlinear problems where the Fourier transform is not applicable...".

A similar statement is found in the book by Lighthill [2, p. 425]:

"... the need to use Fourier decomposition limits us to *homogeneous* [italicized by the author] systems usually described by equations with *constant* coefficients, so that each Fourier component (a sine wave of constant amplitude) individually can be a solution to the equations of motion."

The authors of these and many other monographs (see, for example, [3, 4]) believe that Fourier analysis can be used only in cases where the coefficients of the differential equation are constant and, conversely, it cannot be applied if these coefficients are not constant.

In this paper, we prove that Fourier analysis can be applied in problems containing differential equations with variable coefficients. Moreover, the problem can be two-dimensional and with inseparable variables, but Fourier analysis is still applicable.

Thus, the goal of this study is to expand the boundaries of the field of applicability of Fourier analysis, extending its approaches to problems in inhomogeneous media.

Definition and basic properties of the Fourier transform

The Fourier transform is defined as follows.

Forward transform:

$$\varphi(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-ikx)\Phi(x)\,dx; \tag{1}$$

inverse transform:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(+ikx)\varphi(k)\,dk. \tag{2}$$

The properties of the Fourier transform can be found, for example, in monograph [1]. They are derived by differentiating with respect to the parameter or by integration by parts (see, for example, [5]). In this case, it is assumed that the function $\Phi(x)$ decreases at infinity faster than any degree of $|x|^{-1}$. Let us briefly list the properties of the transform that we will consider below.

Fourier transform of derivative function. We derive this property by integration of the formula of the forward Fourier transform (1) by parts:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-ikx) \frac{\partial \Phi}{\partial x} dx = \frac{ik}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-ikx)\,\Phi(x)\,dx = ik\varphi(k). \tag{3}$$

Note that the same result can be obtained by differentiating with respect to x as a parameter using Eq. (2) of the inverse Fourier transform:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(-ikx\right)\frac{\partial^2\Phi}{\partial x^2}dx=\frac{\left(ik\right)^2}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(-ikx\right)\Phi\left(x\right)dx=-k^2\varphi\left(k\right). \tag{4}$$

For brevity, this property of the Fourier transform can be written as follows:

$$\Phi\to\varphi,\ \Phi_x\to ik\varphi,\ \Phi_{xx}\to -k^2\varphi. \tag{5}$$

Properties (5) are often used in problems with constant coefficients for homogeneous media.

Fourier transform of function with a linear multiplier. To represent this property, let us first integrate the transformation (2) by parts:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(+ikx\right)\frac{\partial\varphi}{\partial k}dk=\left(-ix\right)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(+ikx\right)\varphi\left(k\right)dk=\left(-ix\right)\Phi\left(x\right). \tag{6}$$

An identical expression can be obtained by differentiating by the parameter k of relation (1). We multiply both parts of expression (6) by an imaginary unit and write it as follows:

$$x\Phi\to i\varphi_k. \tag{7}$$

Fourier transform of second derivative with a linear multiplier. We show this property by integration by parts and parametric differentiation of relation (2), repeated twice; then we obtain the following formula:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(+ikx\right)\frac{\partial\left(k^2\varphi\right)}{\partial k}dk=\left(-ix\right)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left(+ikx\right)k^2\varphi\left(k\right)dk=\left(ix\right)\frac{\partial^2\Phi}{\partial x^2}, \tag{8}$$

which we rewrite as

$$x\,\Phi_{xx}\to -i\left(k^2\varphi\right)_k. \tag{9}$$

One-dimensional reference equations

Finding a solution using the Fourier transform is divided into two stages. At the first stage, we construct a formal solution of the differential equation in Fourier space using operator analysis (see the previous section). At the second stage, we solve the question of the conditions under which this formally constructed solution converges. We define the integration path in the complex space and find asymptotic expressions in each sector [6, 7].

Example 1. Let us consider an inhomogeneous differential equation that arises in the analysis of wave processes in inhomogeneous plasma as well as in the study of instability in the Orr–Sommerfeld problem [6], [8, equation (1.28)]:

$$\Phi_{yyyy}+\lambda^2\left[y\,\Phi_{yy}+\gamma\,\Phi\right]=0. \tag{10}$$

Let us construct a formal solution for this example. The transform of equation (10) in Fourier space (denoted as l) has the form

$$l^4\varphi+\lambda^2\left[-i\left(l^2\varphi\right)_l+\gamma\,\varphi\right]=0. \tag{11}$$

Let us rewrite Eq. (11) in the following form:

$$\frac{1}{\lambda^2}l^2P-i\,P_l+\frac{\gamma}{l^2}P=0,\ P=l^2\varphi. \tag{12}$$

It is a homogeneous differential equation of the first order (such equations are called quadrature in the mathematical literature (see, for example, [9]).

Integrating equation (11), we obtain the following expression:

$$\varphi = \frac{1}{l^2}\exp\left(-i\frac{1}{3\lambda^2}l^3 + i\frac{\gamma}{l}\right). \tag{13}$$

The inverse Fourier transform gives the solution:

$$\varphi = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty}\frac{1}{l^2}\exp\left(-i\frac{1}{3\lambda^2}l^3 + i\frac{\gamma}{l} + ily\right)dl. \tag{14}$$

Next, we should perform a change of the variable $t = il$ converting the Fourier integral into the Laplace integral. It is important to note here that many consider the Laplace transform a special case of the Fourier transform (see, for example, [5, Eqs. (1.4.1), (1.4.2)]).

For convergence of integrals, we use Cauchy's theorem on the analytical function and replace the integration limits by some paths in the complex plane. We do not consider the specifics of bypassing the pole and the choice of sectors that the integration path crosses in this study, since these issues are discussed in detail in many monographs, in particular in [12], where the author relies on the Laplace transform, unlike the discussion in [5].

Thus, we obtain the following integral, which is commonly called the Laplace integral in the literature:

$$\varphi = \frac{1}{2\pi i}\int_C\frac{1}{t^2}\exp\left(\frac{1}{3\lambda^2}t^3 - \frac{\gamma}{t} + ty\right)dt. \tag{15}$$

Monograph [6] gives this integral without calculations. In our work, we show how to independently derive this formal solution with the help of the Fourier rather than Laplace transform.

Example 2. Let us consider an equation describing topographic waves on an inhomogeneous continental shelf on the *f*-plane. Monograph [19] claims that the solution of this inhomogeneous equation can be constructed using the Laplace transform, and the reader is invited to do this independently.

We in turn construct a formal solution using the Fourier transform and prove the identity of the approaches to both Fourier and Laplace transformations. Consider the equation

$$x F_{xx} + F_x + \left[\mu k - k^2 x\right]F = 0. \tag{16}$$

We introduce a dimensionless variable $\chi = kx > 0$. Eq. (16) then takes the form

$$\chi F_{\chi\chi} + F_\chi + \left[\mu k - \chi\right]F = 0. \tag{17}$$

The image of equation (17) in Fourier space with respect to the variable χ (we denote it as l) has the form

$$-i\left(l^2\Phi\right)_l + i\,l\,\Phi + \mu\,\Phi - i\Phi_l = 0. \tag{18}$$

Let us introduce a new variable $s = il$. Eq. (18) then takes the form

$$\frac{\Phi_s}{\Phi} = \frac{\mu - s}{s^2 - 1} = \frac{\mu - 1}{2(s-1)} - \frac{\mu + 1}{2(s+1)}. \tag{19}$$

Integrating equation (19) and performing the inverse Fourier transform, we obtain the following formal integral:

$$F(x) = \frac{1}{\sqrt{2\pi i}}\int_C\frac{(s-1)^{(\mu-1)/2}}{(s+1)^{(\mu+1)/2}}\exp(skx)\,ds. \tag{20}$$

The integral constructed here coincides with Eq. (25.20) in monograph [19] up to a multiplier; the monograph also presents analysis of integral (20) with the choice of integration paths.

Results obtained in the section "One-dimensional reference equations". Following the approach outlined in monograph [5], we can change the variable, and then the properties of the Fourier transform are transferred to the properties of the Laplace transform. Thus, according to the statement made in [5], the Fourier transform is a kind of basic transformation, from which other transformations, for example, Laplace and Mellin types, follow. A similar approach is followed by the authors of monograph [9], who constructed fundamental solutions to the thermal conductivity operator, the Laplace and Helmholtz operators, as well as the wave operator in terms of the Fourier transform.

Thus, there is no fundamental difference between Fourier analysis and the Laplace transform in one-dimensional inhomogeneous media. It can be assumed that the solution is constructed in terms of the Laplace transform, but it can also be argued that the solution is derived in terms of the Fourier transform and Cauchy's theorem. Following [5] from now on, we adhere to the second approach.

Unsteady Cauchy problem for Rossby waves in zonal current

As the first example of a two-dimensional Fourier transform in inhomogeneous media, consider the unsteady Cauchy problem for Rossby waves. Yamagata solved this problem in 1976 using convective coordinates [20]. Convective coordinates are a common way for an operator of the type $\partial t + U(y)\partial x$ for the case of a linear velocity profile $U(y) = U_y y$, convective coordinates transform an inhomogeneous differential equation into a homogeneous one, and then the Fourier transform is applied over spatial convective coordinates (two-dimensional transform). Next, an unsteady differential equation with respect to t is obtained.

It is fundamentally important that there is no point in doing the Laplace transform with respect to the variable t, as is customary in some mathematical groups (see, for example, [15]). It is easier to solve the differential equation with respect to t explicitly than to perform an additional transformation. By solving the differential equation with respect to t and taking the inverse Fourier transform in convective coordinates, we can convert convective variables to ordinary ones and get a solution in the form of a two-dimensional Fourier transform.

Next, we can find a solution to an inhomogeneous differential equation and, repeating the calculations, obtain the Yamagata solution, but we propose a different approach. We will not adopt the convective coordinates to eliminate the inhomogeneity of the differential equation to then perform the Fourier transform. We will immediately apply the Fourier transform for an inhomogeneous differential equation using its properties. Thus, on the one hand, we will significantly reduce the number of operations, and on the other hand, we will arrive at a known result and confirm the correctness of mathematical calculations. We will demonstrate this approach for the problem solved above, but we will solve it by a new, shorter technique, which allows to immediately find the Fourier transform of an inhomogeneous differential equation.

Example 3. The linear Cauchy problem for non-divergent barotropic Rossby waves in zonal shear flow is considered in [20], and its generalization to the case of divergent waves is considered in [18]:

$$\left(\partial_t + U_y y \partial_x\right)\left[\Psi_{xx} + \Psi_{yy}\right] + \beta\Psi_x = 0, \tag{21}$$

where Ψ is the function of current; β is the classical parameter, $\beta = \dfrac{df}{dy}$ ($f = 2\Omega\sin\varphi$, Ω is the angular velocity of the Earth's rotation, φ is the latitude); the x axis is directed to the east, the y axis is to the north.

Let there be inhomogeneous zonal shear flow $U(y) = U_y y$, where $U_y = \text{const}$. Let us perform a two-dimensional Fourier transform for inhomogeneous differential equation (22) with respect to two spatial variables x and y (without adopting the convective variables):

$$\Psi(x,y,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\varphi(k,l,t)\exp\left[+i\left(kx+ly\right)\right]dk\ dl. \tag{22}$$

Then Eq. (21) takes the following form in Fourier space

$$\left[\left(-k^2-l^2\right)\varphi\right]_t - U_y k\left[\left(-k^2-l^2\right)\varphi\right]_l + i\beta k\varphi = 0, \tag{23}$$

where the subscripts denote partial derivatives.

This equation is homogeneous, contains only the first partial derivatives and is easily solved.

Let us rewrite Eq. (23) in the following form:

$$P_t - U_y k\, P_l - \frac{i\beta k}{k^2+l^2}P = 0,\ P \equiv \left(k^2+l^2\right)\varphi. \tag{24}$$

Performing a substitution of variables ($\tau = t$, $l' = l + kU_y t$), we obtain the following equation:

$$P_\tau - \frac{i\beta k P}{k^2+\left(l'-kU_y\,\tau\right)^2} = 0. \tag{25}$$

Eq. (25) can be integrated explicitly (the exponent of the arctangent), and then the final solution has the form of a double Fourier integral:

$$\begin{aligned}\Psi(x,y,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_1(k,l)\frac{k^2+l^2}{k^2-\left(l-kU_y t\right)^2}\times \\ \times\exp\left[-i\frac{\beta}{kU_y}\left\{\arctan\left(\frac{l}{k}\right)-\arctan\left(\frac{l}{k}-U_y t\right)\right\}\right]\times \\ \times\exp\left[+i\left(kx+\left(l-kU_y t\right)y\right)\right]dk\,dl,\end{aligned} \tag{26}$$

where the solution is normalized to the initial condition

$$g_1(k,l) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Psi(x,y,t=0)\exp\left[-i\left(kx+ly\right)\right]dx\,dy. \tag{27}$$

Analysis of the double Fourier integral by the stationary phase method and the construction of wave packet trajectories can be found in [20]. It is important to note that these studies do not rely on the assumption that the time variable must be large.

Unsteady Cauchy problem for Rossby waves in meridional current

As a second example, consider the unsteady Cauchy problem for Rossby waves in meridional current. A solution to this problem using convective coordinates can be found in [20].

Example 4. This is the case of a linear velocity profile of meridional current. The linear Cauchy problem for divergent barotropic Rossby waves has the following form [20]:

$$\left(\partial_t + V_x x\partial_y\right)\left[\Psi_{xx}+\Psi_{yy}\right]+\beta\Psi_x = 0, \tag{28}$$

where β is the classical parameter; the x axis is directed to the east, the y axis is directed to the north.

There is inhomogeneous meridional shear flow $V(x) = V_x x$, where $V_x = const$. As before (see Example 3), we perform a two-dimensional Fourier transform for inhomogeneous differential equation (28) with respect to two spatial variables x and y. Then Eq. (28) takes the following form in Fourier space

$$\left[\left(-k^2-l^2\right)\varphi\right]_t - U_x l\left[\left(-k^2-l^2\right)\varphi\right]_k + i\beta k\varphi = 0. \tag{29}$$

This equation is homogeneous, contains only the first partial derivatives and is easily solved. Let us rewrite it in the following form:

$$P_t - U_x l P_k - \frac{i\beta k}{k^2 + l^2} P = 0,\ P \equiv \left(k^2 + l^2\right)\varphi. \tag{30}$$

Performing a change of variables ($\tau = t$, $k' = k + lU_x t$), we obtain the following equation:

$$P_\tau - \frac{i\beta\left(k' - lU_x\tau\right)P}{\left(k' - lU_x\tau\right)^2 + l^2} = 0. \tag{31}$$

Eq. (31) can be integrated explicitly (the exponent of the logarithm), and then the final solution has the form of a double Fourier integral:

$$\Psi\left(x,y,t\right) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} g_2\left(k,l\right)\frac{k^2+l^2}{\left(k - lU_x t\right)^2 + l^2}\exp\left[i\frac{\beta}{2lU_x}\ln\left\{\frac{\left(k - lU_x t\right)^2 + l^2}{k^2 + l^2}\right\}\right]\times \tag{32}$$

$$\times\exp\left[+i\left(\left(k - lU_x t\right)x + ly\right)\right]dk\,dl,$$

where the solution is normalized to the initial condition

$$g_2\left(k,l\right) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\Psi\left(x,y,t=0\right)\exp\left[-i\left(kx + ly\right)\right]dx\,dy. \tag{33}$$

Analysis of the double Fourier integral by the stationary phase method and the construction of wave packet trajectories can be found in [20].

The considered examples are simple in the sense that the obtained double integrals are already known. The novelty of our solution lies in the fact that the solution is formally constructed by a direct Fourier transform of an inhomogeneous differential equation without involving convective coordinates.

Let us now move on to a more complex problem, where the solution in the form of the Fourier integral of the boundary-value problem was not previously known, but it was considered in terms of special functions with respect to complex variables. The very procedure of constructing a certain complex variable for a special hypergeometric function and integration along a certain circle in a complex space suggests that there must be a way to obtain this solution in terms of the direct Fourier transform of the original inhomogeneous differential equation with inseparable variables.

Reference equation for a two-dimensional inhomogeneous medium. Abnormal focusing of internal waves

In the examples discussed above, the solution was sought in the form of a two-dimensional Fourier integral, while the inhomogeneity of the external field (the velocity field of the background flow, or topography) was one-dimensional.

Now let us consider a more complex example of a problem with two-dimensional inhomogeneity of the external field.

Example 5. The theory of anomalous focusing of internal waves in a two-dimensional inhomogeneous fluid introduces the following reference equation of elliptical-hyperbolic type for vertical displacement in the vicinity of the focus [13, Eq. (2.5)]:

$$\Psi_{zz} + \left(\frac{y}{L_y} + \frac{z^2}{L_z^2}\right)\Psi_{yy} + \frac{2}{L_y}\Psi_y = 0, \tag{34}$$

where Ψ is the function of current; (x, y, z) is the rectangular coordinate system; L_y, L_z are the lengths of inhomogeneities along the y and z axes.

We will search for solutions localized in a small neighborhood of a certain level along the vertical coordinate and exponentially attenuating outside this level; here, for the case of internal waves, the following notations are introduced [13]:

$$\frac{1}{L_y} = 2\nabla_y \ln\Omega,\ \frac{1}{L_z^2} = \frac{\nabla_z^2\Omega}{\Omega} - \frac{\nabla_z^2 N}{N}, \tag{35}$$

where $\Omega = \omega - kU$ (ω is the frequency, k is the zonal wavenumber, $U(z,y)$ is the inhomogeneous horizontal background shear flow); $N^2(z) = -g\dfrac{d}{dz}\ln\rho_0(z)$ ($\rho_0(z)$ is the density).

The value of all derivatives is taken at the focal point. Since equation (34) is invariant with respect to the scale transformation $z = az'$, $y = a^2y'$, a certain self-similar variable is introduced. The solution is constructed as a summation of all partial solutions with respect to hypergeometric functions of a complex argument. The procedure for constructing this complex variable is not entirely clear. It is also not entirely clear what functions are considered in [13], how these functions appear, what their physical nature is, and what primary and secondary quantization mean in construction of asymptotic forms of the solution. Note that the asymptotic forms of the two-dimensional function are constructed as one-dimensional only on the waveguide axis.

To interpret these solutions in terms of special functions of complex arguments on the one hand and to represent these classes of solutions using the classical Fourier transform and its properties (presented above) on the other hand, we will independently construct a solution in integral form, find its two-dimensional asymptotic forms and show what primary and secondary quantization mean in terms of the classical Sturm–Liouville problem. To do this, we can use the well-known integral representations of the hypergeometric function as a basis, and then the approach to finding a solution will become more transparent. In a sense, we are using the integral representation as a starting point, but it is better to take steps in the opposite direction to search for the solution.

The solution to equation (34) is sought in the form of a Fourier integral. First we confine ourselves to the upper half of the integral:

$$\Psi(k,y,z,\omega) = \int_0^\infty G(k,l,z,\omega)\exp(ily)\,dl. \tag{36}$$

In fact, a reasonable question arises whether to take the whole integral or only the upper (or lower) part of it. We will further discuss this problem in the final section of the paper.

We use the properties of the Fourier transform again:

$$\Psi \to G,\ \Psi_y \to ilG,\ \Psi_{yy} \to -l^2G,\ y\Psi_{yy} \to -i\left(l^2G\right)_l. \tag{37}$$

The first three formulas in this equation are the properties of the Fourier transform of the derivative, which are widely known. The latter formula is a special case of equality (21) in monograph [9]. Despite the popularity of this formula, it is not used in applied problems. Our work focuses specifically on the practical application of this last formula from Eq. (37).

Substituting integral (36) into equation (34) and taking into account (37), we obtain the following equation for the Fourier transform G:

$$G_{zz} - \frac{l^2z^2}{L_z^2}G - i\frac{l^2}{L_y}G_l = 0. \tag{38}$$

Equality (38) is not an equation with separable variables. To convert it to such form, let us perform the following variable substitution

$$(z,l) \to (\eta,\varphi),$$

where

$$\eta = \frac{z\,l^{1/2}}{L_z^{1/2}},\ \varphi = l. \tag{39}$$

The Jacobian of such a substitution has the form

$$\frac{\partial(\eta,\varphi)}{\partial(z,l)} = l^{1/2}. \tag{40}$$

It should be noted that equations (39) and (40) contain $l^{1/2}$. Technically, this specific fact allows us to consider only one of the parts of the Fourier integral. For simplicity, we first chose the upper, positive part of the integration to resolve the question related to the square root.

This however raises the question of why such a variable substitution should be chosen. The answer is contained in [17], where the solution is constructed in the WKB approximation. In fact, any reasoning about self-similarity turns out to be superfluous, since in a certain sense the entire self-similarity of the solution is reduced to a simple substitution of variables of the form (39).

Equality (38) in terms of new variables (η,φ), takes the form of an equation with separable variables:

$$G_{\eta\eta} - \eta^2 G - i\frac{\eta L_z}{2L_y} G_\eta - i\frac{\varphi L_z}{L_y} G_\varphi = 0. \tag{41}$$

In this case, we search for a solution with separable variables:

$$G(\eta,\varphi) = H(\eta) F(\varphi). \tag{42}$$

We obtain the following equation for the function $H(\eta)$:

$$H_{\eta\eta} - i\frac{\eta L_z}{2L_y} H_\eta - \left(\eta^2 + \mu_0\right) H = 0, \tag{43}$$

where μ_0 is the separation constant.

Next, the term with the first derivative in equation (43) is removed by the following substitution:

$$H(\eta) = P(\eta)\exp\left(i\frac{L_z}{8L_y}\eta^2\right). \tag{44}$$

We obtain the following equation for the function $P(\eta)$:

$$P_{\eta\eta} + \left[-\eta^2\left(1 - \frac{L_z^2}{16L_y^2}\right) - \mu_0 + i\frac{L_z}{4L_y}\right] P = 0. \tag{45}$$

Recall that we are searching for solutions localized in the vicinity of the level $z = 0$. Analysis of equation (45) allows to conclude that the coefficient at η^2 must be positive, so we obtain the following condition for the existence of localized solutions:

$$\left(1 - \frac{L_z^2}{16L_y^2}\right) > 0 \Leftrightarrow 0 < |L_z| < 4|L_y|. \tag{46}$$

Condition (46) means that the branches of the parabola bounding the inner area of transparency from the outer area of shadow, must be practically parallel to each other. Otherwise, the vertical mode does not form and the wave does not approach the critical point for an infinitely long time. It is important to note that if condition (46) is not satisfied, then other modes of solution transformation are formally possible. There is no question of any uniqueness of the solution here.

Evaluation of the parameters for internal waves shows that if we take the scales adopted by the authors of [13], we obtain a very good difference in these values ($L_z < 4\ L_y$), so the concept of a parabolic trap is valid from a physical standpoint.

Let us define the quantum values of the separation variable μ_0 [10, 16]:

$$-(2m+1) = \left(\mu_0 - i\frac{L_z}{4L_y}\right) \Big/ \left(1 - \frac{L_z^2}{16L_y^2}\right)^{1/2},\ m = 0, 1, 2, \ldots . \tag{47}$$

From here, we can find the eigenvalues

$$\mu_0 = \frac{L_z}{L_y}\left[\frac{1}{4}i - \frac{\delta}{2}\left(m+\frac{1}{2}\right)\right];\ \delta \equiv \left(\frac{16L_y^2}{L_z^2}-1\right)^{1/2},\ m = 0, 1, 2, ... \tag{48}$$

and eigenfunctions

$$P(\eta) = \left[\sum_{m=0}^{\infty} H_m\left(\eta\left[1-\frac{L_z^2}{16L_y^2}\right]^{1/4}\right)\right]\exp\left[-\frac{\eta^2}{2}\left(1-\frac{L_z^2}{16L_y^2}\right)^{1/2}\right],\ m = 0,1,2,..., \tag{49}$$

where H_m are Hermite polynomials.

Let us now define the second factor $F(\varphi)$ in solution (42). We obtain the following equation from Eq. (41):

$$-i\frac{\varphi L_z}{L_y}F_\varphi + \mu_0 F = 0. \tag{50}$$

The solution of equation (50) has the following form:

$$F(\varphi) = \varphi^\mu,\ \mu \equiv -i\mu_0\frac{L_y}{L_z}. \tag{51}$$

Finally, we obtain the following eigenvalues:

$$\mu = \frac{1}{4} + i\frac{\delta}{2}\left(m+\frac{1}{2}\right). \tag{52}$$

Substituting all the found composite solutions into the initial integral (36), we find the eigenfunctions:

$$\Psi(k,y,z,\omega) = A(k,\omega)\sum_{m=0}^{\infty}\int_0^\infty l^\mu\left[H_m\left(\frac{z\,l^{1/2}}{L_z^{1/2}} - \left[1-\frac{L_z^2}{16L_y^2}\right]^{1/4}\right)\right]\times$$
$$\times\exp\left[-\frac{z^2 l}{2L_z}\left(1-\frac{L_z^2}{16L_y^2}\right)^{1/2}\right]\cdot\exp\left[il\left(y+\frac{z^2}{8L_y}\right)\right]dl, \tag{53}$$

where $A(k,\omega)$ is some constant that determines the spectral density of the initial state.

Further, the obtained eigenfunctions (53) can be reduced by simple transformations to a degenerate hypergeometric function with respect to some complex argument. Note that it is the integral notation (53) that is preferred for finding the asymptotic forms of eigenfunctions. Despite the fact that the constructed eigenfunctions (53) express the dependence on two physical variables (z and y), the integral for the eigenfunctions is one-dimensional, which makes it possible to use the stationary phase method [16].

Let us write the imaginary part of the integral (53) in the following form:

$$\exp\left[il\left(y+\frac{z^2}{8L_y}\right)+i\frac{\delta}{2}\left(m+\frac{1}{2}\right)\ln l\right]. \tag{54}$$

If we differentiate this expression by the variable l and equate the expression in square brackets to zero, we obtain the equation for the point l_c:

$$y+\frac{z^2}{8L_y} = -\frac{\delta}{2l_c}\left(m+\frac{1}{2}\right). \tag{55}$$

Let us rewrite this relation in the following form:

$$l_c = -\frac{\delta\left(m+\frac{1}{2}\right)}{2\left(y+\frac{z^2}{8L_y}\right)}. \tag{56}$$

The resulting expression (56) is a kind of generalization of the short-wave WKB asymptotic form of the dispersion relation $l_c = y^{-1}$. Then the second derivative of the phase with respect to the wavenumber is proportional to l_c^{-2}, and, therefore, the inverse first power root of this derivative is proportional to l^{1}_{c}.

The asymptotic form of eigenfunctions in the vicinity of the critical point is as follows:

$$\Psi_1(k,y,z,\omega) = A(k,\omega)\sum_{m=0}^{\infty} l_c^{\mu+1}\left[H_m\left(\frac{z\,l_c^{1/2}}{L_z^{1/2}}\left[1-\frac{L_z^2}{16L_y^2}\right]^{1/4}\right)\right]\times$$
$$\times\exp\left[-\frac{z^2 l_c}{2L_z}\left(1-\frac{L_z^2}{16L_y^2}\right)^{1/2}\right]\cdot\exp\left[i\,\frac{\delta}{2}\left(m+\frac{1}{2}\right)\right]. \tag{57}$$

Analysis of this equality allows us to conclude that the asymptotic form of the solution of the reference equation exactly coincides with the WKB solution [17], expressed as a vertical mode in the form of Hermite polynomials majored by a Gaussian function, and gives the classical degree of 5/4 for the amplitude of the vertical velocity. If the authors of [13] describe a certain mode, then we are certain that this is not their vertical mode in the form of a WKB solution along a vertical coordinate, but a completely different one, which is constructed in [17].

The solutions we constructed are not functions with respect to the variables z, y, but rather to some curvilinear variables taking the following form:

$$(y,z)\to\left(\left(y+\frac{z^2}{8L_y}\right),\quad\frac{z}{\left(y+\frac{z^2}{8L_y}\right)^{1/2}}\right). \tag{58}$$

Thus, in a sense, there is a curvature of space in the vicinity of the focal point. However, all this "curvilinearity" was also observed in the solution of the problem in the WKB approximation, where formally the following variable substitution took place:

$$(y,z)\to\left(y,\frac{z}{\sqrt{y}}\right).$$

Therefore, by and large, the asymptotic forms of one-dimensional integrals do not give any qualitatively new results other than WKB solutions, with the exception of condition (47), which is satisfied with a large margin.

Reduction of the Fourier integral to a hypergeometric function of a complex variable. To compare our solution with the solution obtained by Erokhin and Sagdeev [13], we rewrite the eigenfunctions (54) in the following form:

$$\Psi_m(k,y,z,\omega) = \int_0^{\infty} l^{\mu} H_m\left(\frac{z l^{1/2}}{2L_y^{1/2}}\delta^{1/2}\right)\exp\left[-\frac{z^2 l}{8L_y}\delta\right]\cdot\exp\left[il\left(y+\frac{z^2}{8L_y}\right)\right]dl. \tag{59}$$

Next, we perform the substitution of variables ($l \to x$), and the argument

$$\left(\frac{z\, l^{1/2}}{2L_y^{1/2}} \delta^{1/2} \right)$$

of the Hermite polynomial is taken as a new variable

$$x = \frac{z l^{1/2} \delta^{1/2}}{2L_y^{1/2}}. \tag{60}$$

It follows from this that

$$\Psi_m \propto \int_0^\infty \exp\left[-2ax^2\right] \frac{x^{2\mu+1}}{z^{2\mu+2}} H_m(x)\, dx. \tag{61}$$

The complex variable $2a$ appeared in Eq. (60), depending on two spatial physical variables, z and y:

$$2a = \frac{1}{2} - i \frac{1}{2\delta z^2}\left(z^2 + 8yL_y\right). \tag{62}$$

We solved the two-dimensional problem in terms of a one-dimensional integral, but only with respect to a complex argument. Apparently,

$$\frac{1}{2a} = \frac{2\delta z^2}{\delta z^2 - i\left(z^2 + 8\, yL_y\right)} \equiv \tau^*, \tag{63}$$

where τ is the complex variable from the study by Erokhin and Sagdeev [13] (the asterisk corresponds to complex conjugation).

The integral representation of the hypergeometric function in terms of Hermite polynomials has the following form [11, Eqs. 7.37, 7.38]:

$$F\left(-n; \frac{\nu+1}{2}; \frac{1}{2}; \frac{1}{2a}\right) \sim \int_0^\infty \exp\left[-2ax^2\right] x^\nu H_{2n}(x) dx, \tag{64}$$

where Re$a > 0$, Re$\nu > -1$;

in addition [11, Eqs. 7.376.3]:

$$F\left(-n; \frac{\nu}{2}+1; \frac{3}{2}; \frac{1}{2a}\right) \sim \int_0^\infty \exp\left[-2ax^2\right] x^\nu H_{2n+1}(x) dx, \tag{65}$$

where Re$a > 0$, Re$\nu > -2$.

Taking into account the eigenvalues (48), we find:

$$\nu = 2\mu + 1 = \frac{3}{2} + i\delta\left(m + \frac{1}{2}\right). \tag{66}$$

Consequently, the constructed solutions are regular, and the integrals converge. Similarity with the solution from Erokhin and Sagdeev [13] was achieved in three of the four parameters. Let us determine the last parameter of the hypergeometric function:

$$\frac{\nu}{2} + 1 = \frac{7}{4} + i\frac{\delta}{2}\left(m + \frac{1}{2}\right) \equiv \gamma^*, \tag{67}$$

where γ is the quantum parameter from [1, Eq. (2.7)].

Similarly, we find that

$$\frac{\nu+1}{2} = \frac{5}{4} + i\frac{\delta}{2}\left(m + \frac{1}{2}\right). \tag{68}$$

Thus, we obtained a complete agreement of our results with [13]. If we take into account the second part of the Fourier integral for negative wave numbers, then by substitution of the variable it can be reduced to an integral with respect to positive wavenumbers. But then the imaginary unit ($i \to -i$ 1) will be replaced in the studied integral, and this will lead to the appearance of the second part of the solution, where complex conjugate τ^* and γ^* will appear instead of τ and γ.

Thus, the general solution of the problem is the sum of solutions with respect to τ and τ^*, which is physically equivalent to the sum of incident and reflected waves. This means that mathematically there is no prohibition on reflection and the hypothesis of infinite focusing is greatly exaggerated.

Discussion and conclusions

This study provides basic information on the operator method of Fourier transformation, which is necessary for practical solution of specific physical problems in inhomogeneous media. The main properties are formulated by two approaches:

integration by parts, which implies attenuation of functions at infinity;

parametric differentiation of the forward or inverse Fourier transform.

Using five specific examples, we established how Fourier analysis works in inhomogeneous media. In the first four examples, formal integral solutions are constructed, since their further analysis is well known and the reader can consider the references to further explore this. Notice that these integrals (see examples 1 and 2) are typically given without derivation and the reader is invited to independently obtain this derivation using the Laplace transform. In our work, we constructed integral solutions using the Fourier transform and Cauchy's theorem, showing their equivalence with the Laplace transform in one-dimensional inhomogeneous problems.

Examples 2, 3 and 4 consider a two-dimensional problem in which the inhomogeneity of the medium is one-dimensional linear in nature. In Example 2, the solution can be obtained in two ways: in terms of the Fourier transform and in terms of the Laplace transform. In Example 5, we performed a complete analysis of the boundary-value problem. We constructed the Fourier integral, found its two-dimensional asymptotic forms using the stationary phase method and the properties of a parabolic quantum oscillator, and also identified the Fourier integral found, reducing it to a well-known degenerate hypergeometric function with respect to a complex argument. Thus, we proved that the statement about the inefficiency of Fourier analysis in inhomogeneous media is erroneous.

Therefore, in terms of the Fourier integral, we analytically proved the identity of the solution of the reference equation for vertical focusing of a monochromatic wave in the vicinity of the focus with the solution of the reference equation in terms of a degenerate hypergeometric function with respect to a complex variable obtained in previous studies. This mathematical solution is also successfully used in problems of magnetohydrodynamic instability and in the description of internal gravitational waves in two-dimensional inhomogeneous fluid [7, 13].

It is established that the issue of wave absorption in the focal zone is ambiguous and therefore both passage and reflection from a singularity can be observed. Specific estimates for typical parameters of oceanic gradients of hydrophysical density and velocity fields show that localization and, as a rule, amplification of wave movements are quite feasible and take the form of highly localized spatial vortex structures.

These aspects should be taken into account in studies of geophysical fields, in particular when analyzing mesoscale vortex dynamics in the ocean.

The analytical method described in these five examples can be used to solve other problems of mathematical physics.

REFERENCES

1. **Witham G. B.,** Linear and nonlinear waves (Series: Pure and Applied Mathematics), Willey, New York, 1974.

2. **Lighthill J.,** Waves in fluids, 2-nd Ed., Cambridge University Press, New York, 1978.

3. **Pedlosky J.,** Geophysical fluid dynamics, Springer Verlag, New York, 1978.

4. **Wirth V., Riemer M., Chang E. K., Martius O.,** Rossby wave packets on the midlatitude waveguide. A review, Mon. Weather Rev. 146 (7) (2018) 1965–2001.

5. **Titmarsh E. C.,** Introduction to the theory of Fourier integrals, 3d Ed., Chelsea Publishing Company, New York, 1986.

6. Reviews of plasma physics, Vol. 7, Ed. by M. A. Leontovich, Consulting Bureau, USA, 1979.

7. **Erokhin N. S., Moiseev S. S.,** Problems of the theory of linear and nonlinear transformation of waves in inhomogeneous media, Soviet Physics Uspekhi. 16 (1) (1973) 64–81.

8. **Wasow W.,** A study of the solutions of the differential equation $y^{(4)} + \lambda^2 (xy'' + y) = 0$ for large values of λ, Ann. Math. 52 (2) (1950) 350–361.

9. **Vladimirov V. S., Zharinov V. V.,** Uravneniya matematicheskoy fiziki [The equations of mathematical physics], Fizmatlit Publishing, Moscow, 2004 (in Russian).

10. **Gnevyshev V. G., Belonenko T. V.,** The Rossby paradox and its solution, Hydrometeorology and Ecology (Proceedings of the Russian State Hydrometeorological University). (61) (2020) 480–493 (in Russian).

11. **Gradshteyn I. S., Ryzhik I. M.,** Table of integrals, series, and products, Ed. by A. Jeffrey, D. Zwillinger, Academic Press, Cambridge, USA, 2015.

12. **Evgrafov M. A.,** Analytical functions, Dover Publications, New York, 1978.

13. **Erokhin N. S., Sagdeev R. Z.,** [To the theory of anomalous focusing of internal waves in a two-dimensional nonuniform fluid. Part I: A stationary problem], Morskoy Gidrofizicheskiy Zhurnal (Soviet J. Phys. Oceanography). (2) (1985) 15–27 (in Russian).

14. **Kamke E.**, Handbook of exact solutions for ordinary differential equations. Ed. by A. D. Polyanin, V. F. Zaitsev, CRC Press, Boca Raton, New York, London, 1995.

15. **Miropol'sky Yu. Z.,** Dynamics of internal gravity waves in the ocean (Book Series "Atmospheric and Oceanographic Sciences Library"), Springer, New York, 2001.

16. **Badulin S. I., Shrira V. I.,** On the irreversibility of internal-wave dynamics due to wave trapping by mean flow inhomogeneities. P. 1. Local analysis, J. Fluid Mech. 251 (June) (1993) 21–53.

17. **Badulin S. I., Shrira V. I., Tsimring L. Sh.,** The trapping and vertical focusing of internal waves in a pycnocline due to the horizontal inhomogeneities of density and currents, J. Fluid Mech. 158 (Sept.) (1985) 199–218.

18. **Gnevyshev V. G., Badulin S. I., Belonenko T. V.,** Rossby waves on non-zonal currents: Structural stability of critical layer effects, Pure Appl. Geophys. 177 (11) (2020) 5585–5598.

19. **LeBlond P. H., Mysak L. A.,** Waves in the ocean, Elsevier Oceanography Ser. Elsevier Scientific Publishing Company, Amsterdam, 1981.

20. **Yamagata T.,** On trajectories of Rossby wave-packets released in a lateral shear flow, J. Oceanogr. Soc. Japan. 32 (4) (1976) 162–168.

СПИСОК ЛИТЕРАТУРЫ

1. **Уизем Дж.** Линейные и нелинейные волны. Пер. с англ. М.: Мир, 1977. 606 с.

2. **Лайтхилл Д.** Волны в жидкостях. Пер. с англ. М.: Мир, 1981. 598 с.

3. **Педлоски Дж.** Геофизическая гидродинамика. В 2 тт. М.: Мир, 1984. 398 с. (том 1), 416 с. (том 2).

4. **Wirth V., Riemer M., Chang E. K., Martius O.** Rossby wave packets on the midlatitude waveguide. A review // Monthly Weather Review. 2018. Vol. 146. No. 7. Pp. 1965–2001.

5. **Титчмарш Э. Ч.** Введение в теорию интегралов Фурье. Пер. с англ. Москва, Ленинград: Государственное издательство технико-теоретической литературы, 1948. 419 с.

6. Вопросы теории плазмы. Сб. статей. Выпуск 7. Под ред. акад. М. А. Леонтовича, М.: Атомиздат, 1973. 304 с.

7. **Ерохин Н. С., Моисеев С. С.** Вопросы теории линейной и нелинейной трансформации волн в неоднородных средах // Успехи физических наук. 1973. Т. 109. № 2. С. 225–258.

8. **Wasow W.** A study of the solutions of the differential equation $y^{(4)} + \lambda^2 (xy'' + y) = 0$ for large values of λ // The Annals of Mathematics. 1950. Vol. 52. No. 2. Pp. 350–361.

9. **Владимиров В. С., Жаринов В. В.** Уравнения математической физики. 2-е изд. М.: Физматлит, 2004. 400 с.

10. **Гневышев В. Г., Белоненко Т. В.** Парадокс Россби и его решение // Гидрометеорология и экология (Ученые записки РГГМУ). 2020. № 61. С. 480–493.

11. **Градштейн И. С., Рыжик И. М.** Таблицы интегралов, сумм, рядов и произведений. 7-е изд. СПб: БХВ-Петербург, 2011, 1232 с.

12. **Евграфов М. А.** Аналитические функции. -4е изд. СПб.: Изд-во «Лань», 448 .2008 с.
13. **Ерохин Н. С., Сагдеев Р. З.** К теории аномальной фокусировки внутренних волн в двумерно-неоднородной жидкости. Часть 1. Стационарная задача // Морской гидрофизический журнал. 1985. № 2. С. 15–27.
14. **Камке Э.** Справочник по обыкновенным дифференциальным уравнениям. Пер. с нем. -6е изд. М.: Наука: Гл. ред. физ-мат. лит., 2003. 576 с.
15. **Миропольский Ю. З.** Динамика внутренних гравитационных волн в океане. Ленинград: Гидрометеоиздат, 1981. 301 с.
16. **Badulin S. I., Shrira V. I.** On the irreversibility of internal-wave dynamics due to wave trapping by mean flow inhomogeneities. Part 1. Local analysis // Journal of Fluid Mechanics. 1993. Vol. 251. June. Pp. 21–53.
17. **Badulin S. I., Shrira V. I., Tsimring L. Sh.** The trapping and vertical focusing of internal waves in a pycnocline due to the horizontal inhomogeneities of density and currents // Journal of Fluid Mechanics. 1985. Vol. 158. September. Pp. 199–218.
18. **Gnevyshev V. G., Badulin S. I., Belonenko T. V.** Rossby waves on non-zonal currents: Structural stability of critical layer effects // Pure and Applied Geophysics. 2020. Vol. 177. No. 11. Pp. 5585–5598.
19. **LeBlond P. H., Mysak L. A.** Waves in the ocean. Elsevier oceanography series. Amsterdam: Elsevier Scientific Publishing Company, 1981. 602 p.
20. **Yamagata T.** On trajectories of Rossby wave-packets released in a lateral shear flow // Journal of the Oceanographic Society of Japan. 1976. Vol. 32. No. 4. Pp. 162–168.

THE AUTHORS

GNEVYSHEV Vladimir G.
Shirshov Institute of Oceanology, RAS
36 Nakhimovskiy Ave., Moscow, 117997, Russia
avi9783608@gmail.com
ORCID: 0000-0003-4608-7781

BELONENKO Tatyana V.
St. Petersburg State University
7-9 Universitetskaya Emb., St. Petersburg, 199034, Russia
btvlisab@yandex.ru
ORCID: 0000-0003-4608-7781

СВЕДЕНИЯ ОБ АВТОРАХ

ГНЕВЫШЕВ Владимир Григорьевич – *кандидат физико-математических наук, ведущий научный сотрудник Института океанологии имени П. П. Ширшова Российской академии наук.*
117997, Россия, г. Москва, Нахимовский проспект, 36
avi9783608@gmail.com
ORCID: 0000-0003-4608-7781

БЕЛОНЕНКО Татьяна Васильевна – *доктор географических наук, профессор кафедры океанологии Санкт-Петербургского государственного университета.*
199034, Россия, г. Санкт-Петербург, Университетская наб., 7-9
btvlisab@yandex.ru
ORCID: 0000-0003-4608-7781

Received 30.03.2023. Approved after reviewing 02.08.2023. Accepted 02.08.2023.
Статья поступила в редакцию 30.03.2023. Одобрена после рецензирования 02.08.2023. Принята 02.08.2023.

EXPERIMENTAL TECHNIQUE AND DEVICES

Original article
DOI: https://doi.org/10.18721/JPM.16409

AN EXPERIMENTAL APPARATUS FOR STUDYING THE CHARACTERISTICS OF THERMOELECTRIC EFFECT IN NANOSTRUCTURES

K. R. Trofimovich, P. G. Gabdullin, A. V. Arkhipov✉

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ arkhipov@rphf.spbstu.ru

Abstract. We present an experimental setup for the study of thermoelectric effect in point contacts between different materials and in nanostructures. Point contacts of controlled size are formed with the use of an atomic force microscope (AFM), thermopower dependences against the temperature drop and against the contact spot size determined from the force applied to the probe. Computer simulations of heat transport in the system were performed to evaluate influence of atmospheric air and of a liquid layer covering solid surfaces in the atmospheric conditions onto temperature distributions. This influence was found to be insubstantial, which makes it possible to conduct experiments on the atmosphere and not in high vacuum.

Keywords: thermoelectric effect, point contact, nanostructures, atomic force microscope

Funding: The reported study was funded by Russian Science Foundation (Grant No. 23-29-10027 (https://rscf.ru/project/23-29-10027/) and by St. Petersburg Science Foundation (Grant No. 23-29-10027).

Citation: Trofimovich K. R., Gabdullin P. G., Arkhipov A. V., An experimental apparatus for studying the characteristics of thermoelectric effect in nanostructures, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 101–117. DOI: https://doi.org/10.18721/JPM.16409

Научная статья
УДК 538.93, 536.241
DOI: https://doi.org/10.18721/JPM.16409

ЭКСПЕРИМЕНТАЛЬНАЯ УСТАНОВКА ДЛЯ ИССЛЕДОВАНИЯ ОСОБЕННОСТЕЙ ТЕРМОЭЛЕКТРИЧЕСКОГО ЭФФЕКТА В НАНОСТРУКТУРАХ

К. Р. Трофимович, П. Г. Габдуллин, А. В. Архипов✉

Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Россия

✉ arkhipov@rphf.spbstu.ru

Аннотация. Представлены общая схема и конструкция экспериментальной установки для изучения термоэлектрических явлений в точечных контактах разнородных материалов и в наноструктурах. Точеные контакты регулируемого размера формируются с помощью атомно-силового микроскопа, определяются зависимости термоэдс от разности температур и от силы воздействия зонда на образец. Проведено численное моделирование распределения температуры в такой системе для различных условий, оценено влияние воздушной среды и жидкого слоя, формирующегося на поверхностях в условиях естественной атмосферы. Моделирование показало, что это влияние не является существенным, что делает необязательным проведение экспериментов в высоком вакууме.

Ключевые слова: термоэлектрический эффект, точечный контакт, наноструктуры, нанотеплофизика, атомно-силовой микроскоп

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 10027-29-23 (https://rscf.ru/project/23-29-10027/) и гранта Санкт-Петербургского научного фонда № 23-29-10027.

Ссылка для цитирования: Трофимович К. Р., Габдуллин П. Г., Архипов А. В. Экспериментальная установка для исследования особенностей термоэлектрического эффекта в наноструктурах // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 4 № .16. С. 101–117. DOI: https://doi.org/10.18721/JPM.16409

Introduction

The problem of increasing energy efficiency is becoming more and more urgent for electronics. It was reported, in particular, that more than 10% of the world's electricity is already consumed by computers and telecommunications equipment, and this share is growing rapidly [1]. The energy consumed by electronic devices is eventually released as heat, and the reverse conversion of some of this heat into electricity could significantly increase overall energy efficiency. The most natural way to utilize the thermal energy of electronic components is the use of solid-state thermoelectric generators (TEG). The development and improvement of TEG has been underway for many years (at least since the 1940s), and substantial advances have been made. However, there are fundamental physical limitations to increasing the effectiveness of TEG.

Thermoelectric materials are commonly evaluated by the parameter of thermoelectric performance (quality) Z or, more often, the dimensionless quantity of the same name ZT, where T is the absolute temperature. This parameter characterizes the difference between the efficiency of thermoelectric conversion, achievable using the material, and the efficiency of an ideal heat engine. The dependence of the thermoelectric performance on the properties of the material is given by the formula [1–4]:

$$ZT = \frac{S^2 \sigma T}{\kappa_{el} + \kappa_{ph}}, \quad (1)$$

where S is the Seebeck coefficient (defined as the coefficient of proportionality between the thermopower and the temperature drop); σ is electrical conductivity; κ_{el}, κ_{ph} are the electron and lattice (phonon) components of thermal conductivity.

According to available estimates, the widespread use of TEG in the energy sector will be economically justified when ZT = 3–4 is reached. The ZT values obtained so far for the best thermoelectric materials (in particular, bismuth telluride Bi_2Te_3) at room temperature are close to unity. It is obvious from Eq. (1) that to increase the thermoelectric performance, it is necessary to reduce the thermal conductivity of the material and increase its electrical conductivity and the Seebeck coefficient. Metals, due to their high electrical conductivity, are not optimal materials for TEG, since the values of the Seebeck coefficient are small: in most cases they do not exceed10 MV/K in absolute value. In addition, the heat flow in metals is carried mainly by electric charge carriers, so consequently the thermal and electrical conductivity of metals in the denominator and numerator of Eq. (1), respectively, are proportional to each other (Wiedemann–Franz law), which makes it difficult to optimize the ZT parameter. In contrast, semiconductors have low electronic thermal conductivity, and the heat flow in them is carried almost exclusively by phonons. The Seebeck coefficient of semiconducting materials is usually quite large (hundreds of μV/K), which is determined by significant variations in the energy density function of electron states near the Fermi level [3–5]. The thermoelectric properties of semiconductors can be optimized by choosing the degree of their doping [6, 7] or by shifting the Fermi level in other ways [8]. In general, semiconductors are characterized by the best thermoelectric parameters among homogeneous bulk materials, but the possibilities for their radical improvement are still limited.

In recent decades, the prospects for the creation of commercially successful thermoelectric devices have been associated with nanostructures and nanostructured materials. Many experiments (see, for example, [9–12]) have fundamentally confirmed the theoretical predictions made by in [4, 13] about the possibility of using size effects to selectively reduce thermal conductivity due to the scattering of phonons by interfaces and defects, with less influence on electrical conductivity.

It is equally promising to construct thermoelectric devices using nanoparticles, individual molecules or molecular layers [2, 3, 12–16] with a discrete spectrum of allowed electronic states, which can be optimized to achieve high values of S and ZT. Nanostructures developed can provide conditions simultaneously for destructive interference of lattice excitations and constructive interference of electronic waves [1, 2]. Calculating the thermoelectric parameters of devices using nanoscale elements is a difficult problem, which increases the role of experiment in such studies.

Problem statement

This paper presents the results of the first, initial stage of a new research project, whose purpose is to experimentally verify one of the theoretical models predicting the possibility of constructing a high-efficiency TEG based on heat transfer characteristics and thermoelectric phenomena in carbon-based nanostructures.

Carbon in the state of sp^2-hybridization (single-layer and multilayer graphene, carbon nanotubes, graphite) has unique characteristics [4, 17, 18]: low effective mass of charge carriers, high thermal conductivity, high light absorption coefficient. The electronic properties of graphene and carbon nanotubes are easily modified not only by doping and defects, but also by an electric field, which makes carbon in the state of sp^2-hybridization a promising material for electronics. Among other things, the possibility of its use as part of thermoelectric devices was also considered [5, 19].

Eidelman proposed [20, 21] a TEG based on a structure of nanometer layers of diamond-like and graphite-like carbon. The principle of its operation is based on the theoretical model proposed in Eidelman's earlier studies [22, 23], which substantiates the possibility of obtaining increased values of the Seebeck coefficient S and the ZT parameter using the phenomenon of charge-carrier entrainment by ballistic flow of phonons, even at room and higher temperatures. In this case, to achieve a positive result, localization of the electric field and temperature drop in a planar sp^2-carbon nanolayer is required, where the phonon flow remains ballistic. The diamond-like layer plays the role of a cooler, diverting the unused part of this flow. The presence

of sharp (no more than several lattice spacings) interphase boundaries is also seen as a necessary condition. In our opinion, the experimental demonstration of the predicted effect in [21] was not entirely convincing. The reason for this was probably the objective technological complexity of forming the ideal structure required by the theory; the presence of defects, almost inevitable at the specified values of thickness and area, could lead to electrical shorting of the real structure and a decrease in the thermoelectric voltage recorded for it.

Our study is aimed at conducting an experimental verification of the viability of the concept proposed in [20, 21] using a structure that is easier to create, whose main part is, for example, a carbon nanostructure on the surface of a silicon substrate (serving in this case as a cooler or, conversely, a heater). The second thermal and electric contact with the island should be established by means of an atomic force microscope (AFM) probe. We have previously studied island-type carbon films of the required structure in connection with their ability to emit electrons [24–26].

The advantage of the proposed approach is the possibility of conducting quick independent testing of the thermoelectric properties of many islands, which differ in size and properties of the interface with the substrate. This seems necessary because we expect that only a few of the islands will show high thermoelectric performance.

Such expectations are associated with important features introduced by thermal contacts of small lateral dimensions.

The phenomenon of a decrease in thermopower in nanocontacts is known [27–30], which is explained precisely by the suppression of phonon entrainment of charge carriers. Its cause is seen in the scattering of nonequilibrium phonons at the aperture of the nanocontact, which reduces the likelihood of transmitting their momentum to electrons or holes. Therefore, according to the theory outlined in [27], an additional multiplier $d/Л_p$ appears in the formula for thermal entrainment for small contact diameters d, where $Л_p$ is the phonon path length.

The experiments planned are intended for studying islands whose height is of the order of the phonon path length, approximately equal to 5 nm (according to the estimate in [21]). Their lateral size will be several times larger, but the size of the contact with the AFM probe may be small and poorly controlled.

However, according to many researchers, the boundary scattering of phonons is only one of the reasons for the dependence of the Seebeck coefficient and other kinetic coefficients on the size of the contact region. In particular, both theoretic predictions and experimental observations exist for deviation of the kinetic dependences from the linear form upon violation of the condition $L_T > \Lambda_p$, where L_T is the distance at which the temperature changes significantly; it can be defined, for example, as

$$L_T^{-1} = \max\left(\frac{|\text{grad } T|}{T}\right). \tag{2}$$

The reason for this phenomenon is seen in the nonlocality of the interaction of charge carriers with the lattice [31–37]. If the condition $L_T > \Lambda_p$ is violated, it cannot be assumed that the interaction occurs at a point with certain coordinates. Often, the temperature value also cannot be correctly determined, since the distributions of phonons and charge carriers turn out to be significantly nonequilibrium [5, 38]. One of the manifestations of non-locality is the dependence of the thermopower value not only on the applied temperature difference, but also on its distribution profile, on the maximum value of its gradient, i.e., on the characteristics that can hardly be predicted for the thermal contact of the probe with the island.

Another region with properties that vary greatly from island to island may be the interface with the substrate, since the islands were formed on a layer of natural oxide; notably, it is stated in the literature [30, 39, 40] that thin intermediate layers do not affect the results of measuring the thermoelectric parameters of coatings and nanostructures.

For these reasons, reliable detection of high values of the Seebeck coefficient (the estimate of its expected value in [21] is 50 mV/K) for at least a small part of carbon nanostructures (or other similar samples) brought into contact with the AFM probe with an excellent temperature can be considered as confirmation of theoretical predictions of [20–23] and will become an incentive for subsequent efforts on the practical implementation of the concept of TEG proposed in [21].

Experimental setup

An experimental setup for solving the above problem is currently developed based on NanoDST AFM (Pacific Nanotechnology, USA). This approach should not be considered new: setups based on probe microscopes have been successfully used in similar studies, in particular, to map the distributions of thermal and thermoelectric parameters [5, 6, 39–44] and to study the properties of molecular layers and individual molecules [1, 5, 16, 39, 45]. The novel characteristics of the setup developed are minimal improvements to the serial AFM and the use of standard AFM probes in its operation, which is possible since a limited range of problems will be solved at the initial stage of research.

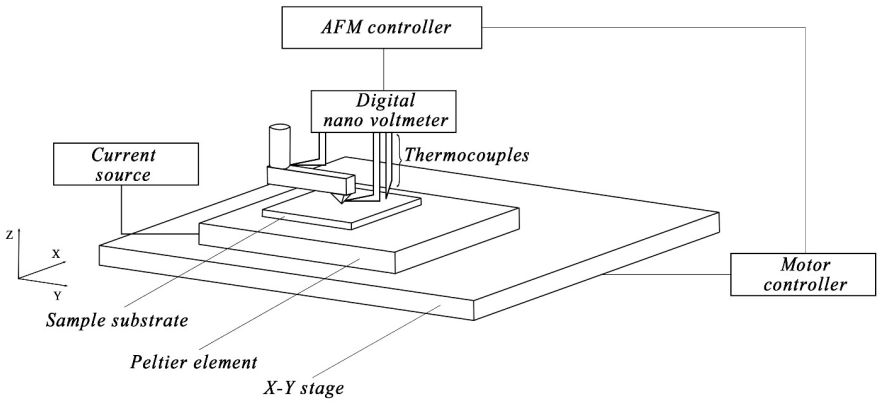

Fig. 1. Schematic of experimental setup

The experimental setup is shown schematically in Fig. 1. The large positioning stage (X-Y stage) and the probe holder play the role of thermostats with the temperature equal to the ambient temperature T_0. The sample substrate considered is placed on the upper surface of the Peltier element, whose temperature is maintained (using the Current source feeding the element) at a set value $T = T_0 + \Delta T$. The temperature values are controlled by thermocouples. During the experiment, the AFM probe, whose initial temperature is T_0, is brought into contact with the point of the sample selected during a preliminary scan of its surface. The thermoelectric voltage formed during contact is recorded by a digital nanovoltmeter synchronized with the AFM controller. Simultaneous recording of its reading and the force-distance curve of the AFM allow to determine the dependence of the thermoelectric voltage on the force of interaction of the probe with the surface, and therefore on the size d of their contact region. To isolate the voltage component associated with the contact potential difference, the experiment is also carried out for $\Delta T = 0$.

The problem of determining purely thermal parameters, i.e., the values of heat flows, thermal resistance, thermal conductance and thermal conductivity, is usually the most difficult from the standpoint of measurement technology [38, 46, 47]. At this stage, we neglected to define them, since to assess the applicability of the TEG concept proposed in [21], it is sufficient to find the maximum values of the Seebeck coefficient S, i.e., temperature and thermopower measurements. If reproducible dependencies $S(\Delta T)$ and $S(d)$ are found, the results obtained will provide additional information about the physics of heat and electric charge transfer processes in the system considered.

In the future, it is planned to expand the capabilities of the setup.

Experimental samples

The following samples were selected for the first experiments (below they are given in sequence from simple to more complex).

Nanocontact of AFM probe with a metal plate or thick (Cu, Au) film. Since the values of the Seebeck coefficient for silicon significantly exceed its values for metals, when a silicon probe comes into contact with a metal plate, it is natural to expect that the value of the thermopower will be mainly determined by the processes in the probe. Thus, the main purpose of such experiments is to determine the contribution of probes of different types to thermopower: with a metal

coating and without a coating, with a different size of the contact spot. Such data are necessary for conducting subsequent experiments with other objects, that is, to take into account the contribution of the AFM probe to the recorded values of thermopower.

Nanocontact of AFM probe with silicon plate. Measurements of the thermoelectric characteristics of nanocontacts formed between AFM probes and substrates are very important for the correct interpretation of other experimental results, since carbon nanostructures (these are the main objects of experiments at this stage of the study) are formed on silicon substrates. However, there is an additional motivation for conducting such experiments.

Island films of carbon and metals (molybdenum, zirconium, tungsten) formed on silicon substrates were found to be capable of low-voltage field emission in previous experiments [24, 25, 48]. It was suggested that thermoelectric potentials play an important role in the physical mechanism of the emission process [26, 48, 49]. At the same time, the absence of significant differences in the emission parameters of carbon and metal islands indicated that thermoelectric potentials are formed not in the islands themselves, but in substrates near the islands. A theoretical consideration of such a process using a model of charge carrier entrainment by ballistic phonon flow [21] made it possible to evaluate the effectiveness of thermoelectric conversion by such a structure. The results of this analysis will be published later, and they may be important in connection with the objectives of this study. This is reason for additional interest in the results of the planned experimental testing of thermoelectric characteristics of nanocontacts of metallized probes with silicon substrates.

Admittedly, the properties of silicon point contacts have already been studied before, from the 1980s to the 1990s. However, these early experiments were carried out mainly with contacts of micron and submicron sizes (20–0.3 μm) [29, 30, 50], which were formed between the pointed edges of silicon wedges pressed against each other with a force of 1–100 mN [50]. Mechanical compression led to the appearance of significant deformations in the near-contact region, which, according to the authors [30], were capable of causing additional scattering of phonons and influencing the propagation of heat flows. The size of the contact between the AFM probe and the planar surface and the force of their interaction can be significantly reduced to more correctly simulate the contact of the substrate with the nanostructure formed on it. The nominal value of the elastic constant of the CSG01 AFM probe (NT MDT, Russia), designed for contact mode measurements, is 0.03 N/m. The transition from attractive forces to repulsive forces on a typical force-distance curve of an AFM probe (when direct contact of the probe with the surface is established) occurs at a distance of about 100 nm [41, 42]. These values can be used to estimate the strength of the mechanical interaction of the probe with the sample surface as $F \approx 3 \cdot 10^{10}$ N. The diameter of the contact spot d can be calculated by solving the problem of the elastic interaction of a sphere of radius R and a plane [41]:

$$d = \left(\frac{6FR}{E} \right)^{1/3}, \tag{3}$$

where E is Young's modulus (E = 109 GPa for silicon), R is the probe tip radius.

For the nominal value R = 10 nm, this gives an estimate of the minimum diameter of the contact spot $d \approx 1.2$ nm. When the pressure reaches the plastic limit (for example, in the case of a metal plate), this value may increase slightly, but it certainly will not exceed the radius of the tip [41]. Thus, the size of the contact spot between the AFM probe and the silicon or metal plate can be made significantly smaller than the corresponding values achieved in classical studies [29, 30, 50]. Comparing their results may be of considerable interest.

Carbon nanostructures and graphene sheets (sp^2-carbon). As already mentioned above, the thermoelectric properties of carbon nanostructures are considered as the main object of this study. Another object are sheets of multilayer graphene with a large interface area with the substrate and representing a closer equivalent of the proposed TEG prototype structure. The use of AFM makes it possible to establish thermal and electrical contact with the local area of a separate sheet of multilayer graphene and vary the force with which it is pressed against the surface of the substrate.

Images and surface profiles of samples of naturally oxidized silicon plates with carbon nanowires and graphene sheets of different areas are shown in Fig. 2, *a*–*c*. They were obtained using NanoDST AFM (Pacific Nanotechnology), which is planned to be used in thermoelectric experiments.

Metal nanoparticles. A feature of nanoscale particles in comparison with bulk materials is the discrete nature of the spectrum of resolved states. This feature is favorable in terms of the possibility of achieving high thermoelectric performance. In accordance with the well-known Mott formula, the Seebeck coefficient is determined by the value of the derivative of the energy density of states at the Fermi level [4]:

$$S = \frac{\pi^2 k_B^2 T}{3e} \left\{ \frac{d\left[\ln\left(\sigma(\varepsilon)\right)\right]}{d\varepsilon} \right\}_{\varepsilon=\varepsilon_F}, \qquad (4)$$

where k_B is the Boltzmann constant, ε_F is the Fermi energy, $\sigma(\varepsilon)$ is the value of the differential contribution of charge carriers with energy ε to electrical conductivity.

It should be borne in mind that Eq. (4), generally speaking, refers to metals and degenerate semiconductors and may not be completely correct for sp^2-carbon. However, it is useful for understanding general trends: in the vicinity of a discrete energy level, all energy derivatives are large, which, at the optimal position of the Fermi level, can provide a large value of the Seebeck coefficient [3]. Varying the parameters, namely, particle size, electric potential, etc., allows to "adjust" the relative position of the Fermi level and the permitted energy levels, optimizing it to achieve high thermoelectric performance.

A comparison of the thermoelectric properties of carbon and metal nanoparticles can make it possible to separate the influence of size effects from the influence of the electronic structure of specific materials. 2,*d* shows the AFM image and the surface profile of tungsten nanoparticles for a sample formed on a silicon substrate and prepared for study; the substrate is identical to those used in other cases.

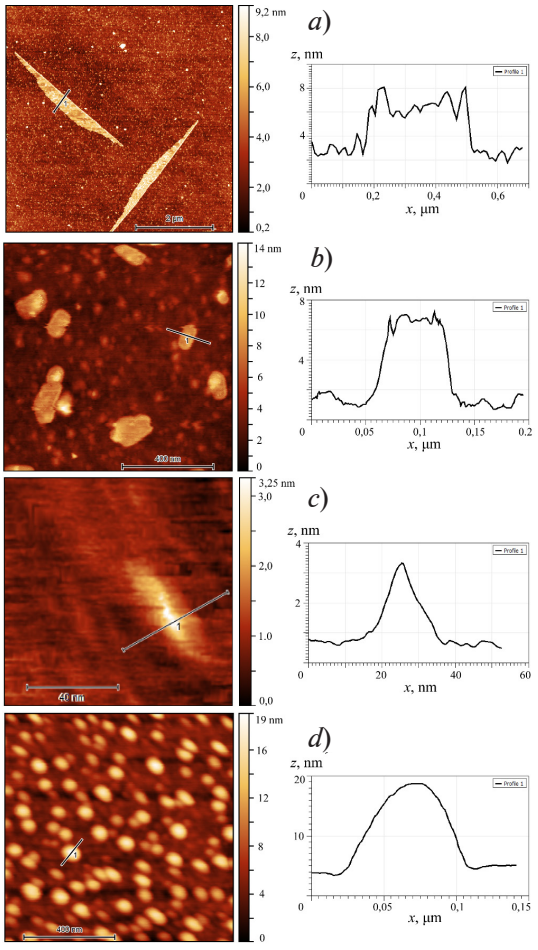

Fig. 2. AFM images (left) and surface topography profiles of nano-objects (right) on naturally oxidized silicon substrates. Nano-objects: sheets of multilayer graphene of large (*a*) and small (*b*) sizes, small carbon nanostructures (*c*) and metal nanoparticles of tungsten (*d*)

Numerical simulation

Numerical simulation of the temperature distribution over the contact region of the AFM probe with a planar substrate was carried out during the general planning of the experiment. The COMSOL Multiphysics software package implementing the finite element method was used. The package generally allows searching for numerical solutions to systems of differential equations of almost any kind with a given geometry and a set of boundary/initial conditions; this includes systems that take into account size effects and nonlocality; for example, equations from theoretical studies [32–37]. However, only the temperature distribution (but not the thermoelectric potential) was modeled at this stage of the research, and standard equations of the macroscopic theory of heat transfer and tabular values of thermal parameters of materials were used. This approach used because the modeling problems at this stage (experiment planning) were limited.

Let us present the basic requirements for these problems.

First, it was necessary to determine whether local temperature measurements near the

nanocontact region were necessary or whether it was sufficient to determine the temperatures of the sample substrate and the massive part of the AFM probe holder. To do this, it was actually necessary to estimate how much of the total temperature drop between these parts (whose temperature is easy to measure) falls on the cantilever of a standard probe under typical experimental conditions.

Secondly, it was necessary to determine whether measurements in vacuum conditions were required or an experiment in atmospheric conditions was acceptable. For this purpose, we had to assess the degree of influence of the ambient air and the equilibrium adsorbate layer on free surfaces on the temperature distribution in the nanocontact region.

Thirdly, it was necessary to determine whether it was possible to use a standard metallized probe as a metal electrode of a metal/sp^2 carbon or metal/semiconductor nanocontact.

Fourthly, it was necessary to estimate the time to establish the temperature distribution after the formation of the nanocontact under the conditions of the planned experiments.

To solve these problems, we estimated the temperature drops primarily on the auxiliary elements of the experimental device (primarily on the AFM probe cantilever), whose characteristic dimensions are large enough for calculations according to standard theory to provide sufficient accuracy. Significant calculation errors could be expected for the nanocontact region itself (most likely, towards lower thermal conductivity), which was taken into account in analysis of the results. We chose an experimental configuration with the requirement that most of the temperature drop created was concentrated in the nanocontact region. In this case, the maximum estimate of the expected value of the thermopower can be obtained by simply multiplying the temperature drop ΔT by the effective value of the thermoelectric coefficient S given by the theory (according to [21], it can reach 50 mV/K for the structure considered there).

Calculations of the temperature distribution were carried out by solving a standard equation of thermal conductivity of the form

$$\rho C_p \frac{\partial T}{\partial t} + \rho C_p \mathbf{u} \nabla T + \nabla \mathbf{q} = Q, \tag{5}$$

where ρ, C_p are the mass density and heat capacity of the substance; $\mathbf{u}$ is its local velocity (assumed to be identically equal to zero); $\mathbf{q}$ is the heat flow density; Q is the density of heat sources.

At this stage of the study, a linear relationship of the quantity $\mathbf{q}$ with the temperature gradient (Fourier's law) was postulated:

$$\mathbf{q} = -\kappa \nabla T, \tag{6}$$

where κ is the thermal conductivity.

The values of the material parameters given in Table 1 were used. The "Heat Transfer in Solids and Fluids" module of the COMSOL Multiphysics package was used, and the "Laminar Flow" module was also used to calculate heat transfer through the ambient air.

Table 1

Thermal parameters of materials used for simulation

Parameter	Parameter value for material			
	Si	Cu	Pt	Adsorbate (water)
Heat capacity, J/(kg·K)	700	375	133	4,200
Density, kg/m^3	2,329	8,960	21,450	1,000
Thermal conductivity, W/(m·K)	130	394	71.6	0.56

The contact of the AFM silicon probe with a copper substrate (thick plate) was simulated. The geometry of the tip and cantilever (Fig. 3,*a*, Table 2) was set in accordance with the parameters and image of the NSG10 probe provided on the manufacturer's website (NT-MDT, Russia). To simplify the calculations, a 2D problem was solved with the axis of symmetry passing through the center of the probe's contact region with a planar substrate. The shape of the tip of the probe was set as a cone conjugate to a sphere (Fig. 3,*b*). The radius of the sphere was taken to be 10 nm.

The same spherical surface was considered to be the interface of the tip with the substrate; it was assumed that the mechanical pressing action of the probe was strong enough for plastic deformation of copper in the contact region. The deviation of the 2D geometry from the true 3D geometry was compensated by setting the parameters of the cantilever material: its thermal conductivity and heat capacity in the left part (see Fig. 3,*a*) were set by tabular data, and then decreased along the radial coordinate according to the law $1/r$.

Table 2

Geometric parameters of AFM probe used in the simulation

Cantilever size, μm			Tip size, nm		Angle at cone vertex, degrees
Length	Width	Thickness	Height	Curvature radius	
125	27	3	$1.5 \cdot 10^4$	10	50

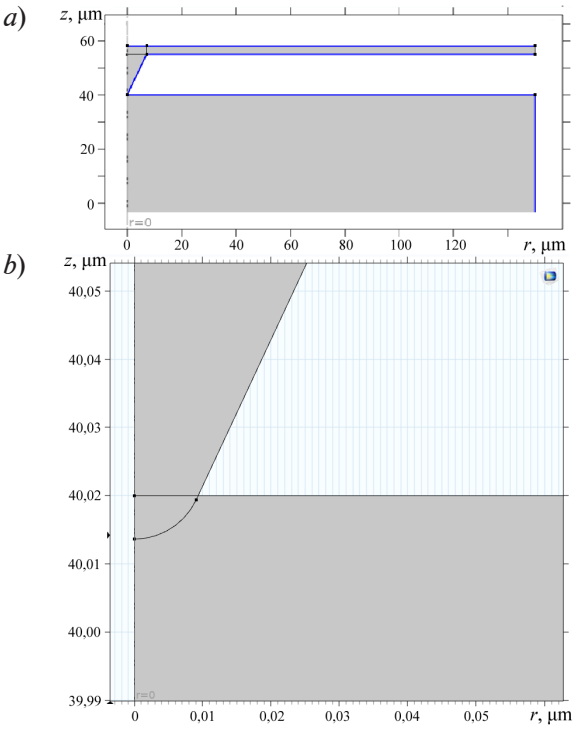

Fig. 3. Geometry of the problem of numerical simulation of temperature distribution over the nanocontact region of an AFM probe with a planar substrate: *a* is the general view, *b* is the nanocontact region

At the initial time, the substrate temperature was set to 0 °C, and the temperature of all parts of the probe was set to 100 °C (the specific values of these parameters do not affect the type of distribution if linear equations are used). Dirichlet boundary conditions were set for the lower boundary of the substrate (0 °C) and for the right end of the cantilever (100 °C). The problem of thermal conductivity was solved by simulating the steady-state temperature distribution, as well as the evolution dynamics of this steady state.

Fig. 4 shows the simulation results. The temperature distribution (see Fig. 4,*a*) was obtained for a silicon probe in contact with the clean surface of a copper plate in vacuum; heat transfer by electromagnetic radiation was considered insignificant [41]. As expected, under these conditions, the larger part of the temperature drop falls on the probe, whose material is characterized by lower thermal conductivity; the geometric factor (conical shape) also slightly increases its thermal resistance. The temperature changes significantly only in the probe region near the nanocontact, up to distances of the order of several magnitudes of its radius. The temperature drop on the cantilever is negligible: it obviously does not exceed 1% of the total temperature difference.

The simulation results for the steady-state temperature distribution in the presence of air are shown in Fig. 4,*b*. Heat flows through the air due to thermal conductivity and convection were taken into account. We can observe a continuous and approximately linear temperature variation in the gap between the plate and the cantilever. It is known from the literature [41, 51] that the total heat flow between the AFM probe and the substrate at atmospheric pressure is largely determined by convection. However, the calculation showed that the thermal conductivity of the cantilever is sufficient to prevent the convective heat flow from noticeably distorting the temperature distribution near the nanocontact and changing the temperature of the base (wide part) of the probe by more than a fraction of a percent of its total drop.

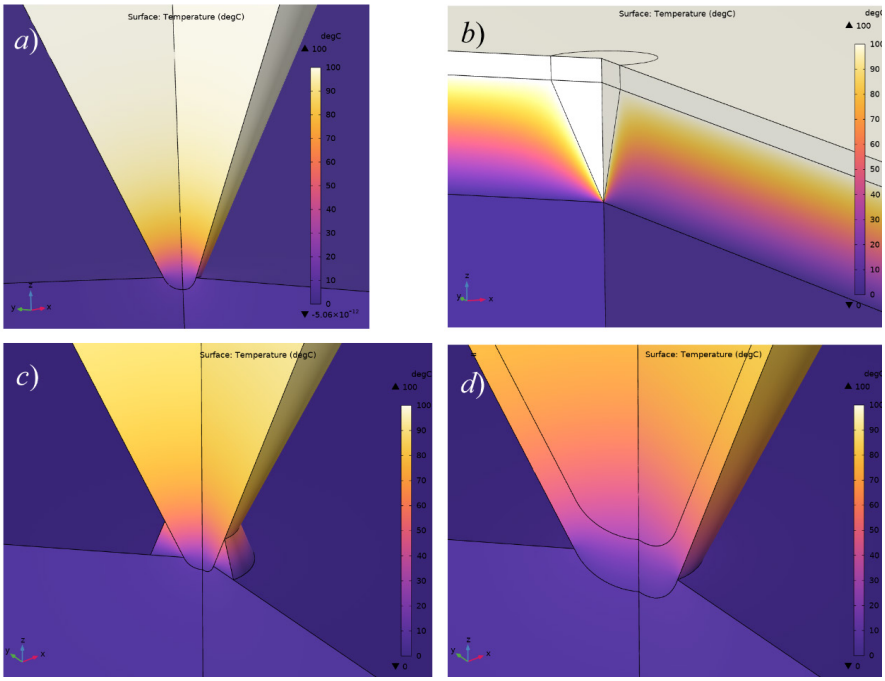

Fig. 4. Solution of the problem of thermal conductivity in the nanocontact region of AFM silicon probe with copper substrate.
Temperature distributions under vacuum conditions (*a*) are presented, taking into account thermal conductivity through air (*b*) and the presence of a liquid layer on the surface (*c*), as well as in the presence of a platinum metallization layer of the probe (*d*)

The above also applies to the influence of a layer of water and adsorbed gases covering free surfaces under normal conditions [41, 42]. To assess the degree of influence of the adsorbate on the temperature distribution, a circular layer (20 nm thick) with thermal conductivity of water was introduced into the model in the nanocontact region (see Fig. 4,*c*). Comparing the calculation result shown in Fig. 4,*c* with the data in Fig. 4,*a*, we can see that the influence of the liquid layer on the temperature distribution in the nanocontact is minimal.

Finally, the impact of metallization of the probe was assessed. The calculated results of the temperature distribution for a probe coated with a 30 nm thick platinum layer are shown in Fig. 4,*d*. Evidently, a significant part of the temperature drop falls precisely on the platinum layer, which has a relatively low thermal conductivity. Therefore, using platinum-coated probes in planned experiments is undesirable. It is preferable to use probes with a different coating, for example, made of gold (its thermal conductivity is 317 W/(m·K)) or another metal with high thermal conductivity (for example, copper or silver).

The simulation results for the dynamics of the steady-state temperature distribution evolving after the probe was brought into contact with the surface showed that in the absence of air, the characteristic time of such a process does not exceed several microseconds (Table 3). Such a delay can be neglected in experiments recording the AFM force-distance curve. As for the simulation of heat transfer through air, the stabilization time for the temperature distribution in the air to establish turned out to be significantly longer, amounting to milliseconds. However, this, apparently, should not interfere with measurements, since the heat flow through the air itself should have little effect on the temperature of the probe and the substrate, and temperature stabilization time for the contact area still amounts to microseconds. This conclusion is consistent with the literature data [45].

Thus, the numerical simulation carried out in a simplified COMSOL model allowed solving its main problem: to obtain a positive answer to the question of the possibility of conducting experiments to identify the features of the thermoelectric effect in atmospheric conditions using NanoDST AFM and standard probes.

Table 3

Simulated times for steady-state temperature distribution to establish under varying conditions

Condition	Stabilizatiion time, μs	
	Uncoated probe	Probe with platinum coating
Basic model (vacuum)	3.0	4.0
Atmospheric air	1.3	4.0
Liquid layer on surface	$4.0 \cdot 10^3$	$4.0 \cdot 10^3$
Air + liquid	$4.0 \cdot 10^3$	–

Conclusion

The paper presents the results of the initial stage of the research conducted at the Higher School of Engineering and Physics at Peter the Great St. Petersburg Polytechnic University and dedicated to the study of nanoscale features of the thermoelectric effect. The ultimate goal is to create thermoelectric generators with improved performance characteristics.

An experimental verification of the theory known from the literature describing the possibility of achieving high values of thermoelectric performance by using the phenomenon of entrainment of electric charge carriers by ballistic flow of phonons in a film nanostructure is chosen as the primary objective. We believe that the verification of this concept can be carried out by relatively simple means, namely, by measuring the thermoelectric characteristics of nanocarbon nanoislands and graphene sheets formed on a silicon substrate.

The results of numerical simulation indicate that such measurements can be carried out using an atmospheric atomic force microscope with standard probes after its retrofitting with systems for controlling and measuring the sample temperature.

REFERENCES

1. **Sadeghi H.,** Quantum and phonon interference-enhanced molecular-scale thermoelectricity, J. Phys. Chem. C. 123 (20) (2019) 12556–12562.

2. **Karlström O., Linke H., Karlström G., Wacker A.,** Increasing thermoelectric performance using coherent transport, Phys. Rev. B. 84 (11) (2011) 113415.

3. **Mahan G. D., Sofo J. O.,** The best thermoelectric, Proc. Natl. Acad. Sci. USA. 93 (15) (1996) 7436–7439.

4. **Dmitriyev A. S.,** Vvedeniye v nanoteplofiziku, 2-e izdaniye, elektronnoye [Introduction to nanothermophysics, Second electronic edition], "Laboratoriya Znaniy" Publishing, Moscow, 2020 (in Russian).

5. **Zevalkink A., Smiadak D. M., Blackburn J. L., et al.,** A practical field guide to thermoelectrics: Fundamentals, synthesis, and characterization, Appl. Phys. Rev. 5 (2) (2018) 021303.

6. **Lyeo H. K., Khajetoorians A. A., Shi L., et al.,** Profiling the thermoelectric power of semiconductor junctions with nanometer resolution, Science. 303 (5659) (2004) 816–818.

7. **Ikeda H., Salleh F.,** Influence of heavy doping on Seebeck coefficient in silicon-on-insulator, Appl. Phys. Lett. 96 (1) (2010) 012106.

8. **Salleh F., Suzuki Y., Miwa K., Ikeda H.,** Modulation of Seebeck coefficient for silicon-on-insulator layer induced by bias-injected carriers, Appl. Phys. Lett. 103 (6) (2013) 062107.

9. **Tretiakov O. A., Abanov Ar., Sinova J.,** Holey topological thermoelectrics, Appl. Phys. Lett. 99 (11) (2011) 113110.

10. **Krali E., Durrani Z. A. K.,** Seebeck coefficient in silicon nanowire arrays, Appl. Phys. Lett. 102 (14) (2013) 143102.

11. **Taniguchi T., Terada T., Komatsubara Y., et al.,** Phonon transport in the nano-system of Si and SiGe films with Ge nanodots and approach to ultralow thermal conductivity, Nanoscale. 13 (9) (2021) 4971–4977.

12. **Bergfield J. P., Solis M. A., Stafford C. A.,** Giant thermoelectric effect from transmission supernodes, ACS Nano. 4 (9) (2010) 5314–5320.

13. **Khvesyuk V. I., Scriabin A. S.,** Thermal conductivity of nanostructures, High Temperature. 55 (3) (2017) 428–450.

14. **Dubi Y., Di Ventra M.,** Colloquium: Heat flow and thermoelectricity in atomic and molecular junctions, Rev. Mod. Phys. 83 (1) (2011) 131–155.

15. **Sadeghi H., Sangtarash S., Lambert C. J**., Oligoyne molecular junctions for efficient room temperature thermoelectric power generation, Nano Lett. 15 (11) (2015) 7467–7472.

16. **Cui L., Miao R., Wang K., et al.,** Peltier cooling in molecular junctions, Nat. Nanotechnol. 13 (2) (2018) 122–127.

17. **Liu C., Lu P., Chen W., et al.,** Phonon transport in graphene based materials, Phys. Chem. Chem. Phys. 23 (46) (2021) 26030–26060.

18. **Varlamov A. A., Kavokin A. V., Luk'yanchuk I. A., Sharapov S. G.,** Anomalous thermoelectric and thermomagnetic properties of graphene, Phys.– Usp. 55 (11) (2012) 1146–1151.

19. **Wu Q., Sadeghi H., Garcна-Suбrez V. M., et al.,** Thermoelectricity in vertical graphene-C_{60}-graphene architectures, Sci. Rep. 7 (15 Sept) (2017) 11680.

20. **Eidelman E. D.,** On a carbon nanostructure-based thermoelectric converter with record parameters, Semicond. 51 (7) (2017) 906–908.

21. **Eidelman E. D.,** Thermoelectric effect and a thermoelectric generator based on carbon nanostructures: Achievements and prospects, Phys.–Usp. 64 (6) (2021) 535–557.

22. **Eydelman E. D., Vul' A. Ya.,** The strong thermoelectric effect in nanocarbon generated by ballistic phonon drag of electrons, J. Phys. Condens. Matter. 19 (7) (2007) 266210–266223.

23. **Koniakhin S.V., Eidelman E. D.,** Phonon drag thermopower in graphene in equipartition regime, Europhys. Lett. 103 (3) (2013) 37006.

24. **Andronov A., Budylina E., Shkitun P., et al.,** Characterization of thin carbon films capable of low-field electron emission, J. Vac. Sci. Technol. B. 36 (2) (2018) 02C108.

25. **Gabdullin P., Zhurkin A., Osipov V., et al.,** Thin carbon films: Correlation between morphology and field-emission capability, Diam. Relat. Mater. 105 (May) (2020) 107805.

26. **Eidelman E. D., Arkhipov A. V.,** Field emission from carbon nanostructures: Models and experiment, Phys.–Usp. 63 (7) (2020) 648–667.

27. **Bogachek E. N., Kulik I. O., Omel'yanchuk A. N., Shkorbatov A. G.,** Drag-related thermo-emf in metallic systems containing a point contact, JETP Lett. 41 (12) (1985) 633–636.

28. **Shklyarevskii O. I., Jansen A. G. M., Hermsen J. G. H., Wyder P.,** Thermoelectric voltage between identical metals in a point-contact configuration, Phys. Rev. Lett. 57 (11) (1986) 1374–1377.

29. **Trzcinski R., Gmelin E., Queisser H. J.,** Quenched phonon drag in silicon microcontacts, Phys. Rev. Lett. 56 (10) (1986) 1086–1089.

30. **Weber L., Lehr M., Gmelin E.,** Reduction of the thermopower in semiconducting point contacts, Phys. Rev. B. 46 (15) (1992) 9511–9514.

31. **Mahan G. D., Claro F.,** Nonlocal theory of thermal conductivity, Phys. Rev. B. 38 (3) (1988) 1963–1969.

32. **Mahan G. D.,** The Benedicks effect: Nonlocal electron transport in metals, Phys. Rev. B. 43 (5) (1991) 3945–3951.

33. **Grigorenko A. N., Nikitin P. I., Jelski D. A., George T. F.,** Thermoelectric phenomena in metals under large temperature gradients, J. Appl. Phys. 69 (5) (1991) 3375–3377.

34. **Vermeersch B., Shakouri A.,** Nonlocality in microscale heat conduction, https://api.semanticscholar.org/CorpusID:118469205. (Accessed September 11, 2023).

35. **Koh Y. K., Cahill D. G., Sun B.,** Nonlocal theory for heat transport at high frequencies, Phys. Rev. B. 90 (20) (2014) 205412.

36. **Ezzahri Y., Joulain K., Ordonez-Miranda J.,** Heat transport in semiconductor crystals: Beyond the local-linear approximation, J. Appl. Phys. 128 (10) (2020) 105104.

37. **Baratifarimani R., Shomali Z.,** Implementation of nonlocal non-Fourier heat transfer for semiconductor nanostructures, https://api.semanticscholar.org/CorpusID:259317109 (Accessed September 11, 2023).

38. **Cahill D. G., Ford W. K., Goodson K. E., et al.,** Nanoscale thermal transport, J. Appl. Phys. 93 (2) (2003) 793–818.

39. **Wang C., Chen F., Sun K., et al.,** Instruments for measuring Seebeck coefficient of thin film thermoelectric materials: A mini-review, Rev. Sci. Instrum. 89 (10) (2018) 101501.

40. **Kim S. J., We J. H., Kim G. S., Cho B. J.,** Simultaneous measurement of the Seebeck coefficient and thermal conductivity in the cross-sectional direction of thermoelectric thick film, J. Appl. Phys. 112 (10) (2012) 104511.

41. **Majumdar A.,** Scanning thermal microscopy, Annu. Rev. Mater. Sci. 29 (1) (1999) 505–585.

42. **Shi L., Plyasunov S., Bachtold A., et al.,** Scanning thermal microscopy of carbon nanotubes using batch-fabricated probes, Appl. Phys. Lett. 77 (26) (2000) 4295–4297.

43. **Fletcher P. C., Lee B., King W. P.,** Thermoelectric voltage at a nanometer-scale heated tip point contact, Nanotechnology. 23 (3) (2011) 035401.

44. **Nakamoto G., Nakabayashi Y.,** Development of a two dimensional scanning Seebeck coefficient measurement system by a micro-probe method, Intermetallics. 32 (Jan) (2013) 233–238.

45. **Tan A., Sadat S., Reddy P.,** Measurement of thermopower and current-voltage characteristics of molecular junctions to identify orbital alignment, Appl. Phys. Lett. 96 (1) (2010) 013110.

46. **Cahill D. G., Braun P. V., Chen G., et al.,** Nanoscale thermal transport. II. 2003–2012, Appl. Phys. Rev. 1 (1) (2014) 011305.

47. **De Boor J., Müller E.,** Data analysis for Seebeck coefficient measurements, Rev. Sci. Instrum. 84 (6) (2013) 065102.

48. **Bizyaev I., Gabdullin P., Chumak M., et al.,** Low-field electron emission capability of thin films on flat silicon substrates: Experiments with Mo and general model for refractory metals and carbon, Nanomaterials. 11 (12) (2021) 3350.

49. **Arkhipov A. V., Eidelman E. D., Zhurkin A. M., et al.,** Low-field electron emission from carbon cluster films: Combined thermoelectric/hot-electron model of the phenomenon, Fuller. Nanotub. Car. N. 28 (4) (2020) 286–294.

50. **Weber L., Gmelin E.,** A new device for transport measurements on point contacts, Rev. Sci. Instrum. 63 (1) (1992) 211–217.

51. **Williams C. C., Wickramasinghe H. K.,** Scanning thermal profiler, Appl. Phys. Lett. 49 (23) (1986) 1587–1589.

СПИСОК ЛИТЕРАТУРЫ

1. **Sadeghi H.** Quantum and phonon interference-enhanced molecular-scale thermoelectricity // The Journal of Physical Chemistry C. 2019. Vol. 123. No. 20. Pp. 12556–12562.

2. **Karlström O., Linke H., Karlström G., Wacker A.** Increasing thermoelectric performance using coherent transport // Physical Review B. 2011. Vol. 84. No. 11. P. 113415.

3. **Mahan G. D., Sofo J. O.** The best thermoelectric // Proceedings of the National Academy of Sciences of the USA. 1996. Vol. 93. No. 15. Pp. 7436–7439.

4. **Дмитриев А. С.** Введение в нанотеплофизику. 2-е изд., электронное. М.: Лаборатория знаний, 2020. 793 с.

5. **Zevalkink A., Smiadak D. M., Blackburn J. L., et al.** A practical field guide to thermoelectrics: Fundamentals, synthesis, and characterization // Applied Physics Reviews. 2018. Vol. 5. No. 2. P. 021303.

6. **Lyeo H. K., Khajetoorians A. A., Shi L., Pipe K. P., Ram R. J., Shakouri A., Shih C. K.** Profiling the thermoelectric power of semiconductor junctions with nanometer resolution // Science. 2004. Vol. 303. No. 5659. Pp. 816–818.

7. **Ikeda H., Salleh F.** Influence of heavy doping on Seebeck coefficient in silicon-on-insulator // Applied Physics Letters. 2010. Vol. 96. No. 1. P. 012106.

8. **Salleh F., Suzuki Y., Miwa K., Ikeda H.** Modulation of Seebeck coefficient for silicon-on-insulator layer induced by bias-injected carriers // Applied Physics Letters. 2013. Vol. 103. No. 6. P. 062107.

9. **Tretiakov O. A., Abanov Ar., Sinova J.** Holey topological thermoelectrics // Applied Physics Letters. 2011. Vol. 99. No. 11. P. 113110.

10. **Krali E., Durrani Z. A. K.** Seebeck coefficient in silicon nanowire arrays // Applied Physics Letters. 2013. Vol. 102. No. 14. P. 143102.

11. **Taniguchi T., Terada T., Komatsubara Y., Ishibe T., Konoike K., Sanada A., Naruse N., Mera Y., Nakamura Y.** Phonon transport in the nano-system of Si and SiGe films with Ge nanodots and approach to ultralow thermal conductivity // Nanoscale. 2021. Vol. 13. No. 9. Pp. 4971–4977.

12. **Bergfield J. P., Solis M. A., Stafford C. A.** Giant thermoelectric effect from transmission supernodes // ACS (American Chemical Society) Nano. 2010. Vol. 4. No. 9. Pp. 5314–5320.

13. **Хвесюк В. И., Скрябин А. С.** Теплопроводность наноструктур // Теплофизика высоких температур. 2017. Т. 55. № 3. С. 447–471.

14. **Dubi Y., Di Ventra M.** Colloquium: Heat flow and thermoelectricity in atomic and molecular junctions // Reviews of Modern Physics. 2011. Vol. 83. No. 1. Pp. 131–155.

15. **Sadeghi H., Sangtarash S., Lambert C. J**. Oligoyne molecular junctions for efficient room temperature thermoelectric power generation // Nano Letters. 2015. Vol. 15. No. 11. Pp. 7467–7472.

16. **Cui L., Miao R., Wang K., Thompson D., Zotti L. A., Cuevas J. C., Meyhofer E., Reddy P.** Peltier cooling in molecular junctions // Nature Nanotechnology. 2018. Vol. 13. No. 2. Pp. 122–127.

17. **Liu C., Lu P., Chen W., Zhao Y., Chen Y.** Phonon transport in graphene based materials // Physical Chemistry, Chemical Physics. 2021. Vol. 23. No. 46. Pp. 26030–26060.

18. **Варламов А. А., Кавокин А. В., Лукьянчук И. А., Шарапов С. Г.** Аномальные термоэлектрические и термомагнитные свойства графена // Успехи физических наук. 2012. Т. 182. № 11. С. 1229–1234.

19. **Wu Q., Sadeghi H., Garcна-Suбrez V. M., Ferrer J., Lambert C. J.** Thermoelectricity in vertical graphene-C_{60}-graphene architectures // Scientific Reports. 2017. Vol. 7. 15 September. P. 11680.

20. **Эйдельман Е. Д.** Термоэлектрический преобразователь с рекордными параметрами на основе углеродных наноструктур: разработка научных основ // Физика и техника полупроводников. 2017. Т. 51. № 7. С. 944–947.

21. **Эйдельман Е. Д.** Термоэлектрический эффект и термоэлектрический генератор на основе углеродных наноструктур: достижения и перспективы // Успехи физических наук. 2021. Т. 191. № 6. С. 561–585.

22. **Eydelman E. D., Vul' A. Ya.** The strong thermoelectric effect in nanocarbon generated by ballistic phonon drag of electrons // Journal of Physics: Condensed Matter. 2007. Vol. 19. No. 7. Pp. 266210–266223.

23. **Koniakhin S.V., Eidelman E. D.** Phonon drag thermopower in graphene in equipartition regime // Europhysics Letters. 2013. Vol. 103. No. 3. P. 37006.

24. **Andronov A., Budylina E., Shkitun P., Gabdullin P., Gnuchev N., Kvashenkina O., Arkhipov A.**

Characterization of thin carbon films capable of low-field electron emission // Journal of Vacuum Science & Technology B. 2018. Vol. 36. No. 2. P. 02C108.

25. **Gabdullin P., Zhurkin A., Osipov V., Besedina N., Kvashenkina O., Arkhipov A.** Thin carbon films: Correlation between morphology and field-emission capability // Diamond & Related Materials. 2020. Vol. 105. May. P. 107805.

26. **Эйдельман Е. Д., Архипов А. В.** Полевая эмиссия из углеродных наноструктур: модели и эксперимент // Успехи физических наук. 2020. Т. 190. № 7. С. 693–714.

27. **Богачек Э. Н., Кулик И. О., Омельянчук А. Л., Шкорбатов А. Г.** Термоэдс увлечения в металлических системах, содержащих микроконтакт // Письма в Журнал экспериментальной и теоретической физики. 1985. Т. 41. № 12. С. 519–521.

28. **Shklyarevskii O. I., Jansen A. G. M., Hermsen J. G. H., Wyder P.** Thermoelectric voltage between identical metals in a point-contact configuration // Physical Review Letters. 1986. Vol. 57. No. 11. Pp. 1374–1377.

29. **Trzcinski R., Gmelin E., Queisser H. J.** Quenched phonon drag in silicon microcontacts // Physical Review Letters. 1986. Vol. 56. No. 10. Pp. 1086–1089.

30. **Weber L., Lehr M., Gmelin E.** Reduction of the thermopower in semiconducting point contacts // Physical Review B. 1992. Vol. 46. No. 15. 9511–9514.

31. **Mahan G. D., Claro F.** Nonlocal theory of thermal conductivity // Physical Review B. 1988. Vol. 38. No. 3. Pp. 1963–1969.

32. **Mahan G. D.** The Benedicks effect: Nonlocal electron transport in metals // Physical Review B. 1991. Vol. 43. No. 5. Pp. 3945–3951.

33. **Grigorenko A. N., Nikitin P. I., Jelski D. A., George T. F.** Thermoelectric phenomena in metals under large temperature gradients // Journal of Applied Physics. 1991. Vol. 69. No. 5. Pp. 3375–3377.

34. **Vermeersch B., Shakouri A.** Nonlocality in microscale heat conduction. https://api.semanticscholar.org/CorpusID:118469205 (Дата обращения 11.09.2023).

35. **Koh Y. K., Cahill D. G., Sun B.** Nonlocal theory for heat transport at high frequencies // Physical Review B. 2014. Vol. 90. No. 20. P. 205412.

36. **Ezzahri Y., Joulain K., Ordonez-Miranda J.** Heat transport in semiconductor crystals: Beyond the local-linear approximation // Journal of Applied Physics. 2020. Vol. 128. No. 10. P. 105104.

37. **Baratifarimani R., Shomali Z.** Implementation of nonlocal non-Fourier heat transfer for semiconductor nanostructures. https://api.semanticscholar.org/ CorpusID:259317109 (Дата обращения 11.09.2023).

38. **Cahill D. G., Ford W. K., Goodson K. E., Mahan G. D., Majumdar A., Maris H. J., Merlin R., Phillpot S. R.** Nanoscale thermal transport // Journal of Applied Physics. 2003. Vol. 93. No. 2. Pp. 793–818.

39. **Wang C., Chen F., Sun K., Chen R., Li M., Zhou X., Sun Y., Chen D., Wang G.** Instruments for measuring Seebeck coefficient of thin film thermoelectric materials: A mini-review // Review of Scientific Instruments. 2018. Vol. 89. No. 10. P. 101501.

40. **Kim S. J., We J. H., Kim G. S., Cho B. J.** Simultaneous measurement of the Seebeck coefficient and thermal conductivity in the cross-sectional direction of thermoelectric thick film // Journal of Applied Physics. 2012. Vol. 112. No. 10. P. 104511.

41. **Majumdar A.** Scanning thermal microscopy // Annual Review of Materials Science. 1999. Vol. 29. No. 1. Pp. 505–585.

42. **Shi L., Plyasunov S., Bachtold A., McEuen P. L., Majumdar A.** Scanning thermal microscopy of carbon nanotubes using batch-fabricated probes // Applied Physics Letters. 2000. Vol. 77. No. 26. Pp. 4295–4297.

43. **Fletcher P. C., Lee B., King W. P.** Thermoelectric voltage at a nanometer-scale heated tip point contact // Nanotechnology. Vol. 23. No. 3. P. 035401.

44. **Nakamoto G., Nakabayashi Y.** Development of a two dimensional scanning Seebeck coefficient measurement system by a micro-probe method // Intermetallics. 2013. Vol. 32. January. Pp. 233–238.

45. **Tan A., Sadat S., Reddy P.** Measurement of thermopower and current-voltage characteristics of molecular junctions to identify orbital alignment // Applied Physics Letters. 2010. Vol. 96. No. 1. P. 013110.

46. **Cahill D. G., Braun P. V., Chen G., et al.** Nanoscale thermal transport. II. 2003–2012 // Applied Physics Reviews. 2014. Vol. 1. No. 1. P. 011305.

47. **De Boor J., Müller E.** Data analysis for Seebeck coefficient measurements // Review of Scientific Instruments. 2013. Vol. 84. No. 6. P. 065102.

48. **Bizyaev I., Gabdullin P., Chumak M., Babyuk V., Davydov S., Osipov V., Kuznetsov A., Kvashenkina O., Arkhipov A.** Low-field electron emission capability of thin films on flat silicon substrates: Experiments with Mo and general model for refractory metals and carbon // Nanomaterials. 2021. Vol. 11. No. 12. P. 3350.

49. **Arkhipov A. V., Eidelman E. D., Zhurkin A. M., Osipov V. S., Gabdullin P. G.** Low-field electron emission from carbon cluster films: Combined thermoelectric/hot-electron model of the phenomenon // Fullerenes, Nanotubes and Carbon Nanostructures. 2020. Vol. 28. No. 4. Pp. 286–294.

50. **Weber L., Gmelin E.** A new device for transport measurements on point contacts // Review of Scientific Instruments. 1992. Vol. 63. No. 1. Pp. 211–217.

51. **Williams C. C., Wickramasinghe H. K.** Scanning thermal profiler // Applied Physics Letters. 1986. Vol. 49. No. 23. Pp. 1587–1589.

THE AUTHORS

TROFIMOVICH Karina R.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
karina-khasanova-2001@mail.ru

GABDULLIN Pavel G.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
gabdullin_pg@spbstu.ru
ORCID: 0000-0002-2519-2577

ARKHIPOV Alexander V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
arkhipov@rphf.spbstu.ru
ORCID: 0000-0002-3321-7797

СВЕДЕНИЯ ОБ АВТОРАХ

ТРОФИМОВИЧ Карина Робертовна – *студентка Института электроники и телекоммуникаций Санкт-Петербургского политехнического университета Петра Великого, Санкт-Петербург, Россия.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
karina-khasanova-2001@mail.ru

ГАБДУЛЛИН Павел Гарифович – *кандидат технических наук, доцент, директор Научно-технологического центра «Нейропрогнозирование материалов и технологий электронной промышленности» Санкт-Петербургского политехнического университета Петра Великого, Санкт-Петербург, Россия.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
gabdullin_pg@spbstu.ru
ORCID: 0000-0002-2519-2577

АРХИПОВ Александр Викторович – *доктор физико-математических наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого, Санкт-Петербург, Россия.*

195251, Россия, г. Санкт-Петербург, Политехническая ул., 29

arkhipov@rphf.spbstu.ru

ORCID: 0000-0002-3321-7797

Received 16.09.2023. Approved after reviewing 01.12.2023. Accepted 01.12.2023.

Статья поступила в редакцию 16.09.2023. Одобрена после рецензирования 01.12.2023. Принята 01.12.2023.

PHYSICAL ELECTRONICS

Original article
UDC 621.385.6
DOI: https://doi.org/10.18721/JPM.16410

ENHANCEMENT OF THE 4-mm WAVELENGTH GYROTRON EFFICIENCY BY MULTISTAGE ENERGY RECOVERY

O. I. Louksha✉, *P. A. Trofimov*, *A. G. Malkin*

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ louksha@rphf.spbstu.ru

Abstract. This study presents the results of a complex physical modeling of a moderate power gyrotron operating at the 4-mm wavelength range. The characteristics of electrodes and magnetic coils in a four-stage recovery collector were optimized taking into account the coordinate and velocity distributions of electrons. These distributions were obtained through a trajectory analysis in the electron optical system and calculation of electron-wave interaction in the gyrotron cavity. To reduce parasitic effects of the bundles of a toroidal solenoid used to create an azimuthal magnetic field in the collector region, a sectioned electron beam was employed. The study demonstrated that the gyrotron's total efficiency of approximately 79 % could be achieved, being close to the maximum efficiency value achievable with separation of electron fractions with different energies, provided that the current of electrons reflected from a collector should not exceed 1% of the total current of an electron beam.

Keywords: microwave electronics, gyrotron, helical electron beam, recuperation, residual electron energy recovery

Funding: The reported study was funded by Russian Science Foundation (Grant No. 22-29-00136).

Citation: Louksha O. I., Trofimov P. A., Malkin A. G., Enhancement of the 4-mm wavelength gyrotron efficiency by multistage energy recovery, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 118–133. DOI: https://doi.org/10.18721/JPM.16410

Научная статья
УДК 621.385.6
DOI: https://doi.org/10.18721/JPM.16410

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ГИРОТРОНА С ДЛИНОЙ ВОЛНЫ 4 мм ЗА СЧЕТ МНОГОСТУПЕНЧАТОЙ РЕКУПЕРАЦИИ

О. И. Лукша ✉, *П. А. Трофимов*, *А. Г. Малкин*

Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Россия

✉ louksha@rphf.spbstu.ru

Аннотация. В работе представлены результаты комплексного физического моделирования гиротрона средней мощности, работающего на длине волны 4 мм. Проведена оптимизация характеристик электродов и магнитных катушек в коллекторе с четырехступенчатой рекуперацией с учетом распределения электронов по координатам и скоростям. Эти распределения были получены путем траекторного анализа в электронно-оптической системе и расчета электронно-волнового взаимодействия в резонаторе гиротрона. Для снижения паразитного воздействия связок тороидального соленоида,

используемого для создания азимутального магнитного поля в области коллектора, использован секционированный электронный пучок. Исследование показало, что можно получить общий КПД гиротрона около 79 %, что близко к максимальному значению, достижимому при идеальном разделении электронных фракций с разными энергиями, при условии, что ток отраженных от коллектора электронов не должен превышать 1 % от общего тока электронного пучка.

Ключевые слова: СВЧ-электроника, гиротрон, винтовой электронный поток, рекуперация, возвращение остаточной энергии электронов

Финансирование: Исследование выполнено при финансовой поддержке Российского научного фонда (грант № 22-29-00136).

Ссылка для цитирования: Лукша О. И., Трофимов П. А., Малкин А. Г. Повышение эффективности гиротрона с длиной волны 4 мм за счет многоступенчатой рекуперации // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. 118–133. DOI: https://doi.org/10.18721/ JPM.16410

Introduction

In recent years, there has been an intensive search for new ways to improve powerful gyrotron-type devices due to the wide possibilities of their practical use. Gyrotrons occupy a leading position among effective sources of powerful microwave radiation in the millimeter and submillimeter wavelength ranges. They are irreplaceable in such an important application as electron-cyclotron plasma heating and current drive (ECH&CD) in magnetic confinement fusion systems designed to produce energy through controlled thermonuclear fusion (see, for example, Refs. [1 – 3]). The requirements to gyrotrons designed for thermonuclear fusion are exceptionally high. The ITER project requires gyrotrons operating at a frequency of 170 GHz, delivering an output power of approximately 1 MW and achieving a total efficiency exceeding 50 % [4]. The development of a new generation of nuclear fusion reactors will require resolving numerous physical and engineering tasks to improve gyrotrons performance. The DEMO project envisions usage of gyrotrons with a frequency exceeding 200 GHz and a total efficiency greater than 60 % at megawatt-level power [5]. Enhancement of device's energy efficiency simplifies dissipation of spent beam energy at collector, which is critical for reliable and long-time operation of high-power gyrotrons operating in the continuous wave (CW) regime. Achieving such high efficiencies is one of the primary objectives for developers of powerful gyrotrons today.

Increasing the gyrotrons' efficiency, as well as other vacuum sources of microwave radiation, can be achieved by recovering residual energy of spent electron beam in the collector. Almost all megawatt-level gyrotrons used in thermonuclear fusion systems are equipped by collectors with single-stage recovery which increase their total efficiency to 50 – 55% [6 – 8]. A further increase in efficiency is possible with multistage energy recovery systems. Such systems require spatial separation of electron beam fractions with different energies and deposition of these fractions on collector sections with different depressing potentials. However, to the best of our knowledge, no experiments have been conducted on gyrotrons with multistage recovery collector systems. The implementation of such systems has proven challenging due to specifics of velocity and coordinate distributions of electrons in helical electron beams (HEBs) of gyrotrons and due to the presence of residual magnetic field in collector region. A promising solution for spatial separation of electrons in gyrotron HEBs is the use of crossed electric and magnetic fields [9 – 11].

At Peter the Great St. Petersburg Polytechnic University (SPbPU), a possibility of electron separation in longitudinal electric and azimuthal magnetic fields has been proposed and investigated theoretically for development of gyrotrons with multistage recovery collectors [12].

The achievement of high gyrotrons' efficiency implies the high efficiency of transformation of electron energy to electromagnetic field energy in the cavity. The efficiency of this transformation is determined by quality of the HEB formed in the electron optical system. The research aimed

at improving the HEB quality was carried out at SPbPU using an experimental gyrotron with a frequency of 74.2 GHz and an output power of approximately 100 kW [13 – 15]. This gyrotron was equipped with unique diagnostic complex capable of measuring the HEB parameters and regulating distributions of electric and magnetic fields in electron optical system. The initial version of a four-stage recovery collector for this gyrotron was described in Ref. [16]. In this work, the collector system geometry was significantly modified and distributions of electric and magnetic fields were optimized, resulting in a noticeable enhancement of residual energy recovery efficiency. The multistage recovery system described in this paper has already been implemented in the SPbPU gyrotron.

This article is organized as follows.

Section I presents the results of trajectory analysis of the HEB in the electron optical system of the gyrotron. The approach that considers initial spread of electron velocities caused by roughness of cathode surface and by thermal velocity spread was implemented [17]. Regulation of electric field distribution in the cathode region allowed to increase the HEB's quality and maximum average pitch factor.

Section II describes the results of the Particle-in-Cell (PIC) simulation in the cavity region of the gyrotron using the input HEB parameters determined from the trajectory analysis in the electron optical system. The output radiation parameters at $TE_{12,3}$ operating mode and the characteristics of spent electron beam entering the collector region were obtained in this simulation.

Section III presents the results of trajectory analysis in the four-stage recovery collector, where parameters of electrodes and magnetic coils were optimized to achieve the maximum recovery efficiency and to minimize the current of electrons reflected from the collector. All calculations were performed using the CST Studio Suite software. Specifics of conducting calculations in CST Studio Suite, such as model construction, meshing, choice of computational parameters, etc., employed in this study, were similar to those described earlier in Refs. [16, 18].

I. Electron optical system

Table 1 shows the parameters of the SPbPU gyrotron operating regime. Fig. 1 shows a schematic drawing of the electron optical system elements in the r–z plane, including calculated electron trajectories. An electron beam in the gyrotron is formed using a three-electrode magnetron-injection gun (MIG). Thus, it is possible to modify parameters of the HEB by varying the voltage U_a between the anode and device's body. The accelerating voltage U_0 between the cathode and body which determines an average electron energy in the HEB remained constant at 30 kV during the calculations described below. The cathode assembly includes a control electrode positioned behind the cathode emissive strip, which can be used to optimize distribution of electric field in the cathode region and to minimize the velocity spread of electrons by adjusting the voltage at control electrode U_{cont} [19]. At the values of U_0, B_0, B_c and I_b indicated in Table 1 and in the case of $U_a = 0$ and $U_{cont} = 0$, the average pitch factor of electrons α was approximately 1.3. All HEB parameters mentioned in this section were determined in the central plane of the cavity $z = 260.5$ mm (see Fig. 1). The objective of trajectory analysis in the electron optical system was to achieve a high pitch factor ($\alpha > 1.5$) with a low velocity spread and minimal electron reflection from the magnetic mirror by regulating the voltages U_a and U_{cont}. Additionally, the magnetic compression coefficient B_0/B_c was varied by adjusting the magnetic field induction near cathode B_c to ensure optimal beam radius in the cavity (see Section II).

The present calculations differ from the previous ones described in Ref. [16] by taking into account initial electron velocity spread caused by roughness of cathode surface and thermal velocity spread. As shown in Ref. [17], the velocity characteristics of electrons in the beam can be made approximately coincided for two cathode models: (a) a rough cathode with inhomogeneities on its surface in the form of micron-sized hemispheres with radius r_0 and (b) a smooth cathode with Maxwell distribution of the initial velocities at an increased effective cathode temperature T^* and a spread of the angle φ between the direction of the initial velocity vector and direction of the normal. The angle φ is uniformly distributed in the range from $-\Delta\varphi$ to $+\Delta\varphi$. For instance, an initial transverse velocity spread is the same for the model (a) at $r_0 = 14$ μm and for the model (b) at $T^* = 67\,000$ K and $\Delta\varphi = \pm 90°$. Therefore, at the appropriate values of T^* and $\Delta\varphi$ the velocity spread factor associated with roughness of the cathode surface can be taken into account when performing a three-dimensional trajectory analysis in the electron optical gyrotron system

with a smooth cathode. It should be noted that at the T^* value being much higher than the actual cathode temperature T_c, the average electron energy in the HEB $\langle W \rangle$ and the energy spread δW additionally increase. An increase in the average energy can be compensated by changing of the accelerating voltage. The additional energy spread δW is much less than the energy spread caused by the beam potential depression [16]. Since $T^* >> T_c$, an initial velocity spread set within the framework of this model can be considered as being caused by the combined effect of roughness of cathode surface and the thermal velocity spread. Further simulations were carried out at $T^* = 67\ 000$ K and $\Delta\varphi = \pm 90°$. At such T^* and $\Delta\varphi$ values, the value of the transverse velocity spread $\delta v_\perp$ obtained in the calculations was found to approximately coincide with the corresponding value of velocity spread determined in experiments with hexaboride lanthanum cathode in the SPbPU gyrotron [17].

An important aspect of the simulation discussed in this study is implementation of a sectioned cathode. Two azimuthal sectors without electron emission were symmetrically located on the emissive strip. This cathode sectioning allowed for a significant reduction of parasitic effects near toroidal solenoid bundles used to create an azimuthal magnetic field in the collector region, which affect efficiency of residual electron energy recovery and electron reflection from the collector (see Section III). In the previous simulations [16], the length of each cathode gap sector in azimuthal direction was $\Delta\theta = 70°$. In this study, $\Delta\theta$ was decreased to 45° as a result of optimization of 4-stage collector geometry and operating regimes compared to the original version described in Ref. [16]. Calculations in the electron optical system were performed using Particle Tracking Solver, with number of emission points at the cathode set to 2700. To minimize parasitic effect of mesh step on parameters of high-energy beam, particularly on electron energy spread, a tetrahedral meshing of calculation domain was used.

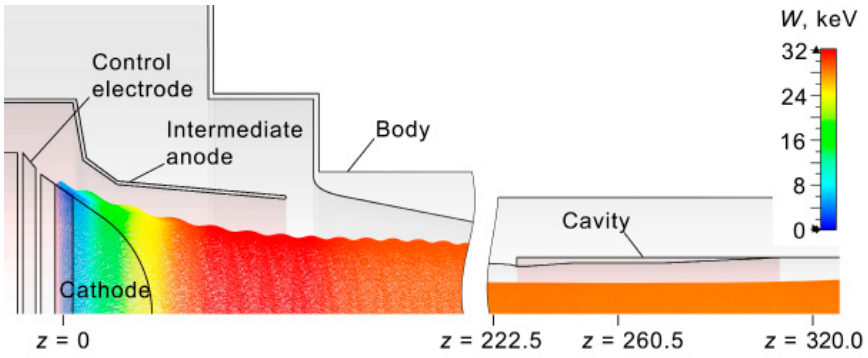

Fig. 1. Schematic drawing of the gyrotron model in the $r - z$ plane with beam trajectories

Table 1

The main geometric parameters of the gyrotron and the characteristics of its operating regime

Parameter, notation, unit	Value
Accelerating voltage U_0, kV	30
Beam current I_b, A	10
Magnetic field induction in the cavity center B_0, T	2.75
Magnetic field induction at the cathode B_c, T	0.152
Operating mode	$TE_{12,3}$
Operating frequency f_0, GHz	74.2
Cavity radius R_0, mm	14.45
Average radius of the cathode emissive strip R_c, mm	35.00

In the experimental gyrotron, magnetic compression coefficient B_0/B_c varies as a result of change in a number of turns of cathode coil [13]. The highest microwave output power in simulations was achieved at $B_0/B_c = 17.01$, which corresponds to 26 turns of cathode coil. In this case, an average beam radius in the cavity was approximately 8.5 mm. In the optimized regime with $T^* = 67\ 000$ K, $\Delta\varphi = \pm 90°$, $\Delta\theta = 45°$, $B_0/B_c = 17.01$, $U_a = 7.85$ kV and $U_{cont} = -14.5$ kV, an average pitch factor value α of 1.56 and a transverse velocity spread $\delta v_\perp$ of 5.32 % were provided. The values of velocity and energy spreads were defined as relative standard deviations from the average value of corresponding quantities. The accelerating voltage U_0, the magnetic field induction in the cavity center B_0 and the beam current I_b had values listed in Table I. In this regime, one of 2700 electron trajectories was reflected from the magnetic mirror, resulting in a reflection coefficient K_{ref} of approximately $4 \cdot 10^{-4}$.

Note that, if we assume a Gaussian distribution of the electron velocities, for $\alpha = 1.56$ and $\delta v_\perp = 5.32\%$, the reflection coefficient from the magnetic mirror is $K_{ref} = 2.1 \cdot 10^{-4}$ [13]. It should be noted that reflection of electrons from the magnetic mirror limits an increase in the average pitch factor in the presence of the electron velocity spread. If the coefficient K_{ref} exceeds threshold value, the parasitic low-frequency oscillations (LFOs) can occur in the electron space charge trapped between the cathode and cavity. These oscillations lead to a degradation of the HEB quality [14, 20 – 22]. Based on the experimental data of the SPbPU gyrotron, the threshold value of the reflection coefficient from the magnetic mirror was determined as approximately equal to $1.7 \cdot 10^{-3}$. In the gyrotron operating regime described above, the reflection coefficient was lower than this threshold value. Values of an average pitch factor α and the transverse velocity spread $\delta v_\perp$ in the case of homogeneous emission from the cathode in the described operating regime were 1.57 and 5.21 %, respectively. Therefore, it can be concluded that there was no significant change in these HEB parameters in the case of transition from a homogeneous to a sectioned distribution of emission from the cathode.

Fig. 2 presents the data characterizing particle distribution in the central plane of cavity at $z = 260.5$ mm. Azimuthal positions of the HEB sectors with no electrons correspond to angle ranges of $115° < \theta < 157°$ and $295° < \theta < 337°$, where $\theta = 0°$ coincides with positive x-axis direction (see Fig. 2,*a*). As a result of the crossed electric and magnetic fields, these sectors experienced an azimuthal shift of approximately 18° in clockwise direction viewed along propagation of the HEB moving from the cathode to the cavity. An average potential depression due to the space charge ΔU is about 1.8 kV, with its minimum value located in the HEB areas adjacent to sectors without electrons (see Fig. 2,*a*). The energy spread δW, which is about 0.5 %, is mainly due to nonuniformity of the ΔU distribution in azimuthal direction. In comparison, the similar regime of the gyrotron operation with the homogeneous HEB is characterized by $\delta W = 0.1$ %. Under the action of the crossed azimuthal electric and longitudinal magnetic fields (diocotron effect), particles in areas close to the sectors without electrons move in the radial direction. This movement

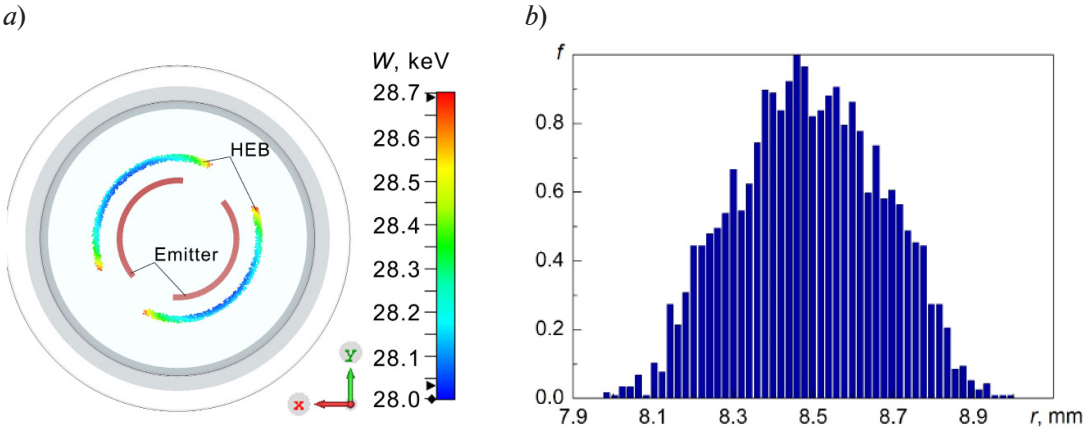

Fig. 2. Simulation results for particle distributions in the central plane of the cavity at $z = 260.5$ mm: positions of HEB particles with energy W in the $x - y$ plane (*a*) and the histogram of the particle radial positions (*b*).
Azimuthal positions of nonemitting sectors on the cathode are shown in Fig. 2,*a*

causes an increase in the HEB wall thickness ΔR_b. As shown in Fig. 2,*b*, ΔR_b is approximately equal to 1 mm if it is determined by the full width of the distribution $f(r)$. In this regime, the radial spread of the leading Larmor centers is $\Delta R_g \approx 0.7$ mm, and the average Larmor radius is $r_1 \approx 0.17$ mm. In Ref. [23], it was shown that ΔR_g in a cavity should not exceed $\lambda/6$, where λ represents a wavelength of microwave radiation. The efficiency of an operating mode generation decreases and there is a possibility of exciting parasitic modes if a value of ΔR_g exceeds this limit.

During the electron-optical system analysis, a special Particle Export Interface Monitor was employed to collect data including the velocities and coordinates of particles in the plane $z = 222.5$ mm (see Fig. 1). Subsequently, this monitor's output was used as an input interface for simulation in the gyrotron cavity.

II. Microwave cavity

The interaction of an electron beam with the electromagnetic field in the cavity of the SPbPU gyrotron was simulated using PIC Solver. Calculation domain was defined by the planes $z = 222.5$ mm and $z = 320$ mm (see Fig. 1). The cavity with a regular part length of 28 mm and a radius of 14.45 mm was designed for the operating $TE_{12,3}$ mode. Simulation results indicated that the maximum output microwave power in the operating mode was achieved at $B_0 = 2.747$ T. The results discussed below were obtained at this magnetic field induction value.

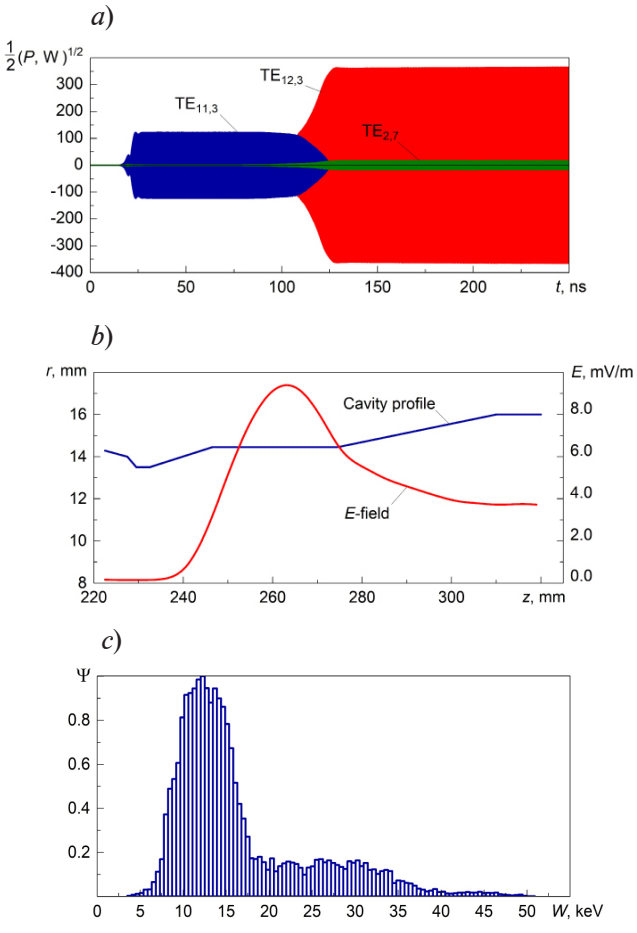

Fig. 3. Simulation results for interaction of an electron beam with the electromagnetic field in the cavity: (*a*) the time dependencies of signals of different modes in the output port ($z = 320$ mm); (*b*) the distributions of the wall radius r and the maximum E-value at 74.5 GHz along the z-axis; (*c*) electron energy spectrum of the spent beam (Ψ – normalized probability density, W – energy)

Fig. 3,*a* illustrates the time variation of the mode signals with the largest amplitude. It can be seen that there are two stable generation regions, namely, the former is with a time interval from 20 to 100 ns and the latter is with a time interval from approximately 130 ns to the simulation end at 250 ns. At $t = 100$ ns and 250 ns, the average output power values of P_{RF} are 15.5 kW and 134.8 kW, respectively. These two regions are distinguished by their mode composition in the cavity. In the first excitation region, the $TE_{11,3}$ mode with an azimuthal index one less than for the operating mode excited at a frequency of 71.5 GHz. Over time, this mode is suppressed simultaneously with excitation of the operating mode $TE_{12,3}$ at a frequency of 74.5 GHz. In the same time interval, the excitation of the parasitic mode $TE_{2,7}$ is also observed. Resonant frequencies of operating $TE_{12,3}$ and parasitic $TE_{2,7}$ modes were calculated using the Eigenmode Solver in CST Studio Suite and were equal to 74.83 and 74.94 GHz, respectively. All modes exhibit two polarization components, and each with approximately the same amplitudes, resulting in the circular wave polarization within the cavity.

At time $t = 250$ ns, the high-frequency power $P_{TE_{12,3}} = 133.9$ kW in the operating mode, and $P_{TE_{2,7}} = 0.3$ kW in the parasitic mode (see Fig. 3,*a*). One can also observe that the power ratio between operating and parasitic modes can be affected by the quality of mesh in the calculation model. We compared the power values of operating and parasitic modes obtained for sectioned

and homogeneous HEBs under the same gyrotron operating regime and mesh settings. For the homogeneous HEB, the power values P_{RF} = 138.7 kW, $P_{TE_{12,3}}$ = 137.6 kW, and $P_{TE_{2,7}}$ = 0.2 kW. Thus, the HEB sectioning has a detrimental effect on the beam quality resulting in a decrease of total output power and an increase in the parasitic mode power. Nonetheless, this change in power is negligible, and the efficiency of the converting electron energy into the high-frequency field energy remains considerably high for the case of sectioned HEB. Specifically, at P_{RF} = 134.8 kW, U_0 = 30 kV, I_b = 10 A, the electronic efficiency η_{el} is determined to be 44.9 %. The data on trajectory analysis and PIC simulation in the cavity for uniform and sectioned electron beams are combined in Table 2.

Table 2

Comparison of the simulation results for homogeneous and sectioned helical electron beams

Parameter, notation, unit	Helical electron beam (HEB)	
	Homogeneous	Sectioned
HEB parameters in the central plane of the cavity (trajectory analysis)		
Beam current I_b, A	10	10
Average pitch factor α	1.57	1.56
Transverse velocity spread $\delta v_\perp$, %	5.21	5.32
Energy spread δW, %	0.1	0.5
Beam wall thickness ΔR_b, mm	1.00	0.75
Coefficient of reflection from magnetic mirror K_{ref}	–	$4 \cdot 10^{-4}$
Output radiation parameters (PIC simulation in the cavity)		
Average output power P_{RF}, kW	138.7	134.8
Power of the operating mode $P_{TE_{12,3}}$, kW	137.6	133.9
Electronic efficiency η_{el}, %	46.2	44.9

Fig. 3,*b* presents a graph depicting the variation of E-field maximum amplitude at a frequency of 74.5 GHz with respect to longitudinal coordinate. The profile of the cavity $r(z)$ is also shown in Fig. 3,*b*. In the cone transition region after regular part of the cavity, where the high-frequency field converts into a traveling electromagnetic wave, an interaction occurs between the electrons and this wave, which is referred to as aftercavity interaction. It is known that this interaction causes an alteration in spent HEB energy spectrum, leading to a decrease in minimum electron energy [24]. The electron energy distribution of spent HEB, in turn, influences the maximum total efficiency that can be achieved through implementation of collector systems with residual energy recovery.

Particle 2D Monitor located in the z = 320 mm plane recorded data on the coordinates, velocities, and macro-charge of particles in the spent HEB required for trajectory analysis in the collector region. The monitor collected particle parameters during time interval $\Delta t = 3 \cdot 10^{-3}$ ns, resulting in an output file containing information on approximately $2 \cdot 10^4$ particles for every moment in time t. The energy spectrum obtained after processing of monitor data for t = 250 ns is presented in Fig. 3,*c*. The minimum electron energy is approximately equal to 15 % of eU_0. There is a noticeable number of accelerated particles with an energy exceeding eU_0 value. The electronic efficiency η_{el} can be estimated by known energy spectrum $f(W)$ using the formula

$$\eta_{el} = 1 - \frac{\int_0^\infty f(W) W dW}{eU_0}, \quad (1)$$

where $\int_0^\infty f(W)dW = 1$.

The efficiency obtained for the spectrum is η_{el} = 44.9 %, which matches the efficiency value obtained from the output microwave power.

III. Collector

Collector design. Recovery of the spent beam energy in the SPbPU gyrotron collector is based on spatial separation of HEB electron fractions with different energies in the crossed longitudinal electric E_z and azimuthal magnetic B_θ fields [12]. Previously, collectors with multistage recovery based on this method were developed for various gyrotrons [16, 25 – 27]. These collectors utilize a toroidal-type solenoid with outer and inner winding for generation of azimuthal magnetic field. Unlike multistage collectors with non-adiabatic fields (see, for example, Refs. [28 – 31]) or collectors with azimuthal electric field (see, for example, Refs. [10, 11]), the developed collectors have extended region with crossed E_z and B_θ fields, which allows to reduce negative influence of velocity and positional spreads of electrons and uncontrolled misalignment of electrodes and magnetic coils on collector efficiency.

The upgraded collector described in this paper differs from its original version presented in Ref. [16] by modified geometry of magnetic coils and collector sections. The data from simulation of similar collector for prototype gyrotron of the DEMO project [27] was used for modernization of the collector for the SPbPU gyrotron. The main elements of the collector for the SPbPU gyrotron are presented in Fig. 4. The cylindrical part of the collector body contains sections S1 – S4 under negative potentials used to create an electric field. Correcting coils C1 – C5, in combination with the gyrotron magnetic system including the main coil, cathode and collector ones, provide required distribution of longitudinal magnetic field. A toroidal solenoid is used to create an azimuthal magnetic field. The end conductors of this solenoid, from the side closest to the cavity, are assembled into two radial bundles located in the tubes providing connection of inner and outer winding. The inner radius of cylindrical part of the collector body is 104.5 mm. The longitudinal coordinate z corresponding to the end of transition from a conical part to cylindrical one of the collector is 667 mm. The coordinate z corresponding to the middle of connecting tubes with bundles is 619 mm. The described collector system has been already equipped in the SPbPU gyrotron. The sectioned emitter for the gyrotron was created by mechanically removing two azimuthal sectors, 45° each, from lanthanum hexaboride emissive strip on the cathode.

Optimization of magnetic field distribution in the collector region. During the process of searching for the optimal distribution of the magnetic field, adjustments were made to geometry and currents of the toroidal solenoid and the correcting coils C1 – C5. Originally, trajectories of "single" electrons starting in the plane z = 320 mm were analyzed. The initial energy and radial coordinate of these electrons were equal to 20 keV and 9 mm respectively. The initial azimuthal coordinate θ was varied. Fig. 5 shows projections of electron trajectories on the r–z plane for

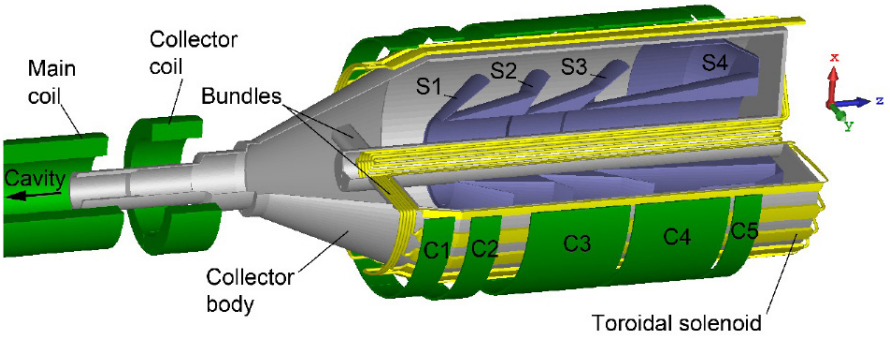

Fig. 4. Schematic drawing of the gyrotron collector region:
S1 – S4 are sections under negative potentials, C1 – C5 are correcting coils

different values of θ from 0° to 180° (Fig. 5,*a*) and longitudinal distributions of different components of magnetic field (Fig. 5,*b*) obtained after optimization. The relative position of the electron emission points with regard to toroidal solenoid bundles can also be seen in Fig. 5,*a*. The bundles are located in the planes θ = 90° and 270°. Due to system symmetry, the electron trajectories in the range of θ from 180° to 360° will coincide with those shown in Fig. 5,*a*. Potentials of the collector sections were set to zero during these calculations.

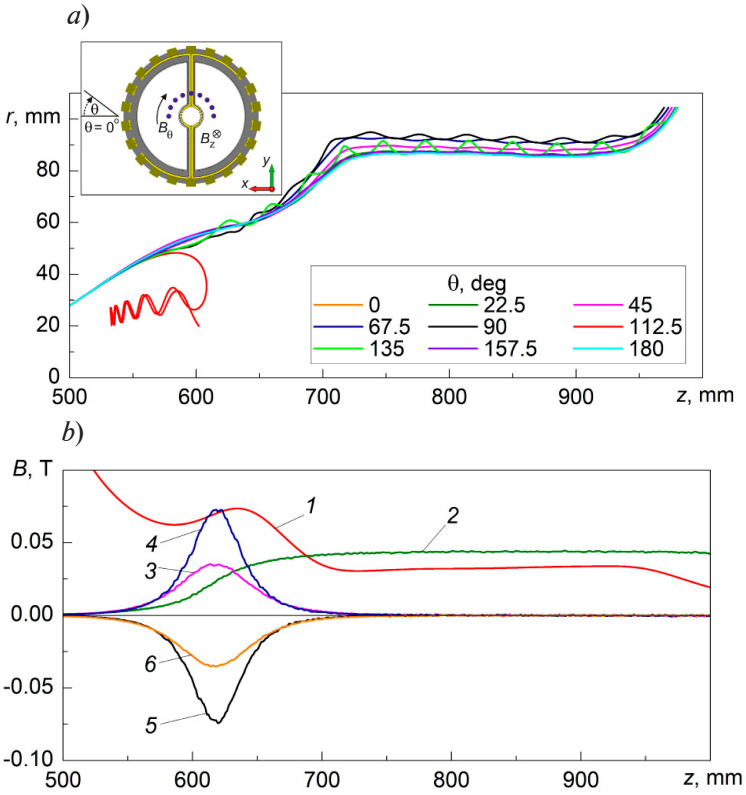

Fig. 5. Simulation results for search for the optimal distribution of the magnetic field *B*: (*a*) projections of electron trajectories with different azimuthal coordinates θ of the starting point; (*b*) the distributions of the *B*-field components (*B*-comps) along the *z*-axis.

Fig 5,*a*: there are azimuthal positions of the starting point of electrons and the bundles of toroidal solenoid in the $x - y$ plane.

Fig 5,*b*: there is data on the axial *B*-comp created by the main gyrotron magnetic system and correcting coils (*1*) as well as azimuthal (*2*) and axial (*3*–*6*) *B*-comps of the toroidal solenoid; the data was obtained at different coordinate values of *r*, mm and θ, degs:

80, 0.0 (*1*, *2*); 60, 45.0 (*3*); 60, 67.5 (*4*); 60, 112.5 (*5*); 60, 135.0 (*6*)

The efficient operation of the multistage collector requires minimizing the spread of radial positions of electron trajectories with different azimuthal coordinates at the collector entry in the absence of voltages on sections. Calculations for the DEMO gyrotron [27] showed that a decrease in the radial spread can be achieved by selecting a positive direction of the azimuthal magnetic field (see Fig. 5,*a*). The decrease can also be achieved by choosing a required magnetic induction value of the main gyrotron magnetic system and correcting coils in the area of toroidal solenoid bundles at $z \approx 620$ mm. The optimized distribution of magnetic field provides the small radial position spread in the recovery region ($z > 700$ mm), where the induction of longitudinal magnetic field B_z is approximately equal to 0.032 T and the azimuthal field B_θ is approximately 0.044 T at $r = 80$ mm.

However, after optimization, the certain number of electrons which propagate near the bundles of the toroidal solenoid are still present, and their trajectories are noticeably perturbed under the action of bundles' parasitic field. These electrons cannot reach the sections with potentials corresponding to their energies and settle on other electrodes of the collector. Alternatively,

they may be reflected from the collector towards the cavity, reducing electronic efficiency. In either scenario, the total gyrotron efficiency is diminished. For example, an electron with the initial coordinate $\theta = 112.5°$ settles on the connecting tube in which the bundle is located (see Fig. 5,*a*). It can be seen that the total longitudinal magnetic field determined by the sum of B_z values at azimuth $\theta = 112.5°$ (curves *1* and *5*) is close to zero in the bundle region (see Fig. 5,*b*). To address this issue, the HEB was sectioned as described in Sections I and II to eliminate negative influence of such electrons on operation of the multistage collector. An additional displacement of the HEB in azimuthal direction during its movement between the planes $z = 260.5$ and 320 mm was insignificant and did not exceed 1°.

For further calculations, the toroidal solenoid was rotated 19° clockwise to achieve a minimum reflection of particles from the collector due to parasitic action of the magnetic field created by bundles' conductors.

Trajectory analysis in the collector with four-stage energy recovery. In the case of ideal separation, each fraction of spent HEB with energy W is deposited on the section under the most negative potential, the modulus of which does not exceed eW (e is the electron charge), and the collector body is under zero potential. The dependency of maximum total efficiency of the SPbPU gyrotron η_{max} achieved with ideal separation on the number of recovery stages N was calculated before the collector modeling. The spectrum of spent HEB shown in Fig. 3,*c* was used. For this calculation, the spectrum was divided into 1000 intervals with different energies. 1 % of the HEB current with electrons having the lowest energy was assumed to be reflected from the collector. The maximum total efficiency was achieved at optimal potentials of the sections U_i ($i = 1, 2, ..., N$) determined through iterations over the values of these potentials with a step of 0.2 kV. As in previous studies presented in [16, 25 – 27], the choice of four stages is dictated by a balance between the achieving maximum total efficiency of the gyrotron and practical difficulties associated with implementing a recovery system with a large number of stages. It should be noted that an increase in the number of collector sections does not substantially complicate the design of the described collector, unlike other designs with nonadiabatic fields [28 – 31].

Four cone-shaped sections are located in the cylindrical part of collector body (see Fig. 4). Changes in geometry of these sections compared with the original design described in Ref. [16] are due to modifications made to the collector magnetic system. Specifically, the direction of azimuthal magnetic field was changed to positive (see Fig. 5,*a*), so the sections were located in the region of smaller radii along the direction of electron drift in the crossed E_z and B_θ fields.

It should be noted that in the regime in the absence of the azimuthal magnetic field and zero voltage on the collector sections, the beam wall thickness in recovery region ($z > 700$ mm) is approximately 10 mm. With optimized distributions of azimuthal and axial magnetic fields and the length of the cathode's sectors without emission of 45° in the absence of voltages on the sections, approximately 94 % of the particles reached the final Section S4, 5 % deposited on the collector body, and less than 1 % deposited on Sections S1 – S3.

During a trajectory analysis in the collector, a Particle Import Interface was placed at the input plane $z = 320$ mm. It contained an array of particles that was determined during the simulation in the gyrotron cavity (as described in Section II). The initial potentials of collector sections $U_{S1} - U_{S4}$ were set equal to the optimal values obtained with ideal separation. Subsequently, through a series of electron trajectory calculations of spent HEB in the collector, these potential values were adjusted to achieve the maximum total efficiency of the gyrotron with an electron reflection coefficient from the collector less than 1 %. At $U_{S1} = -7.1$ kV, $U_{S2} = -10.7$ kV, $U_{S3} = -14.3$ kV, $U_{S4} = -25.2$ kV and $U_{coll} = 0$, the power dissipated over the collector sections and the body was $P_{S1} = 6.2$ kW, $P_{S2} = 6.2$ kW, $P_{S3} = 11.7$ kW, $P_{S4} = 10.2$ kW, and $P_{coll} = 1.2$ kW respectively, with the collector reflection coefficient of 0.99 %. Consequently, the total power P_{diss} dissipated on collector was 35.5 kW. At power $P_{RF} = 134.8$ kW, the total efficiency was

$$\eta_t = \frac{P_{RF}}{P_{RF} + P_{diss}} = 79.2\ \%, \qquad (2)$$

and the collector efficiency was

$$\eta_t = 1 - \frac{P_{diss}}{U_0 I_b - P_{\mathrm{RF}}} = 78.5\ \%. \qquad (3)$$

Thus, through optimization of the magnetic field distribution and the collector sections' geometry, the total efficiency over 79 % was achieved. This value is about 2 % less than the maximum efficiency in a four-stage recovery system with ideal separation of electrons with different energies. For comparison, the total efficiency was 71.8 % in the initial version of the collector for the SPbPU gyrotron [16].

Fig. 6 shows positions of particles in the $x - z$ plane, obtained as a result of intersection of the HEB trajectories with this plane. The picture demonstrates the drift of electrons in the radial direction by the action of crossed E_z and B_θ fields while propagating in the retarding electric field in the recovery region. The direction of field B_θ causes the drift towards smaller radii. Some of the electrons are deposited on the back walls of sections after changing their longitudinal movement direction. As the initial energy of electrons increases, they propagate a greater distance along the z-axis and deposit on sections with a more negative potential.

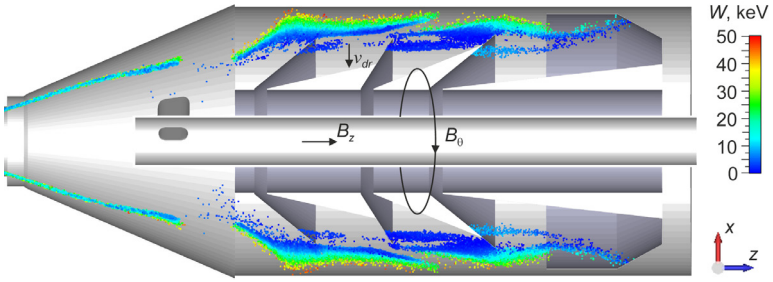

Fig. 6. Positions of particles in the plane x–z (the color corresponds to the particle energy W). The directions of the longitudinal B_z and the azimuthal B_θ magnetic fields and the electron drift velocity v_{dr} are shown

Conclusion

The present study consisted of complex simulation to investigate the possibility of achieving record values of the total efficiency for a moderate-power gyrotron operating in 4 mm wavelength range. High efficiency was achieved by enhancement of the HEB quality in the electron optical system and by recovering the residual electron energy using a multistage collector system. Parameters of the spent HEB were determined through a trajectory analysis in the electron optical system taking into account the spread of initial electron velocities at the cathode and simulation of interaction of formed HEB with a high-frequency field in the gyrotron cavity. Spatial separation of electron fractions with different energies in the collector was achieved using a method based on electron drift in the crossed longitudinal electric and azimuthal magnetic fields. A toroidal solenoid was used as a source of the azimuthal magnetic field. The end conductors of toroidal solenoid were assembled in bundles to increase the number of electrons passing to the recovery area. The negative influence of the magnetic field created by these bundles on the collector efficiency and the electron reflection coefficient was reduced through the sectioning of the electron beam.

It is important to underline the main differences between this study and the one described in Ref. [16] where the former version of the four-stage recovery collector for the SPbPU gyrotron was analyzed. In this study, a trajectory analysis in the electron optical system of the gyrotron has been performed considering the initial velocity spread of electrons on a cathode. As a consequence, characteristics of the beam entering the resonator, output radiation parameters, and, importantly for collector modeling, the spent electron beam parameters have undergone changes. The main distinction from Ref. [16] lies in the modification of the collector system's design. Alteration of the azimuthal magnetic field direction and optimization of the longitudinal magnetic field distribution using correcting coils enabled reducing the radial trajectory spread in the recovery region. Single-electron trajectory calculations in the collector region were employed for this purpose. The collector sections feature a new geometry, and their potentials were selected based on calculations of the maximum total efficiency with ideal separation of electrons of

different energies. Compared to the results of Ref. [16], performed modernization of the collector for the SPbPU gyrotron allowed for an increase in calculated total efficiency from 71.8 to 79.2 % with a reflection coefficient from the collector of less than 1 % and for a reduction of the length of cathode gap sectors where emission is absent, from 70 to 45°. With incorporation of the new simulation data described in this work, an upgraded version of the collector for the SPbPU gyrotron has been manufactured.

Continuation of this work may involve further improvement of the azimuthal magnetic field source, which will simplify the design of the collector.

REFERENCES

1. **Thumm M. K. A., Denisov G. G., Sakamoto K., Tran M. Q.,** High-power gyrotrons for electron cyclotron heating and current drive, Nucl. Fusion. 59 (7) (2019) 073001.

2. **Thumm M.,** State-of-the-art of high-power gyro-devices and free electron masers, J. Infrared Millim. Terahertz Waves. 41 (1) (2020) 1–140.

3. **Litvak A. G., Denisov G. G., Myasnikov V. E., et al.,** Development in Russia of megawatt power gyrotrons for fusion, J. Infrared Millim. Terahertz Waves. 32 (3) (2011) 337–342.

4. **Darbos C., Albajar F., Bonicelli T., et al.,** Status of the ITER electron cyclotron heating and current drive system, J. Infrared Millim. Terahertz Waves. 37 (1) (2016) 4–20.

5. **Jelonnek J., Aiello G., Alberti S., et al.,** Design considerations for future DEMO gyrotrons: A review on related gyrotron activities within EUROfusion, Fusion Eng. Des. 123 (November) (2017) 241–246.

6. **Sakamoto K., Tsuneoka M., Kasugai A., et al.,** Major improvement of gyrotron efficiency with beam energy recovery, Phys. Rev. Lett. 73 (26) (1994) 3532–3535.

7. **Glyavin M. Y., Kuftin A. N., Venediktov N. P., Zapevalov V. E.,** Experimental investigation of a 110 GHz/1 MW gyrotron with the one-step depressed collector, Int. J. Infrared Millim. Waves. 18 (11) (1997) 2129–2136.

8. **Manuilov V. N., Morozkin M. V., Luksha O. I., Glyavin M. Y.,** Gyrotron collector systems: Types and capabilities, Infrared Phys. Technol. 91 (June) (2018) 46–54.

9. **Pagonakis I. Gr., Hogge J.-P., Alberti S., et al.,** A new concept for the collection of an electron beam configured by an externally applied axial magnetic field, IEEE Trans. Plasma Sci. 36 (2) (2008) 469–480.

10. **Wu C., Pagonakis I. G., Avramidis K. A., et al.,** Gyrotron multistage depressed collector based on E×B drift concept using azimuthal electric field. I. Basic design, Phys. Plasmas. 25 (3) (2018) 033108.

11. **Ell B., Wu C., Gantenbein G., et al.,** Toward the first continuous wave compatible multistage depressed collector design for high power gyrotrons, IEEE Trans. Electron Devices. 70 (3) (2023) 1299–1305.

12. **Louksha O. I., Trofimov P. A.,** A method of electron separation for multistep recuperation systems in gyrotrons, Tech. Phys. Lett. 41 (9) (2015) 884–886.

13. **Louksha O. I., Piosczyk B., Sominski G. G., et al.,** On potentials of gyrotron efficiency enhancement: measurements and simulations on a 4-mm gyrotron, IEEE Trans. Plasma Sci. 34 (3) (2006) 502–511.

14. **Louksha O. I., Samsonov D. B., Sominskii G. G., Semin S. V.,** Dynamic processes in helical electron beams in gyrotrons, Tech. Phys. 58 (5) (2013) 751–759.

15. **Louksha O. I., Sominski G. G., Arkhipov A. V., et al.,** Gyrotron research at SPbPU: Diagnostics and quality improvement of electron beam, IEEE Trans. Plasma Sci. 44 (8) (2016) 1310–1319.

16. **Louksha O. I., Trofimov P. A.,** Highly efficient gyrotron with multi-stage recuperation of residual electron energy, Tech. Phys. 64 (12) (2019) 1889–1897.

17. **Louksha O. I., Trofimov P. A., Malkin A. G.,** Trajectory analysis in a gyrotron electron-optical system with allowance for the cathode surface roughness, Radiophys. Quant. El. 65 (3) (2022) 209–218.

18. **Louksha O. I., Trofimov P. A.,** Simulation of non-uniform electron beams in the gyrotron electron-optical system, Tech. Phys. 63 (4) (2018) 598–604.

19. **Louksha O. I., Samsonov D. B., Sominskii G. G., Tsapov A. A.,** Improvement of the helical electron beam quality and the gyrotron efficiency by controlling the electric field distribution near a magnetron injection gun, Tech. Phys. 57 (6) (2012) 835–839.

20. **Tsimring Sh. E.,** Gyrotron electron beams: velocity and energy spread and beam instabilities, Int. J. Infrared Millim. Waves. 22 (10) (2001) 1433–1468.

21. **Manuilov V. N.,** Numerical simulation of low-frequency oscillations of the space charge and potential in the electron-optical system of a gyrotron, Radiophys. Quant. El. 49 (10) (2006) 786–792.

22. **Louksha O. I.,** Simulation of low-frequency collective processes in gyrotron electron beams, Radiophys. Quant. El. 52 (5–6) (2009) 386–397.

23. **Pu R., Nusinovich G. S., Sinitsyn O. V., Antonsen T. M. Jr.,** Effect of the thickness of electron beams on the gyrotron efficiency, Phys. Plasmas. 17 (8) (2010) 083105.

24. **Zapevalov V. E., Moiseev M. A.,** Influence of aftercavity interaction on gyrotron efficiency, Radiophys. Quant. El. 47 (7) (2004) 520–527.

25. **Louksha O. I., Trofimov P. A.,** A multistage depressed collector system for gyrotrons, Proc. 18th Int. Vacuum Electronics Conf. (IVEC), April 24–26, London, UK (2017) 1–2.

26. **Louksha O. I., Trofimov P. A., Manuilov V. N., Glyavin M. Yu.,** Trajectory analysis in a collector with multistage energy recovery for a DEMO prototype gyrotron. Part I. Idealized magnetic field distribution, Tech. Phys. 66 (1) (2021) 118–123.

27. **Louksha O. I., Trofimov P. A., Manuilov V. N., Glyavin M. Yu.,** Trajectory analysis in a collector with multistage energy recovery for a DEMO prototype gyrotron. Part II. Toroidal magnetic field, Tech. Phys. 66 (8) (2021) 992–998.

28. **Read M. E., Lawson W. G., Dudas A. J., Singh A.,** Depressed collectors for high-power gyrotrons, IEEE Trans. Electron Devices. 37 (6) (1990) 1579–1589.

29. **Singh A., Rajapatirana S., Men Y., et al.,** Design of a multistage depressed collector system for 1-MW CW gyrotrons. I. Trajectory control of primary and secondary electrons in a two-stage depressed collector, IEEE Trans. Plasma Sci. 27 (2) (1999) 490–502.

30. **Ling G., Piosczyk B., Thumm M. K.,** A new approach for a multistage depressed collector for gyrotrons, IEEE Trans. Plasma Sci. 28 (3) (2000) 606–613.

31. **Glyavin M. Y., Morozkin M. V., Petelin M. I.,** Separation of energy fractions of an electron beam by a localized nonuniformity of magnetic field in the collector region of gyrodevices, Radiophys. Quant. El. 49 (10) (2006) 811–815.

СПИСОК ЛИТЕРАТУРЫ

1. **Thumm M. K. A., Denisov G. G., Sakamoto K., Tran M. Q.** High-power gyrotrons for electron cyclotron heating and current drive // Nuclear Fusion. 2019. Vol. 59. No. 7. P. 073001.

2. **Thumm M.** State-of-the-art of high-power gyro-devices and free electron masers // Journal of Infrared, Millimeter, and Terahertz Waves. 2020. Vol. 41. No. 1. Pp. 1–140.

3. **Litvak A. G., Denisov G. G., Myasnikov V. E., Tai E. M., Azizov E. A., Ilin V. I.** Development in Russia of megawatt power gyrotrons for fusion // Journal of Infrared, Millimeter, and Terahertz Waves. 2011. Vol. 32. No. 3. Pp. 337–342.

4. **Darbos C., Albajar F., Bonicelli T., et al.** Status of the ITER electron cyclotron heating and current drive system // Journal of Infrared, Millimeter, and Terahertz Waves. 2016. Vol. 37. No. 1. Pp. 4–20.

5. **Jelonnek J., Aiello G., Alberti S., et al.** Design considerations for future DEMO gyrotrons: A review on related gyrotron activities within EUROfusion // Fusion Engineering and Design. 2017. Vol. 123. November. Pp. 241–246.

6. **Sakamoto K., Tsuneoka M., Kasugai A., Imai T., Kariya T., Hayashi K., Mitsunaka Y.** Major improvement of gyrotron efficiency with beam energy recovery // Physical Review Letters. 1994. Vol. 73. No. 26. Pp. 3532–3535

7. **Glyavin M. Y., Kuftin A. N., Venediktov N. P., Zapevalov V. E.** Experimental investigation of a 110 GHz / 1 MW gyrotron with the one-step depressed collector // International Journal of Infrared and Millimeter Waves. 1997. Vol. 18. No. 11. Pp. 2129–2136.

8. **Manuilov V. N., Morozkin M. V., Luksha O. I., Glyavin M. Y.** Gyrotron collector systems: Types and capabilities// Infrared Physics & Technology. 2018. Vol. 91. June. Pp. 46–54.

9. **Pagonakis I. Gr., Hogge J.-P., Alberti S., Avramides K. A., Vomvoridis J. L.** A new concept for the collection of an electron beam configured by an externally applied axial magnetic field // IEEE Transactions on Plasma Science. 2008. Vol. 36. No. 2. Pp. 469–480.

10. **Wu C., Pagonakis I. G., Avramidis K. A., Gantenbein G., Illy S., Thumm M., Jelonnek J.** Gyrotron multistage depressed collector based on E×B drift concept using azimuthal electric field. I. Basic design // Physics of Plasmas. 2018. Vol. 25. No. 3. P. 033108.

11. **Ell B., Wu C., Gantenbein G., Illy S., Misko M. S., Pagonakis I. Gr., Weggen J., Thumm M., Jelonnek J.** Toward the first continuous wave compatible multistage depressed collector design for high power gyrotrons // IEEE Transactions on Electron Devices. 2023. Vol. 70. No. 3. Pp.1299–1305.

12. **Лукша О. И., Трофимов П. А.** Метод сепарации электронов для систем многоступенчатой рекуперации в гиротронах // Письма в Журнал технической физики. 2015. Т. 41. № 18. С. 38–45.

13. **Louksha O. I., Piosczyk B., Sominski G. G., Thumm M. K., Samsonov D. B.** On potentials of gyrotron efficiency enhancement: measurements and simulations on a 4-mm gyrotron // IEEE Transactions on Plasma Science. 2006. Vol. 34. No. 3. Pp. 502–511.

14. **Лукша О. И., Самсонов Д. Б., Соминский Г. Г., Сёмин С. В.** Динамические процессы в винтовых электронных потоках гиротронов // Журнал технической физики. 2013. Т. 83. № 5. С. 132–140.

15. **Louksha O. I., Sominski G. G., Arkhipov A. V., Dvoretskaya N. G., Kolmakova N. V., Samsonov D. B., Trofimov P. A.** Gyrotron research at SPbPU: Diagnostics and quality improvement of electron beam // IEEE Transactions on Plasma Science. 2016. Vol. 44. No. 8. Pp. 1310–1319.

16. **Лукша О. И., Трофимов П. А.** Высокоэффективный гиротрон с многоступенчатой рекуперацией остаточной энергии электронов // Журнал технической физики. 2019. Т. 89. № 12. С. 1988–1996.

17. **Лукша О. И., Трофимов П. А., Малкин А. Г.** Траекторный анализ в электронно-оптической системе гиротрона с учетом шероховатости поверхности катода // Известия высших учебных заведений. Радиофизика. 2022. Т. 65. № 3. С. 226–237.

18. **Лукша О. И., Трофимов П. А.** Моделирование неоднородных электронных потоков в электронно-оптической системе гиротрона // Журнал технической физики. 2018. Т. 88. № 4. С. 614–620.

19. **Лукша О. И., Самсонов Д. Б., Соминский Г. Г., Цапов А. А.** Повышение качества винтового электронного потока и кпд гиротрона при регулировании распределения электрического поля в области магнетронно-инжекторной пушки // Журнал технической физики. 2012. Т. 82. № 6. С. 101–105.

20. **Tsimring Sh. E.** Gyrotron electron beams: velocity and energy spread and beam instabilities // International Journal of Infrared and Millimeter Waves. 2001. Vol. 22. No. 10. Pp. 1433–1468.

21. **Мануилов В. Н.** Численное моделирование низкочастотных колебаний пространственного заряда и потенциала в электронно-оптической системе гиротрона // Известия высших учебных заведений. Радиофизика. 2006. Т. 49. № 10. С. 872–879.

22. **Лукша О. И.** Моделирование низкочастотных коллективных процессов в электронных потоках гиротронов // Известия высших учебных заведений. Радиофизика. 2009. Т. 52. № 5–6. С. 425–437.

23. **Pu R., Nusinovich G. S., Sinitsyn O. V., Antonsen T. M. Jr.** Effect of the thickness of electron beams on the gyrotron efficiency // Physics of Plasmas. 2010. Vol. 17. No. 8. P. 083105.

24. **Запевалов В. Е., Моисеев М. А.** Влияние послерезонаторного взаимодействия на кпд гиротрона // Известия высших учебных заведений. Радиофизика. 2004. Т. 47. № 7. С. 584–592.

25. **Louksha O. I., Trofimov P. A.** A multistage depressed collector system for gyrotrons, Proceedings of the 18th International Vacuum Electronics Conference (IVEC), April 24–26, London, UK (2017) 1–2.

26. **Лукша О. И., Трофимов П. А., Мануилов В. Н., Глявин М. Ю.** Траекторный анализ в коллекторе с многоступенчатой рекуперацией энергии для прототипа гиротрона DEMO. Часть I. Идеализированное распределение магнитного поля // Журнал технической физики. 2021. Т. 91. № 1. С. 125–130.

27. **Лукша О. И., Трофимов П. А., Мануилов В. Н., Глявин М. Ю.** Траекторный анализ в коллекторе с многоступенчатой рекуперацией энергии для прототипа гиротрона DEMO. Часть II. Тороидальное магнитное поле // Журнал технической физики. 2021. Т. 91. № 7. С. 1182–1188.

28. **Read M. E., Lawson W. G., Dudas A. J., Singh A.** Depressed collectors for high-power gyrotrons // IEEE Transactions on Electron Devices. 1990. Vol. 37. No. 6. Pp. 1579–1589.

29. **Singh A., Rajapatirana S., Men Y., Granatstein V. L., Ives R. L., Antolak A. J.** Design of a multistage depressed collector system for 1-MW CW gyrotrons. I. Trajectory control of primary and secondary electrons in a two-stage depressed collector // IEEE Transactions on Plasma Science. 1999. Vol. 27. No. 2. Pp. 490–502.

30. **Ling G., Piosczyk B., Thumm M. K.** A new approach for a multistage depressed collector for gyrotrons // IEEE Transactions on Plasma Science. 2000. Vol. 28. No. 3. Pp. 606–613.

31. **Глявин М. Ю., Морозов М. В., Петелин М. И.** Разделение энергетических фракций электронного пучка локализованной неоднородностью магнитного поля в коллекторной области гироприборов // Известия высших учебных заведений. Радиофизика. 2006. Т. 49. № 10. С. 900–905.

THE AUTHORS

LOUKSHA Oleg I.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
louksha@rphf.spbstu.ru
ORCID: 0000-0002-6402-8112

TROFIMOV Pavel A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
trofpa@yandex.ru
ORCID: 0000-0002-3585-1169

MALKIN Alexander G.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
alexmalkin47@gmail.com
ORCID: 0000-0003-4047-3956

СВЕДЕНИЯ ОБ АВТОРАХ

ЛУКША Олег Игоревич – *доктор физико-математических наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого, Санкт-Петербург, Россия.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
louksha@rphf.spbstu.ru
ORCID: 0000-0002-6402-8112

ТРОФИМОВ Павел Анатольевич – *кандидат физико-математических наук, инженер Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого, Санкт-Петербург, Россия.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
trofpa@yandex.ru
ORCID: 0000-0002-3585-1169

МАЛКИН Александр Геннадьевич – *студент Института электроники и телекоммуникаций Санкт-Петербургского политехнического университета Петра Великого, Санкт-Петербург, Россия.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
alexmalkin47@gmail.com
ORCID: 0000-0003-4047-3956

Received 09.09.2023. Approved after reviewing 18.09.2023. Accepted 18.09.2023.
Статья поступила в редакцию 09.09.2023. Одобрена после рецензирования 18.09.2023. Принята 18.09.2023.

Original article
DOI: https://doi.org/10.18721/JPM.16411

ANALYTICAL STUDY OF OPERATING MODES OF RF ION FUNNELS IN THE GAS DYNAMIC INTERFACES OF TANDEM TRIPLE-QUADRUPOLE MASS-SPECTROMETERS

A. A. Sysoev[1]✉, *A. S. Berdnikov*[2], *S. V. Masyukevich*[2], *K. V. Solovyev*[3],[2], *N. K. Krasnova*[3]

[1] National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow, Russia;

[2] Institute for Analytical Instrumentation of RAS, St. Petersburg, Russia;

[3] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ aasysoyev@mephi.ru

Abstract. The article considers analytical models of high-frequency electric fields which can be used effectively for fast, high-quality simulation of ion flow focusing and transport processes in the radio-frequency funnels. In particular, the use of such devices in the design of a tandem three-quadrupole mass spectrometer increases the amount of ions collected in the forevacuum region of the gas-dynamic interface of the electrospray ion source. The cases of funnels with two- and four-phase electrical voltages (options I and II), as well as with amplitude-modulated electrical voltages providing a pseudopotential mode with an Archimedean wave (III) have been analyzed. As a result, the most preferable design turned out to be option III. The use of such analytical models makes it possible to test effectively promising options and thereby significantly reduce costs for the preliminary selection of a principal scheme of a device with specified characteristics, including similar cases of other mass spectrometric designs.

Keywords: mass spectrometry, gas dynamic interface, radio-frequency trap, tandem mass spectrometer, triple-quadrupole mass-spectrometer

Funding: The research was supported by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-03-2023-097).

Citation: Sysoev A. A., Berdnikov A. S., Masyukevich S. V., Solovyev K. V., Krasnova N. K., Analytical study of operating modes of RF ion funnels in the gas dynamic interfaces of tandem triple-quadrupole mass-spectrometers, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 134–145. DOI: https://doi.org/10.18721/JPM.16411

Научная статья
УДК 537.534.7, 543.51
DOI: https://doi.org/10.18721/JPM.16411

АНАЛИТИЧЕСКОЕ ИССЛЕДОВАНИЕ РЕЖИМОВ РАБОТЫ РАДИОЧАСТОТНЫХ ВОРОНОК В ГАЗОДИНАМИЧЕСКИХ ИНТЕРФЕЙСАХ ТАНДЕМНЫХ ТРЕХКВАДРУПОЛЬНЫХ МАСС-СПЕКТРОМЕТРОВ

А. А. Сысоев[1]✉, *А. С. Бердников*[2], *С. В. Масюкевич*[2], *К. В. Соловьев*[3], [2], *Н. К. Краснова*[3]

[1] Национальный исследовательский ядерный университет МИФИ, Москва, Россия;

[2] Институт аналитического приборостроения РАН, Санкт-Петербург, Россия;

[3] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

✉ aasysoyev@mephi.ru

Аннотация. В статье рассмотрены аналитические модели высокочастотных электрических полей, которые можно эффективно использовать для быстрого качественного моделирования процессов фокусировки и транспорта ионных потоков в радиочастотных воронках. В частности, применение таких устройств в конструкции тандемного трехквадрупольного масс-спектрометра увеличивает количество ионов, собираемых в форвакуумной области газодинамического интерфейса электроспрейного источника ионов. Проанализированы случаи функционирования воронок с двух- и четырехфазными электрическими напряжениями (варианты I и II), а также с амплитудно-модулированными электрическими напряжениями, обеспечивающими режим псевдопотенциала с архимедовой волной (III). В результате проведенного анализа наиболее предпочтительной конструкцией оказался III вариант. Использование подобных аналитических моделей позволяет эффективно проверять перспективные варианты и тем самым существенно снизить трудозатраты на предварительный выбор принципиальной схемы устройства с заданными характеристиками, в том числе и в других масс-спектрометрических разработках.

Ключевые слова: масс-спектрометрия, источник ионов, электрораспыление, газодинамический интерфейс, радиочастотная ловушка, тандемный трехквадрупольный масс-спектрометр

Финансирование: Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках Соглашения № 075-03-2023-097.

Ссылка для цитирования: Сысоев А. А., Бердников А. С., Масюкевич С. В., Соловьев К. В., Краснова Н. К. Аналитическое исследование режимов работы радиочастотных воронок в газодинамических интерфейсах тандемных трехквадрупольных масс-спектрометров // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. № .16 4. С. 134–145. DOI: https://doi.org/10.18721/JPM.16411

Introduction

The paper considers one of the elements of a tandem triple-quadrupole mass spectrometer with electrospray ionization, namely, a radio frequency funnel placed in the forevacuum region of the gas dynamic interface of the ion source. Such a mass spectrometer is currently developed at the National Research Nuclear University MEPhI as part of the Federal Project "Development of domestic instrumentation for civil purposes" supported by the Ministry of Science and Higher Education of the Russian Federation [1–10].

Focusing RF funnels [11], designed to reduce ion losses in the gas-dynamic interface of the ion source, are considered in our study using analytical models. Unfortunately, direct simulation of such devices requires vast computing resources. On the other hand, while analytical models only provide qualitative rather than quantitative results for funnel operation, they allow to effectively select the most promising directions for optimizing the final designs. This way, the labor costs for selecting initial candidates are drastically reduced, so better characteristics of the final design can be achieved.

The paper considers analytical models of high-frequency electric fields of RF funnels, using them to preselect the design for such an element of the gas-dynamic interface of an electrospray ion source. Decisions are made for the potential of different designs. The main results briefly overviewed in this study will be presented in more detail in the future.

Model electric fields for transport channels with circular apertures

SRIG-type (Stacked Ring Ion Guide) RF electric traps, discussed in detail in [12, 13], are a chain of circular apertures (Fig. 1), to which high-frequency electric voltages with opposite phases at adjacent apertures are applied.

The locking effect of the RF electric field increases exponentially for these devices away from the axis and approaching the boundaries of the electrodes, while the locking effect of such an electric field for standard multipole RF traps with long cylindrical electrodes polynomially increases away from the axis and approaching the boundaries of the electrodes. Thus, SRIG traps provide more reliable retention of charged particles (in our case, ions) than classical multipole RF traps.

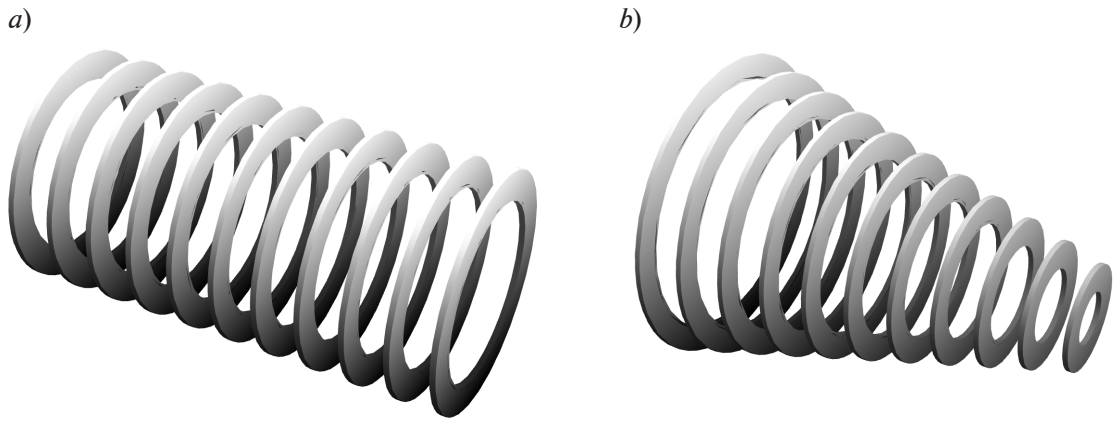

Fig. 1. Electrode configurations of SRIG-type RF traps with cylindrical (*a*) and conical (*b*) channels for ion retention and transport

The electric potential $U(z, r, t)$ near the axis in a classical SRIG-type RF electric trap with apertures of the same radius (see Fig. 1,*a*), where the phase difference of RF voltages applied to adjacent apertures is equal to π, is described with good accuracy by either of the two formulas:

$$U(z,r,t) = U_C(z,r)\cos(\omega t + \varphi),\ U(z,r,t) = U_S(z,r)\cos(\omega t + \varphi), \tag{1}$$

where

$$\begin{aligned} U_C(z,r) &= \frac{U_R}{I_0(\pi R/L)}\cos\left(\frac{\pi z}{L}\right) I_0\left(\frac{\pi r}{L}\right), \\ U_S(z,r) &= \frac{U_R}{I_0(\pi R/L)}\sin\left(\frac{\pi z}{L}\right) I_0\left(\frac{\pi r}{L}\right). \end{aligned} \tag{2}$$

U_R, ω, φ in Eqs. (1), (2) are the amplitude, circular frequency and phase of RF voltages applied to the electrodes; R is the inner radius of circular apertures; L is the distance between adjacent apertures; $r = \sqrt{x^2 + y^2}$ is the distance in the radial direction to the axis from a point with Cartesian spatial coordinates (x, y, z); t is time; I_0 is the modified zero-order Bessel function [14].

The electric potential near the axis for a radio frequency trap with a conically narrowing transport channel (see Fig. 1,*b*), in which the phase difference of RF voltages applied to adjacent apertures is equal to π, is described with good accuracy by either of the two formulas:

$$V(z,r,t) = V_C(z,r)\cos(\omega t + \varphi),\ V(z,r) = V_S(z,r,t)\cos(\omega t + \varphi), \tag{3}$$

where

$$\begin{aligned} V_C(z,r) &= \frac{U_R}{I_0(\pi R/L)}\left[z\cos\left(\frac{\pi z}{L}\right) I_0\left(\frac{\pi r}{L}\right) + r\sin\left(\frac{\pi z}{L}\right) I_1\left(\frac{\pi r}{L}\right)\right], \\ V_S(z,r) &= \frac{U_R}{I_0(\pi R/L)}\left[z\sin\left(\frac{\pi z}{L}\right) I_0\left(\frac{\pi r}{L}\right) - r\cos\left(\frac{\pi z}{L}\right) I_1\left(\frac{\pi r}{L}\right)\right]. \end{aligned} \tag{4}$$

In these formulas, I_1, a modified first-order Bessel function, is added to the previous notation [14].

If it is necessary to achieve an additional effect, the electric field of the funnel should be modified in a non-linear way, using additive correction term, quadratic with respect to the z coordinate. When such field distortions are introduced, the design of the funnel, i.e., the positions and diameters of individual circular apertures, must be changed accordingly. The problem of calculating the position and shape of the electrodes from a given electric field is elementary, unlike the opposite problem, which once again shows the advantages of an analytical approach using model distributions of the electric field.

Any of these formulas can be used for analytical description of the correcting potential $W(z, r,t)$ with a quadratic dependence on the z coordinate (along the axis of the device):

$$W(z,r,t) = W_C(z,r)\cos(\omega t + \varphi),\ W(z,r,t) = W_S(z,r)\cos(\omega t + \varphi), \tag{5}$$

where

$$\begin{aligned} W_C(z,r) &= \frac{U_R}{I_0(\pi R/L)}\left\{ z^2\cos\left(\frac{\pi z}{L}\right) I_0\left(\frac{\pi r}{L}\right) + 2zr\sin\left(\frac{\pi z}{L}\right) I_1\left(\frac{\pi r}{L}\right) - \right. \\ &\left. - \frac{1}{2}r^2\cos\left(\frac{\pi z}{L}\right)\left[I_0\left(\frac{\pi r}{L}\right) + I_2\left(\frac{\pi r}{L}\right)\right]\right\}, \\ W_S(z,r) &= \frac{U_R}{I_0(\pi R/L)}\left\{ z^2\sin\left(\frac{\pi z}{L}\right) I_0\left(\frac{\pi r}{L}\right) - 2zr\cos\left(\frac{\pi z}{L}\right) I_1\left(\frac{\pi r}{L}\right) - \right. \\ &\left. - \frac{1}{2}r^2\sin\left(\frac{\pi z}{L}\right)\left[I_0\left(\frac{\pi r}{L}\right) + I_2\left(\frac{\pi r}{L}\right)\right]\right\}, \end{aligned} \tag{6}$$

I_2 is a modified second-order Bessel function [14].

To check if the functions (2), (4), (6) satisfy the three-dimensional Laplace equation and, therefore, whether they can be considered as electric potentials of some electrostatic field, the Wolfram Mathematica program [15] can be used, which provides an effective tool for symbolic calculations.

Note. The assumption of quasi-static high-frequency electric field is used for formulas of model electric potentials (1), (3), (5). It is valid when the time of the characteristic change in electrical voltages at the electrodes significantly exceeds the time of propagation of an electromagnetic perturbation within the device. Typical sizes of electrode configurations used in mass spectrometer designs are tens of centimeters. Taking into account the equality of the speed of light and the propagation velocity of electromagnetic perturbation, this assumption is obviously fulfilled for the frequencies of electrical voltages used in mass spectrometers (they amount to several gigahertz). In this case, the high-frequency electric potential, which varies in time and space, can be expressed as a product of a time function (it describes the temporal change in electrical voltages) by an electrostatic potential (corresponds to DC voltages at the electrodes). Such a step is, in fact, neglecting the electrodynamic effects, i.e., the accompanying electromagnetic wave.

Pseudopotential model of ion motion in the presence of viscous friction effect

A qualitative description of the motion of charged particles in high-frequency electric fields can be obtained with a pseudo-potential model of motion, according to which motion is divided into the sum of two terms: a «slow» component in some effective quasi-stationary force field and motion as high-frequency oscillations with a small amplitude.

Let us consider the motion of ions in a high-frequency electric field, whose potential has the form [16]:

$$U_{rf}(x,y,z,t)=\sum_k\Big(p_k(t)\cos(\omega_k t+\varphi_k)U^{(k)}(x,y,z)+ \\ +q_k(t)\sin(\omega_k t+\varphi_k)V^{(k)}(x,y,z)\Big), \tag{7}$$

where $p_k(t)$, $q_k(t)$ are «slow» functions of time; ω_k are "fast" frequencies far apart from each other on the frequency scale; $U^{(k)}(x, y, z)$, $V^{(k)}(x, y, z)$ are electrostatic fields corresponding to DC voltages at the electrodes of the device.

It is worth noting here that such concepts as "slow", "fast" and "far away" correspond to the characteristic times T_0 of the motion of ions in the transport channel:

$$dp_k(t)/dt\sim 1/T_0,\ dq_k(t)/dt\sim 1/T_0,\ \omega_k>>1/T_0,\ \forall i\neq j:\ |\omega_i-\omega_j|>>1/T_0),$$

in accordance with the pseudopotential model of ion motion (see [16, 18] and references therein).

In the presence of neutral gas, its action can be replaced by the presence of effective viscous friction, whose strength is set by Stokes' law [17]. Then, e, m are the charge and mass of the ion; $\Omega = \gamma/m$ is the effective frequency of collisions of ions with neutral gas molecules ($\gamma = \gamma(x, y, z, t)$ is the Stokes coefficient for the effective viscosity due to collisions of ions with neutral gas molecules in the vicinity of the considered point in space at the given time; this coefficient varies slowly over time and does not depend (in the first approximation) on the relative velocity of ions.

In further calculations, the subscripts for the functions $U^{(k)}(x, y, z)$ and $V^{(k)}(x, y, z)$ denote partial derivatives of potentials with respect to the corresponding spatial variables, the arguments of potentials are omitted for brevity.

After careful averaging of the equations of motion in a high-frequency electric field, it turns out that the above slow motion of the ion is carried out in a pseudopotential electric field with the pseudopotential $\bar{U}(x,y,z,t)$, which is due to the spatial gradient of the amplitude of the high-frequency electric field. This pseudopotential is expressed as [18]:

$$\bar{U}(x,y,z,t)=\sum_k\frac{e}{4m(\Omega^2+\omega_k^2)}\Big(p_k^2(t)\Big[\Big(U_x^{(k)}\Big)^2+\Big(U_y^{(k)}\Big)^2+\Big(U_{z\backslash}^{(k)}\Big)^2\Big]+ \\ +q_k^2(t)\Big[\Big(V_x^{(k)}\Big)^2+\Big(V_y^{(k)}\Big)^2+\Big(V_{z\backslash}^{(k)}\Big)^2\Big]\Big), \tag{8}$$

In addition, there is a non-potential pseudo-electric field in the equations of slow motion, with components $\left(\overline{E}_X, \overline{E}_Y, \overline{E}_Z\right)$ associated with the presence of viscous friction and with the spatial phase gradient of the high-frequency electric field:

$$\overline{E}_X\left(x,y,z,t\right)=\sum_k \frac{e\left(\Omega/\omega_k\right)}{2m\left(\Omega^2+\omega_k^2\right)} p_k\left(t\right) q_k\left(t\right)\left(U_x^{(k)}V_{xx}^{(k)}-U_{xx}^{(k)}V_x^{(k)}+\right.$$
$$\left.+U_y^{(k)}V_{xy}^{(k)}-U_{xy}^{(k)}V_y^{(k)}+U_z^{(k)}V_{xz}^{(k)}-U_{xz}^{(k)}V_z^{(k)}\right), \quad (9)$$

$$\overline{E}_Y\left(x,y,z,t\right)=\sum_k \frac{e\left(\Omega/\omega_k\right)}{2m\left(\Omega^2+\omega_k^2\right)} p_k\left(t\right) q_k\left(t\right)\left(U_x^{(k)}V_{xy}^{(k)}-U_{xy}^{(k)}V_x^{(k)}+\right.$$
$$\left.+U_y^{(k)}V_{yy}^{(k)}-U_{yy}^{(k)}V_y^{(k)}+U_z^{(k)}V_{yz}^{(k)}-U_{yz}^{(k)}V_z^{(k)}\right), \quad (10)$$

$$\overline{E}_Z\left(x,y,z,t\right)=\sum_k \frac{e\left(\Omega/\omega_k\right)}{2m\left(\Omega^2+\omega_k^2\right)} p_k\left(t\right) q_k\left(t\right)\left(U_x^{(k)}V_{xz}^{(k)}-U_{xz}^{(k)}V_x^{(k)}+\right.$$
$$\left.+U_y^{(k)}V_{yz}^{(k)}-U_{yz}^{(k)}V_y^{(k)}+U_z^{(k)}V_{zz}^{(k)}-U_{zz}^{(k)}V_z^{(k)}\right). \quad (11)$$

In addition to these pseudo-forces, the equations of slow motion also include the viscous friction force (originally present in them) with components (F_X, F_Y, F_Z):

$$\begin{aligned}
\overline{F}_X\left(x,y,z,t\right)&=-\gamma\left(x,y,z,t\right)\left(\dot{x}\left(t\right)-u_X\left(x,y,z,t\right)\right),\\
\overline{F}_Y\left(x,y,z,t\right)&=-\gamma\left(x,y,z,t\right)\left(\dot{y}\left(t\right)-u_Y\left(x,y,z,t\right)\right),\\
\overline{F}_Z\left(x,y,z,t\right)&=-\gamma\left(x,y,z,t\right)\left(\dot{z}\left(t\right)-u_Z\left(x,y,z,t\right)\right),
\end{aligned} \quad (12)$$

where (u_X, u_Y, u_Z) are the components of the gas flow velocity (it changes slowly over time) in the vicinity of the considered point in space at the considered time; $\left(\dot{x},\dot{y},\dot{z}\right)$ are the velocity components of the slow (averaged over fast oscillations) motion of the ion.

Analysis of the properties of a conical funnel with two-phase power supply

We use a pseudopotential model to describe the motion of ions in a high-frequency electric field (3). The pseudopotential is calculated using the general Eq. (8). In this case, there is no spatial phase gradient of the high-frequency electric field, so we are dealing with a non-potential pseudo-electric force. The three-dimensional graph of the pseudopotential has the form of a kind of gutter with the edges that grow sharply with distance away from the axis and approaching the electrodes, and the slope of the edges of the gutter increases as it approaches the outlet from the funnel. This means that such a high-frequency electric field effectively «presses» the ions to the axis of the device, and this pressing force increases significantly along the direction of motion, making the ion beam progressively narrower.

The presence of pseudopotential corrugation on the axis of the system can be considered an unpleasant effect, since it can create parasitic local traps for ions. In addition, in such a design, there is no relying on cooling of the ions (discharging their excess kinetic energy), since the high-frequency electric field on the axis is not zero and therefore the ions are constantly oscillated by this field.

In addition to these disadvantages, the distribution of the pseudopotential function along the funnel axis slowly increases towards its outlet, which slows down the motion of ions and makes it difficult for them to escape through the outlet. For such a funnel, it turns out necessary to apply an additional pulling electric field. Such a measure can be implemented if additional static potentials are applied to the apertures and a constant increase in static electric potential is ensured between adjacent apertures.

Analysis of the properties of a conical funnel with four-phase power supply

For a radio frequency trap with conical electrodes (see Fig. 1,*b*), in which the phase difference of RF voltages applied to adjacent apertures is equal to $\pi/2$, the electric potential near the axis is described with good accuracy by the expression

$$V(z,r,t) = V_C^*(z,r)\cos(\omega t+\varphi) + V_S^*(z,r)\sin(\omega t+\varphi), \tag{13}$$

where

$$\begin{aligned} V_C^*(z,r) &= \frac{U_R}{I_0(\pi R/2L)}\left[z\cos\left(\frac{\pi z}{2L}\right)I_0\left(\frac{\pi r}{2L}\right) + r\sin\left(\frac{\pi z}{2L}\right)I_1\left(\frac{\pi r}{2L}\right)\right], \\ V_S^*(z,r) &= \frac{U_R}{I_0(\pi R/2L)}\left[z\sin\left(\frac{\pi z}{2L}\right)I_0\left(\frac{\pi r}{L}\right) - r\cos\left(\frac{\pi z}{2L}\right)I_1\left(\frac{\pi r}{2L}\right)\right]. \end{aligned} \tag{14}$$

In this case, a gutter also appears on the pseudopotential graph, which effectively «presses» the ions to the axis of the device, and the force of such an impact increases quadratically as it approaches the outlet from the funnel. However, there is no pseudopotential corrugation along the axis, which guarantees the absence of local parasitic traps for ions along the axis of their motion. Nevertheless, the value of the pseudopotential on the axis is still not zero, which means that a high-frequency electric field is formed, which oscillates the ions. Therefore, in such a design, there is also no relying on cooling of the ions (discharging their excess kinetic energy). In addition, the distribution of the pseudopotential function along the funnel axis slowly increases with increasing distance from zero in the z coordinate, which somewhat slows down the motion of ions and prevents their passage through the outlet, similar to the previous case.

In addition to the described features of the model, in the case of a high-frequency electric field of the form (13), there is a spatial gradient of its phase. But it plays a positive role, since the non-potential electric pseudoforce that arises in this case only additionally presses the ions against the axis of the device and provides a constant pulling force directed towards the outlet from the funnel. This factor could make it possible to do without an additional pulling static electric field, however, the pulling electric pseudoforce depends on the mass and, therefore, a new obstacle is possible, additional ion mass discrimination. A situation may arise for too large masses when the pulling of the pseudo-force is unable to overcome the braking of moving ions and an obstacle will prevent them from escaping the funnel.

Analysis of the properties of a conical funnel in pseudopotential mode with an Archimedean wave

Systems with a traveling pseudopotential wave are considered in [19–21]. In this case, the properties of an RF trap with conical electrodes are investigated, for which a high-frequency electric field forms a slowly traveling pseudopotential wave along the axis of the device. At the minima of the pseudopotential wave, the high-frequency electric field is zero, and it is at these points that local ion clusters are formed, which then move along the axis of the device to escape the funnel, simultaneously with the shift of the local minima of the pseudopotential wave. Importantly, the transport of ions is ensured by this, independent of their mass, since the speed with which the minima of the pseudopotential wave move is determined by the parameters of high-frequency voltages applied to the electrodes of the device (and nothing else).

For such an RF trap, the electric potential near the axis is described with good accuracy by the following formula:

$$V(z,r,t)=\left[V_C^*(z,r)\cos(2\pi t/T)+V_S^*(z,r)\sin(2\pi t/T)\right]\cos(\omega t+\varphi), \qquad (14)$$

where T is the period of «slow» time that determines the speed of transport; the potentials $V_C^*(z,r)$ and $V_S^*(z,r)$ are given by Eqs. (14).

In this case, the graph also shows a corrugated gutter of the pseudopotential, and as the process progresses over time, the corrugation slowly moves along with the transporting pseudopotential wave, effectively pressing the ions against the axis of the device. In addition, a slow-moving pseudopotential wave appears on the axis of the system, forcibly transporting ions from the inlet to the outlet.

The traveling pseudopotential wave on the axis of the RF funnel is characterized by a variable maximum amplitude, which increases quadratically as it approaches the outlet from the funnel, but at the points of the minimum, the pseudopotential value is zero. As noted above, ions are trapped and their clusters are formed at the points of the minimum, and with the slow motion of the pseudopotential wave along the axis of the device, synchronized transport of ions is carried out, regardless of their mass. Since the high-frequency electric field in the centers of ion clusters is exactly zero, and the high-frequency electric field turns out to be very small for minor deviations of ions within the volume of the cluster, it is quite possible to count on at least partial cooling of ions during transport through the forevacuum region of the gas-dynamic interface.

The spatial phase gradient of the high-frequency electric field for the electric potential (15) is zero, therefore, there are no additional effects associated with the presence of a non–potential pseudoelectric force (see Eqs. (9)–(11)) in such a system.

Conclusion

Analytical models of the high-frequency electric field were constructed for radio-frequency ion funnels with circular apertures. They were used for qualitative analysis of RF ion funnels in different operating modes. We can conclude that the most promising configuration considered is the RF funnel with a conical transport channel and electric power supply, allowing to generate a pseudopotential Archimedean wave on the axis of the device.

REFERENCES

1. **Ketola R. A., Kiuru J. T., Tarkiainen V., et al.,** Comparison of analytical performance of a micro array quadrupole instrument and a conventional quadrupole mass spectrometer equipped with membrane inlets, Rapid Commun. Mass Spectrom. 17 (7) (2003) 17. 753–756.

2. **Adamov A., Viidanoja J., Kärpänoja E., et al.,** Interfacing an aspiration ion mobility spectrometer to a triple quadrupole mass spectrometer, Rev. Sci. Instrum. 78 (4) (2007) 044101.

3. **Troyan V. I., Borisyuk P. V., Krasavin A. V., et al.,** Multisectional linear ion trap and novel loading method for optical spectroscopy of electron and nuclear transitions, Eur. J. Mass Spectrom. 21 (1) (2015) 1–12.

4. **Borisyuk P. V., Vasiliev O. S., Derevyashkin S. P., et al.,** Trapping, retention and laser cooling of Th^{3+} ions in a multisection linear quadrupole trap, Quant. Electron. 47 (5) (2017) 406–411.

5. **Borisyuk P. V., Derevyashkin S. P., Khabarova K. Y., et al.,** Loading of mass spectrometry ion trap with Th ions by laser ablation for nuclear frequency standard application, Eur. J. Mass Spectrom. 23 (4) (2017) 146–151.

6. **Borisyuk P. V., Derevyashkin S. P., Khabarova K. Y., et al.,** Mass selective laser cooling of $^{229}Th^{3+}$ in a multisectional linear Paul trap loaded with a mixture of thorium isotopes, Eur. J. Mass Spectrom. 23 (4) (2017) 136–139.

7. **Konenkov A. N., Konenkov N. V., Sysoev A. A.,** Modeling dipolar excitation for quadrupole mass filter, Eur. J. Mass Spectrom. 28 (1–2) (2022) 65–72.

8. **Sysoev A. A., Konenkov A. N., Konenkov N. V.,** Balance of the 6th and 10th spatial harmonics amplitudes of a quadrupole mass filter with round rods, Int. J. Mass Spectrom. 482 (December) (2022) 116949.

9. **Bugrov P. V., Sysoev A. A., Konenkov A. N., Konenkov N. V.,** Properties of the multipole fields formed by round electrodes, Int. J. Mass Spectrom. 490 (August) (2023) 117081.

10. **Bugrov P. V., Sysoev A. A., Konenkov N. V.,** Modeling of a quadrupole mass filter with octupole field, Mass-Spektrometriya. 19 (3) (2022) 197–200 (in Russian).

11. **Yavor M. I.,** Optics of charged particle analyzers (Advances of Imaging and Electron Physics. Vol. 157), Academic Press, Amsterdam (2009) 142–168.

12. **Teloy E., Gerlich D.,** Integral cross sections for ion-molecule reactions. Part I. The guided beam technique, Chem. Phys. 4 (3) (1974) 417–427.

13. **Gerlich D., Kaefer G.,** Ion trap studies of association processes in collisions of CH^{3+} and CD^{3+} with *n*-H_2, *p*-H_2, D_2 and He at 80 K, Astrophys. J. 347 (2) (1989) 849–854.

14. **Abramowitz M. A., Stegun I. A. (Eds.),** Handbook of mathematical functions with formulas, graphs, and mathematical tables, Tenth edition, National Bureau of Standards, Washington, 1972.

15. Wolfram Mathematica: the system for modern technical computing, URL: http://wolfram.com/mathematica/

16. **Berdnikov A. S.,** A pseudo potential description of the motion of charged particles in RF fields, Microsc. Microanal. 21 (S4) (2015) 78–83.

17. **Landau L. D., Lifshitz E. M.,** Course of theoretical physics. Vol. 10. **Pitayevskii L. P, Lifshitz E. M.,** Physical kinetics, First edition, Butterworth-Heinemann, Oxford, UK. 1981.

18. **Tolmachev A. V., Chernushevich I. V., Dodonov A. F., Standing K. G.,** A collisional focusing ion guide for coupling an atmospheric pressure ion source to a mass spectrometer, Nucl. Instrum. Meth. Phys. Res. B. 124 (1) (1997) 112–119.

19. **Andreeva A. D., Berdnikov A. S.,** Mass spectrometric devices with Archimedean radio frequency electric fields, J. Anal. Chem. 67 (13) (2012) 1034–1037.

20. **Berdnikov A. S., Andreeeva A. D.,** Ustroystvo dlya manipulirovaniya zaryazhennymi chastitsami, patent Federalnoy sluzhby RF po intellektualnoy sobstvennosti na poleznuyu model RU 113611 (Data prioriteta/podchi 05.05.2011) [A device for charge particle manipulation. Patent of the Federal Service for Intellectual Property for a utility model of the Russian Federation] – RU 113611 (05.05.2011) (in Russian).

21. **Berdnikov A. S., Andreeeva A. D.,** Ustroystvo dlya manipulirovaniya zaryazhennymi chastitsami, patent Federalnoy sluzhby RF po intellektualnoy sobstvennosti na izobreteniye RU 113611 (Data prioriteta/podchi 05.05.2011) [A device for charge particle manipulation. Patent of the Federal Service for Intellectual Property for an invention of the Russian Federation] – RU 2465679 (2012) (in Russian).

СПИСОК ЛИТЕРАТУРЫ

1. **Ketola R. A., Kiuru J. T., Tarkiainen V., Kotiaho T., Sysoev A. A.** Comparison of analytical performance of a micro array quadrupole instrument and a conventional quadrupole mass spectrometer equipped with membrane inlets // Rapid Communications in Mass Spectrometry. 2003. Vol. 17. No. 7. Pp. 753–756.

2. **Adamov A., Viidanoja J., Kärpänoja E., Paakkanen H., Ketola R. A., Kostiainen R., Sysoev A., Kotiaho T.** Interfacing an aspiration ion mobility spectrometer to a triple quadrupole mass spectrometer // Review of Scientific Instruments. 2007. Vol. 78. No. 4. P. 044101.

3. **Troyan V. I., Borisyuk P. V., Krasavin A. V., Vasiliev O. S., Palchikov V. G., Avdeev I. A., Chernyshev D. M., Poteshin S. S., Sysoev A. A.** Multisectional linear ion trap and novel loading method for optical spectroscopy of electron and nuclear transitions // European Journal of Mass Spectrometry. 2015. Vol. 21. No. 1. Pp. 1–12.

4. **Борисюк П. В., Васильев О. С., Деревяшкин С. П. и др.** Захват, удержание и лазерное охлаждение ионов Th^{3+} в многосекционной линейной квадрупольной ловушке // Квантовая электроника. 2017. Т. 47. № 5. С. 406–411.

5. **Borisyuk P. V., Derevyashkin S. P., Khabarova K. Y., et al.** Loading of mass spectrometry ion trap with Th ions by laser ablation for nuclear frequency standard application // European Journal of Mass Spectrometry. 2017. Vol. 23. No. 4. Pp. 146–151.

6. **Borisyuk P. V., Derevyashkin S. P., Khabarova K. Y., et al.** Mass selective laser cooling of $^{229}Th^{3+}$ in a multisectional linear Paul trap loaded with a mixture of thorium isotopes // European Journal of Mass Spectrometry. 2017. Vol. 23. No. 4. Pp. 136–139.

7. **Konenkov A. N., Konenkov N. V., Sysoev A. A.** Modeling dipolar excitation for quadrupole mass filter // European Journal of Mass Spectrometry. 2022. Vol. 28. No. 1–2. Pp. 65–72.

8. **Sysoev A. A., Konenkov A. N., Konenkov N. V.** Balance of the 6th and 10th spatial harmonics amplitudes of a quadrupole mass filter with round rods // International Journal of Mass Spectrometry. 2022. Vol. 482. December. P. 116949.

9. **Bugrov P. V., Sysoev A. A., Konenkov A. N., Konenkov N. V.** Properties of the multipole fields formed by round electrodes // International Journal of Mass Spectrometry. 2023. Vol. 490. August. P. 117081.

10. **Бугров П. В., Сысоев А. А., Коненков Н. В.** Моделирование квадрупольного фильтра масс с октупольным полем // Масс-спектрометрия 2022. Т. 19. № 3. С. 197–200.

11. **Yavor M. I.** Optics of charged particle analyzers. Amsterdam: Academic Press, 2009 (Advances of Imaging and Electron Physics. Vol. 157). Pp. 142–168.

12. **Teloy E., Gerlich D.** Integral cross sections for ion-molecule reactions. Part I. The guided beam technique // Chemical Physics. 1974. Vol. 4. No. 3. Pp. 417–427.

13. **Gerlich D., Kaefer G.** Ion trap studies of association processes in collisions of CH^{3+} and CD^{3+} with n-H_2, p-H_2, D_2 and He at 80 K // The Astrophysical Journal. 1989. Vol. 347. No. 2. Pp. 849–854.

14. **Абрамовиц М., Стиган И.** Справочник по специальным функциям с формулами, графиками и математическими таблицами. М.: Наука, 1979. 832 с.

15. Wolfram Mathematica: the system for modern technical computing; URL: http://wolfram.com/mathematica/

16. **Berdnikov A. S.** A pseudo potential description of the motion of charged particles in rf fields // Microscopy and Microanalysis. 2015. Vol. 21. No. S4. Pp. 78–83.

17. **Ландау Л. Д., Лифшиц Е. М.** Курс теоретической физики. В 10 тт. Т. 10. Физическая кинетика (**Лифшиц Е. М., Питаевский Л. П.**). М.: Физматлит, 2007. 536 с.

18. **Tolmachev A. V., Chernushevich I. V., Dodonov A. F., Standing K. G.** A collisional focusing ion guide for coupling an atmospheric pressure ion source to a mass spectrometer // Nuclear Instruments and Methods in Physics Research B. 1997. Vol. 124. No. 1. Pp. 112–119.

19. **Андреева А. Д., Бердников А. С.** Масс-спектрометрические устройства на основе радиочастотных электрических полей с архимедовыми свойствами // Масс-спектрометрия. 2011. Т. 4 № .8. С. 296–293.

20. **Бердников А. С., Андреева А. Д.** Устройство для манипулирования заряженными частицами. 2011. Патент Федеральной службы Российской Федерации по интеллектуальной собственности на полезную модель RU 113611 (Дата приоритета/подачи заявки: 05.05.2011).

21. **Бердников А. С., Андреева А. Д.** Устройство для манипулирования заряженными частицами. 2012. Патент Федеральной службы Российской Федерации по интеллектуальной собственности на изобретение RU 2465679 (Дата приоритета/подачи заявки: 05.05.2011).

THE AUTHORS

SYSOEV Alexey A.
National Research Nuclear University MEPHI (Moscow Engineering Physics Institute)
31 Kashirskoe HWY, Moscow, 115409, Russia
aasysoyev@mephi.ru
ORCID: 0000-0003-0985-5964

BERDNIKOV Alexander S.
Institute for Analytical Instrumentation, RAS
31–33, Ivana Chernykh St., St. Petersburg, 198095, Russia
asberd@yandex.ru
ORCID: 0000-0003-0985-5964

MASYUKEVICH Sergey V.
Institute for Analytical Instrumentation, RAS
31–33, Ivana Chernykh St., St. Petersburg, 198095, Russia
serg_08@mail.ru
ORCID: 0000-0002-0873-8849

SOLOVYEV Konstantin V.
Peter the Great St. Petersburg Polytechnic University,
Institute for Analytical Instrumentation, RAS
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
k-solovyev@mail.ru
ORCID: 0000-0003-3514-8577

KRASNOVA Nadezhda K.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
n.k.krasnova@mail.ru
ORCID: 0000-0002-6162-9481

СВЕДЕНИЯ ОБ АВТОРАХ

СЫСОЕВ Алексей Александрович – *доктор физико-математических наук, профессор, главный научный сотрудник Национального исследовательского ядерного университета МИФИ.*
115409, Россия, г. Москва, Каширское шоссе, 31
aasysoyev@mephi.ru
ORCID: 0000-0003-0985-5964

БЕРДНИКОВ Александр Сергеевич – *доктор физико-математических наук, ведущий научный сотрудник Института аналитического приборостроения РАН.*
198095, Россия, г. Санкт-Петербург, ул. Ивана Черных, 31–33, лит. А.
asberd@yandex.ru
ORCID: 0000-0003-0985-5964

МАСЮКЕВИЧ Сергей Владимирович – *старший научный сотрудник Института аналитического приборостроения РАН.*
198095, Россия, г. Санкт-Петербург, ул. Ивана Черных, 31–33, лит. А.
serg_08@mail.ru
ORCID: 0000-0002-0873-8849

СОЛОВЬЕВ Константин Вячеславович – *кандидат физико-математических наук, доцент Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого*, младший научный сотрудник Института аналитического приборостроения Российской академии наук.
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
k-solovyev@mail.ru
ORCID: 0000-0003-3514-8577

КРАСНОВА Надежда Константиновна – *доктор физико-математических наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
n.k.krasnova@mail.ru
ORCID: 0000-0002-6162-9481

Received 25.09.2023. Approved after reviewing 16.10.2023. Accepted 16.10.2023.
Статья поступила в редакцию 25.09.2023. Одобрена после рецензирования 16.10.2023. Принята 16.10.2023.

PHYSICAL MATERIALS TECHNOLOGY

Original article
DOI: https://doi.org/10.18721/JPM.16412

COMPARISON OF APPROACHES TO ACCOUNTING FOR IMPERFECT CONTACTS WHEN DETERMINING THE EFFECTIVE PERMEABILITY OF MATERIAL

K. P. Frolova[1]✉, *E. N. Vilchevskaya*[1]

[1] Institute for Problems of Mechanical Engineering, RAS, St. Petersburg, Russia.

✉ fkp@ipme.ru

Abstract. The paper develops a complex approach to accounting for imperfect contacts (IC) when determining effective properties of various nature. The IC are assumed to be caused by various factors (microstructure features, process's specifity and so on). To obtain macroscopic properties, we seek a solution of the homogenization problem for the material containing isolated ellipsoidal inhomogeneities when fields are discontinuous at the interphase boundaries. The paper considers, generalizes and compares two existing approaches to accounting for the IC, namely, an approach where IC is modeled by means of a field jump specified in terms of a ratio of field values on the outer and inner sides of the inhomogeneity boundary, and approach, which introduces inhomogeneity with a surface effect. To take into account IC, we have considered an equivalent inhomogeneity with ideal contacts at the boundary. Working the problem on determining the effective diffusional permeability of material provided an example.

Keywords: effective properties, imperfect contact, equivalent inhomogeneity, effective diffusional permeability, homogenization problem

Funding: The reported study was funded by Russian Science Foundation, Grant No. 23-79-01133 (https://rscf.ru/project/23-79-01133/).

Citation: Frolova K. P., Vilchevskaya E. N., Comparison of approaches to accounting for imperfect contacts when determining the effective permeability of material, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 146–159. DOI: https://doi.org/10.18721/JPM.16412

Научная статья
УДК 539.21
DOI: https://doi.org/10.18721/JPM.16412

СРАВНЕНИЕ ПОДХОДОВ К УЧЕТУ НЕИДЕАЛЬНЫХ КОНТАКТОВ ПРИ ОПРЕДЕЛЕНИИ ЭФФЕКТИВНОЙ ПРОНИЦАЕМОСТИ МАТЕРИАЛА

К. П. Фролова[1]✉, *Е. Н. Вильчевская*[1]

[1] Институт проблем машиноведения РАН, Санкт-Петербург, Россия.

✉ fkp@ipme.ru

Аннотация. В работе развивается комплексный подход к учету неидеальных контактов (НК), появление которых вызвано разнообразными факторами (особенности микроструктуры, специфика процесса на мезоуровне и т. п.), при определении эффективных свойств материала различной природы, представляемых тензорами второго ранга. Макроскопические свойства определяются путем решения задачи гомогенизации для материала, состоящего из матрицы и изолированных эллипсоидальных неоднородностей, на границе которых поля не являются непрерывными. Рассмотрены, обобщены и сопоставлены существующие подходы к учету НК: подход, при котором НК моделируют, вводя скачок поля на границе раздела фаз через задаваемое отношение значений поля по обе стороны границы, а также подход, при котором в рассмотрение вводится неоднородность с поверхностным эффектом. С целью учета НК при нахождении эффективных свойств материала, рассматривается эквивалентная неоднородность с идеальными контактами на границе, вклад которой в макроскопическое свойство эквивалентен вкладу исходной неоднородности. В качестве примера решена задача об определении эффективной диффузионной проницаемости материала.

Ключевые слова: эффективные свойства, неидеальный контакт, эквивалентная неоднородность, эффективная диффузионная проницаемость, задача гомогенизации

Финансирование: Исследование выполнено при финансовой поддержке Российского научного фонда, грант № 23-79-01133 (https://rscf.ru/project/23-79-01133/).

Ссылка для цитирования: Фролова К. П., Вильчевская Е. Н. Сравнение подходов к учету неидеальных контактов при определении эффективной проницаемости материала // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. .16 4 №. С. 146–159. DOI: https://doi.org/10.18721/JPM.16412

Introduction

The properties of a material that is inhomogeneous at the microlevel directly depend on its structure and can be determined within the framework of continuum theory using homogenization methods. Physical fields are introduced into consideration, which, as a rule, are assumed to be continuous at the interphase boundaries. From a physical standpoint, this means that there are "ideal" contacts at the internal boundaries. At the same time, a number of phenomena should be described taking into account the presence of imperfect contacts, which can occur both due to the peculiarities of the microstructure of the material and in connection with the specifics of the described process [1–4].

As a rule, the issues of taking into account imperfect contacts to determine the effective properties are considered in the literature separately, in the context of describing processes that are different in nature. For example, some authors have drawn attention to the need to take into account the phenomenon of segregation when determining effective diffusion coefficients. This phenomenon is understood as the sedimentation of impurities in structural defects, which

is characteristic for mass transfer [2, 5–8]. Such a procedure was implemented by introducing a concentration jump in terms of the ratio of concentrations from the outer and inner sides of the interphase boundary (segregation parameter). Using this approach, the authors of [2] obtained the Voigt–Reuss and Hashin–Shtrikman boundaries for effective impurity mobility by expressing the diffusion flux in terms of a gradient of chemical potential (the potential was assumed to be continuous), after which the effective diffusion coefficients were determined directly. A constant segregation parameter in [5, 6] was introduced into the equations of the modified effective medium method. This parameter was introduced into the equations of effective field methods in [7, 8].

Approaches to accounting for imperfect contacts in determining the effective thermal or electrical conductivity of a material inhomogeneous at the microlevel have been considered separately in the literature [9–12]. It was believed that such contacts appear due to the presence of surface defects (roughness, delamination, etc.). Modeling of imperfect contacts was carried out by considering inhomogeneities with a surface effect (it was assumed that such inhomogeneities were covered with a layer with extreme properties, whose thickness tends to zero). The surface effect was taken into account either by determining the magnitude of the field jump from solving the problem of isolated inhomogeneity in an infinite matrix [9], or by approximating expressions for concentration tensors connecting the average fields inside the inhomogeneity with the applied field [10–12].

The similarity of the equations of diffusion, heat and electrical conductivity allows to make an assumption about the possibility of developing a unified approach to modeling imperfect contacts caused by different factors to determine the effective properties of varying nature for materials.

The goal of this study is to generalize and compare the available approaches to accounting for imperfect contacts in determining the effective properties for cases of materials with spheroidal and ellipsoidal inhomogeneities.

Statement of homogenization problem

The effective properties of the material are found by solving the homogenization problem for a representative volume V, which is a particle of a continuous medium at the macro level. Effective properties are expressed using tensor quantities relating the fields that are average in terms of representative volume. As a rule, it is assumed that the homogenized material satisfies the simplest linear governing relations. Due to the similarity of the equations of diffusion, heat and electrical conductivity, below we will limit ourselves to the consideration of the diffusion problem, for which Fick's law holds true:

$$< \mathbf{J} >_V = -\mathbf{D}^{eff} \cdot < \nabla c >_V, \tag{1}$$

where $\mathbf{D}^{eff}$ *is* the effective diffusion permeability tensor (the diffusion tensor of an impurity in a homogenized material), $\mathbf{J}$ is the diffusion flux, c is the concentration, ∇ is the nabla operator, $< ... >_V = \frac{1}{V}\int_V (...) dV$.

To find the fields to be averaged, the stationary diffusion problem is solved. The law of conservation in the absence of internal sources/sinks has the following form:

$$\nabla \cdot \mathbf{J}(\mathbf{x}) = 0, \tag{2}$$

where $\mathbf{x}$ is the radius vector of a point inside volume V.

The flux and concentration gradient at each point of the representative volume are related by a linear governing relation:

$$\mathbf{J}(\mathbf{x}) = -\mathbf{D}(\mathbf{x}) \cdot \nabla c(\mathbf{x}), \tag{3}$$

where $\mathbf{D}(\mathbf{x})$ is the diffusion permeability tensor of the material at point $\mathbf{x}$.

The independence of effective properties from the conditions at the boundary of the representative volume allows to choose them arbitrarily. It is convenient to set a homogeneous Hill condition, which in the case of the diffusion problem has the form $c(\mathbf{x})\big|_\Sigma = \mathbf{G}_0 \cdot \mathbf{x}$. Then the average value of the concentration gradient is completely determined by the boundary condition [13]:

$$< \nabla c >_V = \mathbf{G}_0. \tag{4}$$

The presence of boundaries within volume V (interphase boundaries Γ) requires imposing additional boundary conditions. These conditions will vary depending on the method of accounting for imperfect contacts.

Next, we consider a material consisting of an isotropic matrix characterized by a diffusion permeability tensor $\mathbf{D}_0 = D_0\mathbf{I}$ ($\mathbf{I}$ is the unit tensor), and ellipsoidal inhomogeneities with the volume V_1 with the permeability $\mathbf{D}_1 = D_1\mathbf{I}$, giving the conditions at the interphase boundaries.

One of the simplest methods known from the literature to account for imperfect contact is the introduction of a field jump at the interface between the matrix (+) and the inhomogeneity (−) using a constant ratio of field values from the outer and inner sides of the boundary. In the context of the diffusion problem, either the concentration field or the normal component of the flow can experience a jump. In the first case, the following conditions hold true at the interface of the phases Γ with the external normal $\mathbf{n}_\Gamma$:

$$D_0 \left.\frac{\partial c(\mathbf{x})}{\partial n_\Gamma}\right|_{\mathbf{x}\to\Gamma+} = D_1 \left.\frac{\partial c(\mathbf{x})}{\partial n_\Gamma}\right|_{\mathbf{x}\to\Gamma-}, \; c(\mathbf{x})\big|_{\mathbf{x}\to\Gamma+} = s_c\, c(\mathbf{x})\big|_{\mathbf{x}\to\Gamma-}, \tag{5}$$

where s_c is the segregation parameter; the jump is expressed as $[c] = (s_c - 1)c(\mathbf{x})\big|_{\mathbf{x}\to\Gamma-}$.

If there is a jump in the normal component of the flux J_n, the following conditions can be imposed by introducing the segregation parameter s_f into consideration:

$$D_0 \left.\frac{\partial c(\mathbf{x})}{\partial n_\Gamma}\right|_{\mathbf{x}\to\Gamma+} = s_f\, D_1 \left.\frac{\partial c(\mathbf{x})}{\partial n_\Gamma}\right|_{\mathbf{x}\to\Gamma-}, \; c(\mathbf{x})\big|_{\mathbf{x}\to\Gamma+} = c(\mathbf{x})\big|_{\mathbf{x}\to\Gamma-}. \tag{6}$$

In this case, the jump is defined as $[J_n] = (s_f - 1)\mathbf{n}_\Gamma \cdot \mathbf{J}(\mathbf{x})\big|_{\mathbf{x}\to\Gamma-}$.

Another way of accounting for imperfect contact is used for inhomogeneities with a surface effect. In the general case, inhomogeneities representing confocal ellipsoids are placed in the matrix, for which the conductivity of the inner ellipsoid is $\mathbf{D}_1 = D_1\mathbf{I}$, and the conductivity of the outer layer is $\mathbf{D}_s = D_s\mathbf{I}$.

The semi-major axes of the outer ellipsoid b_1, b_2, b_3 and the inner ellipsoid a_1, a_2, a_3 are related as follows:

$$b_i^2 = a_i^2 + \xi,$$

where $i = 1, 2, 3$; ξ is a constant.

Perfect contacts take place at the inner boundaries Γ_a of the inner ellipsoid of volume V_a with the outer normal $\mathbf{n}_{\Gamma_a}$ and Γ_b of the outer ellipsoid of volume V_b with the outer normal $\mathbf{n}_{\Gamma_b}$:

$$\begin{aligned} &D_0 \left.\frac{\partial c(\mathbf{x})}{\partial n_{\Gamma_b}}\right|_{\mathbf{x}\to\Gamma_b+} = D_s \left.\frac{\partial c(\mathbf{x})}{\partial n_{\Gamma_b}}\right|_{\mathbf{x}\to\Gamma_b-}, \; c(\mathbf{x})\big|_{\mathbf{x}\to\Gamma_b+} = c(\mathbf{x})\big|_{\mathbf{x}\to\Gamma_b-}, \\ &D_s \left.\frac{\partial c(\mathbf{x})}{\partial n_{\Gamma_a}}\right|_{\mathbf{x}\to\Gamma_a+} = D_1 \left.\frac{\partial c(\mathbf{x})}{\partial n_{\Gamma_a}}\right|_{\mathbf{x}\to\Gamma_b-}, \; c(\mathbf{x})\big|_{\mathbf{x}\to\Gamma_a+} = c(\mathbf{x})\big|_{\mathbf{x}\to\Gamma_a-}. \end{aligned} \tag{7}$$

To take into account the surface effect, it is necessary to pass to the limit at

$\xi \to 0$, as well as (in the context of the diffusion problem) either at $D_s \to 0$ or at $D_s \to \infty$. In the first case, corresponding to insulating coating, it is convenient to introduce an equivalent surface resistance into consideration

$$\beta = \frac{V_s}{D_s S_a} = \frac{4\pi\left(a_1^2 a_2^2 + a_1^2 a_3^2 + a_2^2 a_3^2\right)}{6 a_1 a_2 a_3 S_a} \lim_{\xi \to 0, D_s \to 0} \frac{\xi}{D_s}, \tag{8}$$

where $V_s = \lim_{\xi \to 0} V_b - V_a$, S_a is the surface area of the inhomogeneity with the volume V_a.

In the second case, corresponding to conductive coating, it is convenient to introduce an equivalent surface permeability

$$\lambda = \frac{D_s V_s}{S_a} = \frac{4\pi\left(a_1^2 a_2^2 + a_1^2 a_3^2 + a_2^2 a_3^2\right)}{6 a_1 a_2 a_3 S_a} \lim_{\xi \to 0, D_s \to \infty} \xi D_s. \tag{9}$$

The inhomogeneity with imperfect contacts can be formally replaced by an equivalent inhomogeneity with perfect contacts, which affects the effective properties in the same way as the initial one. To carry out such a replacement, it is necessary to determine the properties $\mathbf{D}^*$ that an equivalent inhomogeneity should possess. These properties will vary depending on the method of accounting for imperfect contact.

The introduction of equivalent inhomogeneity has the advantage that it becomes possible to use existing homogenization methods developed under the assumption of continuity of fields at the real interphase boundary. In this case, it is sufficient to substitute the corresponding diffusion coefficients of the impurity inside the inhomogeneity into expressions known from the literature. Since it is sufficient to take into account the presence of imperfect contacts at the stage of determining the diffusion permeability of equivalent inhomogeneity in this manner, here we will limit ourselves to qualitative and quantitative analysis of expressions for $\mathbf{D}^*$.

Note that imperfect contact, modeled by setting a concentration jump or by considering inhomogeneity with *insulating* coating, may occur when an impurity is aggregated at the interphase boundary. On the other hand, imperfect contact, which is modeled by setting another jump, namely, the normal component of the flux, or by considering inhomogeneity with *conductive* coating, may occur when additional diffusion paths are formed along the interphase boundary.

In view of this, it is of interest for each of these cases to compare two approaches to modeling imperfect contacts:

by setting the field jump in terms of the segregation parameter;

by considering an inhomogeneity with a surface effect.

The effective property can be expressed as a function of various microstructural parameters. This article uses the approach developed by Sevostyanov and Kachanov [13], where the role of the microstructural parameter is played by the sum of the tensors of the contribution of inhomogeneities. Below, we give the expressions for these tensors in the presence of imperfect contacts in the material, modeled using the approaches discussed above.

Contribution tensors

The contribution tensors are determined assuming that the inhomogeneities are isolated. If the concentration is set at the boundary of the representative volume, then the average gradient c over the representative volume is fully determined, while the average flux depends on the microstructure; it can be represented as a sum

$$< \mathbf{J} >_V = -\mathbf{D}_0 \cdot \mathbf{G}_0 + \Delta\mathbf{J}, \tag{10}$$

where $\Delta\mathbf{J}$ is the additional flux due to the presence of inhomogeneity.

Such an additional flux is a linear function of the applied field:

$$\Delta\mathbf{J} = -\frac{V_1}{V} \mathbf{H}^D \cdot \mathbf{G}_0, \tag{11}$$

where $\mathbf{H}^D$ is the tensor of the contribution of inhomogeneity to the diffusion permeability.

The contribution tensor can be found by solving the Eshelby problem for diffusion. The latter has an analytical solution only for ellipsoidal inhomogeneity. In this case, the contribution tensor can be expressed in terms of the concentration tensor, which linearly relates the field inside the inhomogeneity with the applied field.

Thus, to find the contribution tensor, it is necessary to solve the problem of averaging fields and find the concentration tensor. The presence of imperfect contacts should be taken into account at both stages.

A brief description of both stages is provided below.

Field averaging

We determine the average fields in the case of perfect contacts at the matrix/inhomogeneity interface, which in the framework of this study corresponds to a material with equivalent inhomogeneity, as well as modeling of imperfect contacts by various approaches.

According to the Ostrogradsky–Gauss theorem,

$$< \nabla c >_V = \frac{1}{V}\int_\Sigma \mathbf{n}_\Sigma c(\mathbf{x})\, d\Sigma, \quad < \mathbf{J} >_V = \frac{1}{V}\int_\Sigma \mathbf{n}_\Sigma \cdot \mathbf{J}(\mathbf{x})\mathbf{x} d\Sigma, \tag{12}$$

where $\mathbf{n}_\Sigma$ is the external normal to the surface Σ of the representative volume V.

Expressions (12) can be conveniently rewritten taking into account the interphase boundaries; in this case, the corresponding surface integrals should be added and subtracted. Then, with perfect contacts at the interface of the inhomogeneity of volume V_1, we obtain the known formulas:

$$< \nabla c >_V = \left(1 - \frac{V_1}{V}\right) < \nabla c >_{V_0} + \frac{V_1}{V} < \nabla c >_{V_1}, \quad < \mathbf{J} >_V = \left(1 - \frac{V_1}{V}\right) < \mathbf{J} >_{V_0} + \frac{V_1}{V} < \mathbf{J} >_{V_1}, \tag{13}$$

where $< ... >_{V_0} = \frac{1}{V_0}\int_{V_0} (...) \mathrm{d}V_0$, $< ... >_{V_1} = \frac{1}{V_1}\int_{V_1} (...) \mathrm{d}V_1$.

The case of a material with inhomogeneity with a coating of finite thickness characterized by finite properties is a particular case of a three-phase material with perfect contacts at its inner boundaries.

In this case, the average fields follow the expressions

$$\begin{gathered} < \nabla c >_V = \left(1 - \frac{V_b}{V}\right) < \nabla c >_{V_0} + \frac{V_a}{V} < \nabla c >_{V_a} + \frac{V_s}{V} < \nabla c >_{V_s}, \\ < \mathbf{J} >_V = \left(1 - \frac{V_b}{V}\right) < \mathbf{J} >_{V_0} + \frac{V_a}{V} < \mathbf{J} >_{V_a} + \frac{V_s}{V} < \mathbf{J} >_{V_s}, \end{gathered} \tag{14}$$

where $< ... >_{V_a} = \frac{1}{V_a}\int_{V_a} (...) \mathrm{d}V_a$ and $< ... >_{V_s} = \frac{1}{V_s}\int_{V_s} (...) \mathrm{d}V_s$.

In the presence of a concentration jump at the interface, the average concentration gradient should be determined as follows [6]:

$$< \nabla c >_V = \left(1 - \frac{V_1}{V}\right) < \nabla c >_{V_0} + \frac{V_1}{V} < \nabla c >_{V_1} + \frac{1}{V}\int_\Gamma \mathbf{n}_\Gamma [c]\, d\Gamma, \tag{15}$$

whereas the average flux is calculated using Eq. (13).

If the concentration jump is set based on the segregation parameter under condition (5), then it is convenient to rewrite Eq. (15) in the following form:

$$< \nabla c >_V = \left(1 - \frac{V_1}{V}\right) < \nabla c >_{V_0} + s_c \frac{V_1}{V} < \nabla c >_{V_1}. \tag{16}$$

The presence of a jump in the normal flux component leads to the need to use the following formula for the average flux [6]:

$$< \mathbf{J} >_V = \left(1 - \frac{V_1}{V}\right) < \mathbf{J} >_{V_0} + \frac{V_1}{V} < \mathbf{J} >_{V_1} + \frac{1}{V}\int_\Gamma [J_n]\mathbf{x} d\Gamma, \tag{17}$$

in this case, the average concentration gradient is determined by Eq. (13).

In the particular case, when the jump of the normal component of the flux is given in accordance with condition (6), the average flux is determined by the expression

$$< \mathbf{J} >_V = \left(1 - \frac{V_1}{V}\right) < \mathbf{J} >_{V_0} + s_f \frac{V_1}{V} < \mathbf{J} >_{V_1} . \tag{18}$$

Expressing $< \nabla c >_{V_0}$ in terms of $\mathbf{G}_0$, we obtain the following representations for the average flux:
for the material with equivalent inhomogeneity,

$$< \mathbf{J} >_V = -\mathbf{D}_0 \cdot \mathbf{G}_0 - \frac{V_1}{V}\left(\mathbf{D}^* - \mathbf{D}_0\right) \cdot < \nabla c >_{V_1}, \tag{19}$$

for the material with inhomogeneity and coating of finite thickness (characterized by finite properties),

$$< \mathbf{J} >= -\mathbf{D}_0 \cdot \mathbf{G}_0 - \frac{V_a}{V}\left(\mathbf{D}_1 - \mathbf{D}_0\right) \cdot < \nabla c >_{V_a} - \frac{V_s}{V}\left(\mathbf{D}_s - \mathbf{D}_0\right) \cdot < \nabla c >_{V_s} . \tag{20}$$

The following representations hold true for a material with a inhomogeneity with a field jump determined by the segregation parameter occurring at the interface:
if there is a jump in concentration,

$$< \mathbf{J} >_V = -\mathbf{D}_0 \cdot \mathbf{G}_0 - \frac{V_1}{V}\left(\mathbf{D}_1 - s_c \mathbf{D}_0\right) \cdot < \nabla c >_{V_1}, \tag{21}$$

if there is a jump in the normal flux component,

$$< \mathbf{J} >_V = -\mathbf{D}_0 \cdot \mathbf{G}_0 - \frac{V_1}{V}\left(s_f \mathbf{D}_1 - \mathbf{D}_0\right) \cdot < \nabla c >_{V_1} . \tag{22}$$

Representation of the contribution tensors in terms of the concentration tensors

The average concentration gradients included in expressions (19)–(22) can be expressed for the case of ellipsoidal inhomogeneity in terms of the applied field $\mathbf{G}_0$, for which the concentration tensors Λ_*, Λ_a, Λ_s, Λ_c, Λ_f are introduced, satisfying the equalities
$< \nabla c >_{V_1} = \mathbf{\Lambda}_* \cdot \mathbf{G}_0$ (for equivalent inhomogeneity),

$< \nabla c >_{V_a} = \mathbf{\Lambda}_a \cdot \mathbf{G}_0$ and $< \nabla c >_{V_s} = \mathbf{\Lambda}_s \cdot \mathbf{G}_0$ (for inhomogeneity with coating),

$< \nabla c >_{V_1} = \mathbf{\Lambda}_c \cdot \mathbf{G}_0$ (for inhomogeneity with a concentration jump, determined in terms of the segregation parameter, at the interface),

$< \nabla c >_{V_1} = \mathbf{\Lambda}_f \cdot \mathbf{G}_0$ (for inhomogeneity with a jump in the normal component of the flux, determined in terms of the segregation parameter, at the interface).

Expressions for these concentration tensors were obtained in [8, 10–14]. Taking into account these expressions and Eqs. (10), (11), we limit ourselves here to giving the final expressions for the contribution tensors of inhomogeneities:

$$\mathbf{H}^D = D_0 \sum_{i=1}^{3} \frac{D_{ii}^* - D_0}{D_{ii}^* A_i + D_0\left(1 - A_i\right)} \mathbf{e}_i \mathbf{e}_i \tag{23}$$

(for equivalent inhomogeneity with perfect contacts [13]);

$$\mathbf{H}^D = D_0 \sum_{i=1}^{3} \frac{D_1 - D_0 - D_0 D_1 \beta \dfrac{S_a}{V_a} A_i}{A_i D_1 + \left(1 - A_i\right) D_0 + \left(1 - A_i\right) D_0 D_1 \beta \dfrac{S_a}{V_a}\left(A_i - \dfrac{F_i}{H}\right)} \mathbf{e}_i \mathbf{e}_i \tag{24}$$

(for inhomogeneity with insulating coating);

$$\mathbf{H}^D = D_0 \sum_{i=1}^{3} \frac{D_1 - D_0 + \lambda \frac{S_a}{V_a}\left(1 - A_i\right)}{A_i D_1 + \left(1 - A_i\right) D_0 + A_i \lambda \frac{S_a}{V_a}\left(1 - A_i + \frac{F_i}{H}\right)} \mathbf{e}_i \mathbf{e}_i \tag{25}$$

(for inhomogeneity with conductive coating)

$$\mathbf{H}^D = D_0 \sum_{i=1}^{3} \frac{D_1 - s_c D_0}{A_i D_1 + s_c D_0 \left(1 - A_i\right)} \mathbf{e}_i \mathbf{e}_i \tag{26}$$

(in the presence of a concentration jump determined in terms of the segregation parameter s_c [8]),

$$\mathbf{H}^D = D_0 \sum_{i=1}^{3} \frac{s_f D_1 - D_0}{A_i s_f D_1 + D_0 \left(1 - A_i\right)} \mathbf{e}_i \mathbf{e}_i \tag{27}$$

(in the presence of a jump in the normal component of the flux, determined in terms of the segregation parameter s_f).

In the case of spheroidal inhomogeneity, for $a_1 = a_2 = a$, $\gamma = a_3/a$, the following equalities hold true:

$$A_1 = A_2 = f_0(\gamma),\ A_3 = 1 - 2 f_0(\gamma),$$

$$F_1 = F_2 = \frac{1}{a^2} \frac{1}{2} \left(\frac{1}{\gamma^2} f_0(\gamma) - \left(1 - 2 f_0(\gamma)\right) \right),$$

$$F_3 = -\frac{1}{a^2} \left(\frac{1}{\gamma^2} f_0(\gamma) - \left(1 - 2 f_0(\gamma)\right) \right),$$

where

$$f_0(\gamma) = \frac{1 - g(\gamma)}{2\left(1 - \gamma^{-2}\right)},\ g = g(\gamma) = \begin{cases} \dfrac{1}{\gamma\sqrt{1-\gamma^2}} \arctan \dfrac{\sqrt{1-\gamma^2}}{\gamma}, \gamma \le 1 \\ \dfrac{1}{2\gamma\sqrt{\gamma^2-1}} \ln \left(\dfrac{\gamma + \sqrt{\gamma^2-1}}{\gamma - \sqrt{\gamma^2-1}} \right), \gamma \ge 1. \end{cases}$$

In the case of spherical inhomogeneity, $F_1 = F_2 = F_3 = 0$, $f_0(\gamma) = 1/3$.

Note that according to the conclusions presented in monograph [9], where only spherical inhomogeneities were considered, the presence of insulating coating leads to a concentration jump at the matrix/inhomogeneity interface, and the presence of a conductive layer leads to a jump in the normal flux component, determined by solving the problem for composite inhomogeneity at the passage to the limit. This corresponds to the physical understanding of the phenomenon modeled, as noted above.

For the correct implementation of the procedure for comparing two approaches to modeling imperfect contact (by setting the field jump with the appropriate segregation parameter and by considering the inhomogeneity with the appropriate type of surface effect), we will determine what diffusion permeability $\mathbf{D}^*$ that an equivalent inhomogeneity whose contribution to the macroscopic property coincides with the contribution of inhomogeneity with imperfect contact modeled within the framework of different approaches should possess.

Equivalent inhomogeneity

Let us start by considering imperfect contact when an impurity is deposited as sediment at the matrix/inhomogeneity interface. When such a contact is modeled by setting a concentration jump in terms of the segregation parameter, it follows from the equality of the contribution tensor defined by expression (26) and the contribution tensor of equivalent inhomogeneity described by Eq. (23) that

$$\mathbf{D}^* = D^*\mathbf{I} = D_1/s_c\,\mathbf{I}, \tag{28}$$

that is, a material of equivalent inhomogeneity is isotropic.

Evidently, the components of the tensor $\mathbf{D}^*$ depend only on the segregation parameter and the diffusion permeability of the inhomogeneity, but do not depend on its shape. An increase in the segregation parameter leads to a decrease in the diffusion permeability of the equivalent inhomogeneity. In the absence of impurity sedimentation (at $s_c = 1$) $D^* = D_1$. Depending on whether the impurity is deposited at the interphase boundary from the outside or inside, the segregation parameter takes the values $s_c > 1$ or $s_c < 1$, respectively.

In the first case, $D^* < D_1$, which reflects the physics of the process, since the impurity penetrates the inhomogeneity to a lesser extent and, in order to achieve the same effect when considering an equivalent inhomogeneity, it is necessary to reduce its permeability.

In the second case, $D_* > D_1$, which is also physically justified, since the equivalent inhomogeneity should be more permeable to the impurity due to its accumulation inside the «real» inhomogeneity with imperfect contact.

At $s_c \to \infty$ $D^* \to 0$; this is due to the fact that the entire impurity accumulates outside the inhomogeneity and it is impermeable to the diffusant.

At $s_c = D_1/D_0$, we have $D^* = D_0$, i.e., a jump in concentration

$$[c] = \left(D_1 - D_0\right)/D_0\, c\left(\mathbf{x}\right)\Big|_{\mathbf{x}\to\Gamma-}$$

allows to ignore the presence of inhomogeneity when finding effective properties.

In the case of using the second approach to modeling imperfect contact, it follows from the equality of the inhomogeneity contribution tensor with equivalent surface resistance (see Eq. (24)) and the equivalent inhomogeneity contribution tensor (see expression (23)) that

$$\mathbf{D}_* = \sum_{i=1}^{3} D_1 \frac{1 - R\dfrac{S_a a_1}{V_a}\dfrac{D_0}{D_1}\dfrac{F_i}{H}\left(1 - A_i\right)}{1 + R\dfrac{S_a a_1}{V_a}\left(A_i - \dfrac{F_i}{H}\left(1 - A_i\right)\right)}\mathbf{e}_i\mathbf{e}_i; \tag{29}$$

Here, for convenience, a dimensionless parameter of equivalent surface resistance $R = D_1\beta/a_1$ is introduced.

The diffusion permeability tensor of equivalent inhomogeneity, expressed by Eq. (29), is generally orthotropic, and its symmetry group is determined by the shape of the inhomogeneity. In the absence of the surface effect (at $R = 0$), the tensor is isotropic and $\mathbf{D}^* = \mathbf{D}_1$ ($D_{11}^* = D_{22}^* = D_{33}^* = D_1$). In general, the diffusion coefficients D_{ii}^* can take values both greater and smaller than D_1. We should note that such features as redirection of the diffusion flux due to negative values of the components of the tensor $\mathbf{D}^*$, as well as the infinite permeability of equivalent inhomogeneity, can formally appear at certain values of structural characteristics (the ratio of the diffusion coefficients of the impurity in the matrix and in the inhomogeneity, the parameters of the inhomogeneity's shape, the value of equivalent surface resistance). Such cases require separate qualitative and quantitative studies, which is beyond the scope of this paper.

Expression (29) is significantly simplified in the case of spherical inhomogeneity: then the equivalent inhomogeneity is characterized by an isotropic tensor

$$\mathbf{D}^* = D_1/\left(1 + R\right)\mathbf{I}.$$

It follows from comparing this expression with expression (28) that two approaches to modeling imperfect contacts at the boundary of spherical inhomogeneities coincide when

$$s_c = 1 + R. \tag{30}$$

Let us turn to the consideration of imperfect contact, when additional diffusion paths are present in the material at the interphase boundary. When such a contact is modeled by setting the jump of the normal flux component in terms of the segregation parameter, it follows from the equality of the contribution tensors defined by expressions (27) and (23) that

$$\mathbf{D}^* = D^*\mathbf{I} = D_1 s_f \mathbf{I}. \tag{31}$$

The diffusion tensor $\mathbf{D}^*$, defined by Eq. (31), depends only on the segregation parameter and on the diffusion permeability of the inhomogeneity and does not depend on its shape.

An increase in the segregation parameter leads to an increase in the diffusion permeability of the equivalent inhomogeneity. In the absence of surface defects (at $s_f = 1$), $D^* = D_1$. At $s_f \to \infty$, the equivalent inhomogeneity is characterized by infinite permeability, regardless of the properties of the inhomogeneity (in this case, the entire impurity will instantly diffuse over the surface). In the case when $s_f = D_0/D_1$, the equality $D^* = D_0$ is satisfied.

When imperfect contact is modeled using the second approach, the equality of the contribution tensors defined by expressions (25) and (23) gives the following result:

$$\mathbf{D}_* = \sum_{i=1}^{3} D_1 \frac{1 + K\dfrac{S_a a}{V_a}\left(1 - A_i + \dfrac{F_i}{H}\right)}{1 + K\dfrac{S_a a}{V_a}\dfrac{D_1}{D_0}\dfrac{F_i}{H}A_i}\mathbf{e}_i\mathbf{e}_i, \tag{32}$$

where a dimensionless parameter of equivalent surface permeability $K = \lambda/(D_1 a_1)$ is introduced.

The diffusion permeability tensor of equivalent inhomogeneity, defined by expression (32), is generally orthotropic. In the absence of the surface effect (at $K = 0$), $\mathbf{D}^* = \mathbf{D}_1$. In the presence of the surface effect, the diffusion coefficients D_{ii}^* can take values both greater and smaller than D_1. In a certain range of values of structural characteristics, as in the case of an insulating coating, components D_{ii}^* can take values less than zero, which means from a physical standpoint that flux is redirected, as well as goes to infinity. Both cases require separate research beyond the scope of this study.

Expression (32) in the case of spherical inhomogeneities has the form

$$\mathbf{D}_* = D_1\left(1 + 2K\right)\mathbf{I},$$

it follows from here, taking into account Eq. (31), that two approaches to modeling imperfect contacts are equivalent when

$$s_f = 1 + 2K. \tag{33}$$

To summarize, the following qualitative differences can be observed between the two approaches to modeling imperfect contacts.

1. Taking into account imperfect contact by setting the field jump in terms of a constant segregation parameter, the symmetry group of the diffusion permeability tensor of equivalent inhomogeneity coincides with that for the initial inhomogeneity (in particular, it was shown above that the isotropy of the tensor $\mathbf{D}_1$ implies the isotropy of the tensor $\mathbf{D}^*$; we presented a more detailed study of the general anisotropic case in [8]). As a result, the components of the diffusion permeability tensor of equivalent inhomogeneity depend only on the physical properties of the inhomogeneity and the segregation parameter. In the case of modeling imperfect contact by considering inhomogeneity with a surface effect, the components of the tensor $\mathbf{D}^*$ depend both on the properties of the coating and the material of the inhomogeneity and on its shape.

Two approaches to modeling imperfect contacts produce the same results only in the case of a material with spherical inhomogeneities, provided that either equality (30) or (33) is satisfied, depending on the type of imperfect contact.

2. When taking into account imperfect contact by setting the field jump in terms of a constant segregation parameter, the components of the diffusion permeability tensor of equivalent inhomogeneity linearly depend either on the quantity $(s_c)^{-1}$, or on the parameter s_f. In the case of modeling imperfect contact by considering an inhomogeneity with a surface effect, the components of the tensor $\mathbf{D}^*$ depend non-linearly on the characteristics of the coating R or K (or their inverse quantities). At the same time, these dependencies, firstly, are different for different directions, and secondly, can take negative values at certain values of the characteristics of the structure, which, apparently, means redirection of the diffusion flux, as well as going to infinity. These cases need to be further investigated for compliance with the physical meaning of the modeled phenomenon.

Simulation results for imperfect contacts

Let us conduct quantitative analysis of the effect of the method of accounting for imperfect contact at the interface on the diffusion permeability of equivalent inhomogeneity using the example of a polycrystal.

A polycrystal is considered a two-phase material consisting of a matrix that models grain boundaries and elongated spheroidal inhomogeneities that model grains of lower diffusion permeability [7, 15]. For certainty, we take the values $D_1/D_0 = 0{,}2$, $\gamma = a_3/a = 100$ ($a_1 = a_2 = a$). In polycrystals, imperfect contacts can occur for various reasons, which should be modeled in different ways. Let us briefly describe them.

1. The phenomenon of segregation, which is characteristic of diffusion and which is understood as the sedimentation of impurities along grain boundaries from the outside, can be modeled either by setting a concentration jump using the segregation parameter s_c (I), or by considering an insulating coating with equivalent resistance R (II). Let us assume that $s_c = 1 + R$, which, on the one hand, is true for the case of a material with spherical inhomogeneities, on the other hand, satisfies the condition $s_c = 1$ at $R = 0$ in the case of perfect contacts in a material with inhomogeneities of arbitrary shape.

2. Due to cracking along grain boundaries, additional accelerated diffusion paths can be formed; they can be taken into account either by setting the jump of the normal flux component in terms of the segregation parameter s_f(III), or by considering conductive coating characterized by equivalent conductivity K (IV). For the same reasons as when choosing the dependence $s_c(R)$, we assume that $s_f = 1 + 2K$.

The dependences of the diffusion permeability of equivalent inhomogeneity in the presence of segregation, i.e., in the case of imperfect contact modeled by methods I and II, are shown in Fig. 1,*a*. An increase in the parameter R leads to a decrease in the components of the tensor $\mathbf{D}^*$. This, in turn, should subsequently (with further application of homogenization methods not considered in this study) lead to a decrease in the effective permeability of the material.

Note that the parameter R can formally take values from zero to infinity. To carry out quantitative analysis, however, we limited ourselves to considering a smaller range in which the flux does not change direction to the opposite, which would be the case with negative values of the coefficients D^*_{ii} and which, as noted above, requires additional analysis.

It is also worth noting that when using approach II, there is a difference in the behavior of the decreasing curves of the diffusion coefficients D^*_{33} along the symmetry axis of the inhomogeneity and the coefficients $D^*_{11} = D^*_{22}$ in the isotropy plane.

With the selected set of structure parameters, the coefficients $D^*_{11} = D^*_{22}$ change in the same way as the components of the isotropic tensor $\mathbf{D}^*$ introduced using approach I.

Fig. 1,*b* shows the dependences of the diffusion permeability of equivalent inhomogeneity in the presence of cracking, i.e., in the case of imperfect contact modeled by methods III and IV.

An increase in the parameter K leads to an increase in the components of the tensor $\mathbf{D}^*$, which vary in different ways, depending on the method of modeling imperfect contact, as well as on the direction in the case of approach IV. In the future, this type of change in the permeability of equivalent inhomogeneity should lead to an increase in the effective permeability of the material. Parameter K, like parameter R, can formally take values from zero to infinity, while at a certain value of K the component D^*_{33} will go to infinity, which, as discussed above, requires additional analysis beyond the scope of this paper.

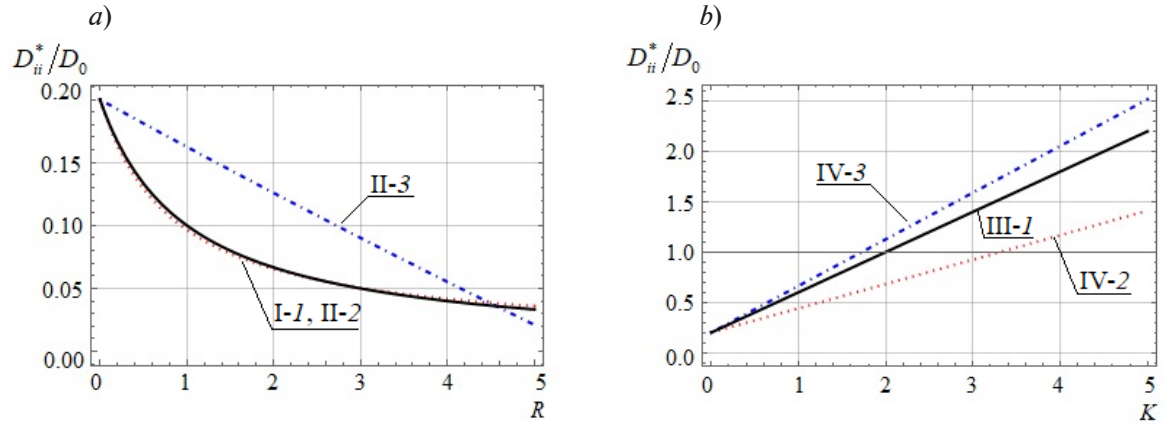

Fig. 1. Dependences of diffusion coefficients of the impurity in equivalent inhomogeneity on parameters $R(a)$ and $K(b)$ for the case of imperfect contact, simulated by methods I, II (a) and III, IV (b) (see explanations in the text).
The following diffusion coefficients are shown: $D_{11}^* = D_{22}^* = D_{33}^*$ (solid lines I-*1* and III-*1*) using approaches I and III; $D_{11}^* = D_{22}^*$ (dashed lines II-*2* and IV-*2*) and D_{33}^* (lines II-*3* and IV-*3*) using approaches II and IV

It is important that the components D_{ii}^* take values both smaller and larger than D_0, depending on the value of K (D_{ii}^*/D_0 can be either smaller or larger than unity).

Thus, the method of accounting for imperfect contact at the interface of the matrix and non-spherical inhomogeneity has a direct impact on the effective properties of the material. To choose the optimal approach, it is necessary to compare the results of numerical simulation with experimental data. The problem of such a comparison, in turn, involves difficulties in identifying the characteristics of the structure and requires separate study.

Conclusion

The paper proposes a generalization of the approaches available in the literature to modeling imperfect contacts at the interphase boundary of a material that is inhomogeneous at the microlevel in determining its effective properties of various nature.

It is taken into account that such contacts can occur in the material for different reasons: due to the particular internal structure and in connection with the specifics of the described process, which affects the physical interpretation of the model, but does not affect the mathematical framework used. A specific example of a diffusion problem was considered. The general case of a material with ellipsoidal inhomogeneities is considered and two approaches to modeling imperfect contacts are compared: by introducing a field jump (concentration or normal flux component) in terms of a constant segregation parameter and by considering inhomogeneity with a surface effect (respectively, with the presence of insulating or conductive coating).

We confirmed that the two approaches are equivalent only in the case of a material with spherical inhomogeneities, while in other cases these methods give qualitatively and quantitatively different results.

REFERENCES

1. **Kaur I., Mishin Yu., Gust W.,** Fundamentals of grain and interphase boundary diffusion, Third Edition, John Wiley & Sons, London, 1995.

2. **Zhang Y., Liu L.,** On diffusion in heterogeneous media, Am. J. Sci. 312 (9) (2012) 1028–1047.

3. **Dumont S., Serpilli M., Rizzoni R., Lebon F. C.,** Numerical validation of multiphysic imperfect interfaces models, Front. Mater. 7 (5 June) (2020) 158.

4. **Costa R., Nobrega J. M., Clain S., Machado G. J.,** Very high-order accurate polygonal mesh finite volume scheme for conjugate heat transfer problems with curved interfaces and imperfect contacts, Comp. Meth. Appl. Mech. Eng. 357 (1 Dec) (2019) 112560.

5. **Kalnin J. R., Kotomin E. A., Maier J.,** Calculations of the effective diffusion coefficient for inhomogeneous media, J. Phys. Chem. Solids. 63 (3) (2002) 449–456.

6. **Belova I. V., Murch G. E.,** Calculation of the effective conductivity and diffusivity in composite solid electrolytes, J. Phys. Chem. Solids. 66 (5) (2005) 722–728.

7. **Knyazeva A. G., Grabovetskaya G. P., Mishin I. P., Sevostianov I.,** On the micromechanical modelling of the effective diffusion coefficient of a polycrystalline material, Philos. Mag. 95 (19) (2015) 2046–2066.

8. **Frolova K. P., Vilchevskaya E. N.,** Effective diffusivity of transversely isotropic material with embedded pores, Mater. Phys. Mech. 47 (6) (2021) 937–950.

9. **Markov K. Z.,** Elementary micromechanics of heterogeneous media. Heterogeneous media: micromechanics modeling methods and simulations, Birkhäuser, Boston, MA, 2000.

10. **Endres A. L., Knight R. J.,** A model for incorporating surface phenomena into the dielectric response of a heterogeneous medium, J. Colloid Interface Sci. 157 (2) (1993) 418–425.

11. **Levin V., Markov M.,** Effective thermal conductivity of micro-inhomogeneous media containing imperfectly bonded ellipsoidal inclusions, Int. J. Eng. Sci. 109 (Dec) (2016) 202–215.

12. **Markov M., Levin V., Markova I.,** Determination of effective electromagnetic parameters of concentrated suspensions of ellipsoidal particles using Generalized Differential Effective Medium approximation, Phys. A: Stat. 492 (15 Febr) (2018) 113–122.

13. **Kachanov M., Sevostianov I.,** Micromechanics of materials, with applications. Ser. “Solid Mechanics and its Applications” (Vol. 249), Springer, Berlin, 2018.

14. **Fricke H.** A mathematical treatment of the electric conductivity and capacity of disperse systems I. The electric conductivity of a suspension of homogeneous spheroids, Phys. Rev. 24 (5) (1924) 575–587.

15. **Pashkovsky D. M., Frolova K. P., Vilchevskaya E. N.,** Effective diffusion properties of a polycrystal, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 15 (3) (2022) 154–168 (in Russian).

СПИСОК ЛИТЕРАТУРЫ

1. **Kaur I., Mishin Yu., Gust W.** Fundamentals of grain and interphase boundary diffusion. Third Edition. London; John Wiley & Sons, 1995. 536 p.

2. **Zhang Y., Liu L.** On diffusion in heterogeneous media // American Journal of Science. 2012. Vol. 312. No. 9. Pp. 1028–1047.

3. **Dumont S., Serpilli M., Rizzoni R., Lebon F. C.** Numerical validation of multiphysic imperfect interfaces models // Frontiers in Materials. 2020. Vol. 7. 05 June. P. 158.

4. **Costa R., Nobrega J. M., Clain S., Machado G. J.** Very high-order accurate polygonal mesh finite volume scheme for conjugate heat transfer problems with curved interfaces and imperfect contacts // Computer Methods in Applied Mechanics and Engineering. 2019. Vol. 357. 1 December. P. 112560.

5. **Kalnin J. R., Kotomin E. A., Maier J.** Calculations of the effective diffusion coefficient for inhomogeneous media // Journal of Physics and Chemistry of Solids. 2002. Vol. 63. No. 3. Pp. 449–456.

6. **Belova I. V., Murch G. E.** Calculation of the effective conductivity and diffusivity in composite solid electrolytes // Journal of Physics and Chemistry of Solids. 2005. Vol. 66. No. 5. Pp. 722–728.

7. **Knyazeva A. G., Grabovetskaya G. P., Mishin I. P., Sevostianov I.** On the micromechanical modelling of the effective diffusion coefficient of a polycrystalline material // Philosophical Magazin. 2015. Vol. 95. No. 19. Pp. 2046–2066.

8. **Frolova K. P., Vilchevskaya E. N.** Effective diffusivity of transversely isotropic material with embedded pores // Materials Physics and Mechanics. 2021. Vol. 47. No. 6. Pp. 937–950.

9. **Markov K. Z**. Elementary micromechanics of heterogeneous media. Heterogeneous media: micromechanics modeling methods and simulations. Boston, USA: Birkhäuser, MA, 2000. 162 p.

10. **Endres A. L., Knight R. J.** A model for incorporating surface phenomena into the dielectric response of a heterogeneous medium // Journal of Colloid and Interface Science. 1993. Vol. 157. No. 2. Pp. 418–425.

11. **Levin V., Markov M.** Effective thermal conductivity of micro-inhomogeneous media containing imperfectly bonded ellipsoidal inclusions // International Journal of Engineering Science. 2016. Vol. 109. December. Pp. 202–215.

12. **Markov M., Levin V., Markova I.** Determination of effective electromagnetic parameters of concentrated suspensions of ellipsoidal particles using Generalized Differential Effective Medium approximation // Physica A: Statistical Mechanics and its Applications. 2018. Vol. 492. 15 February. Pp. 113–122.

13. **Kachanov M., Sevostianov I.** Micromechanics of materials, with applications. Series "Solid Mechanics and its Applications". Vol. 249. Berlin, Germany: Springer, 2018. 712 p.

14. **Fricke H.** A mathematical treatment of the electric conductivity and capacity of disperse systems I. The electric conductivity of a suspension of homogeneous spheroids // Physical Review. 1924. Vol. 24. No. 5. Pp. 575–587.

15. **Пашковский Д. М., Фролова К. П., Вильчевская Е. Н.** Эффективные диффузионные свойства поликристалла // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2022. Т. 15. № 3. С. 154–168.

THE AUTHORS

FROLOVA Ksenia P.
Institute for Problems in Mechanical Engineering, RAS;
61 Bolshoi Ave., V. Isl., St. Petersburg, 199178, Russia
kspfrolova@gmail.com
ORCID: 0000-0003-0376-4463

VILCHEVSKAYA Elena N.
Institute for Problems in Mechanical Engineering, RAS;
61 Bolshoi Ave., V. Isl., St. Petersburg, 199178, Russia
vilchevskaya_en@spbstu.ru
ORCID: 0000-0002-5173-3218

СВЕДЕНИЯ ОБ АВТОРАХ

ФРОЛОВА Ксения Петровна – *кандидат физико-математических наук, старший научный сотрудник Института проблем машиноведения РАН.*
199178, Россия, г. Санкт-Петербург, Большой проспект В. О., 61.
fkp@ipme.ru
ORCID: 0000-0003-0376-4463

ВИЛЬЧЕВСКАЯ Елена Никитична – *доктор физико-математических наук, ведущий научный сотрудник Института проблем машиноведения РАН.*
199178, Россия, г. Санкт-Петербург, Большой проспект В. О., 61.
vilchevskaya_en@spbstu.ru
ORCID: 0000-0002-5173-3218

Received 22.09.2023. Approved after reviewing 23.10.2023. Accepted 23.10.2023.
Статья поступила в редакцию 22.09.2023. Одобрена после рецензирования 23.10.2023. Принята 23.10.2023.

St. Petersburg Polytechnic University Journal. Physics and Mathematics. 2023. Vol. 16. No. 4
Научно-технические ведомости СПбГПУ. Физико-математические науки. 16 (4) 2023

BIOPHYSICS AND MEDICAL PHYSICS

Original article
DOI: https://doi.org/10.18721/JPM.16413

DETERMINING THE SIZE DISTRIBUTION FUNCTION OF IRREGULARLY SHAPED PARTICLES FOR HUMAN BLOOD CELLS AND FINDING THEIR ERYTHROCYTE PARAMETERS (*IN VIVO* CASE)

A. P. Golovitskii[1], *V. G. Kontsevaya*[2],[1]✉, *K. G. Kulikov*[1], *K. T. Koshlan*[3]

[1] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;

[2] Pskov State University, Pskov, Russia;

[3] St. Petersburg State University, St. Petersburg, Russia

✉ nkoncevoi@mail.ru

Abstract. This article continues the authors' research aimed at constructing and developing a mathematical model used both to determine the size distribution function of human blood cells in vivo, and to find blood parameters used in medical practice. At this stage of the work, the nonsphericity of blood particles was taken into account and the convergence of processes describing multiple scattering of laser radiation by blood was optimized through the use of the method of extended boundary conditions, which made it possible to increase the possibilities of using the T-matrix method. The mathematical model for the analysis of biological processes has received material embodiment in a new software package. Regularization parameters are determined automatically based on specified kernel errors and "measured" data using different criteria. It is shown that, using the developed model, it is possible to theoretically predict the number of erythrocytes of abnormal size in a biomaterial based on measuring the width of the found erythrocyte size distribution.

Keywords: laser technologies, Tikhonov regularization, EBCM, erythrocyte index, unspherulated particle

Citation: Golovitskii A. P., Kontsevaya V. G., Kulikov K. G., Koshlan K. T., Determining the size distribution function of irregularly shaped particles for human blood cells and finding their erythrocyte parameters (in vivo case). St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 160–180. DOI: https://doi.org/10.18721/JPM.16413

Научная статья
УДК 517.95+577.3+535.8+519.6
DOI: https://doi.org/10.18721/JPM.16413

НАХОЖДЕНИЕ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ ЧАСТИЦ НЕРЕГУЛЯРНОЙ ФОРМЫ ПО РАЗМЕРАМ ДЛЯ КЛЕТОК ЧЕЛОВЕЧЕСКОЙ КРОВИ И ОПРЕДЕЛЕНИЕ ЕЕ ЭРИТРОЦИТАРНЫХ ПОКАЗАТЕЛЕЙ (*IN VIVO*)

А. П. Головицкий[1], *В. Г. Концевая*[2],[1]✉, *К. Г. Куликов*[1], *Т. В. Кошлан*[3]

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия;

[2] Псковский государственный университет, г. Псков, Россия;

[3] Санкт-Петербургский государственный университет, Санкт-Петербург, Россия

✉ nkoncevoi@mail.ru

Аннотация. Данная статья продолжает исследования авторов, направленные на построение и развитие математической модели, используемой как для определения функции распределения клеток крови человека по размерам in vivo, так и для нахождения показателей крови, используемых в медицинской практике. На данном этапе работы была учтена несферичность частиц крови и оптимизирована сходимость процессов, описывающих многократное рассеяние лазерного излучения кровью за счет использования метода расширенных граничных условий, что позволило увеличить возможности применения Т-матричного метода. Математическая модель анализа биологических процессов получила материальное воплощение в новом программном комплексе. Параметры регуляризации определяются автоматически по заданным погрешностям ядра и «измеренным» данным с использованием разных критериев. Показана возможность, используя разработанную модель, теоретически предсказывать количество аномальных по размеру эритроцитов в биоматериале на основе измерения ширины найденного распределения эритроцитов по размерам.

Ключевые слова: лазерные технологии, многократное рассеяние, метод Т-матриц, метод регуляризации Тихонова, эритроцитарный индекс, несферулированная частица

Ссылка для цитирования: Головицкий А. П., Концевая В. Г., Куликов К. Г., Кошлан Т. В. Нахождение функции распределения частиц нерегулярной формы по размерам для клеток человеческой крови и нахождение ее эритроцитарных показателей (in vivo) // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. 160–180. DOI: https://doi.org/10.18721/JPM.16413

Introduction

Hemorheological and microcirculatory dysfunctions of the human body accompany, as a rule, most diseases and complications. Since 99% of the total volume of blood corpuscles are red blood cells, the study of the functional characteristics of these cells is crucial. The characteristic sizes of erythrocytes, their refractive indices and mechanical properties, as well as the dynamics of changes in such indicators of the state of the body should undoubtedly be investigated in cases of various pathological conditions; such studies are always of major importance.

The human erythrocyte is an elastic cell that has a rather complex discoid shape in its normal mature state. Moreover, under various external influences, with conditions of various kinds, discocyte (mature normal form of erythrocyte) can undergo a transition to other forms, for example platycide, acanocide, etc. [1].

A range of studies (see, for example, [2-5]) have explored the possibilities of theoretical investigation of the optical characteristics of dielectric bodies of different shapes and structures.

The classical problem of scattering light radiation by irregularly shaped particles is solved by direct numerical methods, which make it possible to reduce this problem to solving a system of algebraic equations or to the method of variable separation. In the first case, either an integral equation is constructed, or the expansion of fields is introduced with respect to vector spherical harmonics, i.e., solutions of the Helmholtz wave equation with their subsequent "joining" on the surface of the scatterer.

In our opinion, it is worth listing some successful approximations that allow to obtain rather satisfactory results.

First, it is the Rayleigh–Gans–Debye method [6]. Secondly, it is acceptable to use methods of geometric optics, especially in cases where particles can be considered large enough relative to the wavelength of the incident radiation [7]. Thirdly, these are methods of anomalous diffraction [8, 9]. Further, iterative methods deserve special mention [10]. We should also consider the Wentzel–Kramers–Brillouin method as well as the eikonal approximation [11, 12] as the most well-known method of anomalous diffraction. The latter is, in fact, the implementation of the approximation of short waves or high energies. Another notable approach is the perturbation method [13], which is based on the decomposition of an unknown solution to the scattering problem with respect to a small parameter in the vicinity of the exact solution. When applied to nonspherical particles, this means that the solution is sought as small deviations from the solution, which are caused by small deviations of the shape from the ideal spherical one.

In our opinion, the most convenient and reliable approach to solving the problem of light scattering by bodies of arbitrary shape is the method of integral equations, called the method of extended boundary conditions [14, 15], since it provides an accurate solution to the scattering problem (unlike other methods) by a particle of arbitrary shape; although this solution has the form of infinite series, it is acceptable. The maximum number of expansion terms required to achieve acceptable accuracy depends on the size, shape and refractive index of the scatterer.

In this paper, we investigate some aspects of the problem of light scattering by dispersed elements (blood cells in our case), which are irregular in shape and located in a medium of non-trivial structure (here it is the skin, a multilayered structure).

The problem of modeling scattering by dispersed structures with irregular configuration is set.

The study involves considering light scattering by a dispersed system (blood corpuscles), where the shape of the inhomogeneities is irregular and their orientation is arbitrary. This takes into account the effects of multiple scattering of light incident on a layered medium (human skin).

Such consideration includes several stages.

At the first stage, the problem of light scattering in the system is solved.

At the second stage, the reflection coefficient of a plane wave from a layered surface with a wavy shape is studied (the case of reflection of a Gaussian beam is taken).

At the third and final stage, a search is carried out for the size distribution function of blood corpuscles (scatterers of irregular shape placed in a layered medium). It is important to take into account that the simulated system is assumed to be placed in a layered medium.

Light scattering by the jth individual particle of arbitrary shape (matrix formulation)

Let us start considering the problem with the assumption that only red blood cells are present in the simulated dispersed medium (blood). It is fairly appropriate and does not contradict the problem statement, since the proportion of other blood corpuscles is about 1% of the hematocrit.

In a number of studies, the erythrocyte is considered as a structurally homogeneous sphere [16, 17], which can be taken a first approximation. With a deeper analysis (microscopic level), it is more correct to consider the erythrocyte as a body of irregular shape.

To find the field scattered by an ensemble of particles with irregular shape, we use the T-matrix method. It is rapid compared with most other methods of light diffraction theory based on a rigorous solution of Maxwell′s equations.

A dispersed inhomogeneous medium is considered in a three-dimensional coordinate system, and a linearly polarized plane wave falls on an ensemble of inhomogeneities. It is assumed that the wavelength is smaller than the typical size of red blood cells, that the surface of the dispersed scatterer is regular everywhere, so a continuous normal can be determined for it; Green′s theorem also holds true.

We write a system of Maxwell's equations for the electromagnetic field in the vicinity of a particle with the conditional number j_0, distorted by the presence of other particles:

$$\nabla \times \mathbf{H} = -ik\varepsilon\mathbf{E},\ \nabla \times \mathbf{E} = ik\mu\mathbf{H},\ \nabla \cdot \mathbf{E} = 0,\ \nabla \cdot \mathbf{H} = 0,$$

where k is the wavenumber; ε, μ are the dielectric and magnetic permeabilities of the medium.

At the boundary between the particle with the conditional number j_0 and the medium surrounding it, we require for the following boundary conditions to be satisfied:

$$\mathbf{n} \times \mathbf{E}_i - \mathbf{n} \times \mathbf{E}_s = \mathbf{n} \times \mathbf{E}_I,\ \mathbf{n} \times \mathbf{H}_i - \mathbf{n} \times \mathbf{H}_s = \mathbf{n} \times \mathbf{H}_I, \tag{1}$$

where $\mathbf{E}_i$, $\mathbf{E}_s$, $\mathbf{E}_I$ are the internal, scattered and incident fields, respectively.

The total field can be represented as

$$\mathbf{E}(r') = \mathbf{E}_I(r') + \mathbf{E}_s(r').$$

Let us write corresponding integral equation of the following form [18]:

$$\mathbf{E}_I(r') + \nabla \times \int_S \mathbf{n} \times \mathbf{E}(r) G(r, r')\,\mathrm{ds} + \frac{i}{k\varepsilon} \nabla \times \nabla \times \int_S \mathbf{n} \times \mathbf{H}(\mathrm{r}) \times G(r, r')\,\mathrm{ds} = 0. \tag{2}$$

The Green function in Eq. (2) is defined as follows [18]:

$$G(r, r') = \frac{ik}{\pi} \sum_{n=1}^{\infty} \sum_{m=-n}^{n} (-1)^m E_{mn} [\mathbf{M}^3_{-mn}(kr, \theta, \varphi) \cdot \mathbf{M}^1_{mn}(kr', \theta', \varphi') + \mathbf{N}^3_{-mn}(kr, \theta, \varphi) \cdot \mathbf{N}^1_{mn}(kr', \theta', \varphi')] \tag{3}$$

(for the case $r > r'$),

$$G(r, r') = \frac{ik}{\pi} \sum_{n=1}^{\infty} \sum_{m=-n}^{n} (-1)^m E_{mn} [\mathbf{M}^1_{-mn}(kr, \theta, \varphi) \cdot \mathbf{M}^3_{mn}(kr', \theta', \varphi') + \mathbf{N}^1_{-mn}(kr, \theta, \varphi) \cdot \mathbf{N}^3_{mn}(kr', \theta', \varphi')] \tag{4}$$

(for the case $r' > r$),
where $\mathbf{M}_{mn}$, $\mathbf{N}_{mn}$, $\mathbf{M}_{-mn}$, $\mathbf{N}_{-mn}$ are vector spherical harmonics.

Note that the choice of vector spherical harmonics should be made based on the invariance property (in the sense of closure), namely, that upon rotation of the coordinate system, such harmonics $\mathbf{M}_{mn}$, $\mathbf{N}_{mn}$ should be transformed independently of each other.

The following vector spherical harmonics satisfy the required invariance properties [18]:

$$\mathbf{M}^J_{mn}(rk) = (-1)^m d_n z^J_n(kr) \mathbf{C}_{mn}(\theta) \exp(im\varphi), \tag{5}$$

$$\mathbf{N}^J_{mn}(rk) = (-1)^m d_n [\frac{n(n+1)}{kr} z^J_n(kr) \mathbf{P}_{mn}(\theta) + \frac{1}{kr} \mathrm{z}^J_n \mathbf{B}_{mn}(\theta)] \exp(im\varphi), \tag{6}$$

$$\mathbf{B}_{mn}(\theta) = \mathbf{i}_\theta \frac{d}{d\theta} d^n_{om}(\theta) + \mathbf{i}_\varphi \frac{im}{\sin(\theta)} d^n_{om}(\theta), \tag{7}$$

$$\mathbf{C}_{mn}(\theta) = \mathbf{i}_\theta \frac{im}{\sin(\theta)} d^n_{om}(\theta) - \mathbf{i}_\varphi \frac{d}{d\theta} d^n_{om}(\theta), \tag{8}$$

$$\mathbf{P}_{mn}(\theta) = \mathbf{i}_r d^n_{om}(\theta),\ d_n = \sqrt{\frac{(2n+1)}{4n(\mathrm{n}+1)}}. \tag{9}$$

Any of the four spherical harmonics of the following form can be selected as a z_n^j function

$$j_n(z)=\sqrt{\frac{\pi}{2z}}J_{n+\frac{1}{2}}(z),\ y_n(z)=\sqrt{\frac{\pi}{2z}}Y_{n+\frac{1}{2}}(z),\ h_z^{(1)}=j_n(z)+iy_n(z),\ h_z^{(2)}=j_n(z)-iy_n(z),$$

$$d_{om}^n(\theta)=\frac{(-1)^{n-m}}{2^n n!}[\frac{(n+m)!}{(n-m)!}]^{1/2}(1-\cos^2(\theta))^{-m/2}\frac{d^{n-m}}{d\cos(\theta)^{n-m}}[1-\cos^2(\theta)^n].$$

Let us write the decomposition of the incident wave $\mathbf{E}_I$ on the surface of the jth particle with respect to vector spherical harmonics:

$$\mathbf{E}_I(j)=-\sum_{n=1}^{\infty}\sum_{m=-n}^{n}iE_{mn}[p_{mn}^j\mathbf{N}_{mn}^1+q_{mn}^j\mathbf{M}_{mn}^1]. \tag{10}$$

Similarly, we can write the decomposition with respect to vector spherical harmonics for both the internal field of the jth particle $E_i(j)$ and the scattered field $E_s(j)$:

$$\mathbf{E}_i(j)=-\sum_{n=1}^{\infty}\sum_{m=-n}^{n}iE_{mn}[d_{mn}^j\mathbf{N}_{mn}^1+c_{mn}^j\mathbf{M}_{mn}^1], \tag{11}$$

$$\mathbf{E}_s(j)=\sum_{n=1}^{\infty}\sum_{m=-n}^{n}iE_{mn}[a_{mn}^j\mathbf{N}_{mn}^3+b_{mn}^j\mathbf{M}_{mn}^3]. \tag{12}$$

In accordance with the procedures described in monograph [18], we sequentially substitute expressions (10), (11), (12), taking into account Green's functions (3), (4) and boundary conditions of the form (1), into the integral equation (2); then we obtain:

$$\frac{ik^2}{\pi}\int_s\sum_{n=1}^{\infty}\sum_{m=-n}^{n}(-1)^m[c_{mn}^j\mathbf{n}\times\mathbf{M}_{m'n'}^1+d_{mn}^j\mathbf{n}\times\mathbf{N}_{m'n'}^1]\begin{pmatrix}\mathbf{N}_{-mn}^3\\ \mathbf{M}_{-mn}^3\end{pmatrix}ds+$$

$$+\frac{ik^2}{\pi}\sqrt{\frac{\varepsilon_1}{\mu_1}}\int_s\sum_{n=1}^{\infty}\sum_{m=-n}^{n}(-1)^m[c_{mn}^j\mathbf{n}\times\mathbf{N}_{m'n'}^1+d_{mn}^j\mathbf{n}\times\mathbf{M}_{m'n'}^1]\begin{pmatrix}\mathbf{M}_{-mn}^3\\ \mathbf{N}_{-mn}^3\end{pmatrix}ds=-\begin{pmatrix}p_{mn}^j\\ q_{mn}^j\end{pmatrix}.$$

This expression can be rewritten in matrix form as

$$\begin{pmatrix}I_1^{21}+\tilde{m}\cdot I_1^{12} & I_1^{22}+\tilde{m}\cdot I_1^{11}\\ I_1^{22}+\tilde{m}\cdot I_1^{11} & I_1^{12}+\tilde{m}\cdot I_1^{21}\end{pmatrix}\begin{pmatrix}d^j\\ c^j\end{pmatrix}=-i\begin{pmatrix}p^j\\ q^j\end{pmatrix}, \tag{13}$$

where $\tilde{m}$ is the relative refractive index of the particle.

Furthermore,

$$\frac{ik^2}{\pi}\int_s\sum_{n=1}^{\infty}\sum_{m=-n}^{n}(-1)^m[c_{mn}^j\mathbf{n}\times\mathbf{M}_{m'n'}^1+d_{mn}^j\mathbf{n}\times\mathbf{N}_{m'n'}^1]\begin{pmatrix}\mathbf{N}_{-mn}^1\\ \mathbf{M}_{-mn}^1\end{pmatrix}ds+$$

$$+\frac{ik^2}{\pi}\sqrt{\frac{\varepsilon_1}{\mu_1}}\int_s\sum_{n=1}^{\infty}\sum_{m=-n}^{n}(-1)^m[c_{mn}^j\mathbf{n}\times\mathbf{N}_{m'n'}^1+d_{mn}^j\mathbf{n}\times\mathbf{M}_{m'n'}^1]\begin{pmatrix}\mathbf{M}_{-mn}^1\\ \mathbf{N}_{-mn}^1\end{pmatrix}ds=-\begin{pmatrix}a_{mn}^j\\ b_{mn}^j\end{pmatrix}.$$

This expression is written in matrix form as

$$\begin{pmatrix}a^j\\ b^j\end{pmatrix}=-i\begin{pmatrix}I_1'^{21}+\tilde{m}\cdot I_1'^{12} & I_1'^{22}+\tilde{m}\cdot I_1'^{11}\\ I_1'^{22}+\tilde{m}\cdot I_1'^{11} & I_1'^{12}+\tilde{m}\cdot I_1'^{21}\end{pmatrix}\begin{pmatrix}d^j\\ c^j\end{pmatrix}. \tag{14}$$

Combining expressions (13) and (14) produces the following equation

$$\begin{pmatrix} a^j \\ b^j \end{pmatrix} = -\begin{pmatrix} I_1'^{21} + \tilde{m} \cdot I_1'^{12} & I_1'^{22} + \tilde{m} \cdot I_1'^{11} \\ I_1'^{22} + \tilde{m} \cdot I_1'^{11} & I_1'^{12} + \tilde{m} \cdot I_1'^{21} \end{pmatrix} \begin{pmatrix} I_1^{21} + \tilde{m} \cdot I_1^{12} & I_1^{22} + \tilde{m} \cdot I_1^{11} \\ I_1^{22} + \tilde{m} \cdot I_1'^{11} & I_1^{12} + \tilde{m} \cdot I_1^{21} \end{pmatrix}^{-1} \begin{pmatrix} p^j \\ q^j \end{pmatrix}. \quad (15)$$

We introduce notations for matrices Q_{01}^{11}, Q_{01}^{31} and rewrite expression (15) in a more compact form:

$$\begin{pmatrix} a^j \\ b^j \end{pmatrix} = T_1^j \begin{pmatrix} p^j \\ q^j \end{pmatrix}, \; T_1^j = -Q_{01}^{11}(k, k_1) \cdot [Q_{01}^{31}(k, k_1)]^{-1}, \quad (16)$$

where the elements of matrix T_1^j are expressed as surface integrals.

Consider the normal $\mathbf{n} = n_x\mathbf{i} + n_y\mathbf{j} + n_z\mathbf{k}$.

For a body positioned arbitrarily, we obtain the following expression:

$$\mathbf{n}dS = \frac{\partial(y,z)}{\partial(\theta,\varphi)}\mathbf{i} + \frac{\partial(z,x)}{\partial(\theta,\varphi)}\mathbf{j} + \frac{\partial(\mathrm{x},y)}{\partial(\theta,\varphi)}\mathbf{k},$$

where the components of the vector follow the expressions

$$\begin{aligned} n_x dS &= [r(\theta,\varphi)r(\theta,\varphi)'_\varphi \sin(\varphi) + r(\theta,\varphi)^2 \sin^2(\theta)\cos(\varphi)]d\theta d\varphi - \\ &- r(\theta,\varphi)r(\theta,\varphi)'_\theta \sin^2(\theta)\cos(\varphi)d\theta d\varphi, \\ n_y dS &= [-r(\theta,\varphi)r(\theta,\varphi)'_\varphi \cos(\varphi) + r(\theta,\varphi)^2 \sin^2(\theta)\sin(\varphi)]d\theta d\varphi - \\ &- r(\theta,\varphi)r(\theta,\varphi)'_\theta \sin^2(\theta)\sin(\varphi)d\theta d\varphi, \\ n_z dS &= [r(\theta,\varphi)^2 \sin^2(\theta)\sin(\theta)\cos(\theta) - r(\theta,\varphi)r(\theta,\varphi)'_\theta \sin^2(\theta)]d\theta d\varphi. \end{aligned}$$

The equation of the particle surface in a spherical coordinate system takes the following form:

$$r(\theta,\varphi) = [\sin^2\theta(\frac{\cos^2\varphi}{a^2} + \frac{\sin^2\varphi}{b^2}) + \frac{\cos^2\theta}{c^2}]^{-1/2}. \quad (17)$$

Let us clarify the form of the equation for the ellipsoid of revolution:

$$r(\theta) = [\frac{\sin^2\theta}{a^2} + \frac{\cos^2\theta}{c^2}]^{-1/2} = \frac{ac}{\sqrt{a^2\cos^2\theta + c^2\sin^2\theta}}. \quad (18)$$

Note that a spheroid (ellipsoid of revolution) is obtained by rotating an ellipse around the semi-minor axis (oblate ellipsoid) or a major axis (prolate ellipsoid). Two of the three semi-axes of this ellipsoid have the same length. The aspect ratio of a spheroid is defined as the ratio of the semi-major axis *a* to the semi-minor axis *c* and describes the shape of a particle that varies from a sphere ($a/c = 1$) to a disk for an oblate ellipsoid or a needle for a prolate ellipsoid ($a/c \neq 1$).

For example, the ratio *a*/*c* determines an oblate spheroid, and c/a determines a prolate one. In this case, *a* is the length of the semi-axis along the *x* and *y* axes, and *c is* the length of the semi-axes along the *z* axis, which is the axis of rotation.

Using the formulas for the conversion between Cartesian and spherical coordinates, we obtain the following expressions:

$$\begin{aligned} n_r dS &= [\sin(\theta)\cos(\varphi)n_x + \sin(\theta)\sin(\varphi)n_y + \cos(\theta)n_z]d\theta d\varphi = r(\theta,\varphi)^2 \sin(\theta)d\theta d\varphi, \\ n_\theta dS &= [\cos(\theta)\cos(\varphi)n_x + \cos(\theta)\sin(\varphi)n_y - \sin(\theta)n_z]d\theta d\varphi = -r(\theta,\varphi)r(\theta,\varphi)'_\theta \sin(\theta)d\theta d\varphi, \\ n_\varphi dS &= [-\sin(\varphi)n_x + \cos(\varphi)n_y]d\theta d\varphi = -r(\theta,\varphi)r(\theta,\varphi)'_\theta d\theta d\varphi. \end{aligned}$$

Next, we substitute expressions for $\mathbf{n}dS$, $\mathbf{N}^1_{mn}$, $\mathbf{M}^1_{mn}$, $\mathbf{N}^3_{mn}$, $\mathbf{M}^3_{mn}$ into the surface integrals and obtain explicit expressions for them:

$$I^{11}_{mnmn} = (-1)^{(m+m')}\int_0^{\pi} i[md^n_{om}(\theta)b^{n'}_{om}(\theta) + m'd^n_{om}(\theta)b^{n'}_{om'}(\theta)]\cdot[\int_0^{2\pi} c_{mnmn}(\theta,\varphi)d\varphi]d\theta, \tag{19}$$

$$\begin{aligned} I^{12}_{mnm'n'} &= (-1)^{(m+m')}\int_0^{\pi} -[b^{n'}_{om}(\theta)b^{n'}_{om'}(\theta)\sin(\theta) + mm'd^n_{om}(\theta)d^{n'}_{om'}(\theta)/\sin(\theta)]\times \\ &\times[\int_0^{2\pi} c^2_{mnm'n'}(\theta,\varphi)d\varphi] - \frac{n(n+1)}{x}d^n_{om}(\theta)b^{n'}_{om'}(\theta)\sin(\theta)[\int_0^{2\pi} c^3_{mnm'n'}(\theta,\varphi)d\varphi] - \\ &-i\frac{n'(n'+1)}{x_1}d^n_{om}(\theta)d^{n'}_{om'}(\theta)\sin(\theta)[\int_0^{2\pi} c^4_{mnm'n'}(\theta,\varphi)d\varphi]d\theta, \end{aligned} \tag{20}$$

$$\begin{aligned} I^{21}_{mnm'n'} &= (-1)^{(m+m')}\int_0^{\pi} \frac{n'(n'+1)}{x_1}b^n_{om}(\theta)d^{n'}_{om'}(\theta)\sin(\theta)[\int_0^{2\pi} c^3_{mnm'n'}(\theta,\varphi)d\varphi] - \\ &-im\frac{n'(n'+1)}{x_1}d^n_{om}(\theta)d^{n'}_{om'}(\theta)/\sin(\theta)[\int_0^{2\pi} c^4_{mnm'n'}(\theta,\varphi)d\varphi] + \\ &+mm'd^n_{om}(\theta)d^{n'}_{om'}(\theta)/\sin(\theta) + \\ &+b^{n'}_{om}(\theta)b^{n'}_{om'}(\theta)\sin(\theta)[\int_0^{2\pi} c^5_{mnm'n'}(\theta,\varphi)d\varphi]d\theta, \end{aligned} \tag{21}$$

$$\begin{aligned} I^{22}_{mnm'n'} &= (-1)^{(m+m')}\int_0^{\pi} i[md^n_{om}(\theta)b^{n'}_{om'}(\theta) + m'd^{n'}_{om'}(\theta)b^n_{om}(\theta)]\cdot[\int_0^{2\pi} c^6_{mnm'n'}(\theta,\varphi)d\varphi] + \\ &+\frac{n'(n'+1)}{x_1}b^n_{om}(\theta)d^{n'}_{om'}(\theta)[\int_0^{2\pi} c^7_{mnm'n'}(\theta,\varphi)d\varphi] - \frac{n(n+1)}{x}d^n_{om}(\theta)b^{n'}_{om'}(\theta)\times \\ &\times[\int_0^{2\pi} c^8_{mnm'n'}(\theta,\varphi)d\varphi] + im\frac{n'(n'+1)}{x_1}d^n_{om}(\theta)d^{n'}_{om'}(\theta)[\int_0^{2\pi} c^9_{mnm'n'}(\theta,\varphi)d\varphi] + \\ &+im'\frac{n(n+1)}{x}d^n_{om}(\theta)d^{n'}_{om'}(\theta)[\int_0^{2\pi} c^{10}_{mnm'n'}(\theta,\varphi)d\varphi]d\theta, \end{aligned} \tag{22}$$

$$I'^{11}_{mnm'n'} = (-1)^{(m+m')}\int_0^{\pi} i[md^n_{om}(\theta)b^{n'}_{om}(\theta) + m'd^n_{om}(\theta)b^{n'}_{om'}(\theta)]\cdot[\int_0^{2\pi} f^1_{mnm'n'}(\theta,\varphi)d\varphi]d\theta, \tag{23}$$

$$\begin{aligned} I'^{12}_{mnm'n'} &= (-1)^{(m+m')}\int_0^{\pi} -[b^n_{om}(\theta)b^{n'}_{om'}(\theta)\sin(\theta) + mm'd^n_{om}(\theta)d^{n'}_{om'}(\theta)/\sin(\theta)]\times \\ &\times[\int_0^{2\pi} f^2_{mnm'n'}(\theta,\varphi)d\varphi] - \frac{n(n+1)}{x}d^n_{om}(\theta)b^{n'}_{om'}(\theta)\sin(\theta)[\int_0^{2\pi} f^3_{mnm'n'}(\theta,\varphi)d\varphi] - \\ &-i\frac{n'(n'+1)}{x_1}d^n_{om}(\theta)d^{n'}_{om'}(\theta)\sin(\theta)[\int_0^{2\pi} f^4_{mnm'n'}(\theta,\varphi)d\varphi]d\theta, \end{aligned} \tag{24}$$

$$\begin{aligned} I'^{21}_{mnm'n'} &= (-1)^{(m+m')}\int_0^{\pi} \frac{n'(n'+1)}{x_1}b^n_{om}(\theta)d^{n'}_{om'}(\theta)\sin(\theta)[\int_0^{2\pi} f^3_{mnm'n'}(\theta,\varphi)d\varphi] - \\ &-im\frac{n'(n'+1)}{x_1}d^n_{om}(\theta)d^{n'}_{om'}(\theta)/\sin(\theta)[\int_0^{2\pi} f^4_{mnm'n'}(\theta,\varphi)d\varphi] + \\ &+mm'd^n_{om}(\theta)d^{n'}_{om'}(\theta)/\sin(\theta) + \\ &+b^{n'}_{om}(\theta)b^{n'}_{om'}(\theta)\sin(\theta)[\int_0^{2\pi} f^5_{mnm'n'}(\theta,\varphi)d\varphi]d\theta, \end{aligned} \tag{25}$$

$$I'^{22}_{mnm'n'} = (-1)^{(m+m')}\int_0^{\pi} i[md^n_{om}(\theta)b^{n'}_{om'}(\theta) + m'd^{n'}_{om'}(\theta)b^n_{om}(\theta)]\times$$
$$\times[\int_0^{2\pi} f^6_{mnm'n'}(\theta,\varphi)d\varphi] +$$
$$+\frac{n'(n'+1)}{x_1}b^n_{om}(\theta)d^{n'}_{om'}(\theta)[\int_0^{2\pi} f^7_{mnm'n'}(\theta,\varphi)d\varphi] - \frac{n(n+1)}{x}d^n_{om}(\theta)b^{n'}_{om'}(\theta)\times \quad (26)$$
$$\times[\int_0^{2\pi} f^8_{mnm'n'}(\theta,\varphi)d\varphi] + im\frac{n'(n'+1)}{x_1}d^n_{om}(\theta)d^{n'}_{om'}(\theta)[\int_0^{2\pi} f^9_{mnm'n'}(\theta,\varphi)d\varphi] +$$
$$+im'\frac{n(n+1)}{x}d^n_{om}(\theta)d^{n'}_{om'}(\theta)[\int_0^{2\pi} f^{10}_{mnm'n'}(\theta,\varphi)d\varphi]d\theta,$$

$$c^1_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})h_n(x)j_n(x_1)r^2(\theta,\varphi),\ c^2_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})u_n(x)j_n(x_1)r^2(\theta,\varphi),$$
$$c^3_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})h_n(x)j_n(x_1)\frac{dr(\theta,\varphi)}{d\theta},\ c^4_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})u_n(x)j_n(x_1)\frac{dr(\theta,\varphi)}{d\theta},$$
$$c^3_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})h_n(x)j_n(x_1)\frac{dr(\theta,\varphi)}{d\theta},\ c^4_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})u_n(x)j_n(x_1)\frac{dr(\theta,\varphi)}{d\theta},$$
$$c^5_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})h_n(x)v_n(x_1)r^2(\theta,\varphi),\ c^6_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})u_n(x)v_n(x_1)r^2(\theta,\varphi),$$
$$c^7_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})u_n(x)j_n(x_1)\frac{dr(\theta,\varphi)}{d\varphi},\ c^8_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})h_n(x)v_n(x_1)\frac{dr(\theta,\varphi)}{d\varphi},$$
$$c^9_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})u_n(x)j_n(x_1)\frac{dr(\theta,\varphi)}{d\theta},\ c^{10}_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})h_n(x)v_n(x_1)\frac{dr(\theta,\varphi)}{d\theta},$$
$$f^1_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})j_n(x)j_n(x_1)r^2(\theta,\varphi),\ f^2_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})v_n(x)j_n(x_1)r^2(\theta,\varphi),$$
$$f^3_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})j_n(x)j_n(x_1)\frac{dr(\theta,\varphi)}{d\theta},\ f^4_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})u_n(x)j_n(x)j_n(x_1)\frac{dr(\theta,\varphi)}{d\theta},$$
$$f^5_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})v_n(x)j_n(x_1)r^2(\theta,\varphi),\ f^6_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})v_n(x)u_n(x_1)r^2(\theta,\varphi),$$
$$f^7_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})v_n(x)j_n(x_1)\frac{dr(\theta,\varphi)}{d\varphi},\ f^8_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})j_n(x)v_n(x_1)\frac{dr(\theta,\varphi)}{d\varphi},$$
$$f^9_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})v_n(x)j_n(x_1)\frac{dr(\theta,\varphi)}{d\theta},\ f^{10}_{mnm'n'}(\theta,\varphi) = \exp(i\Delta_{m'm})j_n(x)v_n(x_1)\frac{dr(\theta,\varphi)}{d\theta}.$$

Thus, by using the method of extended boundary conditions, a solution to the scattering problem is obtained for the case of an irregular shape of the scatterer (ellipsoid).

The expansion coefficients of the scattered and incident electromagnetic fields turn out to be related by linear transformations of the T-matrix. The latter depends on a number of parameters (the size of the scatterer with respect to the wavelength, the refractive index, etc.), but it is invariant with respect to the direction of propagation of incident radiation for the selected coordinate system.

It is also necessary to specify the complexity of the application of the *T*-matrix method for biological media with typical optical "softness". This complexity is associated (in the specified cases) with poor convergence of the series corresponding to them in the calculation formulas for the elements of the *T*-matrix. The possibility of highly oscillating behavior of the integrand may also reduce accuracy. Moreover, the numerical inversion of the matrix will be poorly conditioned for scatterers with a zero (or small) imaginary part of the refractive index.

Following [22, 23], convergence can be significantly improved if the so-called LU factorization based on the application of an extended boundary condition is used. The graphs shown in Fig. 1 confirm the validity of this statement.

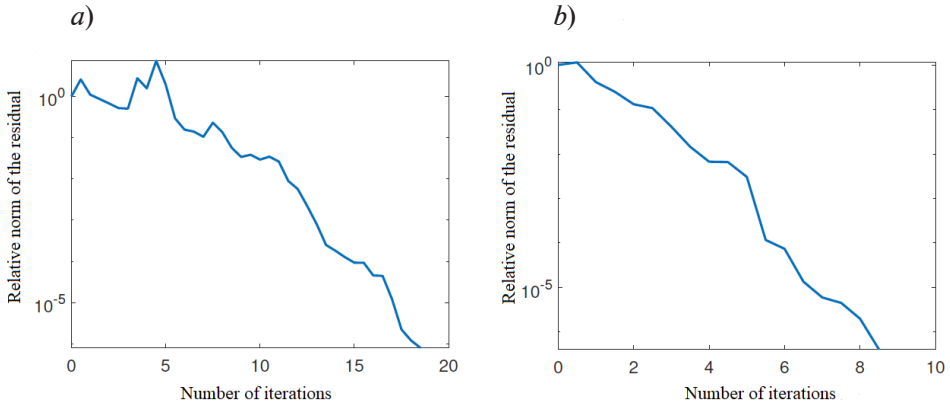

Fig. 1. Dependences of relative residual norm on the iteration number for the distances of 1 μm (*a*) and 2 μm (*b*) between particles
The dependences were btained using the method of biconjugate gradients with preconditioning (see Table 1)

Table 1

Parameters of model medium including 5 particles

Distance between particles, μm	Refractive index for particle		
	m_1	m_2	m_3, m_4, m_5
1	1.37	1.34	1.33
2	1.35	1.33	

Note. a = 18 μm, c = 3 μm for the first three particles, a = 5 μm, c = 5 μm (a are the lengths of the semi-axes of the spheroid along the x, y axes, respectively, and c are the lengths along the z axis).

Multiple scattering by ensembles of nonspherical scatterers

Electromagnetic waves incident on the surface of the jth scatterer form a field $\mathrm{E}_i(j)$, which consists of two components: the field created initially by incident waves and the field created by scattering by an ensemble of particles. The sum of the terms follows the expression

$$\mathbf{E}_i(j) = \mathbf{E}_0(j) + \sum_{i \neq j} \mathbf{E}_s(l, j). \tag{27}$$

A group of fields scattered by the jth particle is contained under the summation sign; (l,j) assigns the conversion from the coordinate system l to the coordinate system j.

Let us write out an expression for the incident field separately:

$$\mathbf{E}_0 = -\sum_{n=1}^{\infty} \sum_{m=-n}^{n} iE_{mn} \left[p_{mn}^{j_0,j} \mathbf{N}_{mn}^1(kr) + q_{mn}^{j_0,j} \mathbf{M}_{mn}^1(kr) \frac{e^{im\phi}}{kr} \right]. \tag{28}$$

Note that the incidence of waves relative to the center of each jth particle in its coordinate system (j-system) is considered.

The expansion coefficients for the given plane electromagnetic wave take the following form [18]:

$$p_{mn}^{j_0,j} = 4\pi(-1)^m i^n d_n \mathbf{C}_{mn}^*(\theta_{ink}) \mathbf{E}_{ink}(\mathbf{k}_{ink}, \mathbf{r}_{j_0,j}) \exp(-im\varphi_{ink}),$$

$$q_{mn}^{j_0,j} = 4\pi(-1)^m i^{n-1} d_n \mathbf{B}^*_{mn}(\theta_{ink})\mathbf{E}_{ink}(\mathbf{k}_{ink}, \mathbf{r}_{j_0,j}) \exp(-im\varphi_{inc}).$$

The complex conjugation is marked with an asterisk as standard, the notation $\mathbf{E}_{ink}(\mathbf{k}_{ink},\mathbf{r}_{j_0,j})$ represents the vector of linear polarization.

The following expression holds true for a field scattered by particles:

$$\mathbf{E}_s = -\sum_{n=1}^{\infty}\sum_{m=-n}^{n} iE_{mn}\left[p_{mn}^{l,j}\mathbf{N}_{mn}^1 + q_{mn}^{l,j}\mathbf{M}_{mn}^1\right], \tag{29}$$

where the expansion coefficients have the form [19].

The next stage consists in constructing an infinite system of algebraic equations based on combining expressions (27)–(29), taking into account expressions (16) for each *j*th particle of arbitrary shape:

$$\begin{pmatrix} a^j \\ b^j \end{pmatrix} = T_{12}^j\left[\begin{pmatrix} p^{i,j} \\ q^{i,j} \end{pmatrix} + \sum_{l\neq j}\begin{pmatrix} A(l,j) & B(l,j) \\ B(l,j) & A(l,j) \end{pmatrix}\begin{pmatrix} a^j \\ b^j \end{pmatrix}\right]. \tag{30}$$

The corresponding coefficients are determined in [19].

To solve the given system, we settled on the reduction method followed by the application of the biconjugate gradient method.

After the coefficients of system (30) are found, it becomes possible to record the total field in the far zone:

$$E_{totalq} = \sum_{n=1}^{\infty}\sum_{m=-n}^{n} iE_{mn}\left[a_{mn}\mathbf{N}_{mn}^3 + b_{mn}\mathbf{M}_{mn}^3\right], \tag{31}$$

$$a_{mn} = \sum_{j=1}^{L}\exp(-i\mathbf{k}_s,\mathbf{r}_j)a_{mn}^j,\ b_{mn} = \sum_{j=1}^{L}\exp(-i\mathbf{k}_s,\mathbf{r}_j)b_{mn}^j. \tag{32}$$

The componentwise notation of the scattered field has the form:

$$E_{s\theta} \sim \sum_{n=1}^{\infty}\sum_{m=-n}^{n}(-i)^n E_{mn} i\left[a_{mn}\tau_{mn} + b_{mn}\pi_{mn}\right]\frac{\exp(ikr)}{ikr}\exp(im\varphi), \tag{33}$$

$$E_{s\varphi} \sim \sum_{n=1}^{\infty}\sum_{m=-n}^{n}(-i)^n E_{mn} i\left[a_{mn}\pi_{mn} + b_{mn}\tau_{mn}\right]\frac{\exp(ikr)}{ikr}\exp(im\varphi), \tag{34}$$

where the functions of the angle follow the expressions

$$\tau_{mn} = \frac{\partial}{\partial\theta}P_n^m(\cos\theta),\ \pi_{mn} = \frac{m}{\sin\theta}P_n^m(\cos\theta).$$

The tilde sign (~) here implies the use of an asymptotic approximation.

Moreover, since we believe that scattering processes are considered at sufficiently large distances from the particle, where the electric vectors of the scattered and incident fields can be considered parallel, we can further simplify expressions (33) and (34) (we believe that only the component θ is nonzero in the far zone).

$$E_{s\theta} \sim E_0\frac{\exp(ikr)}{-ikr}\sum_{n=1}^{\infty}\sum_{m=-n}^{n}\frac{2n+1}{n(n+1)}(a_{mn}\tau_{mn} + b_{mn}\pi_{mn}), \tag{35}$$

$$E_{s\varphi} \sim E_0\frac{\exp(ikr)}{-ikr}\sum_{n=1}^{\infty}\sum_{m=-n}^{n}\frac{2n+1}{n(n+1)}(a_{mn}\pi_{mn} + b_{mn}\tau_{mn}), \tag{36}$$

$$\tau_n = \frac{\partial}{\partial\theta}P_n(\cos\theta),\ \pi_n = \frac{1}{\sin\theta}P_n(\cos\theta).$$

Simulation of the reflection of a plane wave from a nontrivial multilayered structure

Consider a nontrivial layered structure (nontriviality here is understood as waviness of the layers), where each layer has its own refractive index, and use some of the results obtained in [20].

A flat *p*-polarized wave (a similar case of *s*-polarization would be simpler) is incident on the model considered at an angle θ. Our goal is to find the reflected field. Let us write out expressions for the fields formed by light radiation passing through the above layers and reflected from them, assuming that the phases of the waves oscillate rapidly, and the amplitudes change slowly:

$$E_1 = \exp(\frac{i}{\varepsilon}\tau_{inc}(\xi_1,\xi_2,\xi_3)) + \exp(\frac{i}{\varepsilon}\tau_{1ref}(\xi_1,\xi_2,\xi_3))\,\mathrm{A}(\xi_1,\xi_2,\xi_3,\varepsilon_x,\varepsilon_y), \tag{37}$$

$$E_2 = \exp(\frac{i}{\varepsilon}\tau_{2elap}(\xi_1,\xi_2,\xi_3))\,\mathrm{B}^+(\xi_1,\xi_2,\xi_3,\varepsilon_x,\varepsilon_y) + \\ + \exp(\frac{i}{\varepsilon}\tau_{3ref}(\xi_1,\xi_2,\xi_3))\,\mathrm{B}^-(\xi_1,\xi_2,\xi_3,\varepsilon_x,\varepsilon_y), \tag{38}$$

$$E_3 = \exp(\frac{i}{\varepsilon}\tau_{3elap}(\xi_1,\xi_2,\xi_3))\,\mathrm{C}^+(\xi_1,\xi_2,\xi_3,\varepsilon_x,\varepsilon_y) + \\ + \exp(\frac{i}{\varepsilon}\tau_{3ref}(\xi_1,\xi_2,\xi_3))\,\mathrm{C}^-(\xi_1,\xi_2,\xi_3,\varepsilon_x,\varepsilon_y), \tag{39}$$

$$E_4 = \exp(\frac{i}{\varepsilon}\tau_{4elap}(\xi_1,\xi_2,\xi_3))\,\mathrm{D}^+(\xi_1,\xi_2,\xi_3,\varepsilon_x,\varepsilon_y) + \\ + \exp(\frac{i}{\varepsilon}\tau_{5ref}(\xi_1,\xi_2,\xi_3))\,\mathrm{D}^-(\xi_1,\xi_2,\xi_3,\varepsilon_x,\varepsilon_y) + E_{4scat\phi}(\xi_1,\xi_2,\xi_3), \tag{40}$$

$$E_5 = \exp(\frac{i}{\varepsilon}\tau_{5elap}(\xi_1,\xi_2,\xi_3))E(\xi_1,\xi_2,\xi_3,\varepsilon_x,\varepsilon_y). \tag{41}$$

Similar to our earlier study [20], we sequentially found the terms of the series for the required amplitudes, as well as the expression for the Gaussian beam.

Distribution function for particles simulated by ellipsoids of revolution

Let us determine the parameters of the model medium corresponding to normal human skin (Table 2).

Let the incident plane wave propagate in the *x*-axis direction (the semi-minor axis for an oblate ellipsoid) and have polarization in the *z*-axis direction.

Fig. 2 shows images of oblate and prolate ellipsoids and the coordinate system used associated with them.

The model medium is as close as possible to the real parameters of normal human skin.

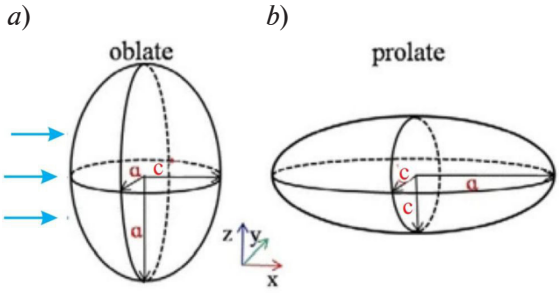

Fig. 2. Images of oblate (*a*) and prolate (*b*) ellipsoids;
a, *c* are the lengths of their semi-axes directed along the corresponding coordinate axes.
Blue arrows indicate the direction of the incident laser beams

Table 2

Adopted characteristics of the model medium [20]

Parameter	Notation	Parameter value for layer i		
		(2)	(3)	(4)
Layer thickness, μm	d_i	65	565	90
Set of distortion parameters	a_i	–0.0024	0.021	0.041
	b_i	0.0200	0.030	0.050
	c_i	0.010		
Refractive index (real part)	n_i^0	1.50	1.40	1.35

Notes. 1. Distortion parameters are represented by the formula $H_i = c_i \sin(a_i x + b_i y)$.
2. Refractive index of ambient air $n_1 = 1{,}000$; we assumed $\chi_2 = \chi_3 = \chi_4 = \chi_5 = 10^{-5}$; $n_5^0 = 1.40$ for the ith layer of the model absorbing medium with $n_i = n_i^0 + i\chi_i$.

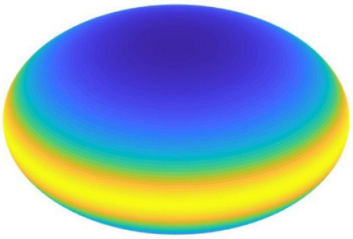

Fig. 3. General view of oblate ellipsoid under consideration with dimensions a = 18 μm, c = 3 μm

A number of well-known facts allow us to consider the erythrocyte as a homogeneous scatterer: the erythrocyte membrane is thin and does not significantly affect the scattering of laser radiation, and cellular organelles are not included in the structure of the erythrocyte. Thus, our calculations are performed for an oblate ellipsoid (Fig. 3).

The mathematical approach we developed allows to detect the aggregation of particles as well as to determine their spectral parameters for the *in vivo* case. The illustrations presented below (Figs. 4–6) demonstrate the capabilities of the software package we created by based on the presented theoretical approach. Evidently, both the numerical parameters and the shapes of the curves change with varying distances between the scatterers.

The results obtained indicate the difference in cell sizes, the diversity of their internal structures, and the effect of interference on the pattern of wave fields scattered by neighboring particles.

Thus, the developed method creates new possibilities, allowing to take into account the effects of cooperative interaction of particles in the case of denser packing of erythrocytes.

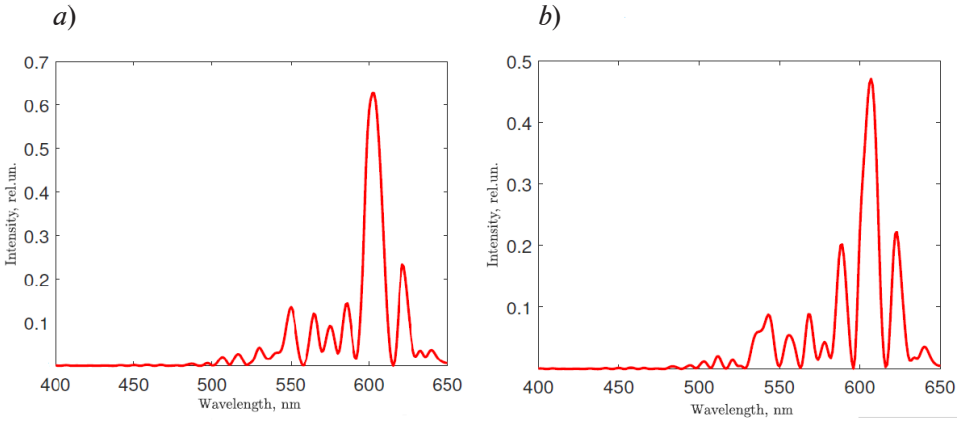

Fig. 4. Functional dependences of intensity versus wavelength for the light scattered by particle ensembles located in the layer; distances between the particles are 1 μm (*a*) and 2 μm (*b*) (see Table 1)

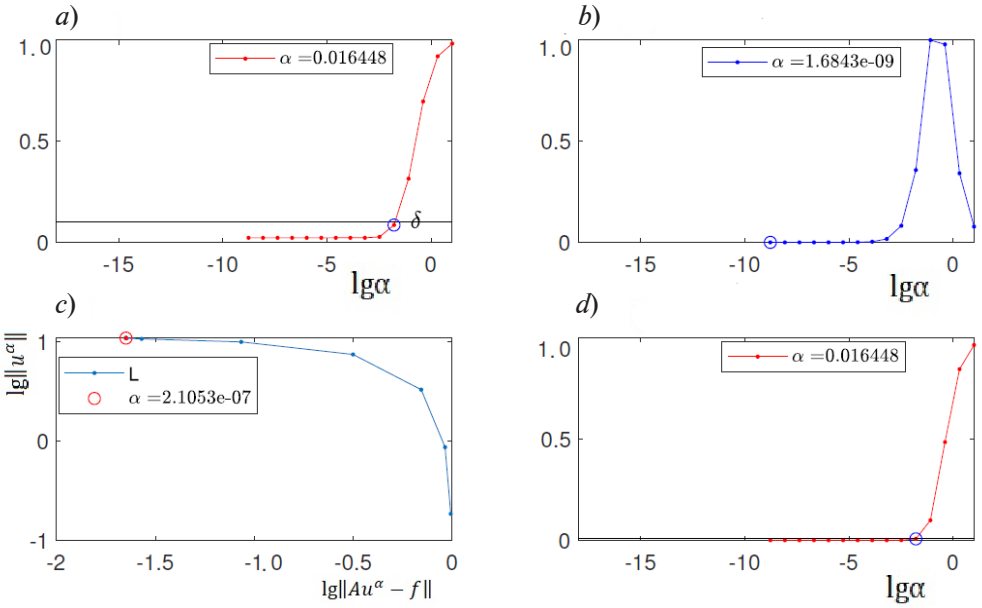

Fig. 5. Results of automatic determination of regularization parameter α from the given kernel errors and "measured" data using different criteria: residual (*a*), quasi-optimality (*b*), *L*-curve (*c*) and generalized residual principle (*d*) for a bimodal distribution (see Fig. 7,*a*)

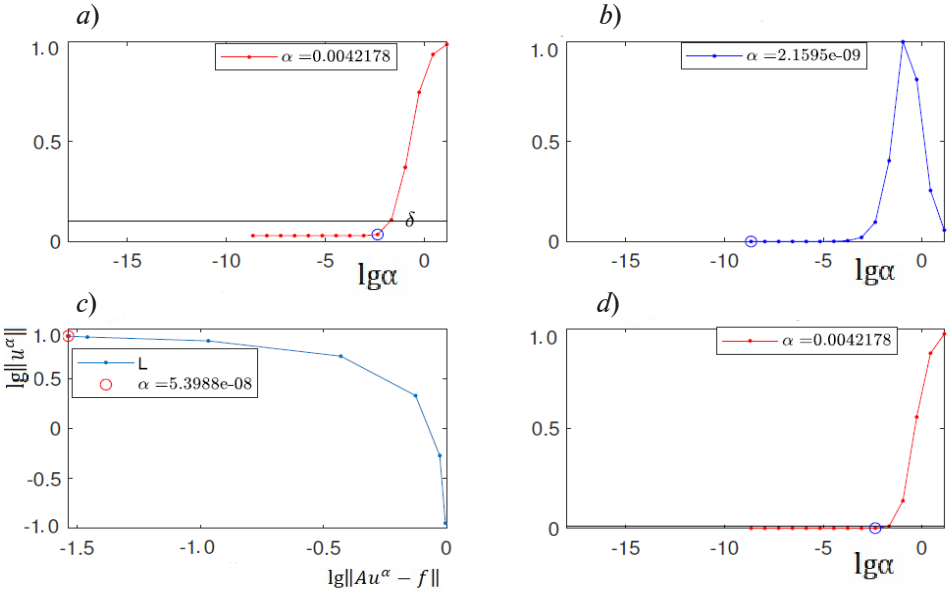

Fig. 6. Results similar to those shown in Fig. 5, but for normal distribution (see Fig. 7,*b*)

The next stage of the study consists of solving the inverse problem: to find the distribution of erythrocytes (assuming them to be spheroids) from the aspect ratio of the spheroid ($\rho = a/c$), based on the known intensity of laser radiation scattering (measured with some error) by an aggregated ensemble of particles located in the layer (*in vivo* case).

Such problems are described by linear first-kind Fredholm integral equations, taking the form

$$\boldsymbol{Au} \equiv \int_{\rho\min}^{\rho\max} I_{scat(\theta)}(\rho,\lambda)u(\rho)d\rho = f(\lambda), \tag{42}$$

where $\boldsymbol{A}$ is the integral operator, $I_{scat(\theta)}(\rho,\lambda)$ is the kernel of the integral equation, $u(\rho)$ is the required cell size distribution, $f(\lambda) \equiv I_{blood}(\theta,\lambda)$ is the scattering intensity.

The kernel $I_{scat(\theta)}(\rho,\lambda)$ is defined as the intensity of light scattered in the direction of the angle θ (the angle is chosen experimentally) by a nonspherical particle (see Eq. (35)). We assume that this kernel is a function continuous in the rectangle $\Omega = ([c,d]\times[a,b])$, and $f(\lambda) \in L_{[c,d]}$ ($a \equiv \rho_{\min}$, $b \equiv \rho_{\max}$, $c \equiv \lambda_{\min}$, $d \equiv \lambda_{\max}$).

The inversion of the integral operator $\boldsymbol{A}$ for the inverse problem (see equation (42)) is unstable, therefore, it is advisable to use the Tikhonov regularization method for the numerical solution [24, 25].

Automatic determination of the regularization parameter based on the given kernel errors and "measured" data is possible within the framework of the software package we developed. (This package includes the methods of relative residual, the generalized residual principle (GRP), the L-curve method and the quasi-optimality criterion.)

Thus, we propose to choose the regularization parameter in accordance with several criteria. In a problem with a known model solution, this allows us to find the range of the best values of the parameter α. It turned out that the residual principle and the GRP gave the same parameter value and when they were used to solve the integral equation (42), a profile close to the model was reconstructed.

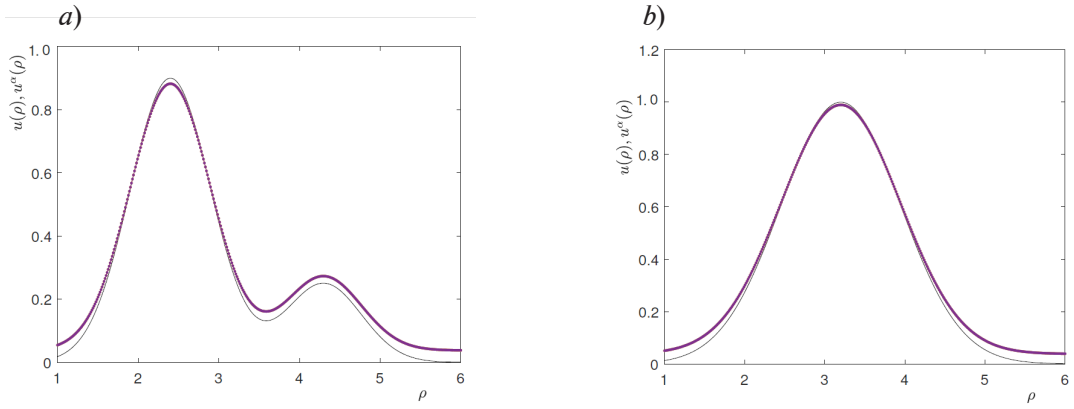

Fig. 7. Bimodal (a) and normal (b) size distributions for spheroidal particles with two values of the distance between the scatterers: 1 μm (a) and 2 μm (b). Graphs of functions from [26] (solid lines) are compared with the results of our numerical solution (dotted lines)

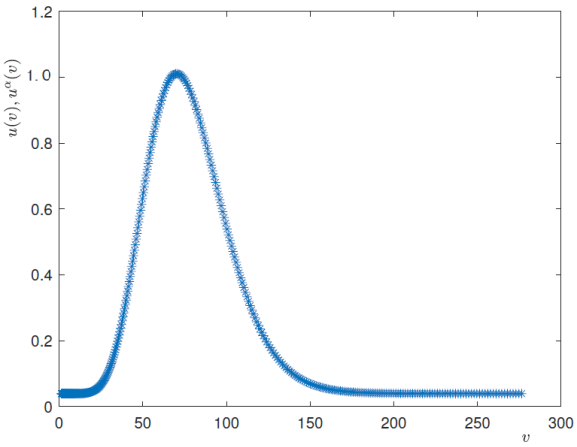

Fig. 8. Case of normal volume distribution of the particles (Price–Jones curve) obtained for distance of 2 μm between the scatterers

Fig. 7,*a* shows a comparison of the two curves. The solid black curve corresponds to an asymmetric bimodal particle size distribution, which is predefined by the function from [26]. The given distribution simulates the presence of fractions of normocytes and macrocytes. The dotted colored curve corresponds to our numerical solution of the problem and demonstrates that both peaks of the size distribution were reconstructed quite satisfactorily. A similar solid curve in Fig. 7,*b* also corresponds to a predefined size distribution, a normal one (see [26]). As a result of our numerical solution of the problem (dotted colored curve), where the noise level in the right-hand side of equation (42) is assumed to be 5%, a completely satisfactory agreement with the given function is also obtained. Thus, the particle size distribution profiles were reconstructed with high accuracy.

The analysis of the graphs in Fig. 7 also allows to conclude that taking into account the nonsphericity of particles accurately reconstructs the Price–Jones curve (Fig. 8), describing the typical volume distribution of human blood corpuscles.

Erythrocyte indices

In this section, we consider the numerical evaluation of erythrocyte indices (which are standard in clinical practice), in particular the mean corpuscular volume (MCV) and the degrees of dispersion of red blood cells by volume. These include the deviations of the relative distribution of erythrocytes over the volume (red cell distribution width, abbreviated as RDW) from the mean (coefficient of variation, or CV) and from the standard (standard deviation, or SD).

In other words, RDW-CV shows a percentage deviation of the erythrocyte volume from the mean, and RDW-SD is the difference between the largest and smallest erythrocyte (measured in femtoliters, like MCV).

Let us first determine the volume of the body formed by revolution around the axis of the shape:

$$V_{rot} = 4\pi\int_0^1 xy(x)dx, \tag{43}$$

where y(x) represents the family of Perseus curves [27]

$$y(x) = \mathrm{c}\sqrt{a^2 - (\sqrt{x^2 - p^2} - d^2)^2}; \tag{44}$$

the following shape parameters are given here: a, b are the semi-axes of the ellipse, $c = a/b$ (a = 0.150662, c = 1.659376); d is the distance from the origin to the center of the ellipse (d = 1.768398); p is the distance from the axis of the torus to the secant plane (p = 1.637922). Expression (44) defines families of Perseus curves. They are lines of intersection of the surface of the torus with planes parallel to its axis, and represent algebraic lines of the 4th order.

The volume of revolution for this shape is

$$V_{rot} = 4\pi\int_0^1 xc\sqrt{a^2 - (\sqrt{x^2 + p^2} - d)^2}\,dx = 1.2799.$$

If we assume that the diameter of a human erythrocyte is 7.55 μm on average, then the relationship between the volume and the radius of the erythrocyte takes the form $V_{\mathrm{MCV}} = V_{rot}R^3$ and the mean volume of the erythrocyte is V_{MCV} = 1.2799·68.8536 μm^3.

The equation of the form (44) is written in a spherical coordinate system:

$$r^4\alpha_1 - 2r^2\alpha_2 - \beta_2 = 0,$$

and the corresponding solution of this biquadratic equation has the form

$$r(\theta,\varphi) = \frac{\sqrt{\alpha_1(-\alpha_2 + \sqrt{\alpha_1\beta_2 + \alpha_2^2})}}{\alpha_1}, \tag{45}$$

where $\alpha_1 = \sin^2\theta\gamma_1,\ \alpha_2 = \sin\theta\beta_1,\ \beta_2 = \frac{2db^2}{a^2}p - \gamma_2;$
here

$$\gamma_1 = \sin^2\varphi + \frac{b^2}{a^2}\cos^2\varphi,\ \gamma_2 = b^2(1+\frac{p}{a^2}) - d^2,\ \beta_1 = \gamma_1\gamma_2 - \frac{db^2}{a^2}\cos^2\varphi.$$

Solving the inverse problem allows to find the volume distribution function (see Fig. 8) taking into account the equation of the surface (45).

We should note that similar results were obtained in [29].

In this case, it is possible to calculate the size heterogeneity index of the erythrocyte based on the obtained theoretical volume distribution.

In medical practice, the RDW-SD indicator is the result of direct measurement of the width of the erythrocyte curve at a 20% level (with the height of the curve taken as 100%) [28].

For example, RDW-SD = 118 − 36 = 82 fl. Then values within the range of 80−100 fl characterize the erythrocyte as a normocyte, below 80 fl as a microcyte, and above 100 fl as a macrocyte.

Notably, the RDW-SD indicator is more sensitive to the appearance of a certain number of micro- and macrocytes in the erythrocyte population, since it is measured at the lower part of the erythrocyte volume distribution curve. If reticulocytosis occurs (an increase in reticulocytes (precursors of erythrocytes) in the process of hematopoiesis), this indicator changes faster, since there is a certain broadening of the erythrocyte curve.

Results and conclusions

The main goal of the study was to develop and refine an electrodynamic model of the interaction of low-power laser radiation with a dispersed medium, including elements of irregular shape (ellipsoid), which are models of blood cells (erythrocytes) located in a medium with a layered structure (*in vivo* case).

Let us overview our main findings.

1. The developed analytical methods for calculating the light scattering characteristics of particles located in a layered medium are described. It is assumed that these particles are arbitrarily oriented and have an irregular shape (nonspherical).

2. Steps were taken to optimize the convergence of processes using the method of extended boundary conditions; this increased the possibilities of using the *T*-matrix method.

3. The mathematical model developed for the analysis of biological processes based on calculated optical characteristics was implemented in a new software package.

4. The developed approach to the problem made it possible to correctly reconstruct the traditionally used indicator of the distribution of blood corpuscles (RDW-SD) for the *in vivo* case, taking into account the aspect ratios and the structural features of the biological aggregate.

5. We found an approach to theoretically predicting the number of abnormally sized erythrocytes in a biological material using the developed model, based on the standard measurement of the width of the size distribution of erythrocytes. For example, in the case of the presence of micro- and macrocytes, it is possible to diagnose the degree of anicytosis, since the width of the distribution is higher than the reference value, and the obtained curves of the volume distribution of erythrocytes clearly indicate the difference in cell size.

Thus, the RDW index is an informative and convenient marker for laboratory diagnostics.

The results obtained serve as the basis for the proposed new method of rapid analysis of whole blood. According to this method, it is necessary to find the distributions of bloody corpuscles by characteristic indices, the RDW index as well as the geometric characteristics of red blood cells related to their volume and shape, for the *in vivo* case.

REFERENCES

1. **Tuchin V. V.,** Handbook of optical biomedical diagnostics, 2nd Edition, Vol. 1: Light-tissue interaction, SPIE Press, Bellingham, WA, USA, 2016.

2. **Eremina E., Eremin Y., Wriedt T.,** Analysis of light scattering by erythrocyte based on the discrete sources method, Opt. Commun. 244 (1–6) (2005) 15–23.

3. **Eremina E., Hellmers J., Eremin Y., Wriedt T.,** Different shape models for erythrocyte: Light scattering analysis based on the discrete sources method, J. Quant. Spectrosc. Radiat. Transfer. 102 (1) (2006) 3–10.

4. **Eremina E., Wriedt T.,** Light scattering analysis by a particle of extreme shape via discrete soused method, J. Quant. Spectrosc. Radiat. Transfer. 89 (1–4) (2004) 67–77.

5. **Mishchenko M. I., Wiscomber W. J., Hovenier J. W., Travis L. D.,** Overview of scattering by nonspherical particles, In book: Mishchenko M. I., Hovenier J. W., Travis L. D. (Eds.), Light scattering by nonspherical particles: Theory, measurements and applications, Academic Press, Cambridge, Massachusetts, USA (1999) 29–60.

6. **Latimer P.,** Light scattering by ellipsoids, J. Colloid Interface Sci. 53 (1) (1975) 102–109.

7. **Cai Q., Liou K.-N.,** Polarized light scattering by hexagonal ice crystals: Theory, Appl. Opt. 21 (19) (1982) 3569–3580.

8. **Hammer M., Schweitzer D., Michel B., et al.,** Single scattering by red blood cells, Appl. Opt. 37 (31) (1998) 7410–7418.

9. **Van de Hulst H. C.,** Light scattering by small particles, John Willey & Sons Inc., New York, 1957.

10. **Shatilov A. V.,** O rasseyanii sveta dielektricheskimi ellipsoidami, sravnimymi s dlinoyvolny. 1. Obshchee vyrazheniye dlya indikatrisy rasseyaniya ellipsoidalnoy chastitsy [About light scattering by dielectric ellipsoids comparable to the wavelength. 1. General expression for the scattering indicatrix of an ellipsoidal particle], Optika i Spektroskopiya. 9 (1) (1960) 86–91 (in Russian).

11. **Newton R. G.,** Scattering theory of waves and particles, Second Edition (Dover Books on Physics), Dover Publications, New York, 2013.

12. **Klett J. D., Sutherland R. A.,** Approximate methods for modeling the scattering properties of nonspherical particles: Evaluation of the Wentzel–Kramers–Brillouin method, Appl. Opt. 31 (3) (1992) 373–386.

13. **Erma V. A.,** An exact solution for the scattering of electromagnetic waves from bodies of arbitrary shape: III. Obstacles with arbitrary electromagnetic properties, Phys. Rev. 179 (5) (1969) 1238–1246.

14. **Waterman P. C.,** Matrix formulation of electromagnetic scattering, Proc. IEEE. 53 (8) (1969) 805–812.

15. **Waterman P. C.,** Symmetry, unitarity and geometry in electromagnetic scattering, Phys. Rev. D. 3 (4) (1971) 825–839.

16. **Steinke J. M., Shepherd A. P.,** Comparison of Mie theory and light scattering of red blood cells, Appl. Opt. 27 (19) (1988) 4027–4033.

17. **Yaroslavsky A. N., Goldbach T., Schwarzmaier H.,** Influence of the scattering phase function approximation on the optical properties of blood determined from the integrating sphere measurements, J. Biomed. Opt. 4 (1) (1999) 47–53.

18. **Tsang L., Kong J. A., Shin R. T.,** Theory of microwave remote sensing, Willey Interscience, New York, 1985.

19. **Kulikov K. G., Radin A. M.,** Study of dispersion and absorption of an ensemble of spherical particles inside an optical cavity and new possibilities of predicting the optical characteristics of biological media by intracavity spectroscopy, Opt. Spectrosc. 92 (2) (2002) 199–206.

20. **Golovitskii A. P., Kontsevaya V. G., Kulikov K. G.,** An electrodynamic model for determining the distribution function of particles by size for blood cells *in vivo*, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (1) (2023) 97–110 (in Russian).

21. **Doicu A., Wriedt T., Eremin Y. A.,** Light scattering by systems of particles. Null-field method with discrete sources: Theory and programs, Springer, Berlin, New York, 2006.

22. **Mishchenko M. I., Travis L. D.,** T-matrix computations of light scattering by large spheroidal particles, Opt. Commun. 109 (1–2) (1994) 16–21.

23. **Mishchenko M. I., Travis L. D.,** Capabilities limitations of a current FORTRAN implementation of the T-matrix method for randomly oriented, rotationally symmetric scatterers, J. Quant. Spectrosc. Radiat. Transfer. 60 (3) (1998) 309–324.

24. **Tikhonov A. N., Arsenin V. A.,** Solution of ill-posed problems, Winston, USA, 1977.

25. **Tikhonov A. N., Goncharsky A. V., Stepanov V. V., Yagola A. G.,** Numerical methods for the solution of ill-posed problems, Book Series: Mathematics and its Applications, Vol. 328, Springer Dordrecht, Netherlands, 1995.

26. **Ustinov V. D.,** On inverse reconstruction problems of the erythrocyte size distribution in laser diffractometry, Math. Models Comput. Simul. 9 (5) (2017) 561–569.

27. **Zub V. V., Kirillov V. Kh., Kuzakon V. M.,** Geometriya eritrotsita [Erythrocyte geometry], Sci. Works. (48) (2015) 182–186 (in Russian).

28. **Caporal F. A., Comar S. R.,** Evaluation of RCW-CV, RDW-SD, and MATH-1SD for the detection of erythrocyte anisocytosis observed by optical microscopy, J. Bras. Pathol. Med. Lab. 49 (5) (2013) 324–331.

29. **Doubrovski V. A., Torbin S. O., Zabenkov I. V.,** Determination of individual and average characteristics of native blood erythrocytes by the static spectral digital microscopy method, Opt. Spectrosc. 130 (6) (2022) 709–719.

СПИСОК ЛИТЕРАТУРЫ

1. **Тучин В. В.** Оптическая биомедицинская диагностика. В 2 тт. Т. 1. М.: Ай Пи Ар Медиа, 2021.549 с.

2. **Eremina E., EreminY., Wriedt T.** Analysis of light scattering by erythrocyte based on the discrete sources method // Optics Communications. 2005. Vol. 244. No. 1–6. Pp. 15–23.

3. **Eremina E., Hellmers J., EreminY., Wriedt T.** Different shape models for erythrocyte: Light scattering analysis based on the discrete sources method // Journal of Quantitative Spectroscopy and Radiative Transfer. 2006. Vol. 102. No. 1. Pp. 3–10.

4. **Eremina E., Wriedt T.** Light scattering analysis by a particle of extreme shape via discrete soused method // Journal of Quantitative Spectroscopy and Radiative Transfer. 2004. Vol. 89. No.1–4. Pp. 67–77.

5. **Mishchenko M.I., Wiscomber W. J., Hovenier J. W., Travis L. D.** Overview of scattering by nonspherical particles // Mishchenko M. I., Hovenier J. W., Travis L. D. (Eds.). Light scattering by nonspherical particles: Theory, measurements and applications. Cambridge, Massachusetts, USA: Academic Press, 1999. Pp. 29–60.

6. **Latimer P.** Light scattering by ellipsoids // Journal of Colloid and Interface Science. 1975. Vol. 53. No. 1. Pp. 102–109.

7. **Cai Q., Liou K.-N.** Polarized light scattering by hexagonal ice crystals: Theory // Applied Optics. 1982. Vol. 21. No. 19. Pp. 3569–3580.

8. **Hammer M., Schweitzer D., Michel B., Thamm E., Kolb A.** Single scattering by red blood cells // Applied Optics. 1998. Vol. 37. No. 31. Pp. 7410–7418.

9. **Ван де Хюлст Г.** Рассеяние света малыми частицами: Пер. с англ. М.: Изд-во иностр. лит-ры, 1961. 536 с.

10. **Шатилов А. В.** О рассеянии света диэлектрическими эллипсоидами, сравнимыми с длиной волны. 1. Общее выражение для индикатрисы рассеяния эллипсоидальной частицы // Оптика и спектроскопия. 1960. Т. 9. № 1. С. 86–91.

11. **Ньютон Р.** Теория рассеяния волн и частиц. Пер. с англ. М.: Мир, 1969. 600 с.

12. **Klett J. D., Sutherland R. A.** Approximate methods for modeling the scattering properties of nonspherical particles: Evaluation of the Wentzel –Kramers – Brillouin method // Applied Optics. 1992. Vol. 31. No. 3. Pp. 373–386.

13. **Erma V. A.** An exact solution for the scattering of electromagnetic waves from bodies of arbitrary shape: III. Obstacles with arbitrary electromagnetic properties // Physical Review. 1969. Vol. 179. No. 5. Pp.1238–1246.

14. **Waterman P. C.** Matrix formulation of electromagnetic scattering // Proceedings of the IEEE. 1969. Vol. 53. No. 8. Pp. 805–812.

15. **Waterman P. C.** Symmetry, unitarity and geometry in electromagnetic scattering // Physical Review D. 1971. Vol. 3. No. 4. Pp. 825–839.

16. **Steinke J. M., Shepherd A. P.** Comparison of Mie theory and light scattering of red blood cells // Applied Optics. 1988. Vol. 27. No. 19. Pp. 4027–4033.

17. **Yaroslavsky A. N., Goldbach T., Schwarzmaier H.** Influence of the scattering phase function approximation on the optical properties of blood determined from the integrating sphere measurements // Journal of Biomedical Optics. 1999. Vol. 4. No. 1. Pp. 47–53.

18. **Tsang L., Kong J. A., Shin R. T.** Theory of microwave remote sensing. New York: Willey Interscience, 1985. 632 p.

19. **Куликов К. Г., Радин А. М.** Исследование дисперсии и спектра поглощения совокупности сферических частиц в полости оптического резонатора и новые возможности прогноза оптических характеристик биологических сред методом внутрирезонаторной лазерной спектроскопии // Оптика и спектроскопия. 2002. Т. 92. № 2. С. 228–236.

20. **Головицкий А. П., Концевая В. Г., Куликов К. Г.** Электродинамическая модель определения функции распределения частиц по размерам для клеток крови (случай *in vivo*) // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 1. С. 97–110.

21. **Doicu A., Wriedt T., Eremin Y. A.** Light scattering by systems of particles. Null-field method with discrete sources: Theory and programs. Berlin, New York: Springer, 2006. 322 p.

22. **Mishchenko M. I., Travis L. D.** T-matrix computations of light scattering by large spheroidal particles // Optics Communications. 1994. Vol. 109. No. 1–2. Pp. 16–21.

23. **Mishchenko M. I., Travis L. D.** Capabilities limitations of a current FORTRAN implementation of the T-matrix method for randomly oriented, rotationally symmetric scatterers // Journal of Quantative Spectroscopy and Radiative Transfer. 1998. Vol. 60. No. 3. Pp. 309–324.

24. **Тихонов А. Н., Арсенин В. А.** Методы решения некорректных задач. М.: Наука, 1979. 288 с.

25. **Тихонов А. Н., Гончарский А. В., Степанов В. В., Ягола А. Г.** Численные методы решения некорректных задач. М.: Наука, 1990. 232 с.

26. **Устинов В. Д.** Об обратных задачах восстановления распределения эритроцитов по размерам в лазерной дифрактометрии // Математическое моделирование. 2017. Т. 29. № 3. С. 51–62.

27. **Зуб В. В., Кириллов В. Х., Кузаконь В. М.** Геометрия эритроцита // Scientific Works. 2015. No. 48. Pp. 182–186.

28. **Caporal F. A., Comar S. R.** Evaluation of RCW-CV, RDW-SD, and MATH-1SD for the detection of erythrocyte anisocytosis observed by optical microscopy // Brazilian Journal of Pathology and Laboratory Medicine. 2013.Vol. 49. No. 5. Pp. 324–331.

29. **Дубровский В. А., Торбин С. О., Забенков И. В.** Определение индивидуальных средних характеристик эритроцитов нативной крови методом статистической цифровой спектральной микроскопии // Оптика и спектроскопия. 2022. Т. 130. № 6. С. 894–905.

THE AUTHORS

GOLOVITSKII Alexander P.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
alexandergolovitski@yahoo.com
ORCID: 0000-0003-4292-0959

KONTSEVAYA Vera G.
Pskov State University
Peter the Great St. Petersburg Polytechnic University
2 Lenin Sq., Pskov, 180000, Russia
nkoncevoi@mail.ru
ORCID: 0000-0002-1434-5056

KULIKOV Kirill G.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
kulikov.kirill.g@gmail.com
ORCID: 0000-0002-4610-7394

KOSHLAN Tatiana V.
St. Petersburg State University
7–9UniversitetskayaEmb., St. Petersburg, 199034, Russia
Koshlan.tetiana@gmail.com
ORCID: 0000-0002-0238-2909

СВЕДЕНИЯ ОБ АВТОРАХ

ГОЛОВИЦКИЙ Александр Петрович – *доктор физико-математических наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого,* Санкт-Петербург, Россия.
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
alexandergolovitski@yahoo.com
ORCID: 0000-0003-4292-0959

КОНЦЕВАЯ Вера Геннадьевна – *старший преподаватель кафедры математики и теории игр Псковского государственного университета, г. Псков, инженер Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого,* Санкт-Петербург, Россия.
180000, Россия, г. Псков, пл. Ленина, 2
nkoncevoi@mail.ru
ORCID: 0000-0002-1434-5056

КУЛИКОВ Кирилл Геннадьевич – *доктор физико-математических наук, профессор Высшей школы биомедицинских технологий Санкт-Петербургского политехнического университета Петра Великого,* Санкт-Петербург, Россия.
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
kulikov.kirill.g@gmail.com
ORCID: 0000-0002-4610-7394

КОШЛАН Татьяна Викторовна – *кандидат физико-математических наук, научный сотрудник Санкт-Петербургского государственного университета, Санкт-Петербург, Россия.*
199034, Россия, Санкт-Петербург, Университетская наб., 7–9
Koshlan.tetiana@gmail.com
ORCID: 0000-0002-0238-2909

Received 12.09.2023. Approved after reviewing 05.10.2023. Accepted 05.10.2023.
Статья поступила в редакцию 12.09.2023. Одобрена после рецензирования 05.10.2023. Принята 05.10.2023.

NUCLEAR PHYSICS

Original article
DOI: https://doi.org/10.18721/JPM.16414

A GENERATOR OF DEEP INELASTIC LEPTON-PROTON SCATTERING BASED ON THE GENERATIVE-ADVERSARIAL NETWORK (GAN)

A. A. Lobanov✉, *Ya. A. Berdnikov*

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ lobanov2.aa@edu.spbstu.ru

Abstract. The paper considers the application of a Generative Adversarial Network (GAN) for the development of a generator of deep inelastic lepton-proton scattering. The difficulty of effective training of the generator based on GAN is noted. It is associated with the use of complex schemes of distributions of physical properties (energies, momentum components, etc.) of particles in the process of deeply inelastic lepton-proton scattering. It is shown that the GAN makes it possible to faithfully reproduce the distributions of lepton physical properties in the final state at different initial energies of the center of mass in the range between 20 and 100 GeV.

Keywords: inclusive deep inelastic scattering, neural network, generative adversarial network, lepton-proton scattering

Citation: Lobanov A. A., Berdnikov Ya. A., A generator of deep inelastic lepton-proton scattering based on the Generative-Adversarial Network (GAN), St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 181–188. DOI: https://doi.org/10.18721/JPM.16415

Научная статья
УДК 539.12
DOI: https://doi.org/10.18721/JPM.16414

ГЕНЕРАТОР ГЛУБОКО НЕУПРУГОГО РАССЕЯНИЯ ЛЕПТОНОВ НА ПРОТОНЕ НА ОСНОВЕ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНОЙ НЕЙРОННОЙ СЕТИ

А. А. Лобанов✉, *Я. А. Бердников*

Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

✉ lobanov2.aa@edu.spbstu.ru

Аннотация. В работе рассмотрено применение генеративно-состязательной сети (ГСС) для создания генератора глубоко неупругого лептон-протонного рассеяния. Отмечена сложность эффективного обучения генератора на основе ГСС, которая связана с использованием сложных схем распределения физических характеристик (энергий, компонентов импульсов и т. п.) частиц в процессе глубоко неупругого лептон-протонного рассеяния. Показано, что ГСС позволяет точно воспроизводить распределения физических характеристик лептона в конечном состоянии.

Ключевые слова: инклюзивное глубоко неупругое рассеяние, нейронная сеть, генеративно-состязательная сеть, лептон-протонное рассеяние

Ссылка для цитирования: Лобанов А. А., Бердников Я. А. Генератор глубоко неупругого рассеяния лептонов на протоне на основе генеративно-состязательной нейронной сети // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. 181–188. DOI: https://doi.org/10.18721/JPM.16414

Introduction

The results of experimental studies of deep inelastic lepton-proton scattering are generally processed and analyzed by modeling both the actual process of particle interaction and the operation of detector setups; the Monte Carlo method is the most convenient for this purpose. The problem is that simulation involves complex physical models, requiring high computational costs and much time.

Machine learning methods provide an alternative, allowing to build event generators. The advantage of these methods is that they can be trained on heterogeneous data, i.e., both experimental results and data obtained by modeling the entire process under consideration (for example, inclusive deep inelastic scattering). The resulting event generator can be capable to collect the necessary data quickly and with minimal computational costs.

In this paper, we consider one of these machine learning models, the generative-adversarial network (GAN) [1].

The advantage of the considered model is its ability to faithfully reproduce the real data on which it was trained.

The GAN model includes two neural networks: a generator and a discriminator. The first network is intended for generating some quantities, such as particle characteristics. The second network identifies the differences between the values obtained by the generator and the real values.

The discriminator tries to distinguish the real values from those created by the generator, thus training it. The generator gets better at producing data with each training iteration, which in turn trains the discriminator [1].

While the GAN method has been successful for diverse applications (for example, generating photos and videos that are indistinguishable from real ones [2, 3]), it has certain drawbacks associated with complications in the training process of the model.

The reason for these complications is the strong dependence on the parameters of the model, often causing the following issues:

instabilities during training,
discrepancies,
parameter variations,
retraining of models.

There are many approaches to solving these problems, for example, those outlined in [4].

In this paper, we used the approach proposed in [5], described in detail below in the following section.

Applying GANs in high energy physics and elementary particle physics comes with additional difficulties. The most crucial are the multiple strict constraints dictated by conservation laws. Consequently, not every generation output can be considered suitable.

The prediction accuracy is also important; otherwise, the relationships between the derived quantities may be violated, which is also unacceptable. Similar problems are described, for example, in [6].

Conservation laws can produce significant irregularities in the distributions of physical quantities (for example, angles, momenta, energies, etc.) characterizing the interaction of particles. An example is the distribution of the p_z momentum component of the final-state lepton (Fig. 1,*a*). Multiplicity is understood (in Fig. 1 and below) as the number of events in the bin normalized by the total number of events, i.e., a dimensionless quantity. As evident from Fig. 1,*a*, the distribution has a sharp edge associated with the laws of conservation of energy–momentum: energy (or momentum) in the final state cannot exceed the level of energy (or momentum) in the initial state. The existence of such an irregularity negatively affects the training of GAN, as discussed in [6].

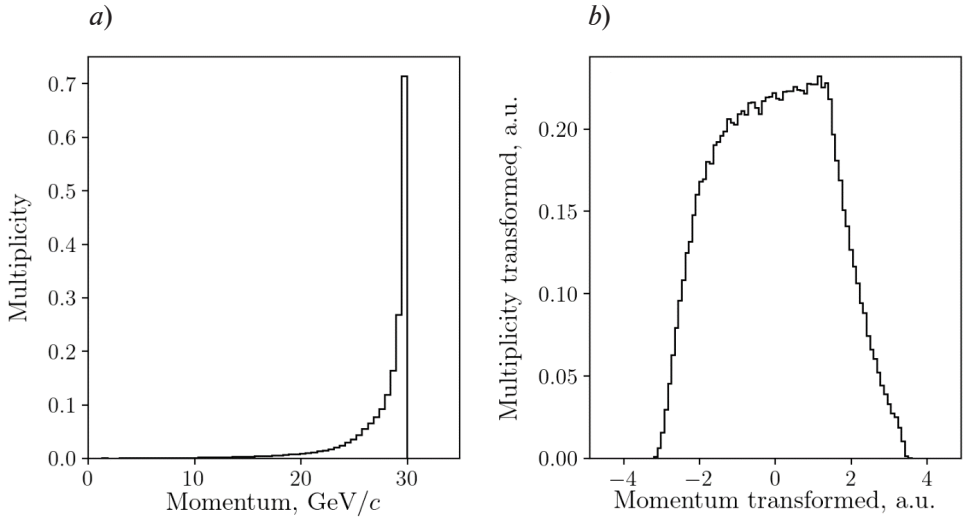

Fig. 1. Distributions of momentum component p_z of final-state lepton (*a*) and transformed quantity $T(p_z)$ (*b*)
Initial electron energy E_0 = 30 GeV

It was proposed in [7] that the problems associated with irregularities in the distributions of quantities with respect to specific physical parameters can be solved by generating transformed 'twins' of the quantities rather than the quantities themselves, modified in such a way that the new distribution becomes smoother.

The following transformation is used in this study for the p_z momentum component of the final-state lepton [7]:

$$T(p_z) = \log[(E_0 - p_z)/(1\ \text{GeV}/c)].$$

As a result, a smoother distribution is obtained (see Fig. 1,*b*).

A similar transformation was applied for the total energy of the scattered lepton E_l:

$$T(E_l) = \log[(E_0 - E_l)/(1\ \text{GeV}/c)].$$

Methodology

Since this paper considers inclusive scattering of charged leptons (e^+, e^-, μ^+, μ^-) by protons, the scattered lepton is characterized by four-momentum in the lepton–proton center-of-mass frame:

$$p_l = (E_l, \mathbf{p}),$$

where **p** is the lepton's three-dimensional momentum vector, given by the components p_x, p_y, p_z; E_l is the total energy of the scattered lepton.

Additional parameters are the total energy E_0 of the incident lepton in the lepton–proton center of mass frame and the type of lepton (e^+ or e^- or μ^+ or μ^-). These parameters allow the GAN to predict the final state of various leptons at different initial energies considered.

The energy E_0 is defined as

$$E_0 \approx \sqrt{s_{lN}}/2,$$

where $\sqrt{s_{lN}}$ is the initial energy in the lepton–proton center of mass frame.

The initial energies E_0 = 10, 20, 30, 40, 50 GeV were considered for training.

The PYTHIA8 program was used to obtain the final states of leptons [8]. 100,000 events were generated at initial energies $\sqrt{s_{lN}}$ = 20, 40, 60, 80 and 100 GeV for each type of lepton: (e^+, e^-, μ^+, μ^-). The four-momenta of the final-state lepton were recorded in each event (referred to as the real values).

Using the quantities $T(p_z)$ and $T(E_l)$ (the transformed quantities) allows the generator to avoid predicting unphysical values, and the discriminator to distinguish the real data from the generated ones.

The following quantities are fed to the discriminator input to increase its accuracy:

$$p_z,\ E_l,\ p_T = \sqrt{p_x^2 + p_y^2},\ \varphi = \arctan\left(p_z \middle/ p_T\right),\ \theta = \arctan\left(p_y \middle/ p_x\right)$$

(these are referred to as additional quantities).

A 128-dimensional noise vector (a vector of values obtained from a Gaussian distribution with the mean equal to 0 and the variance equal to 1), energy E_0 and lepton type are fed to the generator input. The generator network consists of 4 hidden layers of 512 neurons each with a Leaky ReLU activation function and a dropout of 0.2 [9]. The output layer consists of 4 neurons with a linear activation function. The output is four main predicted quantities: p_x, p_y, $T(p_z)$ and $T(E_l)$. In addition to these, the model includes the prediction of additional quantities: p_z, E_l, p_T, φ, θ, obtained based on the predicted ones. The main and additional quantities are then fed to the discriminator input.

The discriminator network consists of 4 hidden layers with 512 neurons each, a Leaky ReLU activation function and a dropout of 0.2 [9]. A so-called dropout layer with a rate of 10% [10] is applied to each of the layers, randomly dropping 10% of the layer weights. This helps prevent overfitting in classification procedures [11]. Spectral normalization is also applied to each layer [12], allowing to achieve a 1-Lipschitz mapping for the discriminator [13]. The output layer consists of a single neuron with a linear activation function. The higher the value obtained, the more confident the discriminator is in identifying the given values as realistic.

The paper uses the type of generative-adversarial network with a least square loss function.

The following expressions are valid for the loss functions of the discriminator (L_D) and the generator (L_G) in such networks [5]:

$$L_D = \frac{1}{2}E_{\boldsymbol{x}\sim p_{data}(\boldsymbol{x})}[(D(\boldsymbol{x}\,|\,\boldsymbol{y}) - b)^2] + \frac{1}{2}E_{z\sim p(z)}[(D(G(\boldsymbol{z}\,|\,\boldsymbol{y})) - a)^2], \quad (1)$$

$$L_G = \frac{1}{2}E_{z\sim p(z)}[(D(G(\boldsymbol{z}\,|\,\boldsymbol{y})) - c)^2], \quad (2)$$

where $D(...)$ is the discriminator network; $G(...)$ is the generator network; $\boldsymbol{x}$ are the real data; $\mathbf{z}$ is the noise vector; $D(\boldsymbol{x})$ are the values obtained by the discriminator based on the real data; $D(G(\mathbf{z}))$ are the values found by the discriminator based on the data obtained by the generator; E is the expected value; a, b are the hyperparameters of this loss function, equal to 0 and 1, respectively [5].

GAN was trained for 400 epochs in our study. RMSProp was used for gradient descent optimization, with $\rho = 0.9$ [14], $1\cdot10^{-4}$ training steps for the generator and $5\cdot10^{-5}$ for the discriminator. Using different training steps contributes to better training convergence, as shown in [15].

Simulation results

Due to the large number of possible scattering configurations (different types of leptons and different initial energies E_0), only some configurations are given below to illustrate the operation of the GAN.

Fig. 2 shows the distributions of the momentum components for the muon μ^+ and the electron e^- in the final states, obtained by GAN and the PYTHIA8 program. It can be seen that the model generates quantities with virtually identical distributions, as evidenced by the χ^2 values in the graphs and the corresponding momenta (p-value) [17].

Fig. 3 shows the distributions of the p_z momentum components of final-state electrons at different energies, obtained by GAN and the PYTHIA8 program. Analyzing the results obtained, we can conclude that the model can predict the correct distributions both at the energies at which the network was trained (10, 20, 30, 40, 50 GeV), and at interpolated energies (15, 25, 35, 45 GeV). Notably, the model can also predict the p_z values at high energies E_0 (60, 70, 80, 90 GeV).

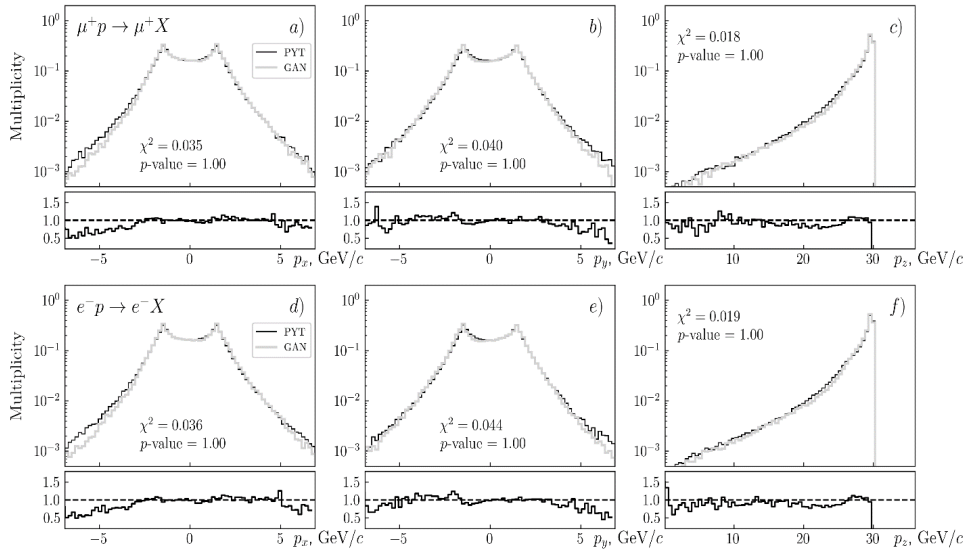

Fig. 2. Predicted distributions over momentum components p_x, p_y, p_z for the muon μ^+ (*a*, *b*, *c*) and electron e^- (*d*, *e*, *f*) at the same initial energy $E_0 = 30$ GeV, obtained using GAN (gray curves) and PYTHIA8 (black curves). The corresponding values of χ^2 and the graphs for the ratio of GAN to PYTHIA8 (GAN/PYT) predictions are given for each distribution

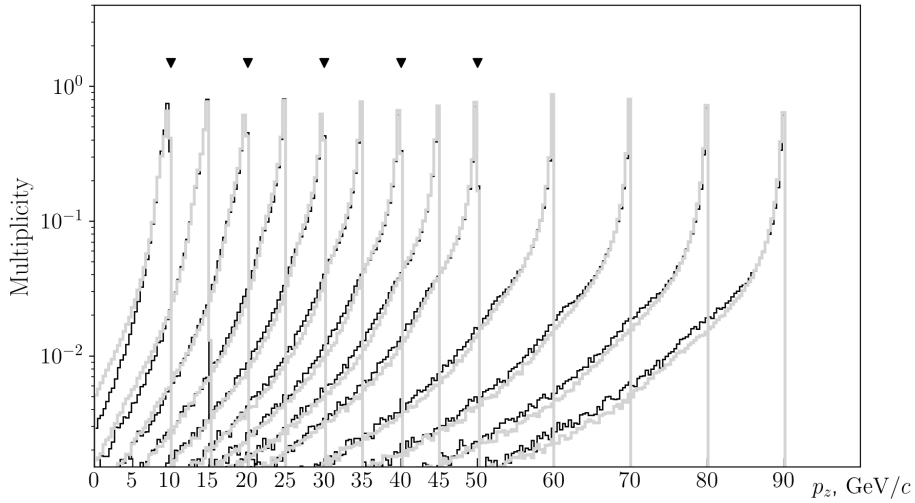

Fig. 3. Distributions of p_z momentum component of the electron, predicted using the PYTHIA8 program (gray curves) and using GAN (black), at different initial energies E_0 Triangles indicate the energies at which the model was trained

Aside from lepton momenta and energies, let us consider the quantities derived from them, used to characterize scattering. Such quantities include the squared momentum transfer $Q^2 = -q^2$ (q is the momentum of the virtual photon) and the Bjorken variable $x_{\text{Bj}} = Q^2/2Pq$ (P is the momentum of the incident proton).

Fig. 4 shows the joint distributions of Q^2 and x_{Bj} at energies $E_0 = 10$ and 40 GeV, obtained based on data from PYTHIA8 and GAN. Comparing the distributions in Fig. 4, *a* and *b* and those in Fig. 4, *c* and *d*, obtained by two approaches at two values of E_0 (10 and 40 GeV), we can see good agreement between the distributions obtained using PYTHIA8 and GCC. The χ^2 values calculated for all distribution bins are given as a quantitative assessment of this agreement.

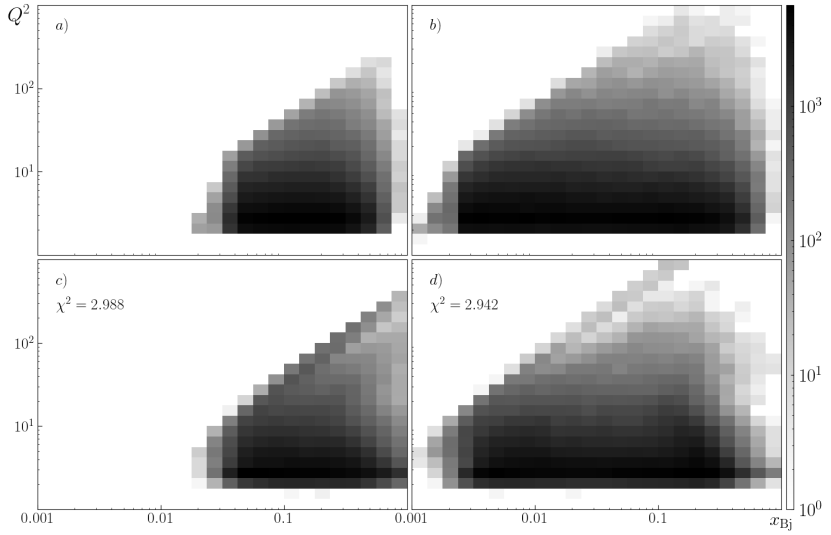

Fig. 4. Joint distribution of Q^2 and x_{Bj} for electrons at initial energies E_0 = 10 GeV (*a*, *c*) and 40 GeV (*b*, *d*), predicted using PYTHIA8 (*a*, *b*) and GAN *(c*, *d*)
The values of χ^2 are given to characterize the accuracy of GAN predictions for each value of E_0

Conclusion

The paper considers a generative-adversarial network (GAN) to generate the final state of leptons in inclusive deep inelastic lepton–proton scattering in the 20–100 GeV center-of-mass energy range.

We confirmed that the developed model can generate the distributions of various characteristics of different final-state leptons, including the quantities calculated based on the initially generated ones. The GAN can generate distributions not only at initial center-of-mass energies on which it was trained but also at interpolated energies (GeV): 15, 25, 35, 45.

In addition, we found that the model can generate the required distributions at extrapolated initial energies (GeV): 120, 140, 160 and 180.

In the future, there is a clear interest in considering semi-inclusive deep inelastic scattering, generating the characteristics of an additional particle, in particular a pion.

REFERENCES

1. **Goodfellow I., Pouget-Abadie J., Mirza M., et al.,** Generative Adversarial Networks, Commun. ACM. 63 (11) (2020) 139–144.

2. **Karras T., Laine S., Aila T.,** A style-based generator architecture for Generative Adversarial Networks, Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, June 15–20 (2019) 4401–4410.

3. **Clark A., Donahue J., Simonyan K.,** Adversarial video generation on complex datasets, arXiv: 1907.06571v2, 2019. https://doi.org/10.48550/arXiv. 1907.06571.

4. **Gulrajani I., Ahmed F., Arjovsky M., et al.,** Improved training of Wasserstein GANs, arXiv:1704.00028v3, 2017. https: //doi.org/ 10.48550/arXiv. 1704.00028.

5. **Mao X., Li Q., Xie H., et al.,** On the effectiveness of least squares Generative Adversarial Networks, IEEE Trans. Pattern Anal. Mach. Intell. 41 (12) (2019) 2947–2960.

6. **Hashemi B., Amin N., Datta K., et al.,** LHC analysis-specific datasets with Generative Adversarial Networks. arXiv: 1901.05282v1, 2019. https: //doi.org/10.48550/arXiv.1901.05282.

7. **Alanazi Y., Sato N., Liu T., et al.,** Simulation of electron-proton scattering events by a Feature-Augmented and Transformed Generative Adversarial Network (FAT-GAN). arXiv: 2001.11103v2, 2019. https: //doi.org/10.48550/arXiv.2001. 11103.

8. **Sjöstrand T., Mrenna S., Skands P.,** A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. 178 (11) (2008) 852–867.

9. **Sharma O.,** A new activation function for deep neural network, Proc. Int. Conf. Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), IEEE, Faridabad, India, Febr. 14–16 (2019) 84–86.

10. **Srivastava N., Hinton G., Krizhevsky A., et al.,** Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.

11. **Hawkins D. M.,** The problem of overfitting, J. Chem. Inf. Comput. Sci. 44 (1) (2003) 1–12.

12. **Miyato T., Kataoka T., Koyama M., Yoshida Y.,** Spectral normalization for Generative Adversarial Networks, arXiv: 1802.05957/v1, 2018. https://doi.org/10.48550/arXiv.1802.05957.

13. **Qin Y., Mitra N., Wonka C.,** How does Lipschitz regularization influence GAN training? Computer Vision – ECCV 2020, Springer Int. Publ. (2020) 310–326. https://doi.org/10.48550/arXiv.1811.09567

14. **Xu D., Zhang Sh., Zhang H., Mandic D. P.,** Convergence of the RMSProp deep learning method with penalty for nonconvex optimization, Neural Netw. 139 (July) (2021) 17–23.

15. **Heusel M., Ramsauer H., Unterthiner T., et al.,** GANs trained by a two time-scale update rule converge to a local Nash equilibrium, arXiv: 1706.08500v6, 2017. https://doi.org/10.48550/arXiv.1706.08500.

16. **McHugh M. L.,** The chi-square test of independence, Biochem. Med. 23 (2) (2013) 143–149.

СПИСОК ЛИТЕРАТУРЫ

1. **Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y.** Generative adversarial networks // Communications of the ACM. 2020. Vol. 63. No. 11. Pp. 139–144.

2. **Karras T., Laine S., Aila T.** A style-based generator architecture for generative adversarial networks // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA, June 15–20, 2019. Pp. 4401–4410.

3. **Clark A., Donahue J., Simonyan K.** Adversarial video generation on complex datasets. arXiv: 1907.06571v2, 2019. https://doi.org/10.48550/arXiv. 1907.06571.

4. **Gulrajani I., Ahmed F., Arjovsky M., Dumoulin V., Courville A.** Improved training of Wasserstein GANs. arXiv: 1704.00028v3, 2017. https: //doi.org/ 10.48550/arXiv.1704.00028.

5. **Mao X., Li Q., Xie H., Lau R. Y. K., Wang Zh., Smolley S. P.** On the effectiveness of least squares generative adversarial networks // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019. Vol. 41. No. 12. Pp. 2947–2960.

6. **Hashemi B., Amin N., Datta K., Olivito D., Pierini N.** LHC analysis-specific datasets with Generative Adversarial Networks. arXiv: 1901.05282v1, 2019. https: //doi.org/10.48550/arXiv.1901.05282.

7. **Alanazi Y., Sato N., Liu T., et al.** Simulation of electron-proton scattering events by a Feature-Augmented and Transformed Generative Adversarial Network (FAT-GAN). arXiv: 2001.11103v2, 2019. https: //doi.org/10.48550/arXiv.2001. 11103.

8. **Sjöstrand T., Mrenna S., Skands P.** A brief introduction to PYTHIA 8.1 // Computer Physics Communications. 2008. Vol. 178. No. 11. Pp. 852–867.

9. **Sharma O.** A new activation function for deep neural network // Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE, Faridabad, India, February 14–16, 2019. Pp. 84–86.

10. **Srivastava N., Hinton G., Krizhevsky A., Slutskever I., Salakhutdinov R.** Dropout: A simple way to prevent neural networks from overfitting // The Journal of Machine Learning Research. 2014. Vol. 15. Pp. 1929–1958.

11. **Hawkins D. M.** The problem of overfitting // Journal of Chemical Information and Computer Sciences. 2003. Vol. 44. No. 1. Pp. 1–12.

12. **Miyato T., Kataoka T., Koyama M., Yoshida Y.** Spectral normalization for Generative Adversarial Networks. arXiv: 1802.05957/v1, 2018. https://doi.org/

10.48550/arXiv.1802.05957.

13. **Qin Y., Mitra N., Wonka C.** How does Lipschitz regularization influence GAN training? // Computer Vision – ECCV 2020. Springer International Publishing, 2020. Pp. 310–326. https://doi.org/10.48550/arXiv.1811.09567

14. **Xu D., Zhang Sh., Zhang H., Mandic D. P.** Convergence of the RMSProp deep learning method with penalty for nonconvex optimization // Neural Networks. 2021. Vol. 139. July. Pp. 17–23.

15. **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. arXiv: 1706.08500v6, 2017. https://doi.org/10.48550/arXiv.1706. 08500.

16. **McHugh M. L.** The chi-square test of independence // Biochemia Medica. 2013. Vol. 23. No. 2. Pp. 143–149.

THE AUTHORS

LOBANOV Andrey A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
lobanov2.aa@edu.spbstu.ru
ORCID: 0000-0002-8910-4775

BERDNIKOV Yaroslav A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
berdnikov@spbstu.ru
ORCID: 0000-0003-0309-5917

СВЕДЕНИЯ ОБ АВТОРАХ

ЛОБАНОВ Андрей Александрович – *студент Физико-механического института Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
lobanov2.aa@edu.spbstu.ru
ORCID: 0000-0002-8910-4775

БЕРДНИКОВ Ярослав Александрович – *доктор физико-математических наук, профессор Высшей школы фундаментальных физических исследований Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
berdnikov@spbstu.ru
ORCID: 0000-0003-0309-5917

Received 20.07.2023. Approved after reviewing 31.07.2023. Accepted 31.07.2023.
Статья поступила в редакцию 20.07.2023. Одобрена после рецензирования 31.07.2023. Принята 31.07.2023.

Original article
DOI: https://doi.org/10.18721/JPM.16415

SIMULATION OF SEMI-INCLUSIVE DEEP INELASTIC LEPTON SCATTERING ON A PROTON AT ENERGIES OF 20–100 GEV ON THE BASIS OF A GENERATIVE-ADVERSARIAL NEURAL NETWORK

A. A. Lobanov✉, *Ya. A. Berdnikov*

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ lobanov2.aa@edu.spbstu.ru

Abstract. This paper continues a series of articles devoted to developing the capabilities of a deep inelastic lepton-proton scattering event generator based on the generative adversarial network (GAN). The investigation has focused on semi-inclusive reactions of deep inelastic scattering and, particularly, on hadron registration. The results confirmed that GAN could accurately generate distributions of physical properties of leptons and hadrons. It worked for different types of leptons and hadrons in the range of initial energies from 20 to 100 GeV in the center-of-mass system. The GAN demonstrated to preserve the inherent correlation between the characteristics of leptons and protons.

Keywords: semi-inclusive deep inelastic scattering, machine learning, neural network, generative-adversarial network

Citation: Lobanov A. A., Berdnikov Ya. A., Simulation of semi-inclusive deep inelastic lepton scattering on a proton at energies of 20−100 GeV on the basis of a generative-adversarial neural network, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 189−197. DOI: https://doi.org/10.18721/JPM.16415

Научная статья
УДК 539.12
DOI: https://doi.org/10.18721/JPM.16415

МОДЕЛИРОВАНИЕ ПОЛУИНКЛЮЗИВНОГО ГЛУБОКО НЕУПРУГОГО РАССЕЯНИЯ ЛЕПТОНА НА ПРОТОНЕ ПРИ ЭНЕРГИЯХ 20 – 100 ГЭВ НА ОСНОВЕ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНОЙ НЕЙРОННОЙ СЕТИ

А. А. Лобанов✉, *Я. А. Бердников*

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

✉ lobanov2.aa@edu.spbstu.ru

Аннотация. Данная работа продолжает цикл статей, посвященных развитию возможностей генератора событий глубоко неупругого лептон-протонного рассеяния на основе генеративно-состязательной сети (ГСС). Здесь рассмотрены полуинклюзивные реакции глубоко неупругого рассеяния с регистрацией адрона. Показано, что ГСС позволяет с высокой точностью генерировать распределения физических характеристик конечных лептона и адрона в диапазоне начальных энергий 100 − 20 ГэВ в системе центра масс.

Ключевые слова: полуинклюзивное глубоко неупругое рассеяние, машинное обучение, нейронная сеть, генеративно-состязательная сеть

Ссылка для цитирования: Лобанов А. А., Бердников Я. А. Моделирование полуинклюзивного, глубоко неупругого рассеяния лептона на протоне при энергиях 20 −100 ГэВ на основе генеративно-состязательной нейронной сети // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. 189−197. DOI: https://doi.org/10.18721/JPM.16415

Introduction

Modern experimental research in high energy physics deals with increasingly large datasets [1], collected from large-scale experiments or simulation results. Processing these data requires high computational costs and much time.

Machine learning methods offer an approach to solving the above-mentioned problems [2], allowing to construct computer simulation software (called event generators) with the following new capabilities:

using the experimental results on interactions of particles and nuclei at discrete points to predict the characteristics of secondary particles at any energies in the given range based on interpolation (and possibly extrapolation) quickly and without high computational costs;

the above-mentioned software can be developed even without experimental results, using the simulation results for interactions of particles and nuclei obtained by the Monte Carlo method [3].

A generative-adversarial network (GAN) was described in [4] to create a generator for inclusive deep inelastic lepton−proton scattering.

This paper continues the research in this direction, extending the capabilities of the given event generator [4] to semi-inclusive deep inelastic scattering with hadron production.

The goal of the study was to build a generator that can be trained on experimental data (or those obtained by computer simulation), allowing to collect intermediate data based on interpolation and extrapolation, since the experiment cannot be carried out at arbitrary initial energies.

There are several reasons for the interest in semi-inclusive processes.

First, the production of an additional hadron allows to learn more about the structure of the proton. Thus, the type of hadron produced by lepton−proton interaction depends on the flavor of the quark in the proton that the virtual photon emitted by the charged lepton interacted with [5].

Secondly, the characteristics of an additional hadron can carry information about the processes of parton hadronization [5].

Thirdly, various spin and azimuthal asymmetries can be measured during semi-inclusive processes, allowing to gain an understanding on the spin structure of the proton [6].

Methodology

The characteristics of the final state of the charged lepton (e^+, e^-, μ^+, μ^-) and hadron (π^0, π^+, π^-, K^+, K^-) are their four-momenta $p_l = (E_l, \mathbf{p}_l)$ and $p_h = (E_h, \mathbf{p}_h)$ respectively, where E_l is the total energy of the scattered lepton; p_l, $\mathbf{p}_l$ are the four- and three-dimensional momentum vectors of the lepton, the latter determined in terms of its components p_{xl}, p_{yl}, p_{zl}; E_h is the total energy of the hadron, p_h, $\mathbf{p}_h$ are four- and three–dimensional momentum vectors of the hadron, and also the components of the latter, p_{xh}, p_{yh}, p_{zh}.

For GAN to predict the four-momenta of various hadrons (π^0, π^+, π^-, K^+, K^-), their types (as well as the types of lepton) are fed to the input of the GAN as additional parameters along with the initial energy E_0, defined as $E_0 \approx \sqrt{s_{lN}}/2$, where $\sqrt{s_{lN}}$ is the initial energy in the lepton–proton center of mass frame [4].

Since it is currently impossible to experimentally obtain the characteristics of final-state leptons and hadrons (due to the lack of experiments), the finite states of leptons and hadrons were obtained using the PYTHIA8 software package [7].

For each type of lepton (e^+, e^-, μ^+, μ^-) and hadron (π^0, π^+, π^-, K^+, K^-), 100,000 events were generated at initial energies $\sqrt{s_{lN}}$ = 20, 40, 60, 80 and 100 GeV. The four-momentum values of the final-state lepton and hadron were obtained from each event (real data).

Following the approach in [4], we solved the problems associated with irregularities in the distributions of the quantities E_l, E_h and p_{zl} by generating, instead of the actual quantities E_l, E_h, p_{zl}, the quantities obtained by their transformation (transformed quantities):

$$T(p_{zl}) = \log[(E_0 - p_{zl})/(1\ \text{GeV}/c)],$$

$$T(E_l) = \log[(E_0 - E_l)/(1\ \text{GeV}/c)],$$

$$T(E_h) = \log[(E_h)/(1\ \text{GeV}/c)].$$

As established in [4], the distribution over the transformed quantities becomes smoother, preventing predictions of unphysical values.

Also similarly to [4], the event generator in this study is based on GAN with a least square loss function [8].

The generator consists of 5 layers of 512 neurons each with a Leaky ReLU activation function and a dropout of 0.2 [9]. A 128-dimensional noise vector (a vector of values obtained from a Gaussian distribution with the mean equal to 0 and the variance equal to 1), energy E_0, lepton type and hadron type are fed to the generator input. The generator outputs 8 characteristics:

$$p_{xl}, p_{yl}, T(p_{zl}), T(E_l), p_{xh}, p_{yh}, p_{zh} \text{ and } T(E_h),$$

corresponding to lepton and hadron.

Based on these characteristics, the model calculates additional values used to increase the accuracy of GAN predictions [4]:

$p_{Tl} = \sqrt{p_{xl}^2 + p_{yl}^2}$, $p_{Th} = \sqrt{p_{xh}^2 + p_{yh}^2}$ are the lepton and hadron transverse momenta, respectively;

$\varphi_l = \arctan(p_{zl}/p_{Tl})$, $\varphi_h = \arctan(p_{zh}/p_{Th})$ are the lepton and hadron azimuthal angles, respectively;

$\theta_l = \arctan(p_{yl}/p_{xl})$, $\theta_h = \arctan(p_{yh}/p_{xh})$ are the lepton and hadron polar angles, respectively.

All additional quantities are then fed to the discriminator input during training.

The discriminator also consists of 5 layers of 512 neurons each with a Leaky ReLU activation function and a dropout of 0.2 [9]. A dropout layer with a rate of 10% [11] is applied to each of the layers to prevent overfitting of the discriminator, randomly dropping 10% of the layer weights.

Spectral normalization is additionally applied to all layers for more stable training [12]. The output layer consists of a single neuron with a linear activation function. The higher the value obtained, the more confident the discriminator is in identifying the given values as realistic.

The model was trained for 400 epochs. RMSProp was used for gradient descent optimization, with $\rho = 0.9$ [13], $1 \cdot 10^{-4}$ training steps for the generator and $5 \cdot 10^{-5}$ for the discriminator. Using different training steps contributes to better training convergence, as shown in [14].

The Kullback–Leibler (KL) divergence was used as a measure of the divergence between the real data and those generated by GAN [15]. This measure was used to compare the histograms of the obtained distributions. In this case, the Kullback–Leibler divergence D_{KL} is defined as follows [15]:

$$D_{KL}(P \| Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i},$$

where P, Q are the distributions of the real and generated data, respectively; p_i, q_i are the probabilities of the ith bins of histograms for real and generated data; n is the number of bins.

Simulation results

Since there is a wide range of scattering scenarios (different types of leptons and hadrons as well as different initial energies E_0), only some of the individual cases are given below to illustrate GAN's predictive capabilities.

Fig. 1 shows the distributions of p_T, θ, φ for the positron e^+ and the negative kaon K^-, obtained using GAN and PYTHIA8. Multiplicity is understood (in Fig. 1 and below) as the number of events in the bin normalized by the total number of events. Evidently, the model generates quantities whose distributions are almost identical, as indicated by the values of the Kullback–Leibler divergence shown in the graphs as well as the logarithmic ratios of GAN to PYTHIA8 predictions given for each graph.

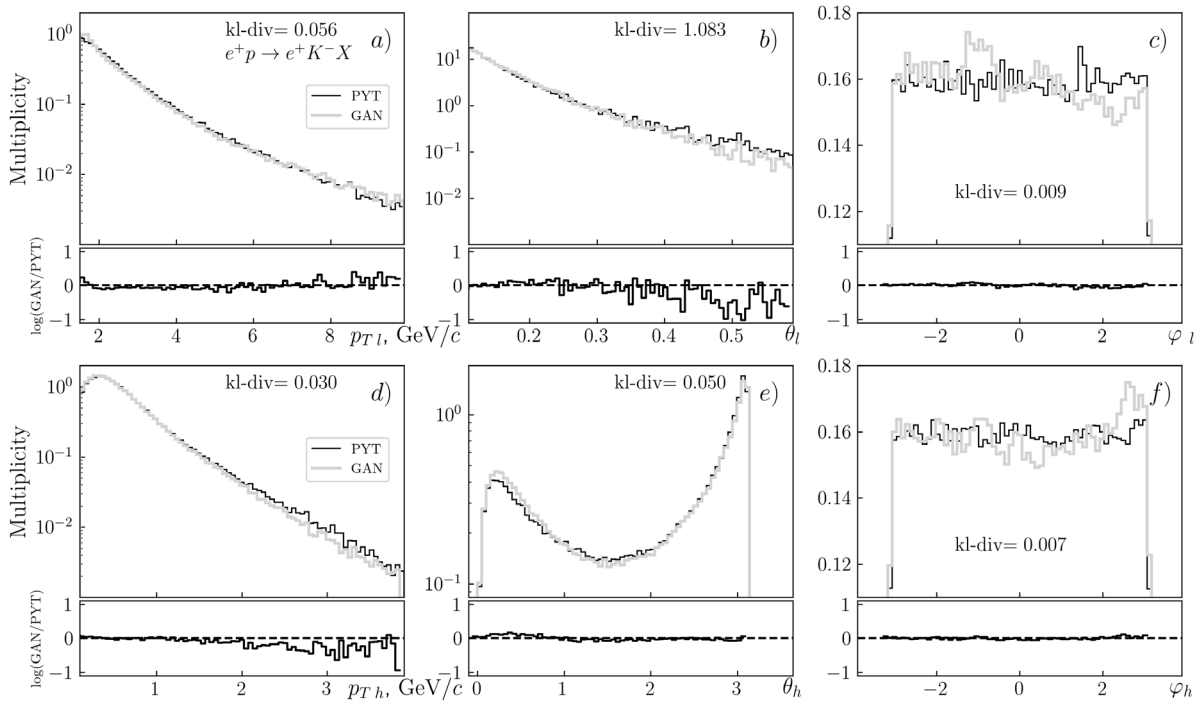

Fig. 1. Distributions of quantities p_T, θ, φ for positrons e^+ (*a*, *b*, *c*) and negative kaons K^- (*d*, *e*, *f*) at initial energy $E_0 = 50$ GeV.
The data were obtained using GAN (gray curves) and PYTHIA8 (black).
The corresponding values of KL divergence (kl-div) and graphs of the logarithmic ratio of GAN to PYTHIA8 (GAN/PYT) predictions are given for each distribution.

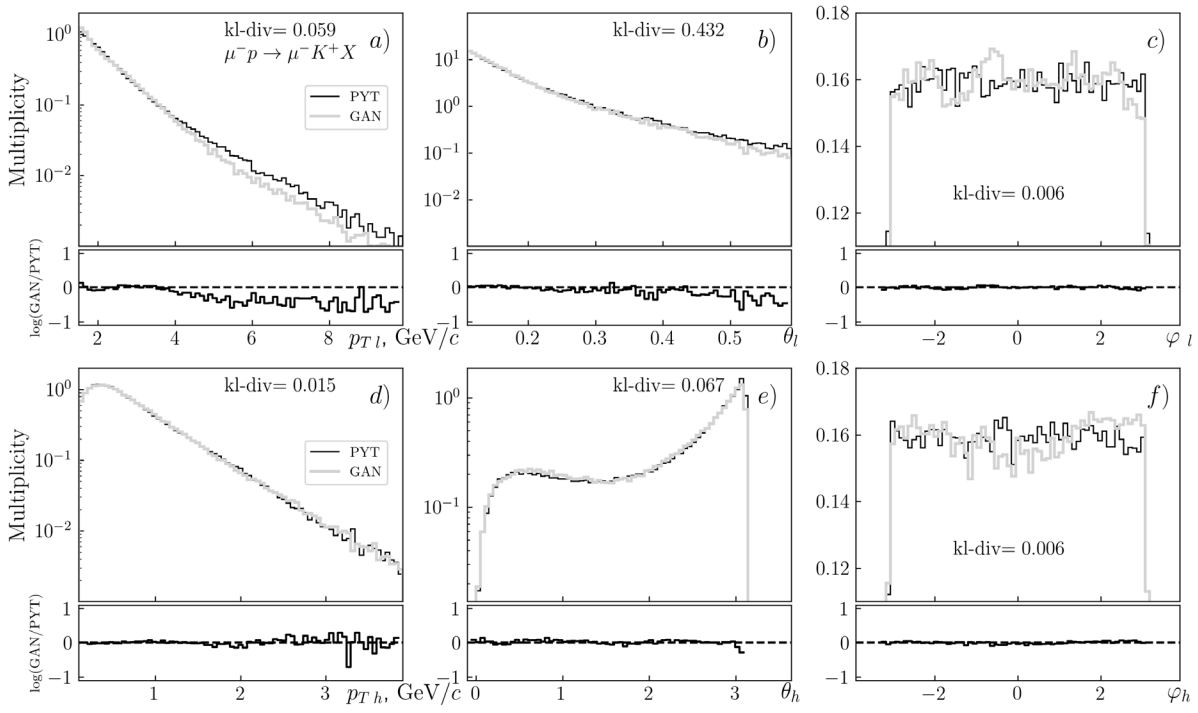

Fig. 2. Graphs similar to those shown in Fig. 1, but for muons μ^- (*a*, *b*, *c*) and positive kaons K^+ (*d*, *e*, *f*) at the initial energy $E_0 = 20$ GeV

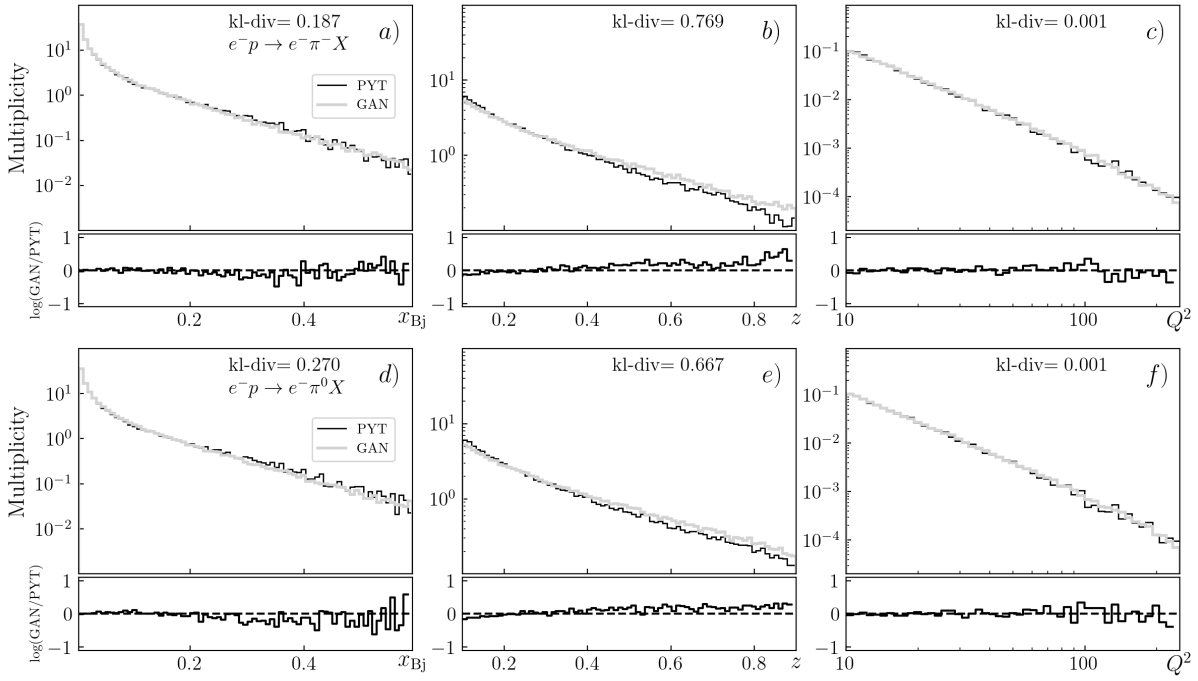

Fig. 3. Distributions of quantities x_{Bj}, z, Q^2 for the reactions $e^- p \to e^- \pi^- X$ (*a*, *b*, *c*) and $e^- p \to e^- \pi^0 X$ (*d*, *e*, *f*), respectively, at initial energy $E_0 = 40$ GeV.
The corresponding values of KL divergence (kl-div) and graphs of the logarithmic ratio of GAN to PYTHIA8 (GAN/PYT) predictions are given for each distribution.

Fig. 2 shows the distributions of the quantities p_T, θ, φ for the muon μ^- and the positive kaon K^+, obtained using GCC and PYTHIA8. These data demonstrate that the model can operate just as accurately with different leptons and hadrons at different initial energies.

Fig. 3 shows the distributions of squared momentum transfer $Q^2 = -q^2$ (q is the momentum of the virtual photon), as well as the Bjorken variable $x_{Bj} = Q^2/2Pq$ (P is the momentum of the incident proton) and the fraction of the energy of the virtual photon transferred to the hadron, $z = P{\cdot}P_h/P{\cdot}q$ (P_h is the momentum of the proton) for nuclear reactions $e^-p \to e^-\pi^-X$ and $e^-p \to e^-\pi^0X$, where X denotes all other reaction products.

It follows from the presented results that the distributions generated by the model only differ slightly, as indicated by the values of the KL divergence obtained for each distribution.

Fig. 4 shows the distributions of the quantities x_{Bj}, z, Q^2 for the reactions $e^+p \to e^+\pi^+X$ and $e^+p \to e^+K^-X$. Analyzing the obtained data, we can conclude that the accuracy of GAN predictions is preserved relative to real data from PYTHIA8 for different types of leptons and hadrons and different initial energies.

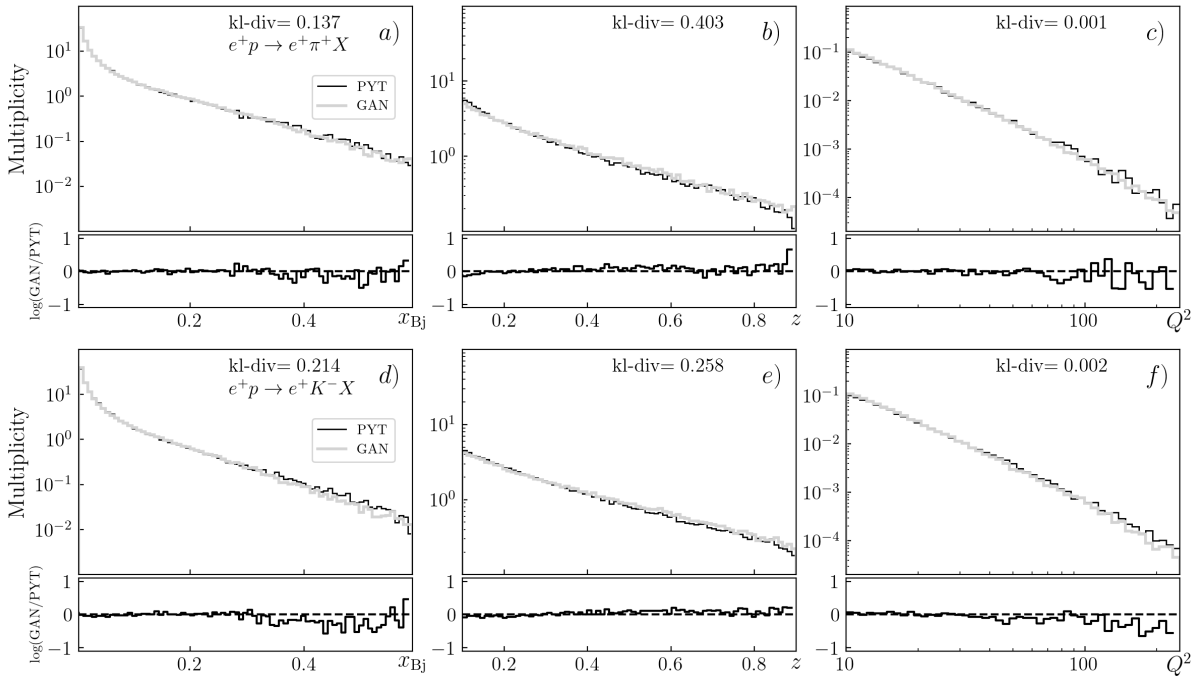

Fig. 4. Distributions of quantities x_{Bj}, z, Q^2 for reactions $e^-p \to e^-\pi^-X$ (*a*, *b*, *c*) and $e^-p \to e^-\pi^0X$ (*d*, *e*, *f*) at initial energy $E_0 = 30$ GeV
The corresponding values of KL divergence (kl-div) and graphs of the logarithmic ratio of GAN to PYTHIA8 (GAN/PYT) predictions are given for each distribution.

Conclusion

We developed a generative-adversarial network model that can predict the characteristics of final-state leptons (e^+, e^-, μ^+, μ^-) and hadrons (π^0, π^+, π^-, K^+, K^-) in semi-exclusive deep inelastic lepton–proton scattering in the initial energy range of 20–100 GeV.

We established that the above-mentioned GAN model is capable of faithfully reproducing four-momentum components of final-state leptons and hadrons.

It was confirmed that the model constructed can calculate the distributions for particles with high accuracy based on the transverse momentum p_T of the particles, the azimuthal (φ) and polar (θ) angles, the Bjerken variable x_{Bj}, the energy fractions $z$of the virtual photon and the square momentum Q^2 transferred by the lepton to the hadron. The distributions of these quantities show high accuracy relative to the real data, proving that the model is capable of preserving the internal relationships between the values.

We also established that the GAN model accurately predicts the characteristics of leptons and hadrons both for the initial energies at which the model was trained and for the interpolated energies (intermediate values).

REFERENCES

1. **Cremonesi M., Bellini C., Bian B., et al.,** Using big data technologies for HEP analysis. arXiv: 1901.07143, 2019. https://doi.org/10.48550/arXiv.1901. 07143

2. **Jordan M. I., Mitchell T. M.,** Machine learning: Trends, perspectives, and prospects, Science. 349 (6245) (2015) 255–260.

3. **Weinzierl S.,** Introduction to Monte Carlo methods. arXiv: 0006269, 2000. https://doi.org/10.48550/arXiv.hep-ph/0006269.

4. **Lobanov A. A., Berdnikov Ya. A.,** A generator of deep inelastic lepton-proton scattering based on the Generative-Adversarial Network (GAN), St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) ...–.... (in Russian).

5. **Aschenauer E. C., Borsa I., Sassot R., Van Hulse C.,** Semi-inclusive deep-inelastic scattering, parton distributions, and fragmentation functions at a future electron-ion collider, Phys. Rev. D. 99 (9) (2019) 094004.

6. **Barone V., Boglione M., Hernandez J. O. G., Melis S.,** Phenomenological analysis of azimuthal asymmetries in unpolarized semi-inclusive deep inelastic scattering, Phys. Rev. D. 91 (7) (2015) 074019.

7. **Sjöstrand T., Mrenna S., Skands P.,** A brief introduction to PYTHIA 8.1, Comput. Phys. Commun. 178 (11) (2008) 852–867.

8. **Mao X., Li Q., Xie H., et al.,** On the effectiveness of least squares Generative Adversarial Networks, IEEE Trans. Pattern Anal. Mach. Intell. 41 (12) (2019) 2947–2960.

9. **Sharma O.,** A new activation function for deep neural network, Proc. Int. Conf. Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), IEEE, Faridabad, India, Febr. 14–16 (2019) 84–86.

10. **Hawkins D. M.,** The problem of overfitting, J. Chem. Inf. Comput. Sci. 44 (1) (2003) 1–12.

11. **Srivastava N., Hinton G., Krizhevsky A., et al.,** Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.

12. **Miyato T., Kataoka T., Koyama M., Yoshida Y.,** Spectral normalization for Generative Adversarial Networks, arXiv: 1802.05957/v1, 2018. https://doi.org/10.48550/arXiv.1802.05957.

13. **Xu D., Zhang Sh., Zhang H., Mandic D. P.,** Convergence of the RMSProp deep learning method with penalty for nonconvex optimization, Neural Netw. 139 (July) (2021) 17–23.

14. **Heusel M., Ramsauer H., Unterthiner T., et al.,** GANs trained by a two time-scale update rule converge to a local Nash equilibrium, arXiv: 1706.08500v6, 2017. https://doi.org/10.48550/arXiv.1706. 08500.

15. **Shlens J.,** Notes on Kullback–Leibler divergence and likelihood, arXiv: 1404.2000, 2014. https://doi.org/10.48550/arXiv.1404.2000

СПИСОК ЛИТЕРАТУРЫ

1. **Cremonesi M., Bellini C., Bian B., et al.** Using big data technologies for HEP analysis. arXiv: 1901.07143, 2019. https://doi.org/10.48550/ arXiv. 1901. 07143

2. **Jordan M. I., Mitchell T. M.** Machine learning: Trends, perspectives, and prospects // Science. 2015. Vol. 349. No. 6245. Pp. 255–260.

3. **Weinzierl S.** Introduction to Monte Carlo methods. arXiv: 0006269, 2000. https://doi.org/10.48550/ arXiv.hep-ph/0006269.

4. **Лобанов А. А., Бердников Я. А.** Генератор глубоко неупругого рассеяния лептонов на протоне на основе генеративно-состязательной нейронной сети // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. ...–....

5. **Aschenauer E. C., Borsa I., Sassot R., Van Hulse C.** Semi-inclusive deep-inelastic scattering, parton distributions, and fragmentation functions at a future electron-ion collider // Physical Review D. 2019. Vol. 99. No. 9. P. 094004.

6. **Barone V., Boglione M., Hernandez J. O. G., Melis S.** Phenomenological analysis of azimuthal asymmetries in unpolarized semi-inclusive deep inelastic scattering // Physical Review D. 2015. Vol. 91. No. 7. P. 074019.

7. **Sjöstrand T., Mrenna S., Skands P.** A brief introduction to PYTHIA 8.1 // Computer Physics Communications. 2008. Vol. 178. No. 11. Pp. 852–867.

8. **Mao X., Li Q., Xie H., Lau R. Y. K., Wang Zh., Smolley S. P.** On the effectiveness of least squares generative adversarial networks // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019. Vol. 41. No. 12. Pp. 2947–2960.

9. **Sharma O.** A new activation function for deep neural network // Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE, Faridabad, India, February 14–16, 2019. Pp. 84–86.

10. **Hawkins D. M.** The problem of overfitting // Journal of Chemical Information and Computer Sciences. 2003. Vol. 44. No. 1. Pp. 1–12.

11. **Srivastava N., Hinton G., Krizhevsky A., Slutskever I., Salakhutdinov R.** Dropout: A simple way to prevent neural networks from overfitting // The Journal of Machine Learning Research. 2014. Vol. 15. Pp. 1929–1958.

12. **Miyato T., Kataoka T., Koyama M., Yoshida Y.** Spectral normalization for Generative Adversarial Networks. arXiv: 1802.05957/v1, 2018. https://doi.org/10.48550/arXiv.1802.05957.

13. **Xu D., Zhang Sh., Zhang H., Mandic D. P.** Convergence of the RMSProp deep learning method with penalty for nonconvex optimization // Neural Networks. 2021. Vol. 139. July. Pp. 17–23.

14. **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. arXiv: 1706.08500v6, 2017. https://doi. org/10.48550/arXiv.1706. 08500.

15. **Shlens J.** Notes on kullback-leibler divergence and likelihood. arXiv: 1404.2000, 2014. https:// doi.org/10.48550/arXiv.1404.2000.

THE AUTHORS

LOBANOV Andrey A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
lobanov2.aa@edu.spbstu.ru
ORCID: 0000-0002-8910-4775

BERDNIKOV Yaroslav A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
berdnikov@spbstu.ru
ORCID: 0000-0003-0309-5917

СВЕДЕНИЯ ОБ АВТОРАХ

ЛОБАНОВ Андрей Александрович – *студент* Физико-механического института *Санкт-Петербургского политехнического университета Петра Великого*.
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
lobanov2.aa@edu.spbstu.ru
ORCID: 0000-0002-8910-4775

БЕРДНИКОВ Ярослав Александрович – *доктор физико-математических наук, профессор* Высшей школы фундаментальных физических исследований *Санкт-Петербургского политехнического университета Петра Великого*.
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
berdnikov@spbstu.ru
ORCID: 0000-0003-0309-5917

Received 28.09.2023. Approved after reviewing 12.10.2023. Accepted 12.10.2023.
Статья поступила в редакцию 28.09.2023. Одобрена после рецензирования 12.10.2023. Принята 12.10.2023.

RADIOPHYSICS

Original article
DOI: https://doi.org/10.18721/JPM.16416

SPUN FIBERS AND THEIR DESCRIPTION WITHIN THE JONES FORMALISM IN ANALYZING THE PRACTICAL FIBER-OPTIC CIRCUITS

V. S. Temkina✉, *L. B. Liokumovich*, *A. B. Archelkov*, *A. V. Medvedev*, *A. S. Kozlov*, *K. V. Greshnevikov*

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

✉ temkina_vs@spbstu.ru

Abstract. In this paper, an analytical form for the Jones matrix of a real spun fiber has been obtained, taking into account a slight deviation of its properties from an idealized representation of this fiber by the rotation matrix. The derivation was made within the framework of the optical element model with phase anisotropy. The features of using the Jones matrix of a real spun fiber in analysis of practical fiber-optic circuits and modeling their signals were considered. The experiments with the spun fiber revealing the parameter deviations of the polarization modes of the real spun fiber from the idealized model and allowing estimation of this deviation level were performed.

Keywords: Jones matrix formalism, spun fiber, phase anisotropy, polarization state of light

Funding: The reported study was funded by Russian Science Foundation Grant No. 22-19-00513 (https://rscf.ru/en/project/22-19-00513/)

Citation: Temkina V. S., Liokumovich L. B., Archelkov A. B., Medvedev A. V., Kozlov A. S., Greshnevikov K. V., Spun fibers and their description within the Jones formalism in analyzing the practical fiber-optic circuits, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 198–214. DOI: https://doi.org/10.18721/JPM.16416

Научная статья
УДК 535.5, 535-4, 535.012.2
DOI: https://doi.org/10.18721/JPM.16416

ВОЛОКОННЫЕ СВЕТОВОДЫ SPUN-ТИПА И ИХ ОПИСАНИЕ В РАМКАХ ФОРМАЛИЗМА МАТРИЦ ДЖОНСА ПРИ АНАЛИЗЕ ПРАКТИЧЕСКИХ ОПТОВОЛОКОННЫХ СХЕМ

В. С. Темкина✉, *Л. Б. Лиокумович*, *А. Б. Арчелков*, *А. В. Медведев*, *А. С. Козлов*, *К. В. Грешневиков*

Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Россия

✉ temkina_vs@spbstu.ru

Аннотация. В работе получено выражение для матрицы Джонса реального волоконного световода spun-типа, которое учитывает слабое отклонение его свойств от таковых для случая идеализированного представления этого волокна матрицей поворота. Вывод проведен в рамках модели оптического элемента с фазовой анизотропией. Рассмотрены особенности использования полученной матрицы Джонса реального spun-волокна при анализе практических оптоволоконных схем и моделировании их сигналов. Выполнены эксперименты со spun-волокном, демонстрирующие отклонение параметров поляризационных мод реального волокна от идеализированной модели и позволившие оценить уровень этого отклонения.

Ключевые слова: формализм матриц Джонса, spun-волокно, фазовая анизотропия, состояние поляризации света

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 00513-19-22 (https://rscf.ru/project/22-19-00513/).

Ссылка для цитирования: Темкина В. С., Лиокумович Л. Б., Арчелков А. Б., Медведев А. В., Козлов А. С., Грешневиков К. В. Волоконные световоды spun-типа и их описание в рамках формализма матриц Джонса при анализе практических оптоволоконных схем // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. 198–214. DOI: https://doi.org/10.18721/JPM.16416

Introduction

The improvement of fiber-optic technologies has contributed to the active development of various types of specialized optical fibers. One of the directions in this field is the development of a unique class of spun-type fibers, which have a particular internal anisotropy structure. Such fibers have the same internal structure as fibers with linear anisotropy, however, upon shifting along the longitudinal axis of the fiber, the direction of the polarization axes undergoes regular rotation. This is achieved by twisting a preform with a birefringent structure (polarization-maintaining fiber) during fiber drawing.

The type of intrinsic polarization modes of such a fiber depends on the ratio of two key parameters in the resulting fiber structure. The first parameter, V_L, rad/m, is the increment of the phase difference of the linear polarization modes of the local fiber segment, which characterizes the linear anisotropy caused by the transverse deformation of the core induced during manufacture. The second one, V_α, rad/m, is the linear velocity of the longitudinal rotation of the direction of the polarization axes.

Depending on the achieved ratio V_α/V_L, the eigenmodes of the spun fiber may have a different character, but it is important to note that with an increase in the value of V_α/V_L, the proper modes of the spun fiber tend to orthogonal circular polarizations [1].

Two types of spun fibers are known: with low (LoBi) and high (HiBi) birefringence.

The first type is characterized by the fact that a high value of the ratio V_α/V_L is due to a low value of V_L. These fibers are made from preforms without a PM structure twisted under drawing and have polarization eigenmodes with a fairly small phase difference [2–4]. With limited length and small bends, they function as isotropic optical fiber that preserves the polarization state of incident radiation. Spun-LoBi fibers are used, for example, to amplify light in high-power fiber lasers to overcome the drift of the polarization state of light in the fiber due to its heating during pumping [5, 6]. However, these fibers are significantly affected by induced anisotropy during bends and other external perturbations of the fiber.

On the contrary, the second type, spun-HiBi fiber, has a relatively high value of V_L and is made from preforms with a significant transverse PM structure by rapidly rotating the workpiece during fiber drawing [7]. Due to the relatively high V_α/V_L ratio, these fibers also have polarization eigenmodes close to circularly polarized modes [8], and the intrinsic anisotropy of such fibers is slightly distorted by bending, compression and other impacts. The second type of fiber is used for various purposes, for example, to create sensitive elements of high-precision fiber-optic current sensors [9–12]. Such an application is among the most promising, widely known and researched.

The interest in the use of spun fibers, especially with high birefringence, is caused by their potential to preserve circularly polarized modes. However, such polarization eigenmodes correspond only to the limiting case with an increase in the value of V_α/V_L. In practice, this ratio is limited, therefore for a number of reasons, even without taking into account the internal fluctuations of the structure that arise during manufacture and induced during fiber placement, the polarization eigenmodes of real spun fibers only approach circular modes and may differ markedly from the idealized case.

Many studies were dedicated to the analysis of the polarization properties of real spun fibers [1, 4, 8, 10, 12–15]. However, these studies are generally aimed at analyzing the complex mechanisms of regular and random transformation of light polarization during its propagation in inhomogeneous anisotropic fiber structure and are based on the application of the mode coupling formalism and the equations of coupled waves [4, 10, 12, 13]. Models of the formation of the Jones matrix of the spun fiber have also been considered in a number of works, both for the differential matrices of the segment and the resulting integral matrix [1, 8, 14, 15]. Such models have a complex structure in the form of a product of matrices, and they must take into account (even when reduced to an integral matrix) rigorous values of V_α, V_L and fiber lengths, which are usually unknown. In addition, such models do not allow to take into account the influence of possible fluctuations in parameters and anisotropy induced by external perturbations of the fiber for the Jones matrix of spun fiber. Therefore, although the results of such studies describe the properties of spun fibers, they are difficult to apply to the analysis and modeling of practical devices based on these fibers.

The goal of this study is to obtain the structure of the Jones matrix of a real spun fiber in the simplest possible integral form without using the parameters of the internal structure of the fiber, based only on the condition of a small difference in the polarization modes of such a fiber from an idealized representation, and to analyze the properties of the resulting matrix.

It is this case of the Jones matrix of spun fibers that is effective for analyzing and modeling devices based on these fibers; in addition, it is very useful to study the effect of imperfections (differences between real spun fibers and idealized ones) on the operation of these devices.

Jones matrix of idealized spun fiber

First of all, it should be borne in mind that there are different options in the literature for determining a polarized wave with a right or left direction of rotation as well as different options for taking into account the phases of components in Jones vectors and, as a result, in Jones matrices. It is important to understand these features in further analysis, so the Appendix (given at the end of the paper) contains the refinements we adopted.

As noted above, the polarization modes in the idealized representation of the spun fiber are considered to be circular. A device with circular eigenvectors in the linear Cartesian basis of Jones vectors is described by a rotation matrix. Therefore, we assume that the Jones matrix of idealized spun fiber has the form

$$\mathbf{M}_0 = \begin{bmatrix} \cos(\varphi/2) & \sin(\varphi/2) \\ -\sin(\varphi/2) & \cos(\varphi/2) \end{bmatrix}. \quad (1)$$

It is important to emphasize that the concept of idealized spun fiber described by a rotation matrix is not related to the idea of spun fiber with an ideal structure, in which regular rotation of the direction of the axes of linear anisotropy is introduced. In such a structure, even with regular parameters without fluctuations, the form of the matrix will differ from the presented form (1). The idealized spun fiber described by the rotation matrix implies precisely the idealized concept of converting the state of polarization of light in such fiber, when it is preferable to have an element with circularly polarized eigenmodes in the optical circuit.

Under the refining conditions described in the Appendix, matrix (1) rotates the azimuth of the polarization state by an angle $\varphi/2$ clockwise when observed towards the direction of wave propagation. At the same time, the phase difference of the eigenmodes φ is taken into account here, but the general phase shift Φ of the eigenwaves is not included, which is not difficult to take into account by introducing the factor $e^{-j\Phi}$, although this factor is not needed to consider only the transformation of the polarization state of light. From this point of view, matrix (1) belongs to the class of special unitary matrices with a determinant equal to unity.

The eigenvectors $\mathbf{J}_{01}$ and $\mathbf{J}_{02}$ of matrix (1), corresponding to the eigenvalues $\lambda_1 = e^{j\varphi/2}$ and $\lambda_2 = e^{-j\varphi/2}$ (φ here and below is assumed to be positive), correspond to waves with right and left circular polarization, which are generally written as follows [17 ,16]:

$$\mathbf{J}_{01} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ j \end{bmatrix}, \mathbf{J}_{02} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -j \end{bmatrix}. \quad (2)$$

According to the accepted rules, the vectors $\mathbf{J}_{01}$ and $\mathbf{J}_{02}$ for matrix (1) refer respectively to the fast and slow polarization modes of the idealized spun fiber.

It is important to note here that an alternative case of idealized spun fiber can be formulated, which rotates the polarization state of the light passing through the fiber counterclockwise. In practice, this is set by the direction of rotation of the preform during drawing. Such a case of idealized fiber will be described by the matrix $\mathbf{M}_0' = \mathbf{M}_0^T$, which also has eigenvectors (2), but the first will correspond to the slow mode, and the second to the fast one. In general, if this case needs to be considered, then all the expressions listed below can be used by replacing φ with $-\varphi$ (again, it is assumed that φ is positive).

Jones matrix of real spun fiber

Real spun fibers do not correspond to an idealized representation and are described by a Jones matrix different from the rotation matrix. At the same time, there is a fundamental difference from fibers with linear anisotropy (i.e., polarization-maintaining (PM) fibers), where the imperfection of the fiber is associated with fluctuations in the magnitude and direction of core deformations that arose during the manufacture of the fiber or induced by subsequent external perturbations.

Spun fiber with regular rotation of the orientation of the polarization axes differs from the idealized representation discussed above, even without fluctuations in the anisotropy parameters, since circular polarization eigenmodes are achieved only in the case of a limiting value of the ratio of fiber parameters, which cannot be done in practice. The intrinsic and induced fluctuations of the anisotropy parameters additionally distort the final polarization properties of the fiber, but are not the main reason for the deviation from the idealized representation.

In order to formulate a relatively simple representation of the Jones integral matrix for a segment of real spun fiber, we propose to use only the condition of a slight deviation in the polarization properties of such fiber from the idealized representation, i.e., a slight deviation in the polarization eigenstates from circular ones.

We assume that the spun fiber remains an element with phase anisotropy and is described by a unitary Jones matrix. This circumstance can be explained by the low loss of optical power in fibers of relatively short length (in practice, spun fibers up to several tens of meters in length are commonly used), which makes it possible to neglect the possible dichroism. We also do not take into account the total phase shift Φ of eigenmodes, which means we will consider a special unitary matrix.

In view of the above, the Jones matrix of real spun fiber should correspond to the matrix of an elliptical phase plate, whose eigenvectors are close to vectors (2) for circular polarizations.

Due to the importance of the properties of the eigenvectors of the optical element matrix, first consider the properties of the eigenvectors corresponding to the condition of proximity to the vectors (2). In the general case, in terms of the basic parameters of the polarization ellipse (ellipticity angle ε and azimuth Θ), two orthogonal Jones eigenvectors in the Cartesian basis are generally written as follows [16, 18]:

$$\mathbf{J}_1=\begin{bmatrix}\cos\Theta\cos\varepsilon-j\sin\Theta\sin\varepsilon\\ \sin\Theta\cos\varepsilon+j\cos\Theta\sin\varepsilon\end{bmatrix},\ \mathbf{J}_2=\begin{bmatrix}-\sin\Theta\cos\varepsilon+j\cos\Theta\sin\varepsilon\\ \cos\Theta\cos\varepsilon+j\sin\Theta\sin\varepsilon\end{bmatrix}, \tag{3}$$

where the parameters ε and Θ are set directly for vector $\mathbf{J}_1$, and vector $\mathbf{J}_2$ is obtained as orthogonal to $\mathbf{J}_1$.

Form (3) defines normalized vectors with unit length, and in general, orthogonal vectors are written to a constant complex factor, i.e., they can have different values of both length and initial phase.

It is useful to consider the transition to the idealized case with circular polarizations, for which it is necessary to take the value $\varepsilon = \pi/4$. Integrals (13) are then transformed to the form

$$\mathbf{J}_1=\frac{e^{-j\Theta}}{\sqrt{2}}\begin{bmatrix}1\\ j\end{bmatrix},\ \mathbf{J}_2=\frac{je^{j\Theta}}{\sqrt{2}}\begin{bmatrix}1\\ -j\end{bmatrix}. \tag{4}$$

The difference between the obtained expressions (4) and form (2) lies only in the factors depending on the azimuth; the latter give some additional arguments for Jones complex vectors. Such factors do not affect the shape of the polarization ellipses, which in this case are degenerated into a circle and formally do not have a definite azimuth. An additional factor does not change the unit length of the vector, but, strictly speaking, it still has meaning, since it determines the initial position of the end of the electric field strength vector of the wave on a circular hodograph. Thus, it is evident from the analysis of expressions (4), that the common representation of circular polarization vectors (2) formally corresponds to $\varepsilon = \pi/4$ and $\Theta = 0$.

Let us assume that the polarization eigenstate differs little from circular polarization; this is characterized by an elliptical angle $\varepsilon = \pi/4 - \delta$, where the deviation δ is assumed to be small ($\delta << 1$). Then, approximations for trigonometric functions can be applied in the general form of Jones vectors (3) and approximate equalities can be used preserving the components of only first-order smallness:

$$\sin\left(\frac{\pi}{4}-\delta\right)\approx\frac{1}{\sqrt{2}}(1-\delta),\ \ \cos\left(\frac{\pi}{4}-\delta\right)\approx\frac{1}{\sqrt{2}}(1+\delta). \tag{5}$$

If we substitute expressions (5) into form (3) and apply the known trigonometric transformations, we obtain the eigenvectors of the matrix of imperfect spun fiber in the following form:

$$\mathbf{J}_1=\frac{e^{-j\Theta}}{\sqrt{2}}\begin{bmatrix}1+\delta\cdot e^{j2\Theta}\\ j\left(1-\delta\cdot e^{j2\Theta}\right)\end{bmatrix},\ \mathbf{J}_2=\frac{e^{j\Theta}}{\sqrt{2}}\begin{bmatrix}j\left(1-\delta\cdot e^{-j2\Theta}\right)\\ 1+\delta\cdot e^{-j2\Theta}\end{bmatrix}. \tag{6}$$

The polarization states described by vectors (6), taking into account the smallness of δ, are elliptical, although close to circular. Here, Θ has a clear meaning, the direction of the semi-major axis of the polarization ellipse and can have an arbitrary value in the full range of azimuth variation [0; π]. The considered polarization eigenstates of the spun fiber are shown in Fig. 1, illustrating the deviation of the polarization state of the vector $\mathbf{J}_1$ from the idealized representation (from point *A* to some point *B*) on the Poincaré sphere. The vector $\mathbf{J}_2$ corresponds to diametrically opposite points of the sphere.

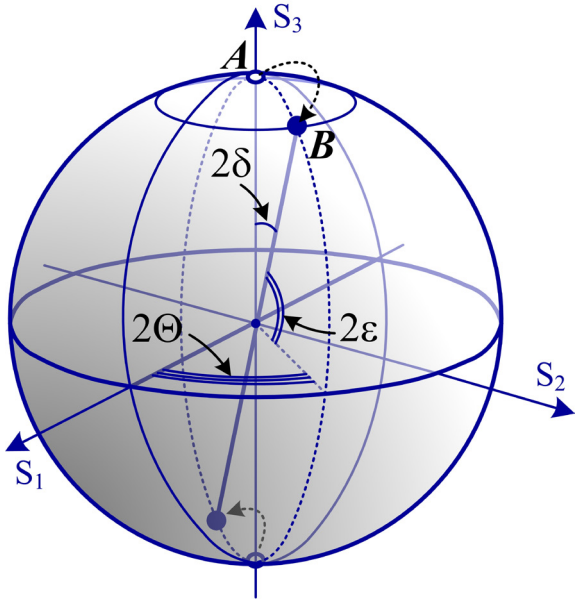

Fig. 1. Displacement of position of polarization eigenmode of spun fiber on the Poincaré sphere, taking into account its real characteristics: point *A* corresponds to polarization with the Jones vector $\mathbf{J}_1$ for the case of idealized spun fiber, point *B* for the case of real spun fiber; Θ, ε are the azimuth and ellipticity angle parameters; 2δ is the angular deviation of the polarization eigenmode from the point of circular polarization

Two approaches can be proposed to obtain the Jones matrix of real spun fiber $\mathbf{M}_{\text{SPUN}}$. The first is to use the general form of the phase anisotropy matrix expressed in terms of the eigenvalues λ_1, λ_2 and eigenvectors [18, 19]:

$$\mathbf{M} = \frac{1}{\mathbf{j}_{1x}\mathbf{j}_{2y} - \mathbf{j}_{1y}\mathbf{j}_{2x}} \begin{bmatrix} \mathbf{j}_{1x}\mathbf{j}_{2y}\lambda_1 - \mathbf{j}_{2x}\mathbf{j}_{1y}\lambda_2 & -(\lambda_1 - \lambda_2)\mathbf{j}_{1x}\mathbf{j}_{2x} \\ (\lambda_1 - \lambda_2)\mathbf{j}_{1y}\mathbf{j}_{2y} & \mathbf{j}_{1x}\mathbf{j}_{2y}\lambda_2 - \mathbf{j}_{2x}\mathbf{j}_{1y}\lambda_1 \end{bmatrix}, \tag{7}$$

where $\mathbf{j}_{1x}$, $\mathbf{j}_{1y}$ are components of the Jones vector $\mathbf{J}_1$; $\mathbf{j}_{2x}$, $\mathbf{j}_{2y}$are components of the vector $\mathbf{J}_2$

If we consider a special unitary matrix that describes a system without losses and has eigenvalues $\lambda_1 = e^{j\varphi/2}$ and $\lambda_2 = e^{-j\varphi/2}$, then we obtain from the general form (7):

$$\mathbf{M} = \frac{1}{\mathbf{j}_{1x}\mathbf{j}_{2y} - \mathbf{j}_{1y}\mathbf{j}_{2x}} \begin{bmatrix} \mathbf{j}_{1x}\mathbf{j}_{2y}e^{j\varphi/2} - \mathbf{j}_{2x}\mathbf{j}_{1y}e^{-j\varphi/2} & -j2\mathbf{j}_{1x}\mathbf{j}_{2x}\sin(\varphi/2) \\ j2\mathbf{j}_{1y}\mathbf{j}_{2y}\sin(\varphi/2) & \mathbf{j}_{1x}\mathbf{j}_{2y}e^{-j\varphi/2} - \mathbf{j}_{2x}\mathbf{j}_{1y}e^{j\varphi/2} \end{bmatrix}. \tag{8}$$

The required matrix of real spun fiber $\mathbf{M}_{\text{SPUN}}$, which has eigenvectors (6), can be obtained by substituting expressions (6) into form (8).

Another way to obtain the required matrix $\mathbf{M}_{\text{SPUN}}$ is to use the expression for the Jones matrix of an arbitrary elliptical phase plate:

$$\mathbf{M}_{\text{EPP}} = \begin{bmatrix} \cos\frac{\varphi}{2} + j\cos 2\Theta\cdot\cos 2\varepsilon\cdot\sin\frac{\varphi}{2} & (\sin 2\varepsilon + j\sin 2\Theta\cdot\cos 2\varepsilon)\sin\frac{\varphi}{2} \\ -(\sin 2\varepsilon - j\sin 2\Theta\cdot\cos 2\varepsilon)\sin\frac{\varphi}{2} & \cos\frac{\varphi}{2} - j\cos 2\Theta\cdot\cos 2\varepsilon\cdot\sin\frac{\varphi}{2} \end{bmatrix}. \tag{9}$$

Expression (9) was obtained in [20] by substituting expressions for orthogonal Jones vectors written in the general form (3) into form (8). To obtain the matrix $\mathbf{M}_{\text{SPUN}}$, it is necessary to take into account $\varepsilon = \pi/4 - \delta$ in matrix (9) and use simplifications (5).

In both cases, the result is a matrix of the form

$$\mathbf{M}_{\text{SPUN}} = \begin{bmatrix} \cos\frac{\varphi}{2} + j2\delta\cos 2\Theta\cdot\sin\frac{\varphi}{2} & \left(1 + j2\delta\sin 2\Theta\right)\cdot\sin\frac{\varphi}{2} \\ -\left(1 - j2\delta\sin 2\Theta\right)\cdot\sin\frac{\varphi}{2} & \cos\frac{\varphi}{2} - j2\delta\cos 2\Theta\cdot\sin\frac{\varphi}{2} \end{bmatrix}. \quad (10)$$

Notably, the determinant Δ of matrix (10) is expressed as

$$\Delta = 1 + 4\delta^2\cdot\sin^2(\varphi/2)$$

and it is real, but not equal to unity.

To obtain a strict correspondence to a normal unitary matrix, a multiplier of $1/\Delta$ can be introduced into expression (10), but in practical calculations it is advisable to neglect the second-order correction for a small parameter δ and use matrix (10) without additional factors.

As mentioned above, for the idealized representation of the fiber, we could choose not the rotation matrix $\mathbf{M}_0$, given by Eq. (1), but the matrix $\mathbf{M}_0' = \mathbf{M}_0^T$ (formally, this can be justified by replacing φ with $-\varphi$), rotating the azimuth of polarization counterclockwise. Both options are equivalent, since they are given by the direction of rotation of the fiber preform during drawing. In this case, the right circular polarization will be the slow mode of the matrix, and the left one will be the fast one.

Expressions for the eigenvectors $\mathbf{J}_1'$ and $\mathbf{J}_2'$ of the matrix $\mathbf{M}'_{\text{SPUN}}$ can be obtained if we assume that $\varepsilon = -\pi/4 + \delta$ in form (3), since the vector $\mathbf{J}_1'$ is close to the left circular polarization. As a result, we obtain the following expressions:

$$\mathbf{J}_1' = \frac{e^{j\Theta}}{\sqrt{2}}\begin{bmatrix} 1 + \delta\cdot e^{-j2\Theta} \\ j\left(\delta\cdot e^{j2\Theta} - 1\right) \end{bmatrix},\ \mathbf{J}_2' = \frac{e^{-j\Theta}}{\sqrt{2}}\begin{bmatrix} j\left(\delta\cdot e^{j2\Theta} - 1\right) \\ 1 + \delta\cdot e^{j2\Theta} \end{bmatrix}, \quad (11)$$

$$\mathbf{M}'_{\text{SPUN}} = \begin{bmatrix} \cos\frac{\varphi}{2} + j2\delta\cos 2\Theta\cdot\sin\frac{\varphi}{2} & -\left(1 - j2\delta\sin 2\Theta\right)\cdot\sin\frac{\varphi}{2} \\ \left(1 + j2\delta\sin 2\Theta\right)\cdot\sin\frac{\varphi}{2} & \cos\frac{\varphi}{2} - j2\delta\cos 2\Theta\cdot\sin\frac{\varphi}{2} \end{bmatrix}. \quad (12)$$

It can be seen from expression (12) that the matrix $\mathbf{M}'_{\text{SPUN}}$, as expected, corresponds to the matrix transposed to $\mathbf{M}_{\text{SPUN}}$, similarly to the matrices $\mathbf{M}_0'$ and $\mathbf{M}_0$ for idealized spun fiber.

Specifics of applying the Jones matrix for real spun fiber in analysis and modeling of fiber optic circuits

The Jones matrix representation of real spun fiber can be used to analyze and model systems containing such fibers. As a rule, such analysis is aimed at clarifying the effect of imperfection of fibers and other polarization elements on the operation of the system as a whole. The models obtained within the framework of the Jones formalism usually contain many parameters characterizing polarization mismatches, which must be varied in analytical or numerical calculations. Therefore, the obtained expression (10), which is a simple explicit form of the Jones matrix of real spun fiber and takes into account the small difference between the polarization eigenmodes of the fiber and their idealized representation using a small parameter δ, is attractive for these calculations.

Matrix (10) contains three parameters: δ, Θ and φ; all of them can affect the transformation of the polarization state when light passes through the spun fiber and, as a result, the formation of signals in the optical circuit. Therefore, when performing analysis or numerical calculations, it is necessary to determine which parameter values to use.

The small parameter δ sets a quantitative measure of the deviation of the real spun fiber from the idealized representation. This deviation can be related both to the limited value of the ratio V_α/V_L, which is provided during the creation of the fiber, and to fluctuations in parameters that occur during the manufacture or placement of the fiber. As a result, the specific value of δ for real fibers can be difficult to predict. The most appropriate approach for analysis is to determine a

certain threshold value of δ_{max} for the fiber in question. Such a value can be obtained by separate theoretical consideration of the specific structure of the spun fiber or determined empirically. Further, the calculations should consider the effect of mismatches by varying the parameter δ in the range from 0 to δ_{max}.

The azimuth of the eigenstates in a certain orientation basis, depending how the spun fiber is connected and using which elements, must be generally considered unknown, uncontrolled and any possible value of the parameter Θ in the range from 0 to π.

The phase difference of the polarization modes φ, formed when light passes through the fiber, also turns out to be a virtually unknown and uncontrolled parameter. Even if the key parameters of spun fiber are known, in the case of sufficiently long (several meters or more) spun fiber with high birefringence, the phase difference of eigenmodes is difficult to calculate or determine precisely; taking into account fluctuations of the parameters and possible significant changes of temperature, the value of φ may actually be random in the range from 0 to 2π. Therefore, in the analysis and calculations, it should also be varied in the specified range.

As for the phase difference φ, we should make one more important remark. We noted above that spun-HiBi fibers are most commonly used in fiber-optic sensors as sensitive elements. The most common example of using such fibers are fiber-optic current sensors, where it is assumed that due to the Faraday effect, the phase difference in spun fiber wound around a current-carrying conductor changes between two circularly polarized orthogonal modes. Thus, when analyzing such schemes, it should be borne in mind that the phase difference φ must also contain a component induced by the measured effect. In this case, the non-reciprocal anisotropy induced by the measured magnetic field as a consequence of the Faraday effect is circular. If the spun fiber corresponded to the idealized representation and was described by the matrix $\mathbf{M}_0$ (or $\mathbf{M}_0'$), then, obviously, during modeling, the phase difference φ should be given as

$$\varphi = \varphi_0 + \varphi(t),$$

where φ_0 is the quasi-stationary component of the phase difference between circularly polarized modes in the fiber.

The value of φ_0, as indicated above, can actually be any in the range of $0-2\pi$. But since the real spun fiber differs a priori from the idealized representation and the eigenmodes of such fiber are not strictly circular, setting the phase difference φ in the form of the above sum will be approximate. Such an approximation may be quite acceptable in practice in the analytical study and numerical modeling of signals in measuring circuits with spun fiber.

Experimental

To analyze schemes with spun fiber based on the obtained form of the Jones matrix, it is necessary to estimate the possible range of values of the main parameter characterizing the deviation of the fiber from the idealized representation, the parameter δ. Such an estimate can be made both based on additional studies of fiber anisotropy factors and experimentally. The following are the results of experiments that allow us to estimate the parameter δ for specific spun fiber and illustrate the analysis presented above.

For measurements, we used the fact that if two polarization modes are excited during propagation through an element with phase anisotropy (for example, through anisotropic optical fiber), then when the phase difference φ of the modes changes by 2π, the evolution of the polarization state at the output from the element on the Poincaré sphere forms a circle [17, 19]. The change in φ leads to the rotation of the sphere around the axis, which is set by the points of the polarization eigenstates, and the angular radius R of the circle is determined by the ratio of the amplitudes of the polarization modes. Therefore, the experimental formation and detection of such an evolution as well as its subsequent analysis with the determination of the parameters Θ_0 and ε_0 of the center of the small circle of the sphere allow to measure the polarization eigenstates of the element. Fig. 2, *a*, *b* illustrates this approach and provides a diagram of the experimental setup for its implementation.

The key issue determining the possibility of the correct implementation of this approach to measuring the fiber polarization eigenstates is the method of organizing changes in the phase difference φ. We used fiber heating for this purpose. Unlike other measures such as longitudinal tension that change the optical length of the fiber, heating has a smaller effect on the inner structure of the fiber determining its anisotropy. In addition, this method can be used with relatively long fibers.

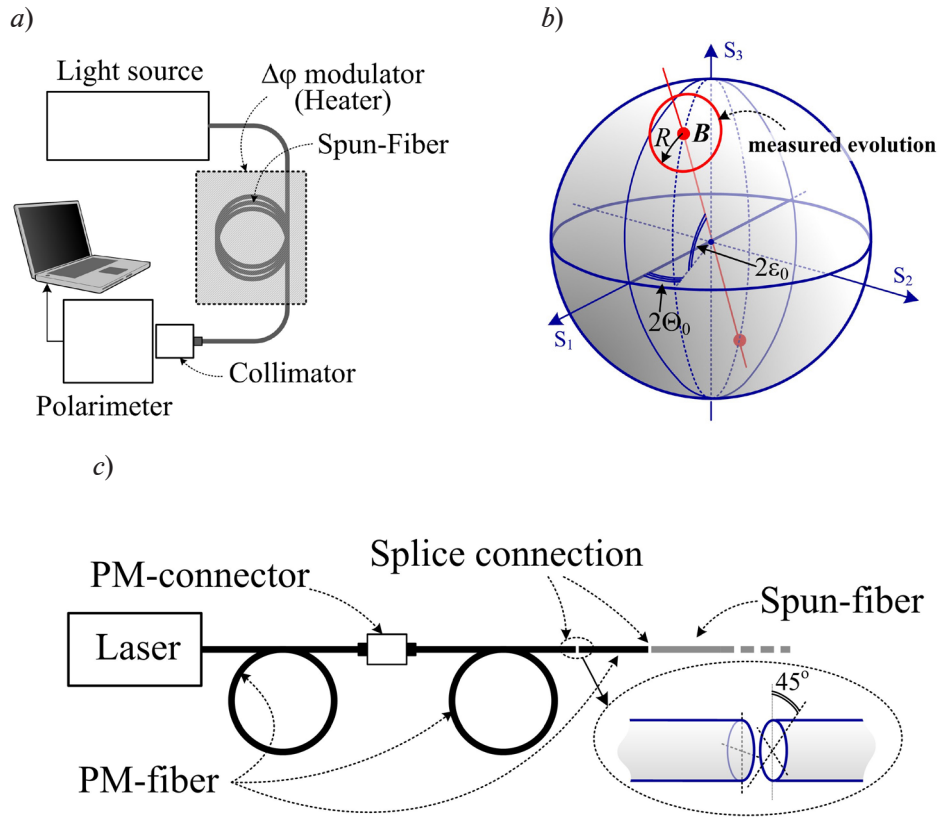

Fig. 2. Schematic of the experiment (*a*) with inset (*c*) illustrating the passage of light to the input into the tested spun fiber as well as the evolution of the polarization state on the Poincaré sphere, recorded during measurements (*b*)

However, the measurement approach used has its own specifics.

Firstly, heating the fiber can still lead not only to a change in the phase difference φ. A change in temperature, due to various mechanisms, can change the ratio of linear and circular anisotropy and transform the nature of its eigenmodes. This should lead to a more complex evolution of the polarization state at the fiber output, since the point on the Poincaré sphere will move along a circle when both the center of the circle and its radius are changed. The latter is due to the fact that if the polarization eigenmodes of the fiber change, then taking into account the fixed radiation parameters of the source, the ratio of the excited polarization modes will also change. However, the change in the parameter φ with an increase in fiber temperature should occur faster than the change in the angular parameters of the eigenmodes. We believe that if a fragment of the evolution of the polarization state observed during measurements corresponds well to the small circle of the Poincaré sphere, then this allows to estimate the values of ε_0 and Θ_0 of the polarization eigenstates of the fiber corresponding to this fragment. As a result of the experiment, our measurements can show not only the parameters of the eigenmodes of real fiber, but also detect their fluctuations when external conditions change.

Secondly, the azimuth of the points recorded by the polarimeter is determined by the position of the polarimeter axis, which is set virtually arbitrarily relative to the end of the fiber. Therefore, the absolute value of the measured azimuth Θ_0 of the polarization mode will not be informative (for the second mode, the azimuth will be shifted by $\pi/2$). However, when analyzing spun fiber, as discussed above, the deviation of the fiber from the idealized representation is characterized not by azimuth, but by the imperfection parameter δ, which is related only to how much the ellipticity angle of the polarization mode ε_0 differs from $\pi/4$. However, if the value of Θ_0 changes during the measurement process, then these changes will indeed characterize changes in their polarization eigenmodes.

The experiments used spun fiber manufactured by Fibercore (model SHB1500(8.9/125)), the length of the test segment was 80 m, the fiber was wound around a standard coil with a diameter of 16 cm. The scheme of the radiation source to which the fiber was connected is shown in Fig. 2,*c*. A DFB laser from Optilab was used (model DFB 1550 PM-20, wavelength 1550 nm, output power 9.5 MW), which had a fiber output (PM fiber with an APC-type connector). Next, a segment of PM fiber of the Bow-Tie type manufactured by Fibercore (model HB1250, the beat length of polarization modes is 3.28 mm) was spliced to the laser output via a connector. A short fiber segment (approximately 0.82 mm long) was formed at the end of the fiber input, rotated 45° relative to the axes of the main segment, after which the spun fiber was spliced. This segment served as a quarter-wave phase plate.

When linearly polarized radiation passes from the laser output through the PM fiber input and the quarter-wave plate rotated by 45°, circularly polarized radiation should be formed, and one polarization mode should be excited in it in the idealized representation of the spun fiber. However, since the real spun fiber has polarization modes other than circular ones, and the formed fiber phase plate is not an ideal quarter-wave, in fact two polarization modes in an unequal amplitude ratio were excited in the tested fiber. This exactly corresponded to the conditions required for measurements: it was possible to directly monitor the correspondence of fragments of the evolution of the polarization state to circular trajectories on the Poincaré sphere and measure the parameters of the polarization modes of the fiber.

The polarization state was recorded with the Thorlabs polarimeter PAX1000IR2 (USA), which allowed measuring the azimuth and elliptical angle of the polarization state with an accuracy of 0.25°. A collimator was used to connect the fiber to the polarimeter.

During the experiment, the tested fiber was slowly heated to 40 °C (in 50 minutes). The evolution of the recorded polarization state of light at the fiber outlet caused by heating is shown in Fig. 3,*a*. Evidently, the trajectory of the point of the output state of polarization on the Poincaré sphere forms many turns covering the pole of the sphere under heating. The radius of the turns varies noticeably, and their shape does not always correspond to circles, which is quite understandable for the reasons mentioned above.

Nevertheless, many turns in the trajectory of the polarization state correspond well to circles. Such fragments illustrate a situation where, with stable polarization eigenstates of the fiber, the phase difference φ changes. For example, Fig. 3,*b* shows three fragments of the observed evolution of the polarization state at the output of the spun fiber, which are consistent with the circles on the sphere. This can be seen by the correspondence between the points measured by the polarimeter and the circles on the sphere approximating these points. Such fragments make it possible to determine the parameters of the polarization eigenmodes in a given segment of the trajectory. The table shows the values of the circle parameters for the three fragments shown in Fig. 3,*b*.

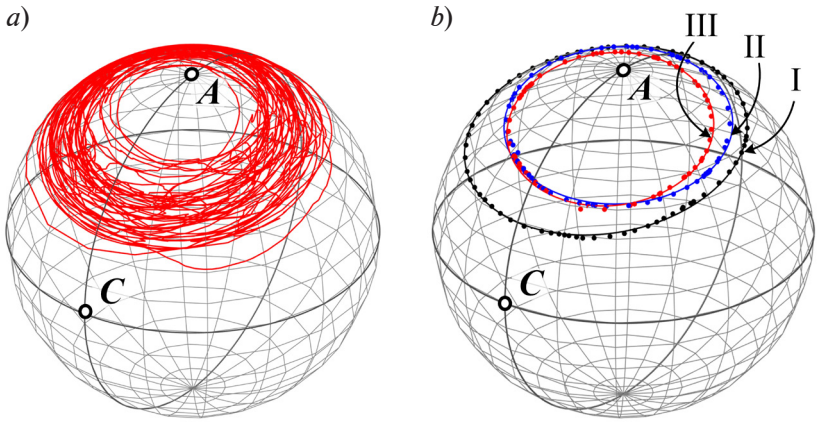

Fig. 3. Complete evolution of polarization state (*a*) and fragments of evolution I, II and III (*b*) at the output from the spun fiber, shown on the Poincaré spheres
Solid lines correspond to the approximation of the fragment points by circles on the sphere. Points *A* and *C* correspond to right circular polarization (ε = 45°) and linear polarization along the axis *X* (Θ = ε = 0), respectively

Additionally, the measured evolution of the polarization state allows to estimate the average normalized temperature sensitivity of the phase difference φ of the polarization modes. This sensitivity was approximately 0.02 rad/(m·°C).

A change in the value of R (see Table) means that when the parameters of the polarization eigenstates change, the ratio of their excitation by radiation at the fiber input also changes. As evident from the examples with three fragments of a fixed trajectory on the Poincaré sphere, the azimuth Θ_0 of the polarization ellipse of its eigenmodes changes most significantly (by almost 20°). The ellipticity angle ε_0, which characterizes the difference between the fiber and the idealized case, varies significantly less. If we convert the value of ε_0 into the parameter δ, then, according to Table, the average value of δ is approximately 8.3°, and the difference between the maximum and minimum values is 2.4°.

Thus, the measurement results indicate for the fiber tested that it is possible to make quite definite estimates for the main parameter (δ) characterizing its imperfection and necessary for the analysis of optical circuits using the obtained Jones matrix of the spun fiber.

Table

Parameters of circles approximating the measured fragment points of evolution in polarization states on the Poincaré sphere (see Fig. 3,*b*)

Angular parameter	Parameter value, degrees, for fragment		
	I	II	III
Radius R	48.4	36.9	32.8
Azimuth Θ_0	2.9	15.7	12.5
Ellipticity angle ε_0	37.6	37.3	35.2

Conclusion

An expression for the Jones matrix of real spun fiber is obtained within the framework of the phase anisotropy model. The expression takes into account a slight deviation of the fiber properties from the idealized case with polarization eigenstates in the form of right and left circular polarizations. For this purpose, a small parameter δ is used, which takes into account the deviation of the ellipticity angle of the polarization eigenstate from π/4. The resulting expression can be used to describe and analyze optical circuits containing spun-type fibers based on the Jones formalism.

The results of the proposed and conducted experiments on measuring the parameters of the polarization eigenstates of the fiber illustrate the deviation of the real spun fiber from the idealized representation and show the difference between the polarization eigenstates and circular polarization. At the same time, measurements for the fiber model used allowed to estimate the value of the imperfection parameter δ in the range of about 7°–10°.

Appendix

Variability of representation of polarization state in the Jones formalism

Although the representation of polarized waves is well-established in the literature and the Jones formalism is widely used to describe transformations of the polarization state, unfortunately, there are conflicting viewpoints on some details of such a description. In general, the choice of certain formulations does not affect the correct result. However, given the importance of these features for the material of this paper, it is preferable to clarify some points of the approaches we use in order to avoid confusion and possible questions.

The first aspect for which there are conflicting viewpoints in the literature is the accounting for phases when constructing Jones vectors and the correspondence of slow and fast polarization eigenmodes to the eigenvalues of the Jones matrix.

Consider the polarization eigenmodes of some optical element with phase anisotropy. Let us assume for the first mode that the X component of the field at the input to the optical element has the form

$$E_x^{\text{in}} = A_1\cos(\omega t + \varphi_0),$$

where ω, φ_0 are the angular frequency and the initial phase of the oscillation.

The Y component with the initial phase shifted by $\delta\varphi$ has the form

$$E_y^{\text{in}} = A_2\cos(\omega t + \varphi_0 + \delta\varphi).$$

Then the Jones eigenvectors $\mathbf{J}_1^{\text{in}}$ and $\mathbf{J}_2^{\text{in}}$ (in the Cartesian basis) are written as follows:

$$\mathbf{J}_1^{\text{in}} = \begin{bmatrix} A_1 \\ A_2 e^{j\delta\varphi} \end{bmatrix}, \mathbf{J}_1^{\text{in}} = \begin{bmatrix} -A_2 e^{-j\delta\varphi} \\ A_1 \end{bmatrix}. \tag{A1}$$

The second vector is represented so that it is orthogonal to the first one.

Jones vectors can also include an overall factor $\exp(j\varphi_0)$, but it is typically omitted because it does not affect the shape and orientation of the polarization ellipse. In such notation, the components of the vector contain complex amplitudes, whose arguments are given as initial phases. In this case, for the vector $\mathbf{J}_1^{\text{in}}$, assuming $\delta\varphi > 0$, it turns out that the X component of the vector is delayed relative to the Y component.

Passing through an optical element with phase anisotropy, in the case when polarization-independent losses are negligibly small, the eigenmodes acquire only a phase delay, each a different one. The first and second modes acquire phase delays Φ_1 and Φ_2, respectively:

$$\Phi_1 = n_1 L/\lambda,\ \Phi_2 = n_2 L/\lambda,$$

where L is the geometric wavelength of the waves in the optical element; λ is the wavelength of light; n_1, n_2 are the effective refractive indices for polarization eigenmodes.

If we include the average refractive index n and the difference Δn of the form

$$n = (n_1 + n_2)/2,\ \Delta n = n_2 - n_1$$

(Δn characterizes the anisotropy of the element), then the phase delays can be written as

$$\Phi_1 = \Phi - \varphi/2,\ \Phi_2 = \Phi + \varphi/2,$$

where $\Phi = nL/\lambda$, $\varphi = \Delta nL/\lambda$.

The value of φ is positive if $\Delta n > 0$. In this case, the first mode propagates faster and has a lower phase delay, while the second one propagates slower and acquires a greater phase delay. Therefore, when $\varphi > 0$, it is logical to call the first and second modes "fast" and "slow", respectively.

Taking into account the Cartesian components of the first mode introduced above at the input to the optical element, they can be written as

$$E_x^{\text{out}} = A_1\cos[\omega t + \varphi_0 - (\Phi - \varphi/2)],$$

$$E_y^{\text{out}} = A_2\cos[\omega t + \varphi_0 + \delta\varphi - (\Phi - \varphi/2)].$$

Similarly, the components of the second mode at the output of the optical element are obtained by adding the terms $-(\Phi + \varphi/2)$ to the phase of the components at the input.

It is easy to prove that if the Jones vectors take into account the initial phases of the field oscillations, then the relationship between the input ($\mathbf{J}_1^{\text{in}}$, $\mathbf{J}_2^{\text{in}}$) and output ($\mathbf{J}_1^{\text{out}}$, $\mathbf{J}_2^{\text{out}}$) vectors of eigenmodes should have the following form:

$$\begin{aligned} \mathbf{J}_1^{\text{out}} &= e^{-j\Phi}\cdot\mathbf{M}\cdot\mathbf{J}_1^{\text{in}} = e^{-j\Phi}\cdot e^{j\varphi/2}\cdot\mathbf{J}_1^{\text{in}}; \\ \mathbf{J}_2^{\text{out}} &= e^{-j\Phi}\cdot\mathbf{M}\cdot\mathbf{J}_2^{\text{in}} = e^{-j\Phi}\cdot e^{-j\varphi/2}\cdot\mathbf{J}_2^{\text{in}}. \end{aligned} \tag{A2}$$

The Jones matrix $\mathbf{M}$ of the optical element is introduced in (A2), which does not take into account the average phase shift and is a special unitary matrix with eigenvalues

$$\lambda_1 = e^{j\varphi/2},\ \lambda_2 = e^{-j\varphi/2}.$$

It is clear from the above arguments that if $\varphi > 0$, then the vectors with eigenvalues λ_1 and λ_2 belong to the fast and slow modes of the anisotropic element, respectively.

The described representation of phases in terms of vectors and Jones matrices is widely used in textbooks, monographs and articles [16, 17, 21]. However, an alternative approach to constructing phases in terms of vectors and Jones matrices can also be found in the literature [19]. It is based on the representation of a harmonic wave propagating along the z axis with the wavenumber k in terms of the function $\cos(\omega t - kz)$. Then the phase shift of the wave relative to the zero initial phase can be interpreted as a phase delay due to the passage of a certain path. It is not the initial phases that are taken into account in the notation for the vectors and Jones matrices, but phase delays, i.e., negative changes in the initial phases are taken into account as positive delays and vice versa.

In this representation, the same E_x^{in} and E_y^{in} can be written as

$$E_x^{\text{in}} = A_1\cos[\omega t - (-\varphi_0)],\ E_y^{\text{in}} = A_2\cos[\omega t - (-\varphi_0 - \delta\varphi)],$$

where the phase delays are now given in parentheses.

In this case, the vectors $\mathbf{J}_1^{\text{in}}$ and $\mathbf{J}_2^{\text{in}}$ are already written as

$$\mathbf{J}_1^{\text{in}} = \begin{bmatrix} A_1 \\ A_2 e^{-j\delta\varphi} \end{bmatrix},\ \mathbf{J}_2^{\text{in}} = \begin{bmatrix} -A_2 e^{j\delta\varphi} \\ A_1 \end{bmatrix}. \tag{A3}$$

If it is necessary to take into account the phase φ_0, then the overall phase factor $\exp(-j\varphi_0)$ must be included into Eq. (A3). The components E_x^{out} and E_y^{out} also do not change, but the output vectors are now represented as

$$\begin{aligned} \mathbf{J}_1^{\text{out}} &= e^{j\Phi}\cdot\mathbf{M}\cdot\mathbf{J}_1^{\text{in}} = e^{j\Phi}\cdot e^{-j\varphi/2}\cdot\mathbf{J}_1^{\text{in}}; \\ \mathbf{J}_2^{\text{out}} &= e^{j\Phi}\cdot\mathbf{M}\cdot\mathbf{J}_2^{\text{in}} = e^{j\Phi}\cdot e^{j\varphi/2}\cdot\mathbf{J}_2^{\text{in}}. \end{aligned} \tag{A4}$$

Here, the optical element with phase anisotropy is also represented by a special unitary matrix $\mathbf{M}$ with the same eigenvalues $e^{j\varphi/2}$ and $e^{-j\varphi/2}$. However, in this representation, a vector with an eigenvalue $\lambda_1 = e^{j\varphi/2}$ corresponds to a slow mode, and a vector with an eigenvalue $\lambda_2 = e^{-j\varphi/2}$ corresponds to a fast mode (at $\varphi > 0$).

In this paper, we adhere to the first case of the representation of Jones vectors, when they take into account the initial phases of Cartesian components, rather than phase delays.

The second aspect on which there are conflicting viewpoints in the literature is the definition of polarized waves with right and left directions of rotation of the electric field intensity vector.

Most textbooks and monographs [16, 22, 23] define a right-polarized wave as the one where the electric field strength vector rotates clockwise if looking towards the direction of wave propagation. Accordingly, the left-polarized wave has a counterclockwise rotation of the electric field intensity vector. In this article, we adhere to this definition. However, an equally valid opposing opinion can be found in the literature for the definition of right- and left-polarized waves [21].

The third aspect important for this paper is the way Jones vectors are written for right and left circular polarizations.

In accordance with the definition of right- and left-polarized waves, which we adhere to, it is not difficult to verify the following. For the right circular polarization, the component E_x is delayed relative to the component E_y by $\pi/2$. For example, when $E_x = A\cos(\omega t + \varphi_0)$, the right circular polarization is followed by $E_y = -A\sin(\omega t + \varphi_0)$. Consequently, the initial phase of the Y component is additionally increased by $\pi/2$. For the left circular polarization, on the contrary, the component E_y is delayed relative to the component E_x. Therefore, taking into account all the conditions we adopted, the Jones vectors for the right and left circular polarizations will be described by vectors (2).

Nevertheless, it is important to note that the comparison of Jones vectors (2) with right and left circular polarizations by some researchers in the available literature may be the opposite, due to differences in the adopted notations for the phases in terms of Jones vectors and matrices, as well as definitions of right- and left-polarized light. For example, in [21], the use of an alternative variant of vectors is associated with an alternative definition of the names of the direction of rotation, and in [19] it is associated with an alternative representation of Jones vectors using phase delays.

REFERENCES

1. **Liu Y., Song H.,** Theoretical analysis on polarization characteristics of spun birefringent optical fiber based on an analytical Jones matrix model, Optik. 228 (Febr) (2021) 166179.

2. **Barlow A. J., Ramskov-Hansen J. J., Payne D. N.,** Birefringence and polarization mode-dispersion in spun single-mode fibers, Appl. Opt. 20 (17) (1981) 2962–2968.

3. **Payne D. N., Barlow A. J., Ramskov-Hansen J. J.,** Development of low- and high-birefringence optical fibers, IEEE Trans. Microw. Theory Tech. 30 (4) (1982) 323–334.

4. **Polynkin P., Blake J.,** Polarization evolution in bent spun fiber, J. Light. Technol. 23 (11) (2005) 3815–3820.

5. **Fedotov A., Ustimchik V., Rissanen J., et al.,** Active tapered double-clad fiber with low birefringence, Opt. Express. 29 (11) (2021) 16506–16519.

6. **Fedotov A., Ustimchik V., Rissanen J., et al.,** Large mode area double-clad ytterbium-doped spun tapered fiber, J. Opt. Soc. Am. B. 38 (12) (2021) F161–F169.

7. **Laming R. I., Payne D. N.,** Electric current sensors employing spun highly birefringent optical fibers, J. of Lightwave Technology. 7 (12) (1989) 2084–2094.

8. **Gubin V. P., Isaev V. A., Morshnev S. K., et al.,** Use of Spun optical fibres in current sensors, Quantum Electron. 36 (3) (2006) 287–291.

9. **Müller G. M., Frank A., Yang L., et al.,** Temperature compensation of interferometric and polarimetric fiber-optic current sensors with spun highly birefringent fiber, J. Light. Technol. 37 (18) (2019) 4507–4513.

10. **Przhiyalkovsky Ya. V., Gubin V. P., Starostin N. I., et al.,** Detection of electric current pulses by a fibre-optic sensor using spun fibre, Quantum Electron. 48 (1) (2018) 62–69.

11. **Temkina V., Medvedev A., Mayzel A.,** Research on the methods and algorithms improving the measurements precision and market competitive advantages of fiber optic current sensors, Sensors. 20 (21) (2020) 5995.

12. **Peng N., Huang Y., Wang S., et al.,** Fiber optic current sensor based on special spun highly birefringent fiber, IEEE Photon. Technol. Lett. 25 (17) (2013) 1668–1671.

13. **Wang Y., Xu C.-Q., Izraelian V.,** Characterization of spun fibers with millimeter spin periods, Opt. Express. 13 (10) (2005) 3841–3851.

14. **Hu H., Huang J., Huang Y, et al.,** Modeling of the birefringence in spun fiber, Opt. Commun. 473 (15 Oct) (2020) 125919.

15. **Yao P., Chen X., Hao P., et al.,** Introduction and measurement of the effective Verdet constant of spun optical fibers, Opt. Express. 29 (15) (2021) 23315–23330.

16. **Azzam R. M. A., Bashara N. M.,** Ellipsometry and polarized light, Third ed., North Holland Publishing Company, Amsterdam, Netherlands, 1999.

17. **Collett E.,** Polarized light in fiber optics, SPIE Press, Bellingham, Washington, USA, 2003.

18. **Ishchenko E. F., Sokolov A. L.,** Polyarizatsionnaya optika [Polarization optics], Third Ed., Fizmatlit Publishing, Moscow, 2019 (in Russian).

19. **Huard S.,** Polarization of light, John Wiley & Sons, Inc., Chichester, UK, 1997.

20. **Temkina V. S., Liokumovich L. B., Archelkov A. B., et al.,** Description of polarization-maintaining fibers in analyzing the practical fiber-optic circuits using the Jones formalism, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (3) (2023) 95–114 (in Russian).

21. **Yariv A., Yeh P.,** Optical waves in crystals: propagation and control of laser radiation, John Wiley & Sons, Inc., New York, Chichester, Brislane, Toronto, Singapore, 1984.

22. **Gerrard A., Burch J. M.,** Introduction to matrix methods in optics, Dover Publications, Inc., New York, USA, 2012.

23. **Born M., Wolf E.,** Principles of optics, Cambridge University Press, Cambridge, United Kingdom, 2019.

СПИСОК ЛИТЕРАТУРЫ

1. **Liu Y., Song H.** Theoretical analysis on polarization characteristics of spun birefringent optical fiber based on an analytical Jones matrix model // Optik. 2021. Vol. 228. February. P. 166179.

2. **Barlow A. J., Ramskov-Hansen J. J., Payne D. N.** Birefringence and polarization mode-dispersion in spun single-mode fibers // Applied Optics. 1981. Vol. 20. No. 17. Pp. 2962–2968.

3. **Payne D. N., Barlow A. J., Ramskov-Hansen J. J.** Development of low- and high-birefringence optical fibers // IEEE Transactions on Microwave Theory and Techniques. 1982. Vol. 30. No. 4. Pp. 323–334.

4. **Polynkin P., Blake J.** Polarization evolution in bent spun fiber // Journal of Lightwave Technology. 2005. Vol. 23. No. 11. Pp. 3815–3820.

5. **Fedotov A., Ustimchik V., Rissanen J., Kolosovskii A., Voloshin V., Vorob′ev I., Gumenyuk R., Chamorovskiy Y., Filippov V.** Active tapered double-clad fiber with low birefringence // Optics Express. 2021. Vol. 29. No. 11. Pp. 16506–16519.

6. **Fedotov A., Ustimchik V., Rissanen J., Noronen T., Gumenyuk R., Kolosovskii A., Voloshin V., Vorob′ev I., Chamorovskii Y., Filippov V.** Large mode area double-clad ytterbium-doped spun tapered fiber // Journal of the Optical Society of America. B. 2021. Vol. 38. No. 12. Pp. F161–F169.

7. **Laming R. I., Payne D. N.** Electric current sensors employing spun highly birefringent optical fibers // Journal of Lightwave Technology. 1989. Vol. 7. No. 12. Pp. 2084–2094.

8. **Губин В. П., Исаев В. А., Моршнев С. К., Сазонов А. И., Старостин Н. И., Чаморовский Ю. К., Усов А. И.** Использование волоконных световодов типа Spun в датчиках тока // Квантовая электроника. 2006. Т. 36. № 3. С. 287–291.

9. **Müller G. M., Frank A., Yang L., Gu X., Bohnert K.** Temperature compensation of interferometric and polarimetric fiber-optic current sensors with spun highly birefringent fiber // Journal of Lightwave Technology. 2019. Vol. 37. No. 18. Pp. 4507–4513.

10. **Пржиялковский Я. В., Губин В. П., Старостин Н. И., Моршнев С. К., Сазонов А. И.** Регистрация импульсов электрического тока волоконно-оптическим датчиком с использованием spun-световодов // Квантовая электроника. 2018. Т. 48. № 1. С. 62–69.

11. **Temkina V., Medvedev A., Mayzel A.** Research on the methods and algorithms improving the measurements precision and market competitive advantages of fiber optic current sensors // Sensors. 2020. Vol. 20. No. 21. P. 5995.

12. **Peng N., Huang Y., Wang S., Wen T., Liu W., Zuo Q., Wang L.** Fiber optic current sensor based on special spun highly birefringent fiber // IEEE Photonics Technology Letters. 2013. Vol. 25. No. 17. Pp. 1668–1671.

13. **Wang Y., Xu C.-Q., Izraelian V.** Characterization of spun fibers with millimeter spin periods // Optics Express. 2005. Vol. 13. No. 10. Pp. 3841–3851.

14. **Hu H., Huang J., Huang Y, Xia L., Yu J.** Modeling of the birefringence in spun fiber // Optics Communications. 2020. Vol. 473. 15 October. P. 125919.

15. **Yao P., Chen X., Hao P., Xiao H., Ding Z., Liu T., Yao X. S.** Introduction and measurement of the effective Verdet constant of spun optical fibers // Optics Express. 2021. Vol. 29. No. 15. Pp. 23315–23330.

16. **Аззам Р., Башара Н.** Эллипсометрия и поляризованный свет. Пер. с англ. М.: Мир, 1981. 584 с.

17. **Collett E.** Polarized light in fiber optics. Bellingham, Washington, USA: SPIE Press, 2003. 540 p.

18. **Ищенко Е. Ф., Соколов А. Л.** Поляризационная оптика. 3-е изд., испр. и доп. М.: Физматлит, 2019. 576 с.

19. **Huard S.** Polarization of light. Chichester, UK: John Wiley & Sons, Inc., 1997. 352 p.

20. **Темкина В. С., Лиокумович Л. Б., Арчелков А. Б., Бучилко И. Р., Медведев А. В., Петров А. В.** Описание волоконных световодов с линейным двулучепреломлением при анализе практических оптоволоконных схем методом векторов и матриц Джонса // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 3. С. 95–114.

21. **Ярив А., Юх П.** Оптические волны в кристаллах. Пер. с англ. М.: Мир, 1987. 616 с.

22. **Джеррард А., Бёрч Дж. М.** Введение в матричную оптику. Пер. с англ. М.: Мир, 1978. 344 с.

23. **Борн М., Вольф Э.** Основы оптики. Пер. с англ. М.: Наука, 1973. 720 с.

THE AUTHORS

TEMKINA Valentina S.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
temkina_vs@spbstu.ru
ORCID: 0000-0003-2083-8989

LIOKUMOVICH Leonid B.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
leonid@spbstu.ru
ORCID: 0000-0001-5988-1429

ARCHELKOV Arseniy B.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
arsarch11@gmail.com
ORCID: 0009-0007-4713-1293

MEDVEDEV Andrei V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
medvedev@rphf.spbstu.ru
ORCID: 0000-0001-7083-9184

KOZLOV Artemy S.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
kozlov_as@spbstu.ru
ORCID: 0000-0002-1722-1964

GRESHNEVIKOV Konstantin V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russia
kgreshnevikov@yandex.ru
ORCID: 0000-0002-6154-2538

СВЕДЕНИЯ ОБ АВТОРАХ

ТЕМКИНА Валентина Сергеевна – *ассистентка Высшей школы прикладной физики и космических технологий Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
temkina_vs@spbstu.ru
ORCID: 0000-0003-2083-8989

ЛИОКУМОВИЧ Леонид Борисович – *доктор физико-математических наук, профессор Высшей школы прикладной физики и космических технологий Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
leonid@spbstu.ru
ORCID: 0000-0001-5988-1429

АРЧЕЛКОВ Арсений Борисович – *студент Института электроники и телекоммуникаций Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
arsarch11@gmail.com
ORCID: 0009-0007-4713-1293

МЕДВЕДЕВ Андрей Викторович – *кандидат физико-математических наук, доцент Высшей школы прикладной физики и космических технологий Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
medvedev@rphf.spbstu.ru
ORCID: 0000-0001-7083-9184

КОЗЛОВ Артемий Сергеевич – *инженер Высшей школы прикладной физики и космических технологий Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
kozlov_as@spbstu.ru
ORCID: 0000-0002-1722-1964

ГРЕШНЕВИКОВ Константин Владимирович – *кандидат физико-математических наук, доцент Высшей школы прикладной физики и космических технологий Санкт-Петербургского политехнического университета Петра Великого.*
195251, Россия, г. Санкт-Петербург, Политехническая ул., 29
kgreshnevikov@yandex.ru
ORCID: 0000-0002-6154-2538

Received 29.09.2023. Approved after reviewing 09.11.2023. Accepted 09.11.2023.
Статья поступила в редакцию 29.09.2023. Одобрена после рецензирования 09.11.2023. Принята 09.11.2023.

MATHEMATICS

Original article
UDC 515.1.
DOI: https://doi.org/10.18721/JPM.16417

FIXED POINT THEOREMS ON ORTHOGONAL METRIC SPACES VIA τ-DISTANCES

Y. Touail[1]✉, *A. Jaid*[2], *D. El Moutawakil*[3]

[1] Sidi Mohamed Ben Abdellah University, Fès, Morocco

[2] Sultan Moulay Slimane University, Beni-Mellal, Morocco

[3] Chouaib Doukkali University, El Jadida, Morocco

✉ youssef9touail@gmail.com

Abstract. In this paper, we prove two fixed point theorems in the setting of orthogonal complete metric spaces via τ-distances. Our theorems generalize and improve many known results in the literature (see, for example Refs. [6, theorem 4.2] and [3, theorem 3]).

Keywords: fixed point, orthogonal generalized E-weakly contractive maps, orthogonal metric space, Hausdorff topological spaces, τ-distance

Citation: Touail Y., Jaid A., El Moutawakil D., Fixed point theorems on orthogonal metric spaces via τ-distances, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 16 (4) (2023) 215–223. DOI: https://doi.org/10.18721/JPM.16417

Научная статья
УДК 515.1.
DOI: https://doi.org/10.18721/JPM.16417

ТЕОРЕМЫ О НЕПОДВИЖНОЙ ТОЧКЕ НА ОРТОГОНАЛЬНЫХ МЕТРИЧЕСКИХ ПРОСТРАНСТВАХ, ДОКАЗАННЫЕ С ПОМОЩЬЮ ПОНЯТИЯ τ-РАССТОЯНИЯ

Ю. Туай[1]✉, *А. Джайд*[2], *Д. Аль-Мутавакиль*[3]

[1] Университет Сиди Мохамеда Бен Абделлы, г. Фес, Марокко;

[2] Университет Султана Мулая Слимана, г. Бени-Меллал, Марокко;

[3] Университет Шуайб Дуккали, г. Эль-Джадида, Марокко

✉ youssef9touail@gmail.com

Аннотация. В этой статье мы доказываем две теоремы о неподвижной точке в задании ортогональных полных метрических пространств, используя понятие τ-расстояния. Выдвинутые и доказанные теоремы позволяют обобщить и улучшить многие известные результаты, опубликованные в литературе (см., например, результаты в статьях [6, теорема 4.2] и [3, теорема 3]).

Ключевые слова: неподвижная точка, ортогональное обобщенное E-слабосжимаемое отображение, ортогональное метрическое пространство, хаусдорфово топологическое пространство, τ-расстояние

Ссылка для цитирования: Туай Ю., Джайд А., Аль-Мутавакиль Д. Теоремы о неподвижной точке на ортогональных метрических пространствах, доказанные с помощью понятия τ-расстояния // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2023. Т. 16. № 4. С. 215–223. DOI: https://doi.org/10.18721/JPM.16417

Introduction

In 2003, M. Aamri and D. El Moutawakil [1] introduced the concept of τ-distance in general topological spaces. This innovation has extended a lot of ideas about known spaces presented in the literature. Moreover, these scientists proved a version of the Banach's fixed point theorem for this general setting.

In 2017, M. E. Gordji et al. [2] defined so-called orthogonal metric spaces as a generalization of the metric spaces. The authors showed in Ref. [2] that this type of spaces is very powerful and applicable to many cases, such as the fixed point theory. Then an important extension of Banach's fixed point theorem was given.

Without using the compactness of the space, the author of Ref. [6] put forward some fixed point theorems for new classes of mappings via τ-distance in general topological spaces (some related results can be found in Refs. [3 − 5, 7]).

In this paper, motivated by Refs. [2, 6], we extend some results proven in Ref. [6]; in other words, we will restrict our studies to the orthogonal elements only, in order to prove the fixed point property for a large class of contractive mappings. Our results will be based specially on some essential notions like orthogonality, τ-distances in the general topological spaces. Some important examples will also be given to support the proven theorems and to show the usability of this new direction of research.

Preliminaries

The aim of this section is to present some concepts and known results used in the paper.

Let (X, τ) be a topological space and $p: X \times X \to [0, +\infty)$ be a function. For any $\varepsilon > 0$ and any $x \in X$, let $B_p(x, \varepsilon) = \{y \in X / p(x, y) < \varepsilon\}$.

Definition I [1, definition 2.1]. The function p is said to be τ-distance if there exists $\varepsilon > 0$ for each $x \in X$ and any neighborhood V of x, such that $B_p(x, \varepsilon) \subset V$.

Definition II. In a Hausdorff topological space X, a sequence $\{x_n\}$ is said to be a p-Cauchy sequence if it satisfies the usual metric condition with respect to p; in other words, if $\lim_{n,m\to\infty} p(x_n, x_m) = 0$.

Definition III [1, definition 3.1]. Let (X, τ) be a topological space with a τ-distance p.

1. X is S-complete if there exists x in X for every p-Cauchy sequence (x_n), such that $\lim p(x, x_n) = 0$.

2. X is considered p-Cauchy complete if there exists x in X for every p-Cauchy sequence (x_n), such that $\lim x_n = x$ with respect to τ.

3. X is said to be p-bounded if $\sup\{p(x, y) / x, y \in X\} < \infty$.

Lemma 1 [1, lemma 3.1]. *Let* (X, τ) *be a Hausdorff topological space with a* τ*-distance p, then*

1) $p(x, y) = 0$ *implies* $x = y$.

2) *Let* (x_n) *be a sequence in X such that lim* $p(x, x_n) = 0$ *and* $\lim p(y, x_n) = 0$, *then* $x = y$.

Lemma 1 was proved in Ref. [1].

Definition IV. [1, definition 2.5]). Ψ is the class of all functions ψ from $[0, +\infty)$ to $[0, +\infty)$ satisfying:

i) ψ is nondecreasing,

ii) $\lim \psi^n(t) = 0$ for all $t \in [0, +\infty)$.

Definition V. Φ is the class of all functions ϕ from $[1, +\infty)$ to $[0, +\infty)$ satisfying:

i) $\phi(t) = 0$ if and only if $t = 1$,

ii) $\inf_{t>1} \phi(t) = 0$.

Theorem 1. [1, theorem 4.1]. *Let* (X, τ) *be a Hausdorff topological space with a* τ*-distance p. Suppose that X is p-bounded and S-complete. Let T be a selfmapping of X such that*

$$p(Tx, Ty) \le \phi(p(x, y)),$$

for all $x, y \in X$. *Then T has a unique fixed point.*

Theorem 1 was proved in Ref. [1].

Theorem 2 [6]. *Let* (X, τ) *be a Hausdorff topological space with a* τ*-distance p. Suppose that X is p-bounded and S-complete. Let T be a p-continuous selfmapping of X such that*

$$p(Tx,Ty) \le \phi\left(\max\left\{p(x,y), p(x,Tx), p(y,Ty)\right\}\right), \tag{1}$$

f*or all* $x, y \in X$*. Then T has a unique fixed point.*

Theorem 2 was proved in Ref. [6].

Theorem 3 [6]. *Let* $T: X \to X$ *be a generalized E-weakly contractive mapping of a bounded complete metric space* (X, d)*. Then T has a unique fixed point.*

Theorem 3 was proved in Ref. [6].

Theorem 4 [6]. *Let* $T: X \to X$ *be a mapping of a bounded complete metric space* (X, d) *such that*

$$\inf_{x \ne y \in X}\left\{\max\left\{d(x,y), d(x,Tx), d(y,Ty)\right\} - d(Tx,Ty)\right\} > 0. \tag{2}$$

Then T has a unique fixed point.

Theorem 4 was proved in Ref. [6].

Now we recall the definition of an orthogonal set and some related basic notions.

Definition VI [2]. *Let* $X \ne \phi$ and let $\perp \subset X \times X$ be a binary relation. If $\perp$ satisfies the following hypothesis:

$$\exists x_0 : (\forall y, y \perp x_0) \text{ or } (\forall y, x_0 \perp y), \tag{3}$$

then it called an orthogonal set (briefly *O*-set); we denote this *O*-set by $(X, \perp)$.

Note that x_0 is said to be an orthogonal element in the Definition VI.

Remark. In general, x_0 is not unique, otherwise, $(X, \perp)$ is called unique orthogonal set and the element x_0 is said to be a unique orthogonal element.

Definition VII [2]. Let$(X, \perp)$ be an *O*-set. A sequence $\{x_n\}$ is called an orthogonal sequence (briefly, *O*-sequence) if

$$(\forall n, x_n \perp x_{n+1}) \text{ or } (\forall n, x_{n+1} \perp x_n).$$

Definition VIII [2]. The triplet $(X, \perp, d)$ is called an orthogonal metric space if (X, d) is a metric space and $(X, \perp)$ is the *O*-set.

Definition IX [2]. Let $(X, \perp, d)$ be an orthogonal metric space. Then, a mapping $T: X \to X$ is said to be orthogonally continuous (briefly $\perp$-continuous) in $x \in X$, if for each *O*-sequence $\{x_n\} \subset X$ such that $x_n \to x$ as $n \to \infty$, we obtain $Tx_n \to Tx$ as $n \to \infty$. *T* is said to be $\perp$-continuous on *X* if *T* is $\perp$-continuous in each $x \in X$ as well.

Definition X [2]. Let $(X, \perp, d)$ be an orthogonal metric space. Then, *X* is said to be orthogonally complete (or $\perp$-complete) if every Cauchy *O*-sequence is convergent.

Definition XI [2]. Let $(X, \perp)$ be the *O*-set. A mapping $T: X \to X$ is said to be $\perp$-preserving if $Tx \perp Ty$ whenever $x \perp y$.

Remark [2]. Every complete metric space (continuous mapping) is *O*-complete metric space ($\perp$-continuous mapping) and the converse is not true.

Theorem 5 [2]. *Let* $(X, \perp, d)$ *O-complete metric space and T a self-mapping on X which is* $\perp$*-preserving and* $\perp$*-continuous. If there exists* $k \in [0.1)$ *such that for all* $x, y \in X$

$$x \perp y \text{ implies } d(Tx, Ty) \le kd(x, y).$$

Then T has a unique fixed point.

Theorem 5 was proved in Ref. [2].

Now, we give some examples of orthogonal spaces.

Example 1 [2]. Let $X = \mathbf{Z}$. Define the binary relation $\perp$ on *X* by $m \perp n$ if there exists $k \in \mathbf{Z}$ such that $m = kn$. It is easy to see that $0 \perp n$ for all $n \in \mathbf{Z}$. Hence, $(X, \perp)$ is the *O*-set.

Example 2 [2]. Let *X* be an inner product space with the inner product. Define the binary relation $\perp$ on *X* by $x \perp y$ if $(x, y) = 0$. It is easy to see that $0 \perp x$ for all $x \in X$. Hence, $(X, \perp)$ is the *O*-set.

For more details, we refer the reader to see Ref. [2].

Main results

In this section, we start with some definitions and lemmas.

Definition XII. The triplet $(X, \perp, d)$ is called an orthogonal Hausdorff topological space with a τ-distance p if (X, τ) is a Hausdorff topological space with a τ-distance p and $(X, \perp)$ is an orthogonal set.

Definition XIII. Let (X, τ) be a topological space with a τ-distance p. Then $T: X \to X$ is said to be orthogonal p-continuous at $x \in X$ if we have for any orthogonal $\{x_n\} \subset X$ such that $\lim p(x, x_n) = 0$.

Lemma 2. *Let $(X, \perp, d)$ be an orthogonal Hausdorff topological space with a τ-distance p such that $p(x, x) = 0$ for all $x \in X$. Suppose that X is p-bounded and S-complete. Let T be a $\perp$-continuous and $\perp$-preserving self-mapping of X such that $x \perp y$ implies*

$$p(Tx,Ty) \le \phi\left(\max\left\{p(x,y), p(x,Tx)\, p(y,Ty)\right\}\right), \tag{4}$$

for all $x, y \in X$, where $\phi \in \Phi$. Then has a unique fixed point.

Proof. Since X is an orthogonal set, there exists at least $x_0 \in X$ such that

$$(\forall y, y \perp x_0) \text{or} (\forall y, x_0 \perp y). \tag{5}$$

This implies that $x_0 \perp Tx_0$ or $Tx_0 \perp x_0$ Consider the iterated sequence $\{x_n\}$ such that $x_n = T^n x_0$ for all $n \in N$. As T is a $\perp$-preserving, we obtain either $T^n x_0 \perp T^{n+1} x_0$ or $T^{n+1} x_0 \perp T^n x_0$ for all $n \in N$. Then $\{x_n\}$ is an O-sequence.

Let $n \in N$

$$p(x_{n+1}, x_{n+2}) \le \phi\left(\max\left\{p(x_n, x_{n+1}), p(x_n, x_{n+1}), p(x_{n+1}, x_{n+2})\right\}\right) \le$$

$$\le \phi\left(\max\left\{p(x_n, x_{n+1}); p(x_{n+1}, x_{n+2})\right\}\right).$$

If there exists $n \in N$ for which $p(x_{n_0}, x_{n_0+1}) < p(x_{n_0+1}, x_{n_0+2})$, then $p(x_{n_0+1}, x_{n_0+2}) < p(x_{n_0+1}, x_{n_0+2})$, this leads to contradiction.

Then $p(x_{n+1}, x_{n+2}) < p(x_n, x_{n+1})$ for all $n \in N$ which implies that

$$p(x_{n+1}, x_{n+2}) < \phi\left(p(x_n, x_{n+1})\right). \tag{6}$$

for every $n \in N$.

Now, let $n, m \in N$, we obtain from formula (5) $x_0 \perp x_n$ or $x_n \perp x_0$, using the fact that T is $\perp$-preserving, we get $x_n \perp x_{n+m}$ or $x_{n+m} \perp x_n$, which implies by inequality (6) that

$$\begin{gathered}
p(x_n, x_{n+m}) = p(\mathrm{T}x_{n-1}, \mathrm{T}x_{n+m-1}) \le \\
\le \phi\left(\max\left\{p(x_{n-1}, x_{n+m-1}), p(x_{n-1}, x_n), p(x_{n+m-1}, x_{n+m})\right\}\right) \le \\
\le \phi\left(\max\left\{p(x_{n-1}, x_{n+m-1}), p(x_{n-1}, x_n)\right\}\right) \le \\
\le \phi\left(\max\left\{\phi\left(\max\left\{p(x_{n-2}, x_{n+m-2}), p(x_{n-2}, x_{n-1})\right\}\right), \phi\left(p(x_{n-2}, x_{n-1})\right)\right\}\right) \le \\
\le \phi^2\left(\max\left\{p(x_{n-2}, x_{n+m-2}), p(x_{n-2}, x_{n-1})\right\}\right) \le \\
\vdots \\
\le \phi^n\left(\max\left\{p(x_0, x_m), p(x_0, x_1)\right\}\right) \le \\
\le \phi^n(M),
\end{gathered} \tag{7}$$

where $M = \sup\{p(x, y) / x, y \in X\}$.

Letting $n \to \infty$ in formula (7), we deduce that $\{x_n\}$ is an orthogonal p-Cauchy sequence. Since X is an orthogonal S-complete space, there exists $u \in X$ such that $\lim p(u, x_n) = 0$.

On the other side, the orthogonal p-continuity of the mapping T implies that $\lim p(Tu, Tx_n) = \lim p(u, x_n) = 0$.

Therefore, Lemma 1 then gives $Tu = u$.

For uniqueness, let $v \in X$ a fixed point of T, hence we have either $x_0 \perp v$ or $v \perp x_0$. From the orthogonality preserving, we get $x_n \perp v$ or $v \perp x_n$, for all $n \in N$. So,

$$p(v, x_n) \le \phi\left(\max\left\{p(v, x_{n-1}), p(x_n, x_{n+1})\right\}\right);$$

then,

$$p(v, x_n) \le \phi^n\left(\max\left\{p(v, x_0), p(x_0, x_1)\right\}\right). \tag{8}$$

Using Lemma 1 and letting $n \to \infty$ in the inequality (8), we obtain: $u = v$.

Note that the inequality $p(Tx, Ty) \le \phi\ (p(x, y))$ implies that T is p-continuous.

Lemma 2 is proved.

Corollary. Let (X, τ) be a Hausdorff topological space with a τ-distance p. Suppose that X is p-bounded and S-complete. Let T be a p-continuous self-mapping of X such that

$$p(Tx, Ty) \le \phi\left(\max\left\{p(x, y), p(x, Tx), p(y, Ty)\right\}\right) \tag{9}$$

for all $x, y \in X$, where $\phi \in \Phi$.

Then T has a unique fixed point.

Lemma 3. *Let (X, d) be a metric space and p from $X \times X$ to $[0, +\infty)$ be a function defined by*

$$p(x, y) = e^{d(x;y)} - 1. \tag{10}$$

Then p is a τ_d-distance on X where τ_d is the metric topology.

Proof. Let (X, τ_d) be the topological space with the metric topology τ_d, let $x \in X$ and V be an arbitrary neighborhood of x, then there exists $\varepsilon > 0$ such that $B_d(x, \varepsilon) \subset V$, where

$$B_d(x, \varepsilon) = \left\{y \in X, d(x, y) < \varepsilon\right\}$$

is the open ball. It is easy to see that $B_p(x, e^{\varepsilon} - 1) \subset B_d(x, \varepsilon)$, indeed:

let $y \in B_p(x, e^d - 1)$, then $p(x, y) < e^{\varepsilon} - 1$, which implies that $e^{d(x, y)} < e^{\varepsilon}$, and hence $d(x, y) < \varepsilon$.

Lemma 3 is proved.

Theorem 6. *Let $(X, d, \perp)$ be an orthogonal metric space and T: $X \to X$ be a mapping such that*

$$\inf_{x \perp y, x \ne y}\left\{\max\left\{d(x, y), d(x, Tx), d(y, Ty)\right\} - d(Tx, Ty)\right\} > 0. \tag{11}$$

Then T has a unique fixed point.

Proof. Let $\alpha = inf_{x \perp y, x \ne y}\left\{\max\left\{d(x, y), d(x, Tx), d(y, Ty)\right\} - d(Tx, Ty)\right\}$,

then for all $x \ne y \in X$, with $x \perp y$, we have

$$d(Tx, Ty) \le \max\left\{d(x, y), d(x, Tx), d(y, Ty)\right\} - \alpha,$$

hence

$$e^{d(Tx, Ty)} \le k e^{\max\left\{d(x, y), d(x, Tx), d(y, Ty)\right\}}, \tag{12}$$

where $k = e^{-\alpha} < 1$.

Moreover, $x \perp y$ implies

$$p(Tx, Ty) \le k \max\left\{p(x, y), p(x, Tx), p(y, Ty)\right\}, \tag{13}$$

for all $x, y \in X$, where

$$p(x, y) = e^{d(x;y)} - 1$$

is the function mentioned in the formulation of Lemma 3, and by the inequality (13) T is an orthogonal p-continuous mapping. We also have $p(x, x) = 0$ for all $x \in X$.

Now, using Lemma 2 by taking $\phi(t) = kt$ for all $t \in [0, +\infty)$, we deduce from the inequality (13) that T has a unique fixed point.

Theorem 6 is proved.

Corollary [6]. Let $T: X \to X$ be a mapping of a bounded complete metric space (X, d) such that

$$\inf_{x \neq y} \left\{ \max\left\{ d(x,y), d(x,Tx), d(y,Ty) \right\} - d(Tx,Ty) \right\} > 0. \tag{14}$$

Then T has a unique fixed point.

Example 3. Let $X = \{-1, 0\} \cup [1, 2]$ be equipped with the usual metric $d(x, y) = |x - y|$. Suppose that $x \perp y$ if and only if $xy \in \{-1, 0\}$; it is easy to see that $(X, \perp)$ is an O-set. Let us define $T: X \to X$ by the following conditions:

$$Tx = \begin{cases} 0, \ \text{if } x \in \{-1, \ 0\}, \\ 2x, \ \text{if } x \in \left[1, \ \frac{3}{2}\right], \\ \frac{x}{2}, \ \text{if } x \in \left(\frac{3}{2}, \ 2\right]. \end{cases}$$

Then T satisfies all conditions of Theorem 6 and 0 is the unique fixed point. Note that T does not satisfy all conditions (14) given by Corollary of Theorem 6; indeed,

$$\max\left\{ d(0, \ 1), d(0, \ T0), d(1, \ T1) \right\} - d(T0, \ T1) = -1.$$

As applications of Theorem 6 we get a result for a new class of weakly contractive maps defined as follows.

Definition XIV. Let $T: X \to X$ be a mapping of a metric space (X, d), T will be said an orthogonal generalized E-weakly contractive map if $x \perp y$ implies

$$\begin{aligned} d(Tx,Ty) \leq \max\left\{ d(x,y), d(x,Tx), d(y,Ty) \right\} - \\ -\phi\left(1 + \max\left\{ d(x,y), d(x,Tx), d(y,Ty) \right\}\right), \end{aligned} \tag{15}$$

for all $x, y \in X$, where $\phi \in \Phi$ is a function for which the equality $\phi(1) = 0$ and inequality $\inf_{t>1} \phi(t) > 0$ hold.

Theorem 7. *Let $T: X \to X$ be an orthogonal generalized E-weakly contractive mapping of a bounded orthogonal complete metric space $(X, d, \perp)$. Then T has a unique fixed point.*

Proof. Let $x \neq y \in X$ and $x \perp y$, then from Definition XIV, we have

$$\begin{aligned} 0 < \inf_{t>1} \phi(t) \leq \phi\left(1 + \max\left\{ d(x,y), d(x,Tx), d(y,Ty) \right\}\right) \leq \\ \leq \max\left\{ d(x,y), d(x,Tx), d(y,Ty) \right\} - d(Tx,Ty), \end{aligned}$$

and hence

$$\inf_{x \perp y, x \neq y} \left\{ \max\left\{ d(x,y), d(x,Tx), d(y,Ty) \right\} - d(Tx,Ty) \right\} > 0.$$

According to Theorem 6, T has a unique fixed point in X.

Theorem 7 is proved.

Corollary [6]. Let $T: X \to X$ be an orthogonal generalized E-weakly contractive mapping of a bounded orthogonal complete metric space $(X, d, \perp)$. Then T has a unique fixed point.

Example 4. Let $X = \{0, 1, 2, 3\}$ endowed with the usual metric $d(x, y) = |x - y|$. Consider the mapping $T: X \to X$ defined as $T0 = 0 = T1$, $T2 = 3$ and $T3 = 2$.

Define a relation $\perp$ on X by

$$x \perp y \text{ if and only if } xy \leq 1.$$

Then $x \perp y$ implies

$$d(Tx,Ty) \le \max\{d(x,y), d(x,Tx), d(y,Ty)\} - \\ -\phi\left(1+\max\{d(x,y), d(x,Tx), d(y,Ty)\}\right),$$

where $\phi \in \Phi$ is a function defined by

$$\phi(t) = \begin{cases} 0, & if\ t = 1, \\ 1, & if\ t > 1. \end{cases}$$

Therefore, all conditions of Theorem 7 are satisfied, and so T has the unique fixed point 0. On the other hand, since

$$d(T2,\ T3) = 1 > 0 = \max\{d(2,\ 3), d(2,\ T2), d(3,\ T3)\} - \\ -\phi\left(1+\max\{d(2,\ 3), d(2,\ T2), d(3,\ T3)\}\right),$$

the Corollary of Theorem 7 does not ensure the existence of the fixed point.

Summary and an open problem

We have established a fixed point for a new class of contractive mappings as an extension of some results (see Refs. [6, Theorem 4.2] and [3, Theorem 3]). This study was carried out only for orthogonal elements. In light of this, an open problem remains for interested researchers: whether we can generalize these results to "generalized orthogonal sets". For more details on this topic see Refs. [8, 9].

REFERENCES

1. **Aamri M., El Moutawakil D.,** τ-distance in general topological spaces with application to fixed point theory, Southwest J. Pure Appl. Math. (2) (Dec) (2003) 1–5.

2. **Gordji M. E., Rameani M., De La Sen M., Cho Y. J.,** On orthogonal sets and Banach fixed point theorem, Fixed Point Theor. & Appl. 18 (2) (2017) 569−578.

3. **Touail Y., El Moutawakil D., Bennani S.,** Fixed point theorems for contractive self-mappings of a bounded metric space, J. Func. Spaces. 2019 (2019) 4175807.

4. **Touail Y., El Moutawakil D.,** Fixed point results for new type of multivalued mappings in bounded metric spaces with an application, Ric. di Mat. 71 (2) (2022) 315–323.

5. **Touail Y., El Moutawakil D.,** New common fixed point theorems for contractive self-mappings and an application to nonlinear differential equations, Int. J. Nonlinear Anal. Appl. 12 (1) (2021) 903–911.

6. **Touail Y., El Moutawakil D.,** Fixed point theorems for new contractions with application in dynamic programming, Vestnik St. Petersb. Univ. Math. 54 (2) (2021) 2006−2012.

7. **Touail Y., El Moutawakil D.,** Some new common fixed point theorems for contractive selfmappings with applications, Asian-Eur. J. Math. 15 (4) (2021) 2250080.

8. **Touail Y., El Moutawakil D.,** $\perp_{\psi F}$-contractions and some fixed point results on generalized orthogonal sets, Rend. Circ. Mat. Palermo, Ser. 2. 70 (3) (2021) 1459−1472.

9. **Touail Y.,** On multivalued $\perp_{\psi F}$-contractions on generalized orthogonal sets with an application to integral inclusions, Probl. Anal. Issues Anal. 11 (29) (3) (2022) 109−124.

СПИСОК ЛИТЕРАТУРЫ

1. **Aamri M., El Moutawakil D.** τ-distance in general topological spaces with application to fixed point theory // Southwest Journal of Pure and Applied Mathematics. 2003. No. 2. December. Pp. 1–5.

2. **Gordji M. E., Rameani M., De La Sen M., Cho Y. J.** On orthogonal sets and Banach fixed point theorem // Fixed Point Theory and Applications. 2017. Vol. 18. No. 2. Pp. 569−578.

3. **Touail Y., El Moutawakil D., Bennani S.** Fixed point theorems for contractive selfmappings of a bounded metric space // Journal of Function Spaces. 2019. Vol. 2019. Article ID 4175807. https://doi.org/10.1155/2019/4175807.

4. **Touail Y., El Moutawakil D.** Fixed point results for new type of multivalued mappings in bounded metric spaces with an application // Ricerche di Matematica. A Journal of Pure and Applied Mathematics. 2022. Vol. 71. No. 2. Pp. 315–323.

5. **Touail Y., El Moutawakil D.** New common fixed-point theorems for contractive self-mappings and an application to nonlinear differential equations // International Journal of Nonlinear Analysis and Applications. 2021. Vol. 12. No. 1. Pp. 903–911.

6. **Туаль Ю., Аль-Мутавакиль Д.** Теоремы о неподвижной точке для новых сжимающих отображений с приложением в динамическом программировании // Вестник СПБГУ. Математика. Механика. Астрономия. 2021. Т. 8 (66). № 2. С. 338–348.

7. **Touail Y., El Moutawakil D.** Some new common fixed point theorems for contractive selfmappings with applications // Asian-European Journal of Mathematics. 2021. Vol. 15. No. 4. P. 2250080.

8. **Touail Y., El Moutawakil D.** $\perp_{\psi F}$-contractions and some fixed point results on generalized orthogonal sets // The Rendiconti del Circolo Matematico di Palermo. 2021. Series 2. Vol. 70. No. 3. Pp. 1459−1472.

9. **Touail Y.** On multivalued $\perp_{\psi F}$-contractions on generalized orthogonal sets with an application to integral inclusions // Проблемы анализа − Issues of Analysis. 2022. Т. 11 (29). № 3. С. 109−124.

THE AUTHORS

TOUAIL Youssef
Sidi Mohamed Ben Abdellah University, Fès, Morocco
Faculté des Sciences Dhar El Mahraz
Route Imouzzer BP 2626, Fès, 30000, Morocco
youssef9touail@gmail.com
ORCID: 0000-0003-3593-8253

JAID Amine
Sultan Moulay Slimane University, Beni-Mellal, Morocco
Av. Med. V, BP 591, Beni-Mellal, 23000, Morocco
aminejaid1990@gmail.com

EL MOUTAWAKIL Driss
Chouaib Doukkali University, El Jadida, Morocco
6GG6+P89, Av. des Facultés, El Jadida, 24000, Morocco
d.elmoutawakil@gmail.com

СВЕДЕНИЯ ОБ АВТОРАХ

ТУАЙ Юсеф − *аспирант факультета науки и технологий Университета Сиди Мохамеда Бен Абделлы, г. Фес, Марокко.*
Route Imouzzer BP 2626, Fès, 30000, Morocco
youssef9touail@gmail.com
ORCID: 0000-0003-3593-8253

ДЖАЙД Амин − *аспирант Университета Султана Мулая Слимана, г. Бени-Меллал, Марокко.*
Av. Med. V, BP 591, Beni-Mellal, 23000, Morocco
aminejaid1990@gmail.com

АЛЬ-МУТАВАКИЛЬ Дрисс – *PhD, профессор Университета Шуайб Дуккали, г. Эль-Джадида, Марокко*
6GG6+P89, Av. des Facultés, El Jadida, 24000, Morocco
d.elmoutawakil@gmail.com

Received 15.11.2022. Approved after reviewing 26.09.2023. Accepted 26.09.2023.
Статья поступила в редакцию 15.11.2022. Одобрена после рецензирования 26.09.2023. Принята 26.09.2023.

Journal

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL: PHYSICS AND MATHEMATICS

Vol. 16, No. 4, 2023

Founder and publisher: Peter the Great St. Petersburg Polytechnic University

The journal is registered with the Federal Service for Supervision of Communications, Information Technology and Mass Media (Roskomnadzor).
Certificate ПИ ФС77-51457 issued 19.10.2012.

Phone 8 (812) 294-22-85

Website https://physmath.spbstu.ru/

E-mail: physics@spbstu.ru

Typesetting by A.S. Kolgatina

Published 30.01.2024. Format 60x84/8. Digital print.
Printer's sheets Print circulation 1000. Order ID