THE MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY

JOURNAL

Physics and Mathematics

VOLUME 13, No.1, 2020

Peter the Great St. Petersburg
Polytechnic University
2020

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) and the **Russian Science Citation Index** (RSCI) databases.

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

Phone: (812) 294-22-85.

http://ntv.spbstu.ru/physics

Contents

DOI: 10.18721/JPM.13101
УДК 539.21–539.219, 538.9–538.915

THE ANOMALOUS *g*-FACTOR VALUE OF PARAMAGNETIC IRON CENTERS IN THE TOPAZ LATTICE WITH STRONG TETRAGONAL DISTORTION

E.G. Apushkinskiy, B.P Popov, V.P. Saveliev, V.C. Sobolevskiy, L.P. Krukovskaya

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

The theoretical and experimental results of analyzing the electron paramagnetic resonance (EPR) spectra of iron impurity paramagnetic centers in the topaz (aluminum fluorosilicate) lattice are presented. Characteristic defects of the system exhibiting some lines with abnormally large values of *g*-factor (4.33 and 2.66) in the EPR spectra have been found. The experimental results were discussed within the framework of a previously developed model describing a defect involving an impurity iron ion replacing the Al^{3+} or Si^{4+} ion. The "Fe^{3+} – an oxygen vacancy" model is a special case of the complexes with strong tetragonal distortion. The *g*-factors were calculated taking into account the covalent nature of the bonds.

Keywords: EPR spectrum, center symmetry, Hamiltonian, *g*-factor, topaz, tetragonal distortion

Citation: Apushkinskiy E.G., Popov B.P., Saveliev V.P., Sobolevskiy V.C., Krukovskaya L.P. The anomalous g-factor value of paramagnetic iron centers in the topaz lattice with strong tetragonal distortion, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (1) (2020) 4–9. DOI: 10.18721/JPM.13101

АНОМАЛЬНОЕ ЗНАЧЕНИЕ *g*-ФАКТОРА ПАРАМАГНИТНЫХ ЦЕНТРОВ ЖЕЛЕЗА В РЕШЕТКЕ ТОПАЗА С СИЛЬНЫМ ТЕТРАГОНАЛЬНЫМ ИСКАЖЕНИЕМ

Е.Г. Апушкинский, Б.П. Попов, В.П. Савельев, В.К. Соболевский, Л.П. Круковская

Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Российская Федерация

Представлены результаты теоретических и экспериментальных исследований спектров электронного парамагнитного резонанса (ЭПР) примесных центров железа в решетке фтороксиликата алюминия $Al_2SiO_4(OH,F)_2$ – топаза. Обнаружены характерные дефекты системы, приводящие к появлению линий с аномально большими значениями *g*-факторов (4,33 и 2,66) в спектрах ЭПР. Результаты эксперимента обсуждаются в рамках ранее разработанной модели с дефектом при участии примесного иона железа, замещающего ион Al^{3+} или Si^{4+}. Модель «Fe^{3+}– кислородная вакансия» является частным случаем модели комплексов с сильным тетрагональным искажением. В работе приведен расчет *g*-факторов с учетом ковалентного характера химической связи.

Ключевые слова: спектр ЭПР, симметрия центров, топаз, спин-гамильтониан, *g*-фактор, тетрагональное искажение

Ссылка при цитировании: Апушкинский Е.Г., Попов Б.П., Савельев В.П., Соболевский В.К., Круковская Л.П. Аномальное значение *g*-фактора парамагнитных центров железа в решетке топаза с сильным тетрагональным искажением // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2020. Т. 13. № 1. С. 7–14. DOI: 10.18721/JPM.13101

Introduction

Topaz is an aluminum fluorosilicate with the chemical formula $Al_2SiO_4(OH,F)_2$. The structure of topaz consists of SiO_4 groups connecting $Al[O_4(F,OH)_2]$ octahedral chains. Four of the six anions surrounding the Al^{3+} ion belong to oxygen (O^{2-}), and the remaining two to the fluoride ion (F^-) or the hydroxyl group (OH^-).

Topaz has the following lattice parameters, Å:

$$a = 4.6499,\ b = 8.7969,\ c = 8.3909.$$

The color of the crystals can be changed by irradiating them or adding transition metal impurities [1]. However, the coloring mechanisms of topaz are not entirely clear. Irradiation induces complex defects that are unstable. Iron group transition elements present in aluminosilicates prevents the formation of centers generated by ionizing radiation. For this reason, aluminum fluorosilicates appear to be promising materials for radiation dosimetry and radiation-resistant coatings. Since Al_2SiO_4 compounds have good luminescent properties, aluminosilicates with iron group impurities are also interesting as novel materials for laser devices [2]. Furthermore, study of impurity centers in topaz is potentially valuable for fundamental research. Impurity ions can take different charge states due to strong internal electric fields [3, 4]. Considering intrinsic defects in aluminosilicates, we earlier observed an unusual spectrum of electron paramagnetic resonance (EPR) [1, 5]. We discovered three types of iron centers: a Fe(I) center in state S ($3d^5$ electron configuration) with $g = 2.004$ and two Fe(II) and Fe(III) centers with anomalous values of g equaling 4.33 and 2.66. EPR spectra obtained at room temperature for the X band (the frequency $\nu \approx 9.4$ GHz) using a Bruker ER 220D spectrometer are given in [1, 5]. The high intensity of the spectra collected for the samples at room temperature pointed to high concentration of iron impurities ($n \approx 10^{19}\,\text{cm}^{-3}$). A model of the centers that can form with the participation of iron was proposed.

This is primarily the center with the g-factor equal to 2.004. Such an iron ion substitutes aluminum, occupying an octahedral site coordinated by oxygen (Fe(I) center). The iron atom donates its three electrons to bond formation, acquiring the electron configuration $3d^5$ (Fe^{3+}), ground state ^6S. The position of energy levels, their angular dependence and calculation of the $g$factor for this center are given in [5]. The angular dependences of EPR spectra for Fe(II) and Fe(II1) centers suggest their tetrahedral symmetry. The centers are formed when silicon ions are substituted by iron ions. The Fe(II) center with the g-factor equal to 2.66 was an iron–oxygen vacancy complex: $Fe^5 - V_O$. Interacting with an oxygen vacancy, the substituting Fe^{3+} ion ($3d^5$) is shifted from its equilibrium position by $d = 0.544L\cdot \text{tg}\varphi$ in the <110> direction. It was found from analysis of the angular dependence of the EPR spectrum for Fe(II) that the angle φ equals 6°. As a result, the Fe(II) center is shifted by 0.17 E from the center of the tetrahedron.

The Fe(III) center with $g = 4.33$ is formed by the Fe^{4+} ion in $3d^4$ state, substituting silicon at the Si^{4+} site. However, theoretical calculations of anomalous values of the g-factor were not performed in [5].

Our study presents calculations of anomalous values of the g-factor for Fe(II) and Fe(III) centers in a strong crystal field, taking into account bond covalence.

Theoretical calculation of EPR spectra

Following Abragam and Bleaney's classical work [6, 7], let us consider the theory of paramagnetic resonance of iron ions in a cubic field. The angular dependence of EPR spectra [1] indicates that local paramagnetic Fe(II) and Fe(III) centers are located in a crystal field with tetrahedral symmetry. Experimental results confirm that CF splitting exceeds the electron interaction energy. Hund's rule is violated in this case, and the ion is in a low spin configuration. Splitting diagrams for iron ions in a tetrahedral crystal field, taking into account the spin-orbit coupling and tetragonal distortion of the crystal lattice, are shown in Fig. 1.

We use the equivalent spin Hamiltonian to describe the EPR spectrum. In contrast to the spin Hamiltonian used in [5], we take into account the distortion of cubic symmetry of the crystal field due to its axial distortion [8, 9] along the tetragonal axis; the degree of distortion depends on the fine-structure parameter D. In this case, the spin Hamiltonian H is written as follows:

$$H = \beta(\mathbf{HgS}) + \frac{1}{6}a\left\{S_x^4 + S_y^4 + S_z^4 - \right.$$
$$\left. -\frac{1}{5}S(S+1)(3S^2+3S-1)\right\} + \qquad (1)$$
$$+D\left\{S_z^2 - \frac{1}{3}S(S+1)\right\} + \lambda\mathbf{LS},$$

where $\mathbf{H}$ is the applied magnetic field; $\mathbf{S}$ is the full spin of the center, S is its quantum number; $\mathbf{L}$ is the orbital angular momentum, L is the quantum number of the total orbital momentum; β is the Bohr magneton; a, D are the crystal field parameters determining the fine structure of the EPR spectrum; λ is the spin-orbit coupling constant.

The energy levels of allowed transitions were calculated in [5]. Accounting for tetragonal distortion, characterized by the parameter D, generates a change in the energies by $\pm 2D$, $\pm D$. The parameters of the spin Hamiltonian are given in the table.

The structure of the Fe(II) center (includes the Fe^{3+} ion) is determined by the fact that the energy level of d electrons in the crystal field with tetrahedral symmetry is split into a lower doublet (e states) and an upper triplet (t_2 states) with the energy difference denoted as $10Dq$ (see Fig. 1). The positions of resonance transitions in the EPR spectrum of the Fe(II) center indicate that the CF splitting is greater than the spin-spin interaction energy, i.e., $D >> g\beta H$. The lower doublet in the ligand field with tetragonal distortion splits into an orbital triplet with an effective angular momentum $l = 1$, $S = 1/2$. Spin-orbit coupling causes the triplet to split into a series of levels with an effective total momentum

$$J_{eff} = S + 1,\ J,\ S - 1.$$

The diagram for energy level splitting in a crystal field with tetrahedral symmetry taking into account axial distortion is shown in Fig. 1, *a*. It is assumed for iron group transition elements in a strong crystal field [1] that

$$\lambda^2/D \approx 1\ \text{cm}^{-1}.$$

In this case, the energy difference between sublevels with the effective momentum J_{eff} is described by the effective g-factor g_{eff}, which is determined by the expression equivalent to the Landé g-factor [6]:

Table

Parameters of spin Hamiltonian of paramagnetic iron centers in crystal lattice of topaz

Center	g-factor	a, 10^{-2} cm^{-1}	D, 10^{-2} cm^{-1}
Fe(II) $\{3d^5 - V_O\}$	2.66	6.2	3.2
Fe(III) $\{3d^4\}$	4.33	7. 0	3.5

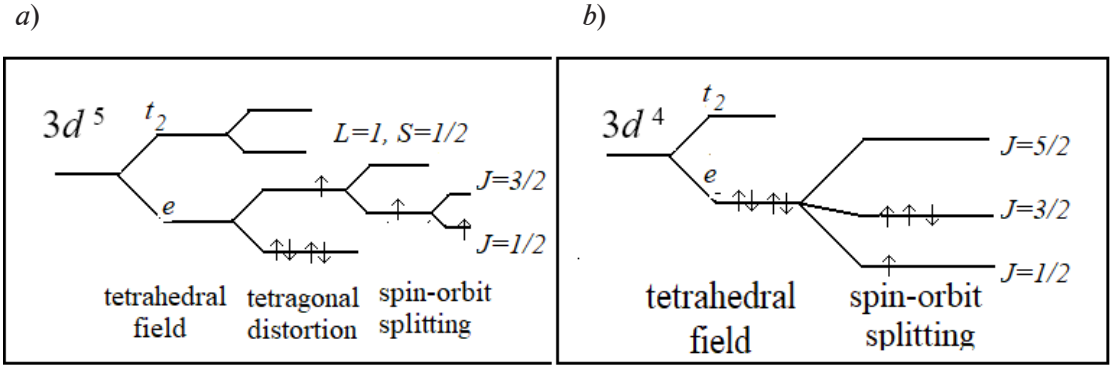

Fig. 1. Energy levels of Fe^{3+} ions in $3d^5$ configuration (*a*) and Fe^{4+} ions in $3d^4$ configuration (*b*) in strong crystal fields with tetrahedral symmetry in the presence of tetragonal lattice distortion (*a*) and spin-orbit interaction (*a*, *b*)

$$g_{eff} = \frac{1}{2}\left(g_s + g_l\right) + \frac{l(l+1) \pm s(s+1)}{2J(J+1)(g_s - g_l)}. \quad (2)$$

Substituting the values $g_l = 1$ and $g_s = 2$ into the formula

$$g_{eff} = \frac{4}{3}g_l + \frac{2}{3}g_s,$$

we obtain the value $g_{eff} = 2.67$, which is in good agreement with the experiment.

The diagram for energy level splitting in a crystal field with tetrahedral symmetry taking into account spin-orbit interaction is shown in Fig. 1, *a*.

The structure of the F(III) center (Fe^{4+} ion) is determined by the fact that it is energetically favorable for electrons to occupy the lower *e* level for the 3^d4 ion in a strong crystal field, as long as this is allowed by the Pauli principle. Consequently, the Fe^{4+} ion is nonmagnetic and its EPR spectrum should not be observed. However, strong spin-orbit coupling can remove spin degeneracy [9, 10]. Three pseudo-*J*-multiplets form, which are characterized by the effective momenta

$$J_{eff} = 1/2,\ 3/2,\ 5/2.$$

The doublet with $J_{eff} = 1/2$ is the ground state because the parameter αλ (spin-orbit coupling constant accounting for chemical bond covalence) is positive. The energy splitting diagram of the Fe^{4+} ion in the $3d^4$ configuration is shown in Fig. 1, *b*. Accounting for spin-orbit interaction changes the gaps between energy levels by

$$\Delta E(\frac{5}{2} \to \frac{3}{2}) = \frac{5}{2}\alpha\lambda,$$

$$\Delta E(\frac{3}{2} \to \frac{1}{2}) = \frac{3}{2}\alpha\lambda.$$

Using the effective total angular momentum allows to calculate the *g*-factor by expression (2), replacing the orbital value of $g_l = 1$ with $g_l = \alpha = -3/2$ [3, 6]. We have for the ground state with $J_{eff} = 1/2$:

$$g_{eff} = \frac{5g_s - 2g_l}{3} = \frac{13}{3} = 4.33.$$

Conclusion

Studying iron impurity centers of iron in topaz by EPR spectroscopy, we found that strong crystal fields make it possible to observe and identify transition ions in different charge states even at room temperature. Iron ions can substitute both Al^{3+} and Si^{4+} ions. The paramagnetic Fe(I) center substitutes aluminum and is located in an oxygen-coordinated octahedral site; the paramagnetic Fe(II) and Fe(III) centers substitute Si_{4+} ions in SiO^4 tetrahedra. The Fe(II) center with oxygen vacancies (V_O) is formed by substituting $Fe^{3+} \to Si^{4+}$, and the Fe(III) center is formed by substituting $Fe^{4+} \to Si^{4+}$. A fragment of the aluminosilicate lattice with tetrahedral oxygen coordination of the iron center and one oxygen vacancy was considered in [5]. Applying procedure for calculating EPR spectra based on representation of the model spin Hamiltonian and effective angular momenta greatly simplified the calculations, yielding good agreement between the experimental data and the theoretical description.

REFERENCES

1. **Apushkinskaya D., Apushkinskiy E., Popov B., et al.,** Analysis of paramagnetic centers for threevalent iron in aluminosilicates, Journal of Physics: Conference Series. 633 (2015) 012115.

2. **Sokolov K., Zhurikhina V., Lipovskii A., et al.,** Spatially periodical poling of silica glass, Journal of Applied Physics. 111 (10) (2012) 104307.

3. **Asatryan G.R., Babunts R.A., Badalyan A.G., et al.,** EPR ionov Yb^{3+} v kristallakh alyuminiyevogo granata [ESR of Yb^{3+} ions in the aluminium garnet crystals], Proceedings of the 24-th International Conference: "Optics and Spectroscopy of Condensed Matter", Kuban State University, Krasnodar, September 22–28 2018, Pp. 47–51.

4. **Yakubenya S.M., Shtel'makh K.F.,** About anomalous *g*-factor value Mn-related defects in GaAs : Mn, Applied Magnetic Resonance. 47 (7) (2016) 671–684.

5. **Apushkinskiy E., Popov B., Romanov V., et al.,** Identification of environment symmetry for iron centers in aluminosilicates, Journal of Physics: Conference Series. 936 (2017) 012011.

6. **Abragam A., Bleaney B.,** Electron paramagnetic resonance of transition ions, Clarendon Press, Oxford, 1970.

7. **Azamat D.V., Basun S.A., Bursian V.É., et al.,** EPR of a non-Kramers iron in $KTaO_3$, Physics of the Solid State. 41 (8) (1999) 1303–1306.

8. **Pleshakov I.V., Klekhta N.S., Kuzmin Yu. I.,** The effect of a pulsed magnetic field on the nuclear spin echo signal in ferrite, Technical Phys. Letters. 38 (9) (2012) 853–855.

9. **Baimova Yu.A., Rysaeva L.Kh., Rudskoi A.I.,** Deformation behavior of diamond-like phases: Molecular dynamics simulation, Diamond and Related Materials. 81 (January) (2018) 154–160.

10. **Politova G.A., Pankratov N.Yu., Vanina P.Yu., et al.,** Magnetocaloric effect and magnetostrictive deformation in Tb-Dy-Gd-Co-Al with Laves phase structure, Journal of Magnetism and Magnetic Materials. 470 (15 January) (2019) 50–54.

Received 06.03.2020, accepted 20.03.2020.

THE AUTHORS

Apushkinskiy Evgeniy G.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
apushkinsky@hotmail.com

Popov Boris P.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
popov_bp@spbstu.ru

Saveliev Vladimir P.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
saveliev@tuexph.stu.neva.ru

Sobolevskiy Vladimir C.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
sobolevskiy@physic.spbstu.ru

Krukovskaya Lidia K.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
lidia.ks@mail.ru

СПИСОК ЛИТЕРАТУРЫ

1. **Apushkinskaya D., Apushkinskiy E., Popov B., Romanov V., Saveliev V., Sobolevskiy V.** Analysis of paramagnetic centers for threevalent iron in aluminosilicates // Journal of Physics: Conference Series. 2015. Vol. 633. P. 012115.

2. **Sokolov K., Zhurikhina V., Lipovskii A., Melehin V., Petrov M.** // Spatially periodical poling of silica glass // Journal of Applied Physics. 2012. Vol. 111. No. 10. P. 104307.

3. **Асатрян Г.Р., Бабунц Р.А., Бадалян А.Г., Единач Е.В., Гурин А.С., Баранов П.Г., Петросян А.Г.** ЭПР ионов Yb^{3+} в кристаллах алюминиевого граната // Материалы XXIV Международной конференции «Оптика и спектроскопия конденсированных сред». Под научной редакцией В.А. Исаева, А.В. Лебедева. Краснодар: Кубанский гос. ун-т, 2018. С. 47–51.

4. **Yakubenya S.M., Shtel'makh K.F.** About anomalous *g*-factor value Mn-related defects in GaAs : Mn // Applied Magnetic Resonance. 2016. Vol. 47. No. 7. Pp. 671–684.

5. **Apushkinskiy E., Popov B., Romanov V., Saveliev V., Sobolevskiy V.** Identification of environment symmetry for iron centers in aluminosilicates // Journal of Physics: Conference Series. 2017. Vol. 936. P. 012011.

6. **Абрагам А., Блини Б.** Электронный парамагнитный резонанс переходных ионов. В 2 тт. Т. 1. М.: Мир, 652 .1972 с.

7. **Азамат Д.В., Басун С.А., Бурсиан В.Э., Раздобарин А.Г., Сочава Л.С., Hesse H., Kapphan S.** ЭПР некрамерсова иона железа в $KTaO_3$ // Физика твердого тела. 1999. Т. 41. Вып. 8. С. 1424–1427.

8. **Плешаков И.В., Клёхта Н.С., Кузьмин Ю.И.** Исследование действия импульсного магнитного поля на сигнал ядерного спинового эха в феррите // Письма в журнал технической физики. 2012. № 18. С. 60 –66.

9. **Baimova Yu.A., Rysaeva L.Kh., Rudskoi A.I.** Deformation behavior of diamond-like phases: Molecular dynamics simulation // Diamond and Related Materials. 2018. Vol. 81. January. Pp. 154–160.

10. **Politova G.A., Pankratov N.Yu., Vanina P.Yu., Filimonov A.V., Rudskoi A.I., Ilyushin A.S., Tereshina I.S.** Magnetocaloric effect and magnetostrictive deformation in Tb-Dy-Gd-Co-Al with Laves phase structure // Journal of Magnetism and Magnetic Materials. 2019. Vol. 470. 15 January. Pp. 50–54.

Статья поступила в редакцию 06.03.2020, принята к публикации 20.03.2020.

СВЕДЕНИЯ ОБ АВТОРАХ

АПУШКИНСКИЙ Евгений Геннадиевич – *доктор физико-математических наук, заведующий кафедрой экспериментальной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
apushkinsky@hotmail.com

ПОПОВ Борис Петрович – *доктор физико-математических наук, профессор кафедры экспериментальной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
popov_bp@spbstu.ru

САВЕЛЬЕВ Владимир Павлович – *старший преподаватель кафедры экспериментальной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
saveliev@tuexph.stu.neva.ru

СОБОЛЕВСКИЙ Владимир Константинович – *кандидат физико-математических наук, доцент кафедры экспериментальной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
sobolevskiy@physic.spbstu.ru

КРУКОВСКАЯ Лидия Петровна – *кандидат физико-математических наук, доцент кафедры экспериментальной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
lidia.ks@mail.ru

DOI: 10.18721/JPM.13102
УДК 536.25

DIRECT NUMERICAL SIMULATION OF THE TURBULENT RAYLEIGH – BÉNARD CONVECTION IN A SLIGHTLY TILTED CYLINDRICAL CONTAINER

S.I. Smirnov, E.M. Smirnov

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

Results of direct numerical simulation of the turbulent convection in a bottom-heated cylindrical container have been presented. The height-to-diameter ratio was equal to 1.0. The calculations were performed for two media: mercury (Pr = 0.025) and water (Pr = 6.4) at Ra = 10^6 and 10^8 respectively. To suppress possible azimuthal movements of the global vortex (large-scale circulation) developing in the container, its axis was tilted a small angle with respect to the gravity vector. Structure of the time-averaged flow pattern symmetrical with respect to the central vertical plane was analyzed. Peculiarities of vortex structures developing in the corner zones were revealed. Representative profiles of the Reynolds stresses and components of the turbulent heat flux vector were obtained for the central vertical plane.

Keywords: Rayleigh – Bénard convection, tilted container, turbulence, direct numerical simulation, large-scale circulation

Citation: Smirnov S.I., Smirnov E.M., Direct numerical simulation of the turbulent Rayleigh – Bénard convection in a slightly tilted cylindrical container, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (1) (2020) 10–20. DOI: 10.18721/JPM.13201

ПРЯМОЕ ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ ТУРБУЛЕНТНОЙ КОНВЕКЦИИ РЭЛЕЯ – БЕНАРА В СЛЕГКА НАКЛОНЕННОМ ЦИЛИНДРИЧЕСКОМ КОНТЕЙНЕРЕ

С.И. Смирнов, Е.М. Смирнов

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

Представлены результаты прямого численного моделирования турбулентной конвекции в подогреваемом снизу цилиндрическом контейнере с высотой, равной диаметру. Расчеты проведены для двух сред: воды (Pr = 6,4) и ртути (Pr = 0,025), при числах Рэлея 10^8 и 10^6 соответственно. Ось контейнера наклонена на небольшой угол по отношению к вектору гравитационного ускорения с целью подавления возможных азимутальных перемещений глобального вихря, развивающегося в контейнере. Анализируется структура осредненного конвективного движения, симметричного относительно центрального вертикального сечения. Выявлены особенности вихревого течения в угловых областях, присущие двум рассмотренным случаям. Получены представительные профили всех ненулевых составляющих тензора рейнольдсовых напряжений и вектора турбулентного теплового потока в центральном сечении.

Ключевые слова: конвекция Рэлея – Бенара, наклоненный контейнер, турбулентность, прямое численное моделирование, крупномасштабная циркуляция

Ссылка при цитировании: Смирнов С.И., Смирнов Е.М. Прямое численное моделирование турбулентной конвекции Рэлея – Бенара в слегка наклоненном цилиндрическом контейнере // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2020. Т. 13. № 1. С. 14–25 DOI: 10.18721/JPM.13201

Introduction

There is much interest in study of natural convection, as it is a phenomenon widely found in nature and technologies. Rayleigh–Bénard convection of fluid in a vertically oriented circular cylindrical container is one of the most attractive model problems in this field.

Diverse experimental and numerical studies found that large-scale circulation (LSC) is a characteristic feature of natural convective flow in a cylindrical container heated from below (see, for example, review [1]). If the height of the container is equal to its diameter or close to it, the LSC is a large-scale vortex covering the entire region of convective flow [1–5]. If the container axis is strictly vertical and axisymmetric boundary conditions are imposed, the problem does not have a preferential azimuthal position, and it is reasonable to assume that the global vortex can occasionally move in the azimuthal direction. Experimental studies on Rayleigh–Bénard convection in a circular cylindrical container confirm this, observing slow (ultra-low frequency) changes in LSC orientation, with irregular behavior (see, for example, [3–9]). Liquid metals [3, 4, 9] and water [5–8] are mainly used for experimental studies. Evidently, the azimuthal behavior of LSC is governed in each case by very small, difficult-to-control deviations from axial symmetry, typical for laboratory models. This feature of LSC was also observed in multiple numerical experiments on transitional and turbulent regimes of Rayleigh–Bénard convection in cylindrical containers at Prandtl numbers (Pr) characteristic for liquid metals [10–13], water [11] and air [14–16].

Random changes in azimuthal orientation of LSC are not the only feature of a global vortex structure of this type. It was found that LSC exhibits two more types of oscillations, *sloshing* and *torsional*. In addition, LSC can disappear for relatively short periods of time and reappear with a pronounced reorientation (this is known as cessation). These features of LSC were studied experimentally in [16–20]. *Sloshing* and *torsional* oscillations are also reproduced in numerical solutions (see, for example, the recent study [21] and references therein).

Azimuthal instability of LSC makes it difficult to obtain the statistical characteristics of turbulent convection in cylindrical containers heated from below, including averaged three-dimensional fields of physical quantities describing relatively small-scale background turbulence. LSC can be locked in a certain azimuthal position by introducing a stabilizing external factor that does not considerably alter the intensity and structure of the flow. For example, slightly tilting the container may act as such a factor. This approach has been repeatedly used in experimental studies conducted at different Rayleigh numbers (Ra) for media with Pr = 0.025 [3, 4], 0.7–0.8 [17, 18] and Pr = 4–6 [17, 19, 20, 22–24]. The effects from slight tilt of a container filled with a medium with Pr = 0.025, from non-uniform heating of the horizontal wall and the structure of the computational grid in the central plane were numerically studied in [25].

One of the most popular numerical approaches used for describing turbulent natural convection in relatively simple geometrical regions is Direct Numerical Simulation (DNS), resolving all components of turbulent motion (see, for example, [26–34] carried out for media with different Prandtl numbers: Pr = 0.005 [30], 0.02 [26, 30], 0.1–1.0 [26, 27, 29, 32–34] and 6.4 [28, 31]). A notable recent work [32] presented DNS for turbulent Rayleigh–Bénard convection at Pr = 1, Ra = 10^8 in regions with different geometric configurations (including cylindrical), focusing on comparing the predictions of integral heat transfer provided by different software packages.

Our study is dedicated to direct numerical simulation of turbulent convection in a slightly tilted cylindrical container, whose height is equal to its diameter, heated from below. Results were obtained for the Rayleigh number Ra = 10^6 at the Prandtl number Pr = 0.025 and Ra = 10^8 at Pr = 6.4.

Problem statement

We considered turbulent convection of a fluid with constant physical properties in the Boussinesq approximation for a circular cylindrical container heated from below with a 1:1

height-to-diameter ratio of the cylinder ($\Gamma = D/H = 1$). The container was tilted by a small angle, $\varphi = 2^{\circ}$, with respect to the gravity vector (Fig. 1, *a*).

Unsteady fluid motion is described by the following system of equations (1)–(3), including the continuity equation, the Navier–Stokes equations and the convection-diffusion equation.

$$\nabla \cdot \mathbf{V} = 0, \tag{1}$$

$$\frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V} \cdot \nabla)\mathbf{V} = -\frac{1}{\rho}\nabla p + \beta(T_0 - T)\mathbf{g} + \nu\nabla^2\mathbf{V}, \tag{2}$$

$$\frac{\partial T}{\partial t} + (\mathbf{V} \cdot \nabla)T = \chi\nabla^2 T. \tag{3}$$

Here $\mathbf{V} = (V_x, V_y, V_z)$ is the velocity vector in the $x'y'z$ coordinate system; t is the time; p, T, and ρ are the pressure, temperature, and density of the fluid; β, ν, and χ are the coefficients of its thermal expansion, kinematic viscosity, and thermal diffusivity; $\mathbf{g}$ is the gravity vector pointing in the opposite direction from the axis y' and making an angle of 2° with it; T_0 is the fluid temperature under hydrostatic equilibrium.

The solution to system (1)–(3) is obtained in the $x'y'z$ coordinate system, whose axis y' coincides with the axis of the container (see Fig. 1, *a*).

No-flow and no-slip conditions are imposed on all boundaries. Constant temperatures are given for the horizontal walls; it is assumed that the temperature of the top wall (T_c) is lower than the bottom (T_h). The side wall is assumed to be adiabatic.

The dimensionless governing parameters of the problem are the Prandtl number $\mathrm{Pr} = \nu/\chi$ and the Rayleigh number, related as

$$\mathrm{Ra} = \mathrm{Pr}\cdot(V_b H/\nu)^2,$$

where V_b is the characteristic (large-scale) flow velocity (buoyant velocity),

$$V_b = (g\beta\Delta T H)^{0.5}$$

(ΔT is the characteristic temperature difference between the hot (T_h) and the cold (T_c) wall),

$$\Delta T = T_h - T_c.$$

Let us also introduce the Grashof number $\mathrm{Gr} = \mathrm{Ra}/\mathrm{Pr}$, whose square root acts as the equivalent of the Reynolds number in natural convection problems.

The computations below were performed for $\mathrm{Pr} = 0.025$, $\mathrm{Ra} = 10^6$ and $\mathrm{Pr} = 6.4$, $\mathrm{Ra} = 10^8$. The Grashof numbers for these two cases are of the same order and equal $4.0\cdot10^7$ and $1.6\cdot10^7$, respectively.

Computational aspects

The computations were carried out using one of the latest versions of the in-house finite-volume code SINF/Flag-S developed at Peter the Great Polytechnic University (the computational algorithms implemented in the code run on unstructured grids). We used a variation of the fractional step method described in [35]. The Crank–Nicolson scheme with second-order accuracy was used for advancing in time. A central difference scheme was used to approximate the convection and diffusion terms in the continuity equations. The computational grid consisted of approximately

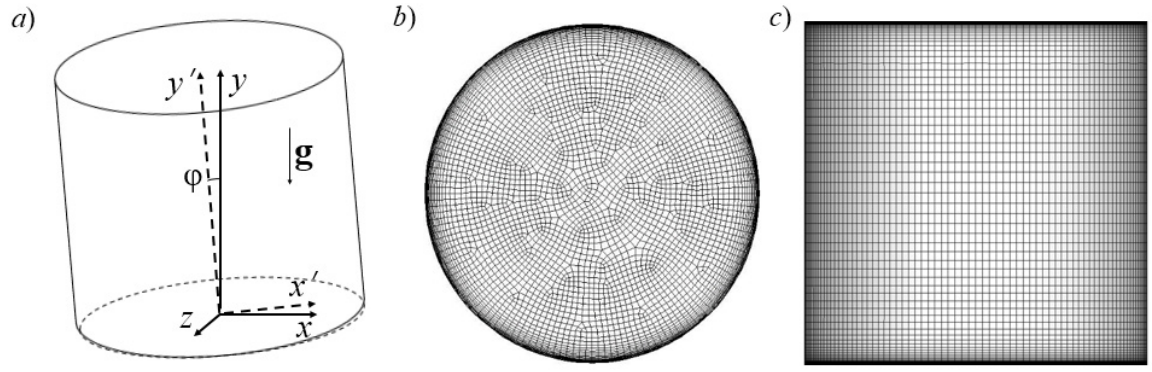

Fig. 1. Geometry of computational domain for tilted container (*a*), grid structure in central horizontal (*b*) and vertical (*c*) planes

$1.5 \cdot 10^7$ hexagonal elements; the grid structure in horizontal and vertical planes is shown in Fig. 1, *b*, *c*. The grid is refined near the walls, while the size of the first near-wall cell was about 10^{-4} *H*. A characteristic feature of the computational grid was a central unstructured (asymmetric) region with a diameter of about $0.8D$ (see Fig. 1, *b*).

The finite-volume computations of Rayleigh–Bénard convection on this grid can be interpreted as direct numerical simulation of turbulence if the local cell size is sufficiently small compared to the size of the smallest vortices in the given region. It is well known the Kolmogorov scale is the smallest scale of turbulent flow if the temperature layers are thicker than the velocity layers (Pr < 1):

$$\delta_K = (\nu^3/\varepsilon)^{0.25},$$

where ε is the dissipation rate of turbulent kinetic energy,

$$\varepsilon = \nu \overline{\frac{\partial V_i'}{\partial x_j} \cdot \frac{\partial V_i'}{\partial x_j}}$$

(V_i' is the fluctuation of the *i*th velocity component, x_j are the Cartesian coordinates).

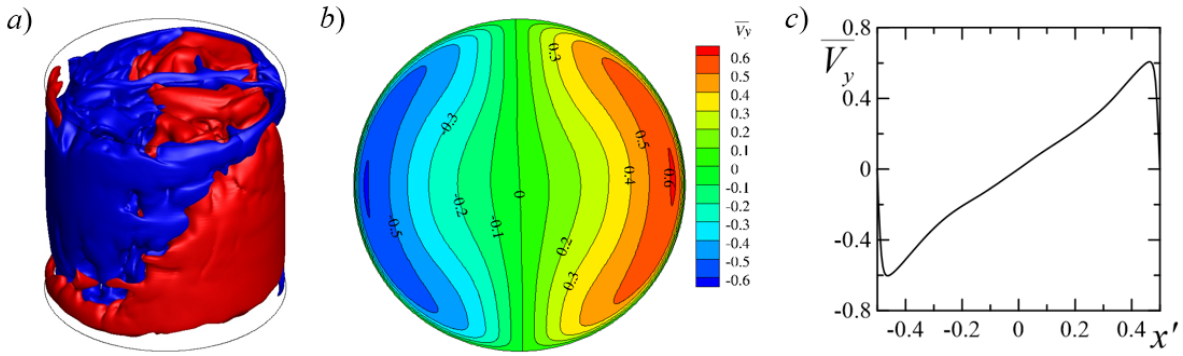

Fig. 2. Distributions of instantaneous (*a*) and averaged (*b*, *c*) vertical velocity components for mercury convection in cylindrical container (CC) heated from below at Pr = 0.025: isosurface components *a*, $|V_y| = 0.14$; distribution *b* in central plane perpendicular to CC axis, $y' = 0.5$; distribution *c* along a line lying in ($x'Oy'$) plane of CC at a height of $y' = 0.5$.
Blue structures correspond to downward flow, red to upward flow.

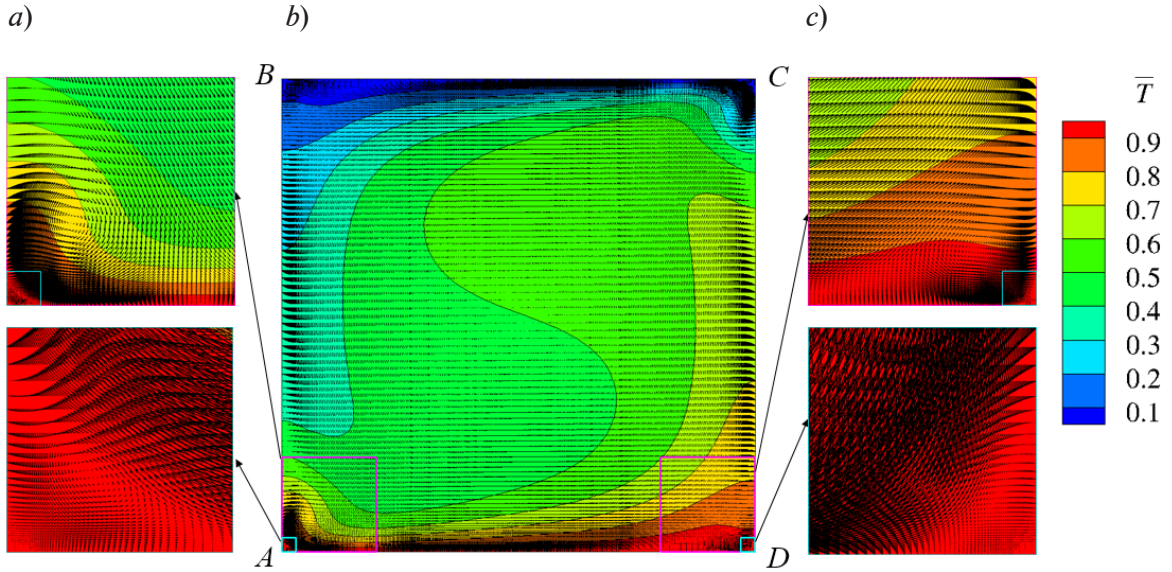

Fig. 3. Averaged temperature field in central vertical plane of container (coinciding with LSC midplane) with superimposed vectors of averaged velocity at Pr = 0.025:
entire convection region of mercury *b*, regions *a*, *c* with corner vortices

If Pr > 1, the smallest scale is the Batchelor scale:

$$\delta_B = \delta_K/\mathrm{Pr}^{0.5}.$$

Accordingly, the quality of the computations can be assessed by comparing the characteristic sizes of the grid elements with different smallest turbulence scales.

The computations started from the zero velocity field and the uniform temperature field, assumed to be equal to $(T_h + T_c)/2$. The time step did not exceed one thousandth of the characteristic time $t_b = H/V_b$, guaranteeing that the local values of the Courant number were less than unity. The computed fields were averaged over time starting after a transient process that lasted about $200t_b$. The samples for averaging were $3000t_b$ for convection of mercury and $4000t_b$ for water.

Computational results and discussion

The quantities **V** and T in the discussion below refer to the velocity and the temperature difference $T - T_c$, related to the corresponding scale (V_b and ΔT), and (x', y', z) refer to the coordinates related to the height of the container.

The quality of grid resolution was assessed after the computations, using the statistics accumulated for the TKE dissipation field. We actually analyzed the fields of the smallest turbulence scales δ_K and δ_B, computed by the above relations and taken relative to the cubic root of the computational cell volume ($V^{1/3}$). It was found that the ratios $\delta_K/V^{1/3}$ and $\delta_B/V^{1/3}$ took values exceeding unity in almost the entire region of the flow. The exceptions were a small area near the side wall, in the layer with the average height, and also the region with the corner vortices, where the smallest values of the ratios $\delta_K/V^{1/3}$ and $\delta_B/V^{1/3}$ were 0.6–0.7.

Results for mercury. Fig. 2, *a* shows an instantaneous distribution for convection of mercury in the cylindrical container heated from below, with pronounced large-scale circulation. Fig. 2, *b*, *c* shows the distribution of the averaged vertical velocity component in the central plane perpendicular to the container axis; the vertical velocity component here and below refers to the velocity component along the axis of a slightly tilted container. Evidently, this distribution has double symmetry, as expected for the case of LSC 'locked' in a certain azimuthal position.

It is of particular interest to explore the characteristic features of convective flow in the central vertical plane of the container, which coincides with the midplane of the LSC (see Fig. 2); this plane is also the $x'Oy'$ plane of the container axis tilt.

ig. 3 shows the vortex structure of the flow and the temperature field in this plane. Notably, aside from LSC, the flow contains several smaller vortices located in the corners of the container. The region occupied by additional

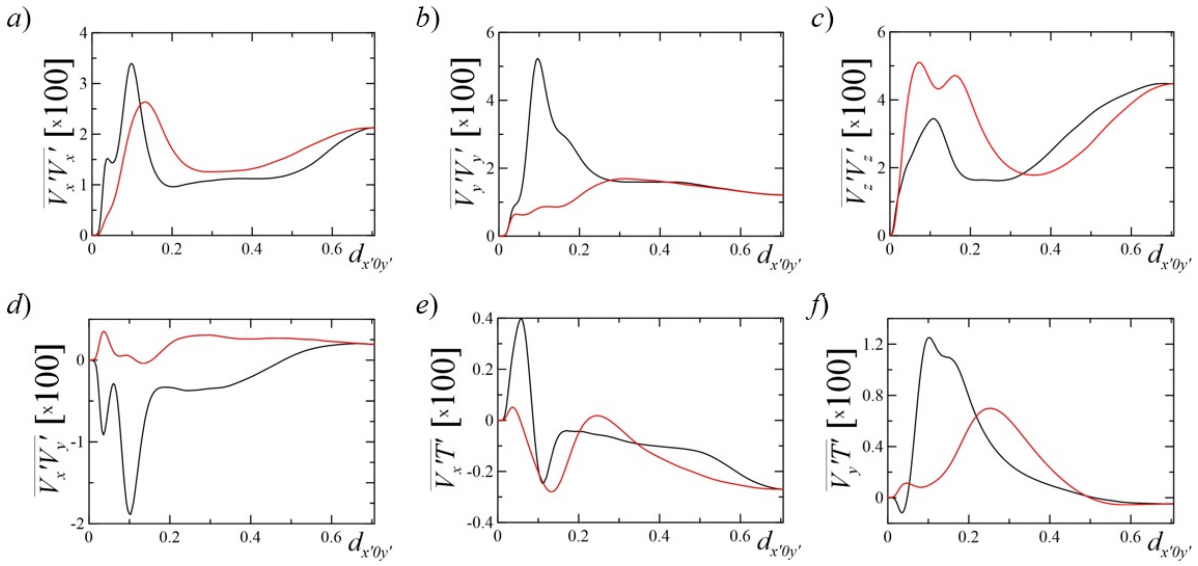

Fig. 4. Distributions of normalized components of Reynolds stress tensor and turbulent heat flux vector along AC (black curve) and BD (red curve) diagonals of container's central plane in case of mercury convection (see Fig. 3), coordinate $d_{x'Oy'} = 0$ at points A and B, respectively

vortices is considerably larger in the corners *A* and *C* (see Fig. 3, *a*) than in the corners *B* and *D* (see Fig. 3, *c*); the intensity of the vortices also differs: it is much higher in the corners *A* and *C*. Gradient layers near the isothermal walls are clearly visible in the temperature field.

We also obtained three-dimensional fields of all components of the Reynolds stress tensor and the turbulent heat flux vector. These data are interesting, in particular, for assessing the capabilities of different second-order turbulence models (Reynolds stress models) used for computations of convective flows based on the Reynolds-averaged Navier–Stokes equations. Data for the central vertical plane of the container are given in this paper. Two of the six components of the Reynolds stress tensor, as well as one of the three components of the turbulent heat flux are equal to zero in this plane due to statistical symmetry of convection.

The distributions of the components of the Reynolds stress tensor and the turbulent heat flux vector are given in Fig. 4 along the diagonals of the central vertical plane (the corresponding coordinate, denoted as $d_{x'Oy'}$, is used). Because the averaged flow is symmetric, distributions are given only for half of the diagonal. Moreover, the given distributions were obtained by averaging over two halves of each of the diagonals (evidently, this technique effectively increases the initial sample for obtaining statistics). Fig. 4 shows that almost all of the given the correlations are close to zero in a small corner region (conditionally, at $d_{x'Oy'} < 0.02$)

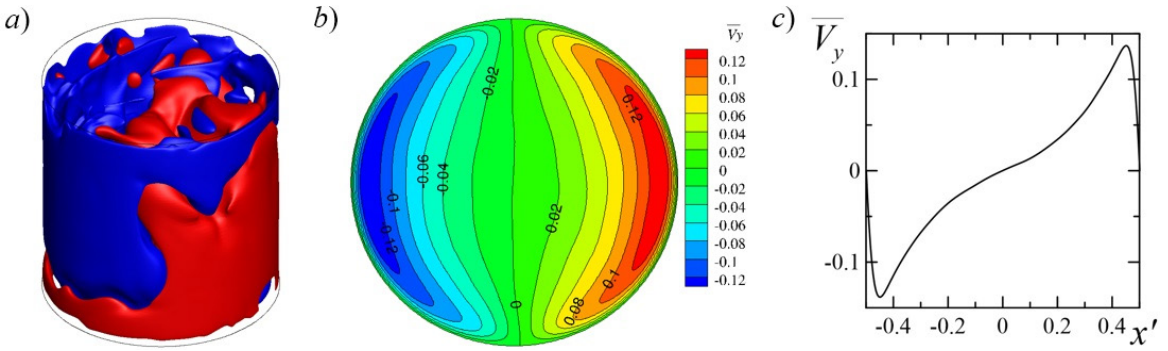

Fig. 5. Distributions of instantaneous (*a*) and averaged (*b*, *c*) vertical velocity components for convection of water in cylindrical container heated from below at Pr = 6.4
The distributions are similar to those shown in Fig. 2 for mercury

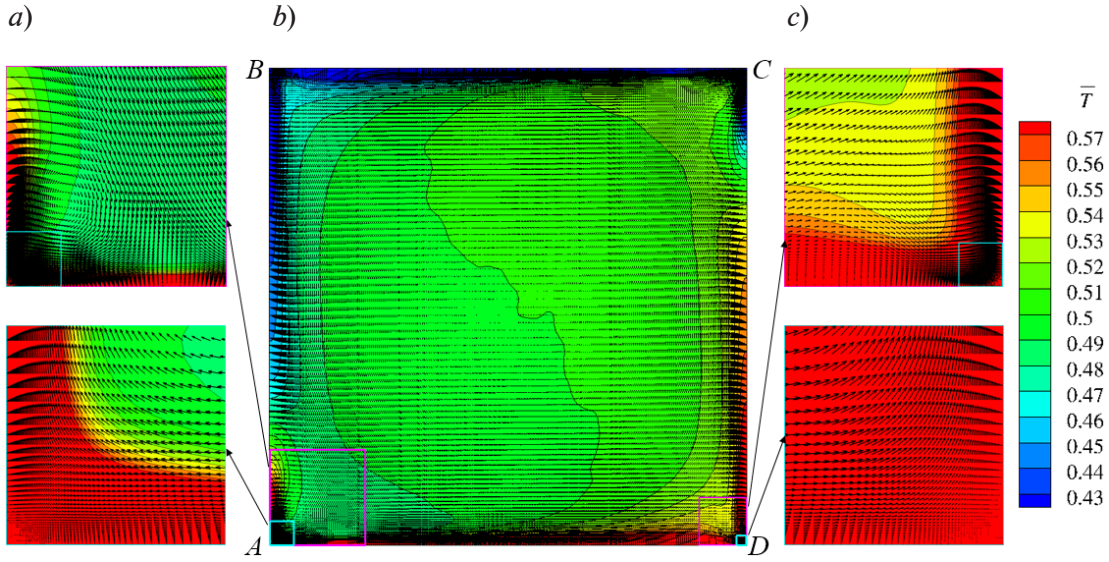

Fig. 6. Averaged temperature field with superimposed vectors of averaged velocity in central vertical plane of container at Pr = 6.4
The distributions are similar to those shown in Fig. 3 for mercury

where there is practically no flow in the medium, with the exception of the Reynolds stress due to fluctuations of the velocity component normal to this plane (see Fig. 4, *c*). With $d_{x'Oy'} > 0.02$, all correlations increase in absolute value to a certain degree, and their variation is essentially nonmonotonic with a further increase in the distance from the corner. This is generally consistent with the picture of the vector velocity field shown in Fig. 3: here, the stable region in the corner is followed by the region where two vortex structures coexist; their presence and interaction determine the nonmonotonic behavior of the distributions shown in Fig. 4. The spans of nonmonotonic segments are somewhat different depending on the choice of the diagonal (AC or *BD*). The region covered by LSC follows the zones occupied by corner vortices ($d_{x'Oy'} > 0.15...0.2$); the correlations change relatively smoothly within this region.

The integral value of the Nusselt number for convection of mercury, obtained as a result of these computations with $\mathrm{Ra} = 10^6$, was $\mathrm{Nu} = 5.64$, which is in good agreement with the results of previous studies carried out in the framework of the implicit LES (ILES) approach: $\mathrm{Nu} = 5.70$ [25], $\mathrm{Nu} = 5.58$ [35], and also with the DNS results, $\mathrm{Nu} = 5.43$ [30].

Results for water. Similar distributions are shown in Figs. 5–7 for convection of water ($\mathrm{Pr} = 6.4$) at $\mathrm{Ra} = 10^8$.

We can conclude from the distributions of the averaged vertical velocity and temperature (Fig. 5) that the solution obtained is also symmetric with respect to the LSC midplane (central vertical plane) in this case. In case of long samples, symmetric statistical characteristics of the flow can be obtained only if the LSC is 'locked' in a certain azimuthal position. Comparing the computational data shown in Figs. 2 and 5, we can establish that the maximum values of the normalized vertical velocity for convection of water are lower than for convection of mercury by approximately five times.

Fig. 6 shows the structure of convective flow of water in the central vertical plane. The same as in in the case of mercury convection considered above (Fig. 3), large-scale circulation of water is complemented by corner vortex structures. However, unlike convection of mercury, there is only one pronounced vortex in each of the corners *A* and *C*, and there are no intense vortices at all in the corners *B* and *D*; the flow turns sharply here, and a small region with very slow motion evolves. These characteristics of water convection in a cylindrical container were earlier discussed in [15]. As expected, high-gradient layers form in the temperature field near the isothermal walls.

Fig. 7 shows the distributions of the nonzero components of the Reynolds stress tensor and the turbulent heat flux vector along the diagonals of the central vertical plane. In

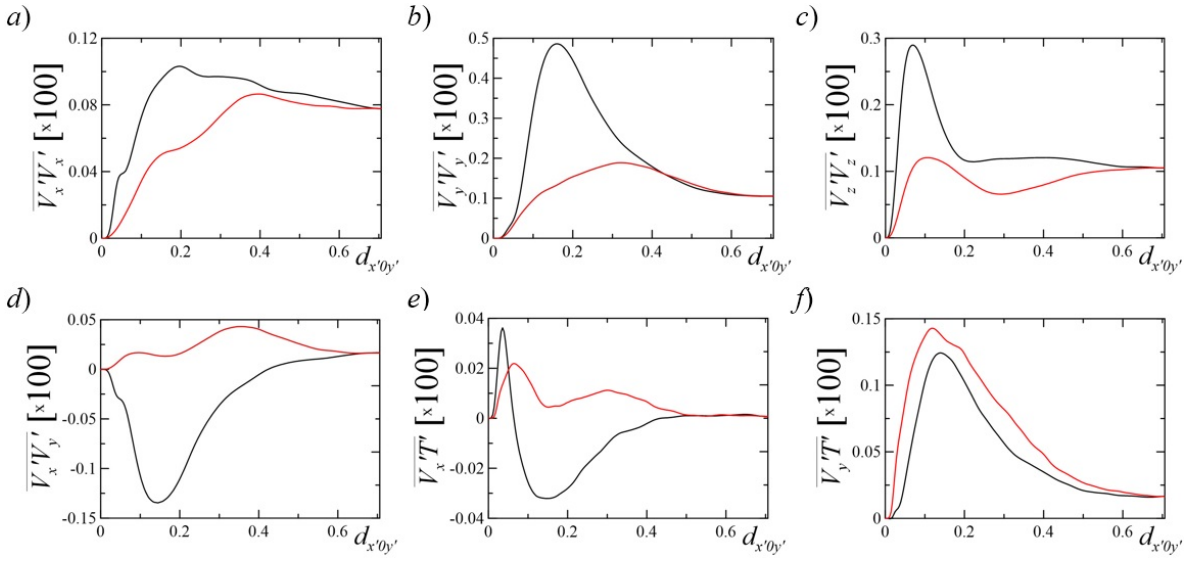

Fig. 7. Distributions of normalized components of Reynolds stress tensor and turbulent heat flux vector along *AC* (black curve) and *BD* (red curve) diagonals of central vertical plane of container for water convection (see Fig. 6), coordinate $d_{x'Oy'} = 0$ at points *A* and *B*, respectively

contrast to convection of mercury, the correlations given for Pr = 6.4 generally change more smoothly; evidently, this is because the corner vortex structures are underdeveloped in case of convection of a fluid with a large Prandtl number. However, segments where the effect of corner vortices can be observed are also visible in this case in the given distributions. Furthermore, compared with the previous case (see Fig. 4), the general level of normalized correlations characterizing the intensity of turbulent transfer is less by about 1 to 1.5 orders of magnitude for convection of water than for a fluid with a small Prandtl number.

The integral Nusselt number obtained for convection of water with $Ra = 10^8$ was Nu = 33.0, which coincides with the results of previous computations [28] performed using the DNS method up to three significant digits.

Conclusion

Direct numerical simulation helped accumulate a large amount of statistical data for essentially three-dimensional turbulent convection in a slightly tilted cylindrical container heated from below, whose height equals the diameter. Computations were carried out for $Ra = 10^6$ at Pr = 0.025 (mercury), and for $Ra = 10^8$ at Pr = 6.4 (water).

We have found that tilting the container axis by 2° relative to the gravity vector allows to reliably 'lock' the global vortex (large-scale circulation (LSC)) in a certain azimuthal position.

The pattern of the averaged flow in the central vertical plane of the container, coinciding with the midplane of the LSC, is characterized by a combination of LSC with corner vortex structures, which are most pronounced for convection of the medium with a small Prandtl number.

We have computed three-dimensional fields of all components of the Reynolds stress tensor and the turbulent heat flux vector. These data can serve, in particular, for assessing the capabilities of different second-order turbulence models (Reynolds stress models) used to compute convective flows based on the Reynolds-averaged Navier–Stokes equations.

The values obtained for the Nusselt integral number are in good agreement with the data given in literature for a container with a vertical axis.

This study was supported by the Russian Foundation for Basic Research (Grant for Vortex-Resolving Numerical Modeling of Turbulent Natural Convection under Conjugate Heat Transfer Conditions no. 17-08-01543).

The computational data were obtained using the resources of the Supercomputer Center at Peter the Great Polytechnic University (www.scc.spbstu.ru).

REFERENCES

1. **Ahlers G., Grossmann S., Lohse D.**, Heat transfer and large scale dynamics in turbulent Rayleigh-Bénard convection, Rev. Mod. Phys. 81 (2) (2009) 503–538.

2. **Takeshita T., Segawa T., Glazier J.A., Sano M.,** Thermal turbulence in mercury, Phys. Rev. Lett. 76 (9) (1996) 1465–1468.

3. **Cioni S., Ciliberto S., Sommeria J.,** Experimental study of high-Rayleigh-number convection in mercury and water, Dyn. Atmos. Oceans. 24 (1) (1996) 117–127.

4. **Cioni S., Ciliberto S., Sommeria J.,** Strongly turbulent Rayleigh–Bénard convection in mercury: comparison with results at moderate Prandtl number, J. Fluid Mech. 335 (1) (1997) 111–140.

5. **Qui X.-L., Tong P.,** Large-scale velocity structures in turbulent thermal convection, Phys. Rev. E. 64 (3) (2001) 036304.

6. **Niemela J.J., Skrbek L., Sreenivasan K.R., Donnelly R.J.,** The wind in confined thermal convection, J. Fluid Mech. 449 (2001) (25 December) 169–178.

7. **Sreenivasan K.R., Bershadskii A., Niemela J.J.,** Mean wind and its reversal in thermal convection, Phys. Rev. E. 65 (5) (2002) 056306.

8. **Brown E., Nikolaenko A., Ahlers G.,** Reorientation of the large-scale circulation in turbulent Rayleigh–Bénard convection, Phys. Rev. Lett. 95 (8) (2005) 084503.

9. **Khalilov R., Kolesnichenko I., Pavlinov A., et al.,** Thermal convection of liquid sodium in inclined cylinders, Phys. Rev. Fluids. 3 (4) (2018) 043503.

10. **Verzicco R., Camussi R.,** Transitional regimes of low-Prandtl thermal convection in a cylindrical cell, Phys. Fluids. 9 (5) (1997) 1287–1295.

11. **Abramov A.G., Ivanov N.G., Smirnov E.M.,** Numerical study of high-Ra Rayleigh–Bénard mercury and water convection in confined enclosures using a hybrid RANS/LES technique, Proc. of the Eurotherm Seminar 74, Eindhoven, TUE. (2003) 33–38.

12. **Schumacher J., Bandaru V., Pandey A., Scheel J.D.,** Transitional boundary layers in low-Prandtl-number convection, Phys. Rev. Fluids. 1 (8) (2016) 084402.

13. **Smirnov S.I., Smirnovsky A.A.,** Numerical simulation of turbulent mercury natural convection in a heated-from-below cylinder with zero and non-zero thickness of the horizontal walls, Thermal Processes in Engineering. 10 (3–4) (2018) 94–100 (in Russian).

14. **Benzi R., Verzicco R.,** Numerical simulations of flow reversal in Rayleigh–Bénard convection, Europhysics Letters. 81 (6) (2008) 64008.

15. **Wagner S., Shishkina O., Wagner C.,** Boundary layers and wind in cylindrical Rayleigh–Bénard cells, J. Fluid Mech. 697 (2012) (25 April) 336–366.

16. **Mishra P.K., De A.K., Verma M.K., Eswaran V.,** Dynamics of reorientations and reversals of large-scale flow in Rayleigh–Bénard convection, J. Fluid Mech. 668 (10 February) (2011) 480–499.

17. **Roche P.-E., Gauthier F., Kaiser R., Salort J.,** On the triggering of the ultimate regime of convection, New J. Phys. 12 (8) (2010) 085014.

18. **He X., van Gils D.P.M., Bodenschatz E., Ahlers G.,** Reynolds numbers and the elliptic approximation near the ultimate state of turbulent Rayleigh–Bénard convection, New J. Phys. 17 (6) (2015) 063028.

19. **Chilla F., Rastello M., Chaumat S., Castaing B.,** Long relaxation times and tilt sensitivity in Rayleigh–Bénard turbulence, Eur. Phys. J. B. 40 (2) (2004) 223–227.

20. **Ahlers G., Brown E., Nikolaenko A.,** The search for slow transients, and the effect of imperfect vertical alignment, in turbulent Rayleigh–Bénard convection, J. Fluid Mech. 557 (25 June) (2006) 347–367.

21. **Zwirner L., Khalilov R., Kolesnichenko I., et al.,** The influence of the cell inclination on the heat transport and large-scale circulation in liquid metal convection, J. Fluid Mech. 884 (10 February) (2020) A18.

22. **Brown E., Ahlers G.,** The origin of oscillations of the large-scale circulation of turbulent Rayleigh–Bénard convection, J. Fluid Mech. 638 (10 November) (2009) 383–400.

23. **Xi H.-D., Zhou S.-Q., Zhou Q., et al.,** Origin of the temperature oscillation in turbulent thermal convection, Phys. Rev. Lett. 102 (4) (2009) 044503.

24. **Weiss S., Ahlers G.,** Effect of tilting on turbulent convection: cylindrical samples with aspect ratio $\Gamma = 0.50$, J. Fluid Mech. 715 (25 January) (2013) 314–334.

25. **Smirnov S.I., Abramov A.G., Smirnov E.M.,** Numerical simulation of turbulent Rayleigh–Bénard mercury convection in a circular cylinder with introducing small deviations from the axisymmetric formulation, J. Phys.: Conf. Ser. 1359 (2019) 012077.

26. **Van der Poel E.P., Stevens R.J.A.M., Lohse D.,** Comparison between two- and three-dimensional Rayleigh–Bénard convection, J. Fluid Mech. 736 (10 December) (2013) 177–194.

27. **Scheel J.D., Schumacher J.,** Local boundary layer scales in turbulent Rayleigh–Bénard convection, J. Fluid Mech. 758 (10 November) (2014) 344–373.

28. **Kooij G.L., Botchev M.A., Geurts B.J.,** Direct numerical simulation of Nusselt number scaling in rotating Rayleigh–Bénard convection, Int. J. Heat Fluid Flow. 55 (October) (2015) 26–33.

29. **Horn S., Shishkina O.,** Toroidal and poloidal energy in rotating Rayleigh–Bénard convection, J. Fluid Mech. 762 (10 January) (2015) 232–255.

30. **Scheel J.D., Schumacher J.,** Global and local statistics in turbulent convection at low Prandtl numbers, J. Fluid Mech. 802 (10 September) (2016) 147–173.

31. **Sakievich P.J., Peet Y.T., Adrian R.J.,** Large-scale thermal motions of turbulent Rayleigh–Bénard convection in a wide aspect-ratio cylindrical domain, Int. J. Heat Fluid Flow. 61 A (October) (2016) 193–196.

32. **Kooij G.L., Botchev M.A., Frederix E.M.A., et al.,** Comparison of computational codes for direct numerical simulations of turbulent Rayleigh–Bénard convection, Computers & Fluids. 166 (30 April) (2018) 1–8.

33. **Zwirner L., Shishkina O.,** Confined inclined thermal convection in low-Prandtl-number fluids, J. Fluid Mech. 850 (10 September) (2018) 984–1008.

34. **Wan Z.-H., Wei P., Verzicco R., et al.,** Effect of sidewall on heat transfer and flow structure in Rayleigh–Bénard convection, J. Fluid Mech. 881 (25 December) (2019) 218–243.

35. **Smirnov S.I., Smirnov E.M., Smirnovsky A.A.,** Endwall heat transfer effects on the turbulent mercury convection in a rotating cylinder, St. Petersburg Polytechnical University Journal. Physics and Mathematics. 3 (2) (2017) 83–94.

Received 20.01.2020, accepted 12.02.2020.

THE AUTHORS

SMIRNOV Sergei I.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
sergeysmirnov92@mail.ru

SMIRNOV Evgueni M.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
smirnov_em@spbstu.ru

СПИСОК ЛИТЕРАТУРЫ

1. **Ahlers G., Grossmann S., Lohse D.** Heat transfer and large scale dynamics in turbulent Rayleigh–Bénard convection // Rev. Mod. Phys. 2009. Vol. 81. No. 2. Pp. 503–538.

2. **Takeshita T., Segawa T., Glazier J.A., Sano M.** Thermal turbulence in mercury // Phys. Rev. Lett. 1996. Vol. 76. No. 9. Pp. 1465–1468.

3. **Cioni S., Ciliberto S., Sommeria J.** Experimental study of high-Rayleigh-number convection in mercury and water // Dyn. Atmos. Oceans. 1996. Vol. 24. No. 1. Pp. 117–127.

4. **Cioni S., Ciliberto S., Sommeria J.** Strongly turbulent Rayleigh–Bénard convection in mercury: comparison with results at moderate Prandtl number // J. Fluid Mech. 1997. Vol. 335. No. 1. Pp. 111–140.

5. **Qui X.-L., Tong P.** Large-scale velocity structures in turbulent thermal convection // Phys. Rev. E. 2001. Vol. 64. No. 3. P. 036304.

6. **Niemela J.J., Skrbek L., Sreenivasan K.R., Donnelly R.J.** The wind in confined thermal convection // J. Fluid Mech. 2001. Vol. 449. 25 December. Pp. 169–178.

7. **Sreenivasan K.R., Bershadskii A., Niemela J.J.** Mean wind and its reversal in thermal convection // Phys. Rev. E. 2002. Vol. 65. No. 5. P. 056306.

8. **Brown E., Nikolaenko A., Ahlers G.** Reorientation of the large-scale circulation in turbulent Rayleigh–Bénard convection // Phys. Rev. Lett. 2005. Vol. 95. No. 8. P. 084503.

9. **Khalilov R., Kolesnichenko I., Pavlinov A., Mamykin A., Shestakov A., Frick P.** Thermal convection of liquid sodium in inclined cylinders // Phys. Rev. Fluids. 2018. Vol. 3. No. 4. P. 043503.

10. **Verzicco R., Camussi R.** Transitional regimes of low-Prandtl thermal convection in a cylindrical cell // Phys. Fluids. 1997. Vol. 9. No. 5. Pp. 1287–1295.

11. **Abramov A.G., Ivanov N.G., Smirnov E.M.** Numerical study of high-Ra Rayleigh-Bénard mercury and water convection in confined enclosures using a hybrid RANS/LES technique // Proc. of the Eurotherm Seminar 74. Eindhoven, TUE, 2003. Pp. 33–38.

12. **Schumacher J., Bandaru V., Pandey A., Scheel J.D.** Transitional boundary layers in low-Prandtl-number convection // Phys. Rev. Fluids. 2016. Vol. 1. No. 8. P. 084402.

13. **Смирнов С.И., Смирновский А.А.** Численное моделирование турбулентной свободной конвекции ртути в подогреваемом снизу цилиндре при нулевой и конечной толщине горизонтальных стенок // Тепловые процессы в технике. 2018. Т. 10. № 3–4. С. 94–100.

14. **Benzi R., Verzicco R.** Numerical simulations of flow reversal in Rayleigh–Bénard convection // Europhysics Letters. 2008. Vol. 81. No. 6. P. 64008.

15. **Wagner S., Shishkina O., Wagner C.** Boundary layers and wind in cylindrical Rayleigh–Bénard cells // J. Fluid Mech. 2012. Vol. 697. 25 April. Pp. 336–366.

16. **Mishra P.K., De A.K., Verma M.K., Eswaran V.** Dynamics of reorientations and reversals of large-scale flow in Rayleigh–Bénard convection // J. Fluid Mech. 2011. Vol. 668. 10 February. Pp. 480–499.

17. **Roche P.-E., Gauthier F., Kaiser R., Salort J.** On the triggering of the Ultimate Regime of convection // New J. Phys. 2010. Vol. 12. No. 8. P. 085014.

18. **He X., van Gils D.P.M., Bodenschatz E., Ahlers G.** Reynolds numbers and the elliptic approximation near the ultimate state of turbulent Rayleigh–Bénard convection // New J. Phys. 2015. Vol. 17. No. 6. P. 063028.

19. **Chilla F., Rastello M., Chaumat S., Castaing B.** Long relaxation times and tilt sensitivity in Rayleigh–Bénard turbulence // Eur. Phys. J. B. 2004. Vol. 40. No. 2. Pp. 223–227.

20. **Ahlers G., Brown E., Nikolaenko A.** The search for slow transients, and the effect of imperfect vertical alignment, in turbulent Rayleigh–Bénard convection // J. Fluid Mech. 2006. Vol. 557. 25 June. Pp. 347–367.

21. **Zwirner L., Khalilov R., Kolesnichenko I., Mamykin A., Mandrykin S., Pavlinov A., Shestakov A., Teimurazov A., Frick P., Shishkina O.** The influence of the cell inclination on the heat transport and large-scale circulation in liquid metal convection // J. Fluid Mech. 2020. Vol. 884. 10 February. P. A18.

22. **Brown E., Ahlers G.** The origin of oscillations of the large-scale circulation of turbulent Rayleigh–Bénard convection // J. Fluid Mech. 2009. Vol. 638. 10 November. Pp. 383–400.

23. **Xi H.-D., Zhou S.-Q., Zhou Q., Chan T.S., Xia K.-Q.** Origin of the temperature oscillation in turbulent thermal convection // Phys. Rev. Lett. 2009. Vol. 102. No. 4. P. 044503.

24. **Weiss S., Ahlers G.** Effect of tilting on turbulent convection: cylindrical samples with aspect ratio $\Gamma = 0.50$ // J. Fluid Mech. 2013. Vol. 715. 25 January. Pp. 314–334.

25. **Smirnov S.I., Abramov A.G., Smirnov E.M.** Numerical simulation of turbulent Rayleigh–Bénard mercury convection in a circular cylinder with introducing small deviations from the axisymmetric formulation // J. Phys.: Conf. Ser. 2019. Vol. 1359. 15–22 September, Yalta, Crimea. P. 012077.

26. **Van der Poel E.P., Stevens R.J.A.M., Lohse D.** Comparison between two- and three-dimensional Rayleigh–Bénard convection // J. Fluid Mech. 2013. Vol. 736. 10 December. Pp. 177–194.

27. **Scheel J.D., Schumacher J.** Local boundary layer scales in turbulent Rayleigh–Bénard convection // J. Fluid Mech. 2014. Vol. 758. 10 November. Pp. 344–373.

28. **Kooij G.L., Botchev M.A., Geurts B.J.** Direct numerical simulation of Nusselt number scaling in rotating Rayleigh–Bénard convection // Int. J. Heat Fluid Flow. 2015. Vol. 55. October. Pp. 26–33.

29. **Horn S., Shishkina O.** Toroidal and poloidal energy in rotating Rayleigh–Bénard convection // J. Fluid Mech. 2015. Vol. 762. 10 January. Pp. 232–255.

30. **Scheel J.D., Schumacher J.** Global and local statistics in turbulent convection at low Prandtl numbers // J. Fluid Mech. 2016. Vol. 802. 10 September. Pp. 147–173.

31. **Sakievich P.J., Peet Y.T., Adrian R.J.** Large-scale thermal motions of turbulent Rayleigh–Bénard convection in a wide aspect-ratio cylindrical domain // Int. J. Heat Fluid Flow. 2016. Vol. 61. Part A. October. Pp. 193–196.

32. **Kooij G.L., Botchev M.A., Frederix E.M.A., Geurts B.J., Horn S., Lohse D., van der Poel E.P., Shishkina O., Stevens R.J.A.M., Verzicco R.** Comparison of computational codes for direct numerical simulations of turbulent Rayleigh–Bénard convection // Computers & Fluids. 2018. Vol. 166. 30 April. Pp. 1–8.

33. **Zwirner L., Shishkina O.** Confined inclined thermal convection in low-Prandtl-number fluids // J. Fluid Mech. 2018. Vol. 850. 10 September. Pp. 984–1008.

34. **Wan Z.-H., Wei P., Verzicco R., Lohse D., Ahlers G., Stevens R.J.A.M.** Effect of sidewall on heat transfer and flow structure in Rayleigh–Bénard convection // J. Fluid Mech. 2019. Vol. 881. 25 December. Pp. 218–243.

35. Смирнов **С.И., Смирнов Е.М., Смирновский А.А.** Влияние теплопереноса в торцевых стенках на турбулентную конвекцию ртути во вращающемся цилиндре // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2017. Т. 10. № 1. С. 31–46.

Статья поступила в редакцию 20.01.2020, принята к публикации 12.02.2020.

СВЕДЕНИЯ ОБ АВТОРАХ

СМИРНОВ Сергей Игоревич – *инженер научно-образовательного центра «Компьютерные технологии в аэродинамике и теплотехнике» Санкт-Петербургского политехнического университета Петра Великого.*

195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
sergeysmirnov92@mail.ru

СМИРНОВ Евгений Михайлович – *доктор физико-математических наук, профессор Высшей школы прикладной математики и вычислительной физики Санкт-Петербургского политехнического университета Петра Великого.*

195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
smirnov_em@spbstu.ru

DOI: 10.18721/JPM.13103
УДК 519.226.2-519.248

A DYNAMIC-STOCHASTIC APPROACH TO THE CONSTRUCTION AND USE OF PREDICTIVE MODELS

Yu.A. Pichugin

Saint-Petersburg State University of Aerospace Instrumentation,
St. Petersburg, Russian Federation

The paper considers two directions of development of the dynamic-stochastic approach to the construction and use of predictive models. The first direction is related to the uncertainty of the initial state of the simulated process, and the second to the stochastic nature of model parameter estimates. In the first case, we consider methods for calculating fast-growing perturbations (FGPs) of the initial state of atmospheric dynamics models and a method for using FGPs in optimizing observation systems based on information ordering. An example of determining the zones of dynamic instability of the Northern hemisphere is given. In the second case, a mathematical apparatus for generating perturbations of model parameters in accordance with their probability distribution is proposed. Based on the data of the USSR economic indices, a numerical example of perturbation of parameter estimates and integration of the Volterra model is given.

Keywords: dynamic model, fast-growing perturbation, distribution of parameter estimates, ensemble of forecasts, economic index.

Citation: Pichugin Yu.A., A dynamic-stochastic approach to the construction and use of predictive models, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (1) (2020) 21–34. DOI: 10.18721/JPM.13103

ДИНАМИКО-СТОХАСТИЧЕСКИЙ ПОДХОД К ПОСТРОЕНИЮ И ИСПОЛЬЗОВАНИЮ МОДЕЛЕЙ ПРОГНОСТИЧЕСКОГО ТИПА

Ю.А. Пичугин

Санкт-Петербургский государственный университет аэрокосмического приборостроения,
Санкт-Петербург, Российская Федерация

В работе рассмотрены два направления развития динамико-стохастического подхода к построению и использованию прогностических моделей. Первое связано с неопределенностью начального состояния моделируемого процесса, а второе – со стохастической природой оценок параметров модели. В первом случае рассмотрены методы вычисления быстрорастущих возмущений начального состояния моделей атмосферной динамики и метод их использования в оптимизации систем наблюдения на основе информационного упорядочивания. Приведен пример определения зон динамической неустойчивости Северного полушария. Во втором случае предложен математический аппарат генерации возмущений параметров модели в соответствии с их вероятностным распределением. На основе данных экономических индексов СССР приведен численный пример возмущения оценок параметров и интегрирования модели Вольтерры.

Ключевые слова: динамическая модель, быстрорастущее возмущение, распределение оценок параметров, ансамбль прогнозов, экономический индекс

Ссылка при цитировании: Пичугин Ю.А. Динамико-стохастический подход к построению и использованию моделей прогностического типа // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2020. Т. 13. № 1. С. 26–41. DOI: 10.18721/JPM.13103

Introduction

The dynamic stochastic approach to forecasting was first developed in meteorology, associated with uncertainty of initial states of predictive models. However, it is well understood that this approach can be easily extended to mathematical modeling in general, since the model parameters estimated using the ordinary least squares method (OLS) have a stochastic nature.

The goals of this study consist, firstly, in developing a method for using fast-growing perturbations of the initial state of a dynamic model to optimize monitoring of any controlled multidimensional process based on information ordering; secondly, in developing a universal method accounting for the stochastic nature of OLS estimates of model parameters to construct a forecasting system allowing to track the dynamics of the probability distribution for the quantities described by the model.

secondly, in developing a universal method accounting for the stochastic nature of OLS estimates of model parameters to construct a forecasting system allowing to track the dynamics of the probability distribution for the quantities described by the model.

These objectives are achieved by solving the following tasks:

describe the key methods for calculating fast-growing perturbations (FGPs) of the initial state of the dynamic model and apply them to the selected optimization;

describe the mathematical tools for generating perturbations based on the probabilistic distribution of their OLS estimates, providing a numerical example for constructing an ensemble of model integrations.

Considering the first task, mainly related to meteorology, we intentionally omit some details of meteorological forecasting so as not to complicate the discussion. For example, considering the errors of measuring the initial state, we do not mention the objective analysis, i.e., interpolation of the measurement data obtained at weather stations to a regular geographic grid. At the same time, we focus closely on mathematical tools, which are not described in sufficient detail in meteorological studies; furthermore, this can allow to transfer these mathematical techniques and methods to predicting other multidimensional processes.

Uncertainty of the initial state, fast-growing perturbations and optimization of observation systems

Judging from the available literature, the dynamic stochastic approach was first applied to constructing prognostic models by Epstein [1], who hypothesized that the stochastic nature of the initial state of the dynamic predictive model, naturally generated by random measurement errors, should be reflected in the result of model integration.

Let $\mathbf{x}(t)$ be the vector of quantities operated by the dynamic model, where t is the time, i.e., $\mathbf{x}(t)$ is the vector used to describe some simulated multidimensional process. Within the purely dynamic approach, we pass from a certain state $\mathbf{x}(t_0)$ to the state $\mathbf{x}(t)$ as a result of integration, where $t = t_0 + \Delta t$, and Δt is the time interval for model integration. In practice, a perturbed initial state always emerges instead of the true initial state $\mathbf{x}(t_0)$,

$$\tilde{\mathbf{x}}(t_0) = \mathbf{x}(t_0) + \Delta\mathbf{x}(t_0),$$

where the perturbation $\Delta\mathbf{x}(t_0)$ is due to measurement errors of the initial state.

Epstein proposed to simulate the spread of perturbations of the initial state $\Delta\mathbf{x}(t_0)$, which would correspond at least to the scale of measurement errors if not to a multidimensional probability distribution. Thus, if there is an ensemble of generated perturbations

$\left\{\Delta\mathbf{x}(t_0)_i\right\}_i^n = 1$, we also have an ensemble

of initial states

$$\left\{\mathbf{x}(t_0)_i = \mathbf{x}(t_0) + \Delta\mathbf{x}(t_0)_i\right\}_{i=1}^n.$$

(see the Remark below).

Integrating a dynamic model from each member of this ensemble of initial states, we obtain a new ensemble:

$\{\mathbf{x}(t)_i\}_{i=1}^{n}$ that is a sample of integration results with the size n. Such a sample makes it possible to estimate the probabilities of certain states of the simulated process, $\mathbf{x}(t)$, if the distribution parameters (presumably normal) are estimated in advance.

Epstein's ideas were first introduced into the practice of meteorological forecasts based on the Monte Carlo method generating perturbations of the initial state [2]. However, very soon, as ensemble forecasts were introduced, fast growing perturbations started to be used. This refers to perturbations which have a (spatial) configuration producing the greatest deviations of the forecast from the result obtained by integration from the measured initial state, while preserving the scale of errors in measurement of the initial state. Using FGPs allows to obtain the largest spread of the forecast ensemble, thus better accounting for the uncertainty due to the error in measuring the initial state. This can be done by several methods.

Let **A** be a real matrix of the model operator linearized in some initial state. It is known from the geometric interpretation of linear operators that the eigenvectors of the matrix $\mathbf{A}^T\mathbf{A}$ (T is the transpose operator) corresponding to the largest eigenvalues of this matrix should be taken as the fastest growing vectors (perturbations). These eigenvectors and the square roots of the eigenvalues of the matrix $\mathbf{A}^T\mathbf{A}$ are known as singular vectors and singular numbers (respectively) of the matrix **A**. It was difficult to apply this to meteorological practices because the dimension of the model (dimension of matrix A) had to be reduced due to limited computational capabilities, at least at the time when this idea was introduced in meteorology. The decrease in dimension naturally leads to smoothing of the initial data, i.e., to inevitable loss of information, which ultimately reduces the effectiveness of this idea [3].

The method based on calculating the eigenvectors of the matrix **A**, corresponding to the eigenvalues largest in magnitude turned out to be relatively easier to implement. There are slight losses here because the largest magnitude of eigenvalues of the matrix **A** does not exceed the largest singular number (see above) of this matrix. The geometrical meaning in this case is that singular numbers can be interpreted as the lengths of the semi-axis of an N dimensional ellipsoid, where a linear operator with the matrix **A** maps an N dimensional sphere of unit radius centered at vector space zero (N is the space dimension). Thus, the singular numbers are the expansion (compression) coefficients along the mutually orthogonal directions of singular vectors; unlike the eigenvectors, singular vectors generally do not preserve their direction, undergoing some rotation in space. The eigenvalue magnitudes are equal to the magnitudes of some segments connecting this ellipsoid with its center.

If the matrix **A** is symmetrical, which happens in case of a self-adjoint operator, then eigenvalues and vectors coincide with the singular values and vectors. Therefore, perturbations proportional to the singular vectors that correspond to the highest singular values can essentially grow faster than the perturbations proportional to the eigenvectors. Therefore, using singular vectors is preferable if the dimension of the model is such that the operator is not self-adjoint and can be linearized without reducing the dimension.

The second approach to calculating FGPs proportional to the eigenvectors of the matrix of the linearized operator of the hydrodynamic model gained great popularity in meteorology. This is because numerical implementation, known as the breeding method [4], is relatively simple. The method is similar to the direct iteration method; the only difference is that multiplication $\mathbf{A}\Delta\mathbf{x}(t_0)$ that this well-known method is based on is replaced by integrating the model over a relatively short time Δt_b (Δt_b is no more than 12 or 24 hours in meteorology), so the operator determined this way can be assumed to be linear. The action of the operator on perturbation $\Delta x_k(t_0)$ in the iterative process of the breeding method is usually formulated as the difference

$$\Delta\mathbf{y}_{k+1}(t_0) = A\left(\mathbf{x}(t_0) + \Delta\mathbf{x}_k(t_0), \Delta t_b\right) - A\left(\mathbf{x}(t_0), \Delta t_b\right) \quad (1)$$

with subsequent normalization

$$\Delta\mathbf{x}_{k+1}(t_0) = \delta\left\|\Delta\mathbf{y}_{k+1}(t_0)\right\|_e^{-1}\Delta\mathbf{y}_{k+1}(t_0), \quad (2)$$

where $A(x(t_0),\Delta t)$ is the result of integration of the model over time t Δ from the initial state $\mathbf{x}(t_0)$; $\|*\|_e$ is the energy norm; δ is the standard perturbation norm (see below); k is the iteration number.

The initial perturbation $\Delta x_0(t_0)$ (if $k = 0$) is chosen arbitrarily but true to scale (the adopted norm).

The scalar product plays an important role in the breeding method. An energy scalar product is commonly used in meteorology.

Let the total energy of the process at time t be expressed in quadratic form with respect to the components of the vector $\mathbf{x}(t)$:

$$E(\mathbf{x}(t)) = \sum_{i=1}^{N} \mu_i x_i^2(t),$$

where μ_i ($i = 1, 2,...,N$, $N = \dim\mathbf{x}$) are the model constants.

Then the energy scalar product of two perturbations $\Delta x'(t)$ and $\Delta x''(t)$ is expressed as [5]

$$\left\langle \Delta\mathbf{x}'(t), \Delta\mathbf{x}''(t) \right\rangle_e = \sum_{i=1}^{N} \mu_i \Delta x_i'(t) \Delta x_i''(t).$$

The magnitude of the perturbation energy norm is

$$\left\| \Delta\mathbf{x}(t) \right\|_e = E^{1/2}(\Delta\mathbf{x}(t)),$$

and the magnitude of the perturbation standard norm δ (see Eq. (2)) is formulated as $\delta = \|\delta\mathbf{x}\|_e$, where the components of the vector $\delta\mathbf{x}$ act as standard measurement errors of the initial state.

Rayleigh relations, which are approximations of eigenvalues and essentially perturbation growth factors, are calculated using the energy scalar product

$$l_{k+1} = \frac{\left\langle \Delta\mathbf{y}_{k+1}(t_0), \Delta\mathbf{x}_k(t_0) \right\rangle_e}{\left\langle \Delta\mathbf{x}_k(t_0), \Delta\mathbf{x}_k(t_0) \right\rangle_e}.$$

When the eigenvectors and eigenvalues of a symmetric real matrix are calculated by direct iterations, Gram–Schmidt orthogonalization should be performed after calculating the first vector (corresponding to the maximum eigenvalue) to calculate subsequent vectors in order to exclude configurations (directions) for eigenvectors already calculated. In case of a symmetric matrix, it is sufficient to perform orthogonalization when each subsequent initial approximation is generated.

Orthogonalization should be performed in each iteration between operating Eqs. (1) and (2) in the breeding method. The subsequent vectors obtained this way can be interpreted as eigenvectors of some self-adjoint approximation of the initial linearized operator, which naturally imposes an additional restriction on the number of growing perturbation vectors. The Rayleigh ratio l_k that stops to grow is taken as the criterion that stops the breeding of perturbations.

Remark. Like most physical, mathematical and natural sciences dealing with real natural processes and phenomena, mathematical modeling has to rely on assumptions in building models, when some obvious discrepancies between the model and the real object have to be neglected. Each of the perturbations is added to the initial state twice with different signs so that the modeled distribution of perturbations is at least symmetric, However, this goal is not fully achieved, since we never have an unperturbed initial state $\mathbf{x}(t_0)$ because simulated perturbations are added to the measurement result already containing errors $\tilde{\mathbf{x}}(t_0)$ (perturbations, see above). Another consideration is that integrating the model over relatively long periods of time (longer than perturbation breeding) is not in fact a linear operator acting on a perturbation. Therefore, simulating an ensemble of normally distributed perturbations (statistically justified perturbations [6]), we do not necessarily obtain an ensemble of normally distributed forecasts.

These issues have to be neglected in meteorology, which is more or less compensated by the fact that the effectiveness of forecasts obtained by averaging over an ensemble of perturbed initial states significantly exceeds the effectiveness of forecasts from the standard initial state. On the other hand, as noted above, the ensemble of forecasts obtained this way allows to estimate the distribution parameters, i.e., to construct the probability distribution and the probability forecast. The dynamic stochastic approach to forecasting implemented in this manner became common practice in meteorological forecasting (in particular, at the Hydrometeorological Center of Russia) in the late 20th century.

Evidently, the fast-growing perturbations (FGPs) calculated by some method depend on the initial state, since the result of linearization of the model operator (see above) depends on the initial state but also significantly depends on the quality of the model used.

Let there be a sample of initial states $\{\mathbf{x}(t_i)\}_{i=1}^n$, obtained by measurements at times $\{t_i\}_{i=1}^n$ covering a sufficiently long period. The sample of FGPs $\{\Delta\mathbf{x}(t_i)\}_{i=1}^n$ with the highest growth coefficient in the breeding interval or (if the dimension allows) corresponding to the largest singular value can be calculated by this sample of initial states. Next, constructing a basis of principal components and a regression of perturbations for this basis by a sample of perturbations (see [7]), we can arrange the

components of the vector $\Delta x(t)$ (i.e., the initial vector $\mathbf{x}(t)$) by decreasing quantity of information (see [8]) relative to the principal components interpreted as hidden factors. If perturbations of only one specific meteorological field, for example, the geopotential H_{500} (the height of the isobaric surface is 500 mbar), surface pressure or surface temperature is considered as $\Delta\mathbf{x}(t)$, then each component of the vector $\Delta\mathbf{x}(t)$ corresponds to a specific point in the geographic grid. Thus, geographical zones where errors of meteorological measurements can lead to significant forecasting errors can be identified. In other words, zones with the largest quantity of information are in fact zones of dynamic instability. This approach was developed in [9] using a hemispheric model of atmosphere circulation by a sample of initial states of volume $n = 216$ and covering a three-year time interval (1999–2001). In this case, the algorithm for calculating FGPs used the formula

$$\Delta\mathbf{y}_{k+1}(t_0) = \\ = A\left(\mathbf{x}(t_0) + \Delta\mathbf{x}_k(t_0), \Delta t\right) - \mathbf{x}(t_0). \quad (1a)$$

The reason why we use Eq. (1a) is that we are interested in the fastest possible deviation from the initial state rather than from the result of integration over a short time Δt_b. Furthermore, using Eq. (1a) accelerates the calculations of fast-growing perturbations, and using the FGPs obtained this way significantly improves the results of ensemble forecasts. The information ordering in [9] was carried out for the perturbations of the field H_{500} as the most important component of atmospheric dynamics. Ref. [9] illustrates the results in [8], published much later. Fig. 1 shows the final result obtained in [9]. The most informative zones marked on the map correspond to known geographical objects (Gulf Stream, Aleutian Islands) commonly believed to significantly impact the atmospheric processes, which, in turn, confirms the validity of the method and the quality of the model used.

Clearly, the technique proposed in [9] can be used to optimize any spatial monitoring systems given a sample of observations and a mathematical model of the controlled process. This information ordering technique (transition from observations to FGPs) is crucial for solving, aside from the problem of control, the problem of forecasts for monitoring systems requiring optimization.

Stochastic nature of model parameter estimates and generation of perturbations corresponding to their probabilistic distribution

Construction of mathematical models of any processes, not necessarily natural, produces the problem of estimating parameters that are not known physical or other constants on the one hand, and are estimated by the OLS method based on the initial data if they are linearly included in the model, on the other hand. Some progress was made in developing the dynamic stochastic approach in mathematical modeling by testing the statistical hypothesis that the true values of the model parameters belong to a region where model integration is Lyapunov-stable (or unstable) [10]. This problem was also solved in [10], subsequently greatly refined and substantiated theoretically in [11].

However, the problem of dynamic stability of the model can be considered from a different standpoint. Instead of checking the statistical hypothesis whether the solution is stable or not with the true values of the parameters, we can assess the degree of possible instability by modeling the spread of parameter estimates in accordance with the distribution obtained. Thus, the next stage in developing the dynamic stochastic approach to constructing predictive models should consist in simulating the distribution of OLS estimates of model parameters, allowing to account for the uncertainty arising from the stochastic nature of these estimates.

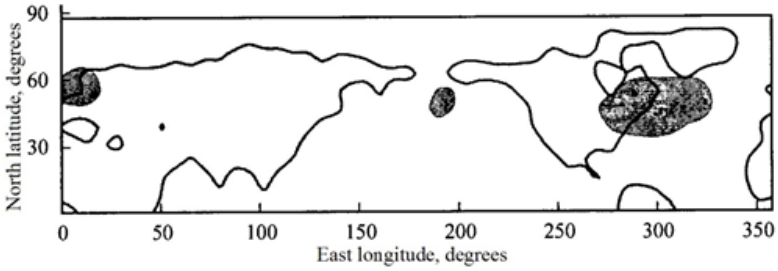

Fig. 1. Zones of Northern Hemisphere associated with dynamic instability

Let us consider the main technical aspects for this approach.

Following [11], we assume that the model parameters are estimated as parameters of a system of regression equations:

$$y_l = \theta_{0l} + \theta_{1l}x_1 + \theta_{2l}x_2 + ... + \theta_{kl}x_k + \varepsilon_l, \quad l = 1,2,...,m, \tag{3}$$

where each equation contains the same set of regressors $\{x_j\}_{j=1}^{k}$ and corresponds to some differential equation of the original model where the parameters to be estimated occur linearly.

In this case, the left-hand sides of equations of system (3) are any expressions, and the variables of the right-hand sides of the system (regressors, see above) are also any expressions whose factors are parameters to be estimated. We will illustrate this below with an example.

Let us assume that there is a sample of all values of variables of size n. By calculating the average values for each of the variables

$$\overline{y}_l = \frac{1}{n}\sum_{i=1}^{n} y_{il},\ l = 1,2,...,m;$$

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij},\ j = 1,2,...,k,$$

we proceed to centered variables

$$y_{il} := y_{il} - \overline{y}_l,\ l = 1,2,...,m;$$

$$x_{ij} := x_{ij} - \overline{x}_j,\ j = 1,2,...,k$$

$$(i = 1,2,...,n),$$

allowing to eliminate the parameters θ_{0l} (l = 1,2,...,m) in system (3)

$$y_l = \theta_{1l}x_1 + \theta_{2l}x_2 + \ldots + \theta_{kl}x_k + \varepsilon_l, \quad l = 1, 2,\ldots, m. \tag{3a}$$

We fill the matrices **Y** and **X** of dimensions $n\times m$ and $n\times k$, respectively, with the centered variables obtained this way. System (3a) takes the following matrix form

$$\mathbf{Y} = \mathbf{X\Theta} + \mathbf{E}, \tag{4}$$

where each lth column of matrix $\mathbf{\Theta}$ is a vector $\boldsymbol{\theta}_l$ of the parameters of an lth equation of centered system (3a); element ε_{il} ($n\times m$) of the matrix **E** is the error of the lth equation substituting ith centered values of the sample, and the OLS estimate of the parameter matrix follows the expression

$$\hat{\mathbf{\Theta}} = \left(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, ..., \hat{\boldsymbol{\theta}}_m\right) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}. \tag{5}$$

The matrices $\mathbf{\Theta}$ and $\hat{\mathbf{\Theta}}$ have the dimension $k\times m$. We use the column of these matrices to construct composite vectors

$$\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, ..., \boldsymbol{\theta}_m^T),$$

$$\hat{\boldsymbol{\theta}}^T = (\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T, ..., \hat{\boldsymbol{\theta}}_m^T).$$

According to OLS theory, if it is assumed that each column of the matrix **E** follows a multidimensional normal distribution, i.e.,

$$\boldsymbol{\varepsilon}_l \sim N(\mathbf{0}, \sigma_l^2\mathbf{I}),$$

where **0** is the zero vector, **I** is the identity matrix, then each column of the matrix $\hat{\mathbf{\Theta}}$ follows the distribution

$$\hat{\boldsymbol{\theta}}_l \sim N(\boldsymbol{\theta}_l, \sigma_l^2(\mathbf{X}^T\mathbf{X})^{-1}),$$

Unbiased estimate σ_l^2 expressed as

$$\hat{\sigma}_l^2 = \frac{1}{n-k-1}(\mathbf{Y}_l - \mathbf{X}\hat{\boldsymbol{\theta}}_l)^T(\mathbf{Y}_l - \mathbf{X}\hat{\boldsymbol{\theta}}_l), \tag{6}$$

where $\mathbf{Y}_l$ is an lth column of the matrix **Y**.

It follows that the composite vector $\hat{\boldsymbol{\theta}}$ also obeys the multidimensional normal distribution $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \mathbf{V}_{\hat{\boldsymbol{\theta}}})$. Therefore, let us consider in detail the construction of matrix $\mathbf{V}_{\hat{\boldsymbol{\theta}}}$.

Let the orthogonal matrix **R** of dimension $m\times m$ produce the diagonal form of the matrix $\mathbf{Y}^T\mathbf{Y}$, i.e., the matrix $\mathbf{R}^T\mathbf{Y}^T\mathbf{Y}\mathbf{R}$ has a diagonal structure. We calculate the matrix

$$\mathbf{Z} = \mathbf{YR},$$

writing a system of regression equations in matrix form, similar to representation (4):

$$\mathbf{Z} = \mathbf{X\Xi} + \mathbf{\Delta}. \tag{7}$$

Next, let us calculate unbiased estimates similar to estimates (5) and (6) ($\mathbf{\Delta}$ is the residual matrix), i.e.,

$$\hat{\mathbf{\Xi}} = \left(\hat{\boldsymbol{\xi}}_1, \hat{\boldsymbol{\xi}}_2, ..., \hat{\boldsymbol{\xi}}_m\right) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Z} \tag{8}$$

and

$$\hat{\delta}_l^2 = \frac{1}{n-k-1}\left(\mathbf{Z}_l - \mathbf{X}\hat{\boldsymbol{\xi}}_l\right)^T\left(\mathbf{Z}_l - \mathbf{X}\hat{\boldsymbol{\xi}}_l\right), \tag{9}$$

where $\mathbf{Z}_l$ corresponds to the lth column of the matrix **Z**, and $\boldsymbol{\xi}_l$ to the lth column of the matrix $\hat{\mathbf{\Xi}}$.

Each column in $\hat{\mathbf{\Xi}}$ follows the multidimensional normal distribution

$$\hat{\xi}_l \sim N(\xi_l, \delta_l^2 (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}),$$

and the composite vector (similar to the vector $\hat{\theta}$) is

$$\hat{\xi} \sim N(\xi, \mathbf{V}_{\hat{\xi}}).$$

The matrix $\mathbf{V}_{\hat{\xi}}$ has the dimension $(mk)\times(mk)$ and block-diagonal structure due to uncorrelated columns **Z**:

$$\mathbf{V}_{\hat{\xi}} = \begin{pmatrix} \delta_1^2(\mathbf{X}^T\mathbf{X})^{-1} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \delta_2^2(\mathbf{X}^T\mathbf{X})^{-1} & ... & \mathbf{0} \\ ... & ... & ... & ... \\ \mathbf{0} & ... & \mathbf{0} & \delta_m^2(\mathbf{X}^T\mathbf{X})^{-1} \end{pmatrix}.$$

Then, following [11], we have:

$$\mathbf{V}_{\hat{\theta}} = \mathbf{R} \otimes \mathbf{I}_{(k)} \mathbf{V}_{\hat{\xi}} (\mathbf{R} \otimes \mathbf{I}_{(k)})^T, \qquad (10)$$

Where $\mathbf{R}\otimes\mathbf{I}_{(k)}$ is the Kronecker product of **R** (see above) and a unit matrix of dimension $k\times k$.

Eq. (10) naturally follows from the equations for estimating composite vectors:

$$\begin{aligned} \hat{\theta} &= \mathbf{R} \otimes \mathbf{I}_{(k)} \hat{\xi}, \\ \hat{\xi} &= (\mathbf{R} \otimes \mathbf{I}_{(k)})^T \hat{\theta}. \end{aligned} \qquad (11)$$

The residuals of regressions (3a) can be composed into a vector

$$\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_m)^T,$$

following the multidimensional normal distribution $\varepsilon \sim N(\mathbf{0}, \mathbf{V}_\varepsilon)$, and a similar vector $\delta \sim N(\mathbf{0}, \mathbf{V}_\delta)$.

Implementations of the vector **ε** are rows of the matrix **E**, and implementations of the vector **δ** are rows of the matrix **Δ** (see Eqs. (4) and (7)). The following equalities hold true for these vectors and their covariance matrices:

$$\begin{aligned} \boldsymbol{\delta} &= \mathbf{R}^T\boldsymbol{\varepsilon},\ \mathbf{V}_\delta = \mathbf{R}^{\mathrm{T}}\mathbf{V}_\varepsilon\mathbf{R}, \\ \boldsymbol{\varepsilon} &= \mathbf{R}\boldsymbol{\delta},\ \mathbf{V}_\varepsilon = \mathbf{R}\mathbf{V}_\delta\mathbf{R}^T. \end{aligned} \qquad (12)$$

It follows from groups of equations (11) and (12) that it is in fact sufficient to calculate estimates (5) and (6), obtaining estimates (8) and (9) using these groups of equations, or, vice versa, calculate only estimates (8) and (9), obtaining (5) and (6) from (11) and (12).

We introduce the following notations for the centered version of model (3a) with calculated OLS estimates for each *l*th equation ($l = 1,2,...,m$)

$$\hat{y}_{il} = \hat{\theta}_{1l}x_{i1} + \hat{\theta}_{2l}x_{i2} + ... + \hat{\theta}_{kl}x_{ik},$$
$$i = 1,2,...,n$$

and calculate the coefficients of determination (squared coefficients of multiple correlation):

$$R_l^2 = \sum_{i=1}^{n} \hat{y}_{il}^2 \Big/ \sum_{i=1}^{n} y_{il}^2. \qquad (13)$$

We can check the hypothesis

$$\mathrm{H}_l: \theta_{l1} = \theta_{l2} = ... = \theta_{lk} = 0$$

using statistics [12]:

$$\gamma_l = \frac{R_l^2(n-k-1)}{\left(1-R_l^2\right)k}, \qquad (14)$$

which, provided that hypothesis H_l is correct, follows the *F* distribution, i.e., $\gamma_l \sim F_{n-k-1,k}$.

Rejecting the hypothesis H_l, we claim that the *l* equation can be included in system (3a), and, therefore, in the initial system (3) ($l = 1,2,...,m$). Notably, adopting matrix formulation (4) for system (3) does not at all require centering of variables. If we did not apply centering, the matrix **X** would have an additional leftmost column filled with units, and the matrix **Θ** would have an additional top row filled with parameters θ_{0l} ($l = 1,2,...,m$). However, the matrix $\mathbf{V}_{\hat{\theta}}$ generally cannot be constructed without regression (7) and estimate (8), which means that centered variables should necessarily be adopted.

The structure of $\mathbf{V}_{\hat{\theta}}$ implies that the orthogonal matrix **Q** reduces $\mathbf{V}_{\hat{\theta}}$ to diagonal form has the following form [11]:

$$\mathbf{Q} = (\mathbf{R} \otimes \mathbf{I}_{(k)})(\mathbf{I}_{(m)} \otimes \mathbf{W}) = \mathbf{R} \otimes \mathbf{W}, \qquad (15)$$

where the orthogonal matrix **W** of dimension $k\times k$ reduces the matrix $\mathbf{X}^T\mathbf{X}$ to diagonal form.

Therefore, we have the equality

$$\mathbf{Q}^T\mathbf{V}_{\hat{\theta}}\mathbf{Q} = \boldsymbol{\Lambda},$$

where **Λ** is a diagonal matrix.

Suppose that the matrix **P** of dimension $h\times s$, where $h = mk$ (see above), contains linearly independent centered rows satisfying the normal distribution test. If **P** is a full-rank matrix, i.e., rank**P** = *h* with $s > h$, then transition to independent (uncorrelated) rows of the matrix does not lead to a decrease in dimension. Therefore, after appropriate transformations and normalization, we can regard the columns of this

matrix $\mathbf{P}_i$ (i = 1,2,...,s) as implementations of the multidimensional normal distribution $\mathbf{P}_i \sim N(\mathbf{0},\mathbf{I})$. We obtain the perturbation ensemble of the parameters $\{\Delta\boldsymbol{\theta}_i\}_{i=1}^{s}$ by the formula

$$\Delta\boldsymbol{\theta}_i = \mathbf{Q}\Lambda^{1/2}\mathrm{P}_i,\ i = 1, 2, \ldots, s, \qquad (16)$$

because the matrix $\mathbf{Q}\boldsymbol{\Lambda}^{1/2}$ transforms the distributions $N(\mathbf{0},\mathbf{I})$ into the distribution $N(\mathbf{0}, \mathbf{V}_{\hat{\boldsymbol{\theta}}})$. In this case, the matrix $\boldsymbol{\Lambda}^{1/2}$ sets the scale of perturbations, and the matrix $\mathbf{Q}$ the dependence corresponding to their distribution.

Returning to the non-centered initial system of equations (3) (to non-centered variables), we need to calculate the estimates of parameters θ_{0l} (l = 1,2,...,m) acting as free terms. Recall that OLS estimates of unperturbed free terms of system (3) satisfy the relations

$$\hat{\theta}_{0l} = \frac{1}{n}\sum_{i=1}^{n}\left(y_{il} - \hat{\theta}_{1l}x_{i1} + \hat{\theta}_{2l}x_{i2} + \ldots + \hat{\theta}_{kl}x_{ik}\right) = \\ = \bar{y}_l - \sum_{j=1}^{k}\hat{\theta}_{jl}\bar{x}_j,\ l = 1,2,\ldots,m, \qquad (17)$$

which can be used for calculations in all cases, including cases of perturbed parameters.

Indeed, let us calculate perturbed values of parameters

$$\tilde{\theta}^i_{jl} = \hat{\theta}_{jl} + \Delta\theta^i_{jl},\quad l = 1,2,\ldots,m, \\ j = 1,2,\ldots,k,\quad i = 1,2,\ldots,s. \qquad (18)$$

It is evident from Eq. (17) that the free terms of equations of system (3) are calculated as arithmetic means. The Statement follows from this.

Statement. For any fixed set of perturbed parameters (see Eq. (18)), substituting these values into Eq. (17) produces an OLS estimate of the free terms of system of equations (3), i.e., the OLS estimate of free terms is

$$\hat{\theta}^i_{0l} = \bar{y}_l - \sum_{j=1}^{k}\tilde{\theta}^i_{jl}\bar{x}_j, \\ l = 1,2,\ldots,m,\ i = 1,2,\ldots,s. \qquad (19)$$

Eq. (19) is necessarily used to calculate the free terms of the model based on the assumption that the perturbations introduced in the estimates of the parameters which are coefficients of the variables satisfactorily account for the stochastic nature of the model.

Reiterating the point made in the Remark to the Section "Uncertainty of initial state, fast-growing perturbations and optimization of observation systems", we should note that what we ultimately simulate is the distribution $N(\hat{\boldsymbol{\theta}}, \mathbf{V}_{\hat{\boldsymbol{\theta}}})$, rather than the distribution $N(\boldsymbol{\theta}, \mathbf{V}_{\hat{\boldsymbol{\theta}}})$, since the true value of the parameter vector $\boldsymbol{\theta}$ remains unknown, and by adding the value of estimate $\hat{\boldsymbol{\theta}}$ to the distribution $N(\boldsymbol{\theta}, \mathbf{V}_{\hat{\boldsymbol{\theta}}})$ (see Eq. (18)), we obtain as a result the distribution $N(\hat{\boldsymbol{\theta}}, \mathbf{V}_{\hat{\boldsymbol{\theta}}})$. Therefore, the term *unperturbed parameters* used below is not quite correct.

Integrating the initial model whose parameters were estimated using regression model (3) for both unperturbed and perturbed parameters, we obtain a time sequence of samples of model elements. Calculating the parameter estimates for the distributions of individual elements or groups of model elements, we take into account the uncertainty associated with estimation of the model parameters, making it possible to estimate the probabilities of certain states of the given process and test certain statistical hypotheses.

Remark. There are two significant problems in implementing the dynamic stochastic approach to construction of predictive models.

The first problem is related to simulating samples belonging to the multidimensional normal distribution $N(\mathbf{0},\mathbf{I})$. Here the dimension is equal to the number of estimated parameters. This requires considerable efforts but is actually achievable.

The second problem is related to the situation when the groups of variables on the right-hand sides in the equations of system (3) are different or only partially coincide, and we cannot use Eqs. (4) and (5), estimating the parameters of each equation separately. This problem is naturally solved when system (3) consists of a single equation or when the left-hand sides of system (3) are independent. In the latter case, $\mathbf{R} = \mathbf{I}$ and $\mathbf{V}_{\hat{\boldsymbol{\theta}}} = \mathbf{V}_{\hat{\boldsymbol{\xi}}}$,, i.e., the mutual covariance matrix of parameter estimates has a block-diagonal structure, where all blocks are different and each block corresponds to one of the equations of system (3). The matrix $\mathbf{Q}$ also has a block-diagonal structure (see Eq. (15)):

$$\mathbf{Q} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \ldots & \mathbf{0} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{W}_m \end{pmatrix}, \qquad (20)$$

where all orthogonal blocks $\mathbf{W}_l$ (l = 1,2, ...,m) are different.

Otherwise (with $\mathbf{R} \neq \mathbf{I}$), the equality $\mathbf{R} = \mathbf{I}$ becomes another assumption that we are forced to adopt.

Finally, we note that the theorem that answers the question whether the true values of the parameters belong to a certain region, formulated and proved in [11], also fully extends to the case of differing right-hand sides of equations of system (3), considered in the above Remark provided that $\mathbf{R} = \mathbf{I}$.

Numerical example

As an example, we consider the construction (estimation of parameters) and integration with the perturbed values of the Volterra model parameters (see [10]), taking a table of indices for output, capital input and labor in the USSR for 1958–1990 ($n = 33$) as initial data for constructing the model; the indices were also used in [11], where all values were given as percentages of the values for 1970 (Table 1).

We assume that the output index is described with sufficient accuracy by the well-known formula of the production function

$$Y = aK^{\alpha 1}L^{\alpha 2}, \tag{21}$$

where the parameters a, α_1 and α_2 were successfully found using the data from Table 1 in [11], and the Volterra model describes the mutual dynamics of capital input and labor, i.e.,

$$\begin{aligned} \frac{\dot{K}}{K} &= \beta_{01} + \beta_{11}K + \beta_{21}L, \\ \frac{\dot{L}}{L} &= \beta_{02} + \beta_{12}K + \beta_{22}L. \end{aligned} \tag{22}$$

This choice was made for several reasons. First, major corrections can be introduced to the problem of statistically correct estimates of parameters α_1 and α_2 of production function (21) considered in [11], related to the estimate of parameter a (see below). Secondly, the Volterra model (22) is an example of minimal dimension, illustrating the proposed mathematical tools. Thirdly, model (22) has the remarkable property that the result of integration never fall beyond positive values (outside the first quadrant) with positive initial values, which is consistent with the nature of the variables included in it. According to model (22) and the data in Table 1, $m = k = 2$ and $n = 33$ for the given example. Notably, the notations for the parameters in model (22) are adopted in accordance with [10] and differ from the notations for models (3) and (3a) in the previous section. However, we left unchanged the rest of the notations given in the previous section, in particular the notations for auxiliary matrices.

Table 1

Output, capital input and labor indices (%) in USSR for 1958–1990 [13]

Year	Y	K	L
1958	43.20	30.83	61.97
1959	46.45	33.94	64.19
1960	50.17	38.10	68.74
1961	53.59	41.59	73.06
1962	56.63	44.97	75.72
1963	58.90	49.79	78.16
1964	64.38	54.31	81.26
1965	68.81	60.16	85.25
1966	74.39	68.11	88.36
1967	80.85	78.00	91.24
1968	87.55	86.68	94.35
1969	91.68	93.01	97.45
1970	100.00	100.00	100.00
1971	105.65	107.84	102.88
1972	109.81	116.64	105.54
1973	119.62	125.98	108.09
1974	125.98	135.32	110.64
1975	131.73	145.69	113.30
1976	139.48	156.78	115.52
1977	145.81	167.69	117.96
1978	153.32	179.45	120.40
1979	157.10	191.56	122.62
1980	164.82	203.74	124.72
1981	173.55	216.58	126.39
1982	186.64	230.20	127.72
1983	195.43	244.73	128.71
1984	203.11	259.80	129.49
1985	206.29	274.32	130.60
1986	211.03	288.73	131.37
1987	214.41	303.50	131.49
1988	223.85	318.14	129.93
1989	229.43	333.94	127.94
1990	220.26	349.49	125.17

Notations: Y is the output of production, K is the capital input, L is the labor resources.
The data for 1970 are taken as 100%.

Recall that the following OLS estimates of the parameters were obtained in [11] after taking the logarithm of Eq. (21): $\hat{\alpha}_1 = 0.631$ and $\hat{\alpha}_2 = 0.260$. To satisfy the conditions

$$\alpha_1 + \alpha_2 = 1, \qquad (23)$$

economists commonly use central projection on a straight line (23) in the plane of the values of these parameters (α_1 and α_2), which gives the values $\alpha_1^e = 0.708$ and $\alpha_2^e = 0.292$.

It was proposed in [11] to take the maximum likelihood point on the straight line (23) for the distribution $N(\hat{\boldsymbol{\alpha}}, \mathbf{V}_{\hat{\boldsymbol{\alpha}}})$, where $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2)^T$, gives the values $\alpha_1^* = 0.585$ and $\alpha_2^* = 0.415$. Verification of the corresponding statistical hypotheses confirmed that the point (the vector $\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*)^T$ of maximum likelihood does not reject the hypothesis H_*: $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ according to any of the standard statistical criteria (χ^2, t, F), and the central projection adopted in economics rejects the hypothesis H_e: $\boldsymbol{\alpha} = \boldsymbol{\alpha}^e$ by all criteria (see [11]).

However, an important point was not discussed in [11], namely, that after corrections are introduced to the estimates of parameters α_1 and α_2, according to the Statement of the previous section, the new value of the coefficient a in Eq. (19) should be calculated, namely, $a = e^{\mu}$, where

$$\hat{\mu} = \frac{1}{33}\sum_{i=1}^{33}(\ln Y_i - \alpha_1^* \ln K_i - \alpha_2^* \ln L_i),$$

giving the following values of the coefficients:

$$\hat{\mu} = 6.13 \cdot 10^{-3},\ a = 1.01.$$

They differ significantly from the initial OLS estimate ($\hat{\mu} = 0.50$, $a = 1.65$).

The time derivatives of investments are approximated by the formulas

$$\dot{K}_1 := \frac{K_2 - K_1}{\Delta t}, \quad \dot{K}_{33} := \frac{K_{33} - K_{32}}{\Delta t},$$

$$\dot{K}_t := \frac{K_{t+1} - K_{t-1}}{2\Delta t}, \quad t = 2,3,\ldots,32.$$

We use the same formulas to approximate the time derivatives of labor L (human resources). In both cases, $\Delta t = 1$ year.

After centering all the values of system (22) by formula (5), we obtain the following values:

$$\begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12} \\ \hat{\beta}_{21} & \hat{\beta}_{22} \end{pmatrix} = 10^{-4} \begin{pmatrix} -1.31 & -1.90 \\ -3.15 & 0.065 \end{pmatrix}, \qquad (24)$$

and the estimates of free terms of system (22), calculated by Eq. (19), are equal to $\hat{\beta}_{01} = 0.129$ and $\hat{\beta}_{02} = 0.050$.

The quality of the estimates obtained is characterized by the determination coefficients (13) $R_1^2 = 0.772$ and $R_2^2 = 0.906$, as well as the statistics (14) $\gamma_1 = 50.7$ and $\gamma_2 = 144.6$, which in both cases significantly exceeds the critical value of F statistics with a significance level $\alpha = 0.01$, equal to $F_{30.2} = 5.39$.

Fig. 2 shows the result of model integration by the Runge–Kutta method with a time step $h = 0.25$ years for unperturbed parameters of system (22).

Next, we confine ourselves to considering the output Y calculated by Eq. (21).

Calculated matrices

$$\mathbf{R} = \begin{pmatrix} 0.760 & -0.650 \\ 0.650 & 0.760 \end{pmatrix},$$

$$\mathbf{W} = \begin{pmatrix} 0.979 & -0.204 \\ 0.204 & 0.979 \end{pmatrix} \qquad (25)$$

confirm that the approach to constructing perturbation parameters discussed in the previous section should be used to the full extent.

Here we omit the parameters of the remaining matrices related to constructing this model. To implement the ensemble of perturbations in Eq. (16), we used a matrix (table of numbers) **P** of dimension 4×25, whose rows are uncorrelated with each other, are normally distributed, and give unbiased estimates of mean value and standard deviation, equal to 0 and 1, respectively.

Fig. 3 shows an ensemble of Y values obtained by integration of our model. The ensemble $\{Y_i(t)\}_{i=0}^{25}$ includes 26 members: 25 for perturbed model parameters ($i = 1, 2,\ldots,25$) and for undisturbed parameter estimates ($i = 0$).

Notably, the ensemble average for 2020, equal to $\overline{Y}(2020) = 174.95\%$, practically coincides with the result of integration for unperturbed parameters, equal to $Y_0(2020) = 173.34\%$ (the relative deviation is less than 1%).

Analysis of the sample of final integration values with perturbed parameters revealed the following. With the number of histogram intervals calculated according to the Sturges rule [14] and equal to five, Pearson's test η for checking the normal distribution is $\eta = 0.650$. The critical value with the significance level $\alpha = 0.05$ and the corresponding number of

degrees of freedom is 9.49. Therefore, we have no reason to reject the normal distribution law. Fig. 4 shows the time dependence of the parameters of the normal distribution $Y(t)$, estimated by the ensemble of all integration results. The figure also shows the time evolution of the boundaries of Student's 95% confidence interval. The growth of the standard deviation corresponds to the fact that wthe degree of uncertainty inevitably increases with an increase in the forecast horizon.

As already noted in the previous section, the parameter estimates calculated by the ensemble of integration results allow to calculate the probabilities of any specific states of the given process and test certain statistical hypotheses. Moreover, we have the time evolution of the estimates obtained, which ultimately, implements the dynamic stochastic approach to constructing predictive models.

Notes regarding model (22)

1. If only the first equation changes in model (22) (see below)

$$\begin{aligned} \dot{K} &= \beta_{01} + \beta_{11}K + \beta_{21}L, \\ \frac{\dot{L}}{L} &= \beta_{02} + \beta_{12}K + \beta_{22}L, \end{aligned} \tag{22a}$$

then, accordingly, the estimates of parameters and statistics of the first equation also change:

$$\hat{\beta}_{01} = -4.083,\ \hat{\beta}_{11} = 0.0175,\ \hat{\beta}_{21} = 0.107,$$
$$R_1^2 = 0.966,\ \gamma_1 = 426.3.$$

Furthermore, the matrix $\mathbf{R}$ changes, which is in this case close to the unit matrix $\mathbf{R} \approx \mathbf{I}$ (coincides with the unit matrix when rounded to the third decimal place). The latter means that $\mathbf{V}_{\hat{\theta}} \approx \mathbf{V}_{\hat{\xi}}$. The results of integration only

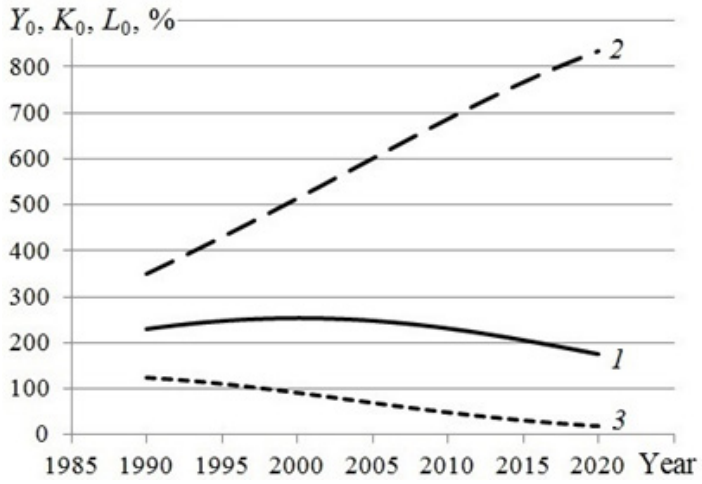

Fig. 2. Model integration by Runge–Kutta method with unperturbed parameters (22):
1 corresponds to production output Y_0, *2* to capital input K_0, *3* to labor resources L_0

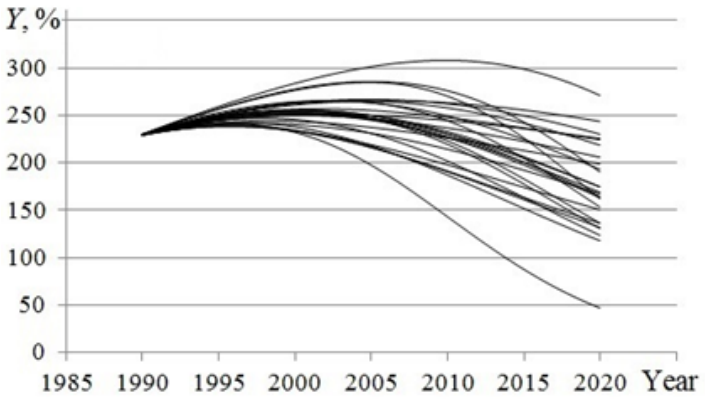

Fig. 3. Ensemble of results for integration of model (22).
The ensemble includes 26 members (curves)

change insignificantly. In this regard, we omit the graphs similar to those shown in Figs. 2–4, giving only a table comparing the results of the integration of these models for 2020 (Table 2).

As follows from Table. 2, transition to model (22a) reduces the standard deviation by 15%. Additionally, this transition worsens the value of Pearson's test checking the normal distribution of the final values of the integration interval, which is in this case $\eta = 3.7$.

2. The approximate equality $\mathbf{R} \approx \mathbf{I}$ mentioned above allows to consider a variant of model (22a) with different right-hand sides, for example, assuming the parameter β_{22} to be zero (excluding L from the right-hand side of the second equation), which indicates that the significance of its estimate is relatively small (see the Remark to the previous section and equality (24)). In this case,

$$\boldsymbol{\beta}^T = (\beta_{11}, \beta_{21}, \beta_{12}),$$

and, according to Eq. (20) and equality (25), the matrix $\mathbf{Q}$ takes the form

$$\mathbf{Q} = \begin{pmatrix} 0.979 & -0.204 & 0 \\ 0.204 & 0.979 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Finally, we should note that models (22) and (22a) can be considered as an alternative to the Solow model well-known in economic science. The final choice of model is left to the researcher.

Conclusion

We have considered the dynamic stochastic approach associated with uncertain initial states of prognostic models in meteorology. We have described all the technical details allowing to apply this approach to forecasting any multidimensional processes.

We have established how fast-growing perturbations (FGPs) of initial states in the dynamic model of a controlled process and the information ordering method can be used by to optimize the monitoring system.

Methods accounting for the stochastic nature of OLS estimates of model parameters have been described. We have proposed an alternative to the previously investigated problem on testing the integration stability hypothesis, which consists in generating a spread of OLS estimates of the parameters with respect to their probability distribution.

The mathematical methods proposed in this study for accounting for the stochastic nature of OLS estimates of dynamic model parameters can be widely used in predicting economic, social, biological, and other processes. A numerical example given confirms the efficiency of this approach.

Table 2

Comparison of integration results for models (22) and (22a) for 2020

Index or estimate	Value, % for model		Relative difference, %
	(22)	(22a)	
K_0	834.00	757.85	–9.13
L_0	19.09	22.99	20.41
Y_0	174.95	178.68	2.13
$\overline{Y}$	173.34	180.29	4.00
$\hat{\sigma}_Y$	47.37	40.21	–15.13

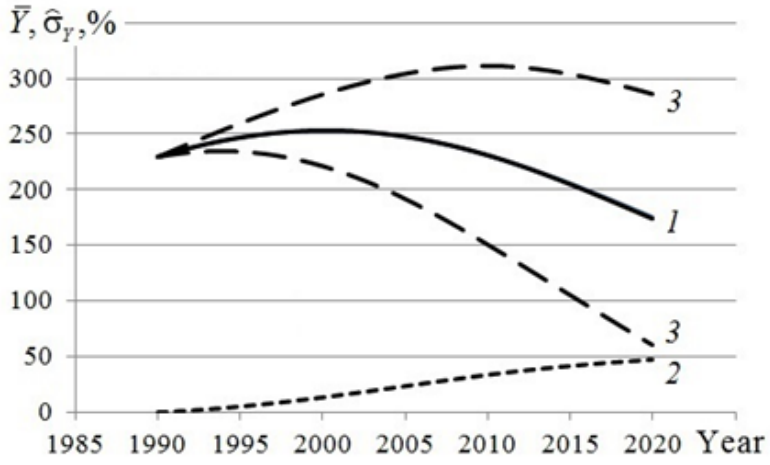

Fig. 4. Normal distribution parameters estimated by ensemble of integrations of model (22): *1* corresponds to the mean value of $\overline{Y}$; *2* to the standard deviation $\hat{\sigma}_Y$; *3* to the boundaries of the 95% confidence interval

REFERENCES

1. **Epstein E.S.**, A scoring system for probability forecast of ranked categories, J. Appl. Meteor. 8 (1969) 985–987.

2. **Leith C.E.,** Theoretical skill of Monte Carlo forecasts, Monthly Weather Review. 102 (6) (1974) 409–418.

3. **Buizza R., Palmer T.N.**, The singular-vector structure of the atmospheric general circulation, J. Atm. Sci. 52 (9) (1995) 1434–1456.

4. **Toth Z., Kalnay E.,** Ensemble forecasting at NCEP and the breeding method, Monthly Weather Review. 125 (12) (1997) 3297–3319.

5. **Pichugin Yu.A., Meleshko V.P., Matyugin V.A., Gavrilina V.M.,** Hydrodynamic long-term weather forecasts with ensemble of initial states, Meteorology and Hydrology. (2) (1998) 5–15.

6. **Astakhova E.D.,** Postroenie ansambley nachalnyh poley dlya sistemy kratko- i srednesrochnogo ansamblevogo prognozirovaniya pogody [Construction of ensembles of initial fields for the system of short-and medium-term ensemble weather forecasting], Proceedings of the Hydrometeorological Center of Russia. (342) (2008) 98–117.

7. **Pichugin Yu.A.,** Notes on using the principal components in the mathematical simulation, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 11 (3) (2018) 74–89.

8. **Pichugin Yu.A.,** The Shannon information quantity in the tasks associated with linear regression: usage pattern, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 12 (3) (2019) 164–176.

9. **Pichugin Yu.A.,** Geografiya dinamicheskoy neustoychivosti tsirkulyatsii atmosfery v Severnom polusharii (modelirovanie i analiz) [Geography of dynamic instability of atmospheric circulation in the Northern hemisphere (simulation and analysis)], Reports of Russian Geographical Society. 137 (3) (2005) 12–16.

10. **Kondrashkov A.V., Pichugin Yu.A.,** On the identification and statistical testing stability of Volterra model, St. Petersburg Polytechnical University Journal. Physics and Mathematics. (1 (189)) (2014) 124–135.

11. **Pichugin Yu.A.,** Geometrical aspects of testing the complex statistical hypotheses in mathematical simulation, St. Petersburg Polytechnical University Journal. Physics and Mathematics. 2 (218) (2015) 123–137.

12. **Seber G.A.F.,** Linear regression analysis, John Wiley & Sons, New York, London, Sydney, Toronto (1977).

13. **Bessonov V.A.,** Problemy postroyeniya proizvodstvennykh funktsiy v rossiyskoy perekhodnoy ekonomike [Problems of construction of production functions in the Russian transitional economy], In the Book.: Bessonov V.A., Tsukhlo S.V., Analiz dinamiki rossiyskoy perekhodnoy ekonomiki [An analysis of the Russian transitional economy], Institute of the Transitional Economy, Moscow, 2002.

14. **Sturges H.,** The choice of a class-interval, J. Amer. Statist. Assoc. 21 (153) (1926) 65–66.

Received 27.01.2020, accepted 25.02.2020.

THE AUTHOR

PICHUGIN Yury A.
Saint-Petersburg State University of Aerospace Instrumentation
61 Bolshaya Morskaya St., St. Petersburg, 190000, Russian Federation
yury-pichugin@mail.ru

СПИСОК ЛИТЕРАТУРЫ

1. **Epstein E.S.** A scoring system for probability forecast of ranked categories // J. Appl. Meteor. 1969. No. 8. Pp. 985–987.

2. **Leith C.E.** Theoretical skill of Monte Carlo forecasts // Monthly Weather Review. 1974. Vol. 102. No. 6. Pp. 409–418.

3. **Buizza R., Palmer T.N.** The singular-vector structure of the atmospheric general circulation // J. Atm. Sci. 1995. Vol. 52. No. 9. Pp.1434–1456.

4. **Toth Z., Kalnay E.** Ensemble forecasting at NCEP and the breeding method // Monthly Weather Review. 1997. Vol. 125. No. 12. Pp. 3297–3319.

5. **Пичугин Ю.А., Мелешко В.П., Матюгин В.А., Гаврилина В.М.** Гидродинамические долгосрочные прогнозы погоды по ансамблю начальных состояний // Метеорология и гидрология. 1998. № 2. С. 5–15.

6. **Астахова Е.Д.** Построение ансамблей начальных полей для системы

кратко- и среднесрочного ансамблевого прогнозирования погоды // Труды Гидрометцентра России. 2008. Вып. 342. С. 98–117.

7. **Пичугин Ю.А.** Замечания к использованию главных компонент в математическом моделировании // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2018. Т. 11. № 3. С. 74–89.

8. **Пичугин Ю.А.** Особенности использования информации по Шеннону в задачах, связанных с линейной регрессией // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2019. Т. 12. № 3. С. 164–176.

9. **Пичугин Ю.А.** География динамической неустойчивости циркуляции атмосферы в Северном полушарии (моделирование и анализ) // Известия Русского географического общества. 2005. Т. 137. Вып. 3. С. 12–16.

10. **Кондрашков А.В. Пичугин Ю.А.** Идентификация и статистическая проверка устойчивости модели Вольтерры // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2014. № 1 (189). С. 124 –135.

11. **Пичугин Ю.А.** Геометрические аспекты проверки сложных статистических гипотез в математическом моделировании // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2015. № 2 (218) С. 123–137.

12. **Себер Дж.** Линейный регрессионный анализ. М.: Мир, 456 .1980 с.

13. **Бессонов В.А.** Проблемы построения производственных функций в российской переходной экономике // Бессонов В.А., Цухло С.В. Анализ динамики российской переходной экономики. М.: Институт экономики переходного периода, 2002. 589 с.

14. **Sturges H.** The choice of a class-interval // J. Amer. Statist. Assoc. 1926. Vol. 21. No. 153. Pp. 65–66.

Статья поступила в редакцию 27.01.2020, принята к публикации 25.02.2020.

СВЕДЕНИЯ ОБ АВТОРЕ

ПИЧУГИН Юрий Александрович – *доктор физико-математических наук, профессор Института инноватики и базовой магистерской подготовки Санкт-Петербургского государственного университета аэрокосмического приборостроения.*

190000, Российская Федерация, Санкт-Петербург, Большая Морская ул., 61.

yury-pichugin@mail.ru

DOI: 10.18721/JPM.13104
УДК 517.51; 517.28; 517.983; 537.213, 537.8

MUTUALLY HOMOGENEOUS FUNCTIONS WITH FINITE-SIZED MATRICES

A.S. Berdnikov[1], *K.V. Solovyev*[2,1], *N.K. Krasnova*[2]

[1] Institute for Analytical Instrumentation of the Russian Academy of Sciences,

St. Petersburg, Russian Federation;

[2] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

This work continues our studies in the properties of the homogeneous Euler' s functions that can be used in the synthesis of electric and magnetic fields for electron and ion-optical systems to carry out spectrographic recording mode. A generalization of a functional general equation for homogeneous functions has been considered. This equation corresponds to linear functional relations with a minimal-sized matrix. A general solution of the obtained functional equation was found assuming of differentiability of the functions in question. The resulting systems of functions were termed mutually homogeneous functions by analogy with the homogeneous Euler's functions and the associated homogeneous Gel'fand's functions.

Keywords: functional equation, associated homogeneous function, mutually homogeneous functions, spectrograph

Citation: Berdnikov A.S., Solovyev K.V., Krasnova, N.K., Mutually homogeneous functions with finite-sized matrices, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (1) (2020) 35–46. DOI: 10.18721/JPM.13104

ВЗАИМНО-ОДНОРОДНЫЕ ФУНКЦИИ С МАТРИЦАМИ КОНЕЧНОГО РАЗМЕРА

А.С. Бердников[1], *К.В. Соловьев*[2,1], *Н.К. Краснова*[2]

[1] Институт аналитического приборостроения Российской академии наук,

Санкт-Петербург, Российская Федерация;

[2] Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Российская Федерация

Данная работа продолжает изучение свойств функций, однородных по Эйлеру, которые можно использовать при синтезе электрических и магнитных полей электронно- и ионно-оптических систем, реализующих спектрографический режим регистрации. Рассматривается обобщение функционального уравнения общего вида для однородных функций, которое соответствует линейным функциональным соотношениям с матрицей минимального размера. В предположении о дифференцируемости рассматриваемых функций найдено общее решение построенного функционального уравнения. Полученные системы функций названы взаимно-однородными по аналогии с однородными функциями Эйлера и присоединенными однородными функциями Гельфанда.

Ключевые слова: функциональное уравнение, присоединенная однородная функция, взаимно-однородные функции, спектрограф

Ссылка при цитировании: Бердников А.С., Соловьев К.В., Краснова Н.К. Взаимно-однородные функции с матрицами конечного размера // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2020. Т. 1 № .13. С. 42–53. DOI: 10.18721/JPM.13104

Introduction

This paper continues a series of studies [1–4] on the properties of homogeneous harmonic functions and applying these functions to synthesis of electric and magnetic fields for electron and ion-optical systems using spectrographic recording [5–8].

Euler-homogeneous functions with the degree of homogeneity equal to p are real functions of several variables satisfying the following identity for any λ [9]:

$$f(\lambda x_1, \lambda x_2,\ldots, \lambda x_n) = \lambda^p f(x_1, x_2,\ldots, x_n). \quad (1)$$

Any Euler-homogeneous function has a one-to-one correspondence taking the form [9]:

$$f(x_1, x_2,\ldots,x_n) = x_1^p g(x_2/x_1, x_3/x_1,\ldots,x_n/x_1), \quad (2)$$

where $g(x_2/x_1, x_3/x_1,..., x_n/x_1) = g(t_2, t_3,..., t_n)$ is a real function of $(n - 1)$ variables.

Accordingly, the only homogeneous function of degree p of one variable is the power function $f(x) = \text{const}\cdot x^p$, and the only homogeneous function of degree zero of one variable is a constant.

If the function $f(x_1, x_2,...,x_n)$ is differentiable, then its partial derivatives with respect to the variables $x_1, x_2,...,x_n$ are homogeneous functions of degree $p - 1$ [9]. Besides, if the function $f(x_1, x_2,...,x_n)$ is differentiable at any point in the space R^n, then the necessary and sufficient condition for this function to be Euler-homogeneous of degree p is that the following condition holds true at any point in the space R^n:

$$fx_1\partial f/\partial x_1 + x_2\partial f/\partial x_2 + \ldots + x_n\partial f/\partial x_n = pf \quad (3)$$

(the Euler theorem on homogeneous functions, also called Euler's criterion for homogeneous functions [9]).

Instead of definition (1), we can consider a functional equation of the form

$$f(\lambda x_1, \lambda x_2,\ldots, \lambda x_n) = a_0(\lambda) f(x_1, x_2,\ldots, x_n) \quad (4)$$

with the previously unknown function $a_0(\lambda)$, which, at first glance, should have more generality than condition (1). However, it soon turns out that if the function $a_0(\lambda)$ is continuous at least at one point, then the only case when Eq. (4) can have non-zero solutions that are of practical interest is a power function $a_0(\lambda) = \lambda^p$. At the same time, although Eq. (4) may have solutions different from the power function $a_0(\lambda) = \lambda^p$ and discontinuous at any point, such solutions are interesting only in the abstract mathematical sense rather than for physical applications.

Indeed, condition (4) implies that the function $a_0(\lambda)$ must satisfy the functional equation

$$\forall\lambda_1,\lambda_2: a_0(\lambda_1\lambda_2) = a_0(\lambda_1)a_0(\lambda_2),$$

since

$$f(\lambda_1\lambda_2 x) = a_0(\lambda_1\lambda_2)f(x) = a_0(\lambda_1) f(\lambda_2 x) =$$

$$= a_0(\lambda_2) f(\lambda_1 x) = a_0(\lambda_1)a_0(\lambda_2) f(x).$$

This equation is a Cauchy multiplicative functional equation. Any solution to this equation has the form of the power function $a_0(\lambda) = \lambda^p$ if the function $a_0(\lambda)$ is continuous at least at one point. The proof of this statement for differentiable functions $a_0(\lambda)$ is obtained in an elementary way after differentiating the relations

$$a_0(\lambda\mu) = a_0(\lambda)a_0(\mu)$$

with respect to μ at the point $\mu = 1$ and the solutions of the corresponding ordinary differential equation.

Homogeneous adjoint Gelfand functions [10, 11], are a generalization of Euler-homogeneous functions; they can be defined as the solution of a semi-infinite system of functional equations:

$$f_0(\lambda x_1, \lambda x_2,\ldots,) = a_0(\lambda) f_0(x_1, x_2,\ldots);$$

$$f_1(\lambda x_1, \lambda x_2,\ldots,) = a_1(\lambda) f_0(x_1, x_2,\ldots) + \quad (5)$$

$$+ a_0(\lambda) f_1(x_1, x_2,\ldots);$$

$$f_2(\lambda x_1, \lambda x_2,\ldots,) = a_2(\lambda) f_0(x_1, x_2,\ldots) +$$

$$+ a_1(\lambda) f_1(x_1, x_2,\ldots,) + a_0(\lambda) f_2(x_1, x_2,\ldots);$$

which must be satisfied for any $\lambda > 0$; at the same time, the functions $a_k(\lambda)$ are unknown in advance.

The general solution for the system of functional equations (5), which has the form of a lower triangular matrix with identical functions $a_k(\lambda)$ along the diagonals, can be rather complex. However, only the so-called main chain of homogeneous adjoint functions can be of practical interest, for which

$$a_k(\lambda) = (1/k!)\,\lambda^p\,(\ln\lambda)^k, \tag{6}$$

$$f_k(x_1, x_2, \dots, x_n) = \\ = (1/k!)(x_1)^p\,(\ln x_1)^k g(x_2/x_1, x_3/x_1, \dots, x_n/x_1), \tag{7}$$

where $g(t_2, t_3, \dots, t_n)$ is an arbitrary real function of $(n-1)$ variables.

The next level of generalization are functions that must satisfy the system of functional equations

$$f_k(\lambda x_1, \lambda x_2, \dots) = \sum a_{kj}(\lambda)\, f_j(x_1, x_2, \dots), \tag{8}$$

where $k = 1, 2, \dots, m$, and the functions $a_{kj}(\lambda)$ are not known in advance.

Such functions, which we called mutually homogeneous, are considered in this paper.

The resulting formulations can have not only theoretical but also practical meaning. In particular, the principle of trajectory similarity, introduced by Golikov [5–8], holds true for Euler-homogeneous electric and magnetic potentials:

if the initial conditions of charged particles are properly scaled, then, provided that the nonrelativistic approximation holds true, article trajectories in such fields are geometrically scaled expressions.

This property allows to synthesize efficient electron and ion-optical systems, for example, such as those obtained in [12–28].

To simplify the calculations, we assume that both the functions $a_{kj}(\lambda)$, and the functions $f_k(x_1, x_2, \dots, x_n)$ are differentiable at any point. Another assumption, valid for Euler-homogeneous functions and for associated homogeneous Gelfand functions, is that imposing the condition that the functions be differentiable at all points can be considerably weakened by replacing it with the condition that the functions be continuous least at one point, producing exactly the same general formulas at the output.

Proving the corresponding theorems is beyond the scope of our study, since the requirement for differentiability at any point is always fulfilled for the scalar potentials of electric and magnetic fields used in electron and ion optics.

Matrix of minimum size

Let us consider a system of functional equations corresponding to a 2×2 matrix (8):

$$f_1(\lambda x_1, \lambda x_2, \dots) = a_{11}(\lambda) f_1(x_1, x_2, \dots) + \\ + a_{12}(\lambda) f_2(x_1, x_2, \dots), \tag{9}$$

$$f_2(\lambda x_1, \lambda x_2, \dots) = a_{21}(\lambda) f_1(x_1, x_2, \dots) + \\ + a_{22}(\lambda) f_2(x_1, x_2, \dots), \tag{10}$$

where the functions $a_{11}(\lambda)$, $a_{12}(\lambda)$, $a_{21}(\lambda)$, $a_{22}(\lambda)$ are not known in advance.

We apply one-to-one substitution of variables:

$$x = \ln x_1,\ t_2 = x_2/x_1,\ t_3 = x_3/x_1, \dots,\ t_n = x_n/x_1.$$

Substituting

$$f_1(x_1, x_2, \dots) = g_1(\ln x_1, x_2/x_1, \dots, x_n/x_1),$$

$$f_2(x_1, x_2, \dots) = g_2(\ln x_1, x_2/x_1, \dots, x_n/x_1)$$

instead of Eqs. (9), (10), we obtain the equivalent functional equations:

$$g_1(x + \ln\lambda, t_2, t_3, \dots, t_n) = \\ = a_{11}(\lambda) g_1(x, t_2, t_3, \dots, t_n) + \\ + a_{12}(\lambda)\, g_2(x, t_2, t_3, \dots, t_n), \tag{11}$$

$$g_2(x + \ln\lambda, t_2, t_3, \dots, t_n) = \\ = a_{21}(\lambda) g_1(x, t_2, t_3, \dots, t_n) + \\ + a_{22}(\lambda) g_2(x, t_2, t_3, \dots, t_n). \tag{12}$$

After differentiating Eqs. (11), (12) with respect to the variable λ at the point $\lambda = 1$, we obtain ordinary linear differential equations with constant coefficients with respect to the variable x:

$$g'_1(x, \dots) = a'_{11}(1) g_1(x, \dots) + \\ + a'_{12}(1) g_2(x, \dots), \tag{13}$$

$$g'_2(x, \dots) = a'_{21}(1) g_1(x, \dots) + \\ + a'_{22}(1) g_2(x, \dots). \tag{14}$$

The form of the analytical solution for Eqs. (13), (14) depends on the class to which the eigenvalues of the matrix $\|a'_{ij}(1)\|$ belong.

Mismatched real eigenvalues. Let the eigenvalues of the matrix (13), (14) be real and not equal to each other. The general solution for system of differential equations (13), (14) has the form

$$g_1(x, t_2, t_3,\ldots, t_n) = c_{11}(t_2, t_3,\ldots, t_n)\exp(p_1 x) + c_{12}(t_2, t_3,\ldots, t_n)\exp(p_2 x),$$

$$g_2(x, t_2, t_3,\ldots, t_n) = c_{21}(t_2, t_3,\ldots, t_n)\exp(p_2 x) + c_{22}(t_2, t_3,\ldots, t_n)\exp(p_1 x),$$

where c_{11}, c_{12}, c_{21}, c_{22} are some functions of $(n-1)$ variables.

In this case, the functions $f_1(x_1, x_2,..., x_n)$ and $f_2(x_1, x_2,..., x_n)$ should have the form

$$f_1(x_1, x_2,\ldots, x_n) = x_1^{p1}c_{11}(x_2/x_1,\ldots, x_n/x_1) + x_1^{p2}c_{12}(x_2/x_1,\ldots, x_n/x_1), \quad (15)$$

$$f_2(x_1, x_2,\ldots, x_n) = x_1^{p1}c_{21}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) + x_1^{p2}c_{22}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1). \quad (16)$$

Because the functions x_1^{p1} and x_1^{p2} are linearly independent, substituting expressions (15) and (16) into conditions (9) and (10) yields the relations

$$\lambda^{p1}c_{11} = a_{11}(\lambda)c_{11} + a_{12}(\lambda)c_{21}, \quad (17)$$

$$\lambda^{p2}c_{12} = a_{11}(\lambda)c_{12} + a_{12}(\lambda)c_{22}, \quad (18)$$

$$\lambda^{p1}c_{21} = a_{21}(\lambda)c_{11} + a_{22}(\lambda)c_{21}, \quad (19)$$

$$\lambda^{p2}c_{22} = a_{21}(\lambda)c_{12} + a_{22}(\lambda)c_{22}. \quad (20)$$

Linear algebraic equations (17), (18) for unknown functions $a_{11}(\lambda)$ and $a_{12}(\lambda)$ cannot be linearly dependent (proportional to each other), except for the degenerate case

$$c_{11} = c_{12} = c_{21} = c_{22} = 0,$$

which is of no practical interest, since the functions λ^{p1} and λ^{p2} are linearly independent.

Similarly, linear algebraic equations (19), (20) for unknown functions $a_{21}(\lambda)$ and $a_{22}(\lambda)$ are also linearly independent.

Therefore, we can assume without loss of generality that

$$\Delta = c_{11}c_{22} - c_{12}c_{21} \neq 0.$$

In this case,

$$a_{11}(\lambda) = \lambda^{p1}(c_{11}c_{22}/\Delta) + \lambda^{p2}(-c_{12}c_{21}/\Delta), \quad (21)$$

$$a_{12}(\lambda) = \lambda^{p1}(-c_{11}c_{12}/\Delta) + \lambda^{p2}(c_{11}c_{12}/\Delta), \quad (22)$$

$$a_{21}(\lambda) = \lambda^{p1}(c_{21}c_{22}/\Delta) + \lambda^{p2}(-c_{21}c_{22}/\Delta), \quad (23)$$

$$a_{22}(\lambda) = \lambda^{p1}(-c_{12}c_{21}/\Delta) + \lambda^{p2}(c_{11}c_{22}/\Delta). \quad (24)$$

Since the functions $a_{11}(\lambda)$, $a_{12}(\lambda)$, $a_{21}(\lambda)$ and $a_{22}(\lambda)$ should not depend on the set of variables $x_1, x_2,...,x_n$, and the functions

$$c_{11}(x_2/x_1,\ldots,x_n/x_1),\ c_{12}(x_2/x_1,\ldots,x_n/x_1),$$

$$c_{21}(x_2/x_1,\ldots,x_n/x_1),\ c_{22}(x_2/x_1,\ldots,x_n/x_1)$$

should not depend on λ, the factors

$$c_{11}c_{22}/\Delta,\ c_{12}c_{21}/\Delta,\ c_{11}c_{12}/\Delta,\ c_{21}c_{22}/\Delta$$

are constants that do not depend on the given set of variables or on λ.

Therefore, the expressions

$$c_{22} : c_{12} = (c_{11}c_{22}/\Delta) : (c_{11}c_{12}/\Delta);$$

$$c_{11} : c_{21} = (c_{11}c_{22}/\Delta) : (c_{21}c_{22}/\Delta)$$

must also be constants.

As a result,

$$c_{11}(x_2/x_1, x_3/x_1, \ldots, x_n/x_1) = s_{11}h_1(x_2/x_1, x_3/x_1, \ldots, x_n/x_1),$$

$$c_{12}(x_2/x_1, x_3/x_1, \ldots, x_n/x_1) = s_{12}h_2(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{21}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) = s_{21}h_1(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{22}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) = s_{22}h_2(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

where the values s_{11}, s_{12}, s_{21}, s_{22} are arbitrary constants; $h_1(t_2, t_3,...,t_n)$, $h_1(t_2, t_3,...,t_n)$ are arbitrary functions of $(n - 1)$ variables.

The final form that the general solution for functional equations (9) and (10) takes is

$$a_{11}(\lambda) = \lambda^{p1} + (\lambda^{p2} - \lambda^{p1})(-s_{12}s_{21}/\Delta^*), \quad (25)$$

$$a_{12}(\lambda) = (\lambda^{p2} - \lambda^{p1})(s_{11}s_{12}/\Delta^*), \quad (26)$$

$$a_{21}(\lambda) = (\lambda^{p2} - \lambda^{p1})(-s_{21}s_{22}/\Delta^*), \quad (27)$$

$$a_{22}(\lambda) = \lambda^{p2} + (\lambda^{p2} - \lambda^{p1})(s_{12}s_{21}/\Delta^*), \quad (28)$$

$$f_1(x_1, x_2,\ldots, x_n) = x_1^{p1}s_{11}h_1(x_2/x_1, x_3/x_1,\ldots,x_n/x_1) + x_1^{p2}s_{12}h_2(x_2/x_1, x_3/x_1, \ldots,x_n/x_1), \quad (29)$$

$$f_2(x_1, x_2,\ldots, x_n) = x_1^{p1}\, s_{21}h_1(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) + x_1^{p2}s_{22}h_2(x_2/x_1, x_3/x_1,\ldots, x_n/x_1), \quad (30)$$

where

$$\Delta^* = s_{11}s_{22} - s_{12}s_{21} \neq 0,$$

and s_{11}, s_{12}, s_{21}, s_{22} are arbitrary constants; $h_1(t_2, t_3,...,t_n)$ and $h_2(t_2, t_3,...,t_n)$ are arbitrary functions of $(n - 1)$ variables.

In the general case, some of the constants in Eqs. (25)–(30) are redundant because, for example, the constant s_{11} can be combined with the function

$$h_1(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

and the constant s_{22} with the function

$$h_2(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

however, then the cases $s_{11} = 0$ or $s_{22} = 0$ have to be considered separately. In particular, we can assume without loss of generality that $s_{11} = s_{22} = 1$ in the general formulas, and regard cases when $s_{11} = s_{22} = 0$ or $s_{11} = 0$, $s_{22} = 1$ as degenerate.

Notably, Eqs. (25)–(30) hold true even with $p_1 = p_2 = p$, when they take the form

$$a_{11}(\lambda) = \lambda^p,\ a_{12}(\lambda) = 0,$$

$$a_{21}(\lambda) = 0,\ a_{22}(\lambda) = \lambda^p,$$

$$f_1(x_1, x_2,\ldots, x_n) = x_1^p h_1(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$f_2(x_1, x_2,\ldots, x_n) = x_1^p h_2(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

i.e., the solution splits into two independent homogeneous functions of the same degree in this case.

Equal real eigenvalues. Let the eigenvalues of matrix (13), (14) be real and equal to each other. The general solution for system of differential equations (13), (14) has the form

$$g_1(x, t_2, t_3,\ldots, t_n) = c_{11}(t_2, t_3,\ldots, t_n)\exp(px) + c_{12}(t_2, t_3,\ldots, t_n)\, x\exp(px);$$

$$g_2(x, t_2, t_3,\ldots, t_n) = c_{21}(t_2, t_3,\ldots, t_n)\exp(px) + c_{22}(t_2, t_3,\ldots, t_n)\, x\exp(px),$$

where c_{11}, c_{12}, c_{21}, c_{22} are some functions of $(n - 1)$ variables.

In this case, the functions

$$f_1(x_1, x_2,\ldots, x_n), f_2(x_1, x_2,\ldots, x_n)$$

should have the form

$$f_1(x_1, x_2,\ldots, x_n) = x_1^p c_{11}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) + x_1^p(\ln x_1)c_{12}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1), \quad (31)$$

$$f_2(x_1, x_2,\ldots, x_n) = x_1^p c_{21}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) + x_1^p\,(\ln x_1)c_{22}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1). \quad (32)$$

Because the functions x_1^p and $x_1^p\ (\ln x_1)$ are linearly independent, substituting expressions (31) and (32) into conditions (9) and (10) yields the relations

$$a_{11}(\lambda)c_{11} + a_{12}(\lambda)c_{21} = \lambda^p c_{11}, \quad (33)$$

$$a_{11}(\lambda)c_{12} + a_{12}(\lambda)c_{22} = \lambda^p(\ln\lambda)c_{12}, \quad (34)$$

$$a_{21}(\lambda)c_{11} + a_{22}(\lambda)c_{21} = \lambda^p c_{21}, \quad (35)$$

$$a_{21}(\lambda)c_{12} + a_{22}(\lambda)c_{22} = \lambda^p(\ln\lambda)c_{22}. \quad (36)$$

Since the functions λ^p и λ^p $(\ln\lambda)$ are linearly independent, linear algebraic equations (33), (34) for unknown functions $a_{11}(\lambda)$ и $a_{12}(\lambda)$, as well as linear algebraic equations (35), (36) for unknown functions $a_{21}(\lambda)$ and

$a_{22}(\lambda)$ cannot be linearly dependent (proportional to each other), with the exception of the degenerate case $c_{12} = c_{22} = 0$, considered separately.

Let

$$\Delta = c_{11}\, c_{22} - c_{12}\, c_{21} \neq 0.$$

In this case,

$$a_{11}(\lambda) = \lambda^p(1 + (1 - \ln\lambda)(c_{12}c_{21}/\Delta)), \quad (37)$$

$$a_{12}(\lambda) = \lambda^p(1 - \ln\lambda)(-c_{11}c_{12}/\Delta), \quad (38)$$

$$a_{21}(\lambda) = \lambda^p(1 - \ln\lambda)(c_{21}c_{22}/\Delta), \quad (39)$$

$$a_{22}(\lambda) = \lambda^p(1 + (1 - \ln\lambda)\,(-c_{11}c_{22}/\Delta)). \quad (40)$$

Since the functions $a_{11}(\lambda)$, $a_{12}(\lambda)$, $a_{21}(\lambda)$ and $a_{22}(\lambda)$ should not depend on the set of variables x_1, x_2,...,x_n, and functions

$$c_{11}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{12}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{21}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{22}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1)$$

should not depend on λ, then the factors $c_{12}c_{21}/\Delta$, $c_{11}c_{12}/\Delta$, $c_{21}c_{22}/\Delta$, $c_{11}c_{22}/\Delta$ are constants that do not depend on either the given set of variables or on λ.

Therefore, the expressions $c_{21} : c_{11} = (c_{12}c_{21}/\Delta) : (c_{11}c_{12}/\Delta)$ and $c_{12} : c_{22} = (c_{11}c_{12}/\Delta) : (c_{11}c_{22}/\Delta)$ are also constants, so that

$$c_{11}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) = s_{11}$$

$$h_1(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{21}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) = s_{21}$$

$$h_1(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{12}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) = s_{12}$$

$$h_2(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{22}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) = s$$

$$_{22}\, h_2(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

where s_{11}, s_{12}, s_{21} and s_{22} are constants that are not simultaneously equal to zero; $h_1(x_2/x_1,\ x_3/x_1,\ldots,\ x_n/x_1)$ and $h_2(x_2/x_1,\ x_3/x_1,\ldots,\ x_n/x_1)$ are some functions of $(n - 1)$ variables.

The solution takes the final form

$$a_{11}(\lambda) = \lambda^p(1 + (1 - \ln\lambda)(s_{12}s_{22}/\Delta^*)), \quad (41)$$

$$a_{12}(\lambda) = \lambda^p(1 - \ln\lambda)(-s_{11}s_{12}/\Delta^*), \quad (42)$$

$$a_{21}(\lambda) = \lambda^p(1 - \ln\lambda)(s_{21}s_{22}/\Delta^*), \quad (43)$$

$$a_{22}(\lambda) = \lambda^p(1 + (1 - \ln\lambda)(-s_{11}s_{22}/\Delta^*)), \quad (44)$$

$$f_1(x_1, x_2,\ldots, x_n) =$$

$$= x_1^p s_{11} h_1(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) + \quad (45)$$

$$+ x_1^p(\ln x_1) s_{12} h_2(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$f_2(x_1, x_2,\ldots, x_n) =$$

$$= x_1^p s_{21} h_1(x_2/x_1, x_3/x_1,\ldots, x_n/x_1) + \quad (46)$$

$$+ x_1^p(\ln x_1) s_{22} h_2(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

where $\Delta^* = s_{11}s_{22} - s_{12}s_{21} \neq 0$, s_{11}, s_{12}, s_{21}, s_{22} are arbitrary constants, and $h_1(t_2, t_3,...,t_n)$ and $h_2(t_2, t_3,...,t_n)$ are arbitrary functions of $(n - 1)$ variables.

Complex conjugate eigenvalues. Let the eigenvalues of matrix (13), (14) be conjugate complex numbers taking the form $p \pm i\omega$.

The general solution for system of differential equations (13), (14) has the form

$$g_1(x, t_2, t_3,\ldots, t_n) =$$

$$= c_{11}(t_2, t_3,\ldots, t_n)\cos(\omega x)\exp(px) +$$

$$+ c_{12}(t_2, t_3,\ldots, t_n)\sin(\omega x)\exp(px);$$

$$g_2(x, t_2, t_3,\ldots, t_n) =$$

$$= c_{21}(t_2, t_3,\ldots, t_n)\cos(\omega x)\exp(px) +$$

$$+ c_{22}(t_2, t_3,\ldots, t_n)\sin(\omega x)\exp(px),$$

where c_{11}, c_{12}, c_{21}, c_{22} are some functions of $(n - 1)$ variables.

In this case, the functions f_1 и f_2 should have the following form:

$$f_1(x_1, x_2,\ldots, x_n) =$$

$$= x_1^p c_{11}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1)\cos(\omega\ln x_1) + \quad (47)$$

$$+ x_1^p c_{12}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1)\sin(\omega\ln x_1);$$

$$f_2(x_1, x_2,\ldots, x_n) =$$

$$= x_1^p c_{21}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1)\cos(\omega\ln x_1) + \quad (48)$$

$$+ x_1^p c_{22}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1)\sin(\omega\ln x_1).$$

Because the functions $x_1^p\cos(\omega\ln x_1)$ and $x_1^p\sin(\omega\ln x_1)$ are linearly independent, substituting expressions (47) and (48) into conditions (9) and (10) yields the relations

$$c_{11}\,a_{11}(\lambda) + c_{21}a_{12}(\lambda) = \\ = \lambda^p c_{11}\cos(\omega\ln\lambda) + \lambda^p c_{12}\sin(\omega ln\lambda), \quad (49)$$

$$c_{12}\,a_{11}(\lambda) + c_{22}a_{12}(\lambda) = \\ = \lambda^p c_{12}\cos(\omega ln\lambda) - \lambda^p c_{11}\sin(\omega ln\lambda), \quad (50)$$

$$c_{11}\,a_{21}(\lambda) + c_{21}a_{22}(\lambda) = \\ = \lambda^p\, c_{21}\cos(\omega\ln\lambda) + \lambda^p c_{22}\sin(\omega\ln\lambda), \quad (51)$$

$$c_{12}a_{21}(\lambda) + c_{22}a_{22}(\lambda) = \\ = \lambda^p c_{22}\cos(\omega\ln\lambda) - \lambda^p c_{21}\sin(\omega\ln\lambda). \quad (52)$$

Linear algebraic equations (49), (50) for unknown functions $a_{11}(\lambda)a_{12}(\lambda)$ and linear algebraic equations (51), (52) for unknown functions $a_{21}(\lambda)a_{22}(\lambda)$ cannot be linearly dependent, except the degenerate case

$$c_{11} = c_{12} = c_{21} = c_{22} = 0,$$

which is of no practical interest.

Indeed, the functions $\lambda^p\cos(\omega\ln\lambda)$ and $\lambda^p\sin(\omega\ln\lambda)$ are linearly independent, while the proportionality relations

$$c_{11} : c_{12} = c_{12} : (-c_{11}),$$

$$c_{21} : c_{22} = c_{22} : (-c_{21})$$

for the right-hand sides of Eqs. (49)–(52) cannot be satisfied with nonzero values of c_{11}, c_{12}, c_{21}, c_{22}.

Therefore, without loss of generality, we can assume that

$$\Delta = c_{11}\,c_{22} - c_{12}\,c_{21} \neq 0.$$

In this case,

$$a_{11}(\lambda) = \lambda^p\cos(\omega\ln\lambda) + \\ + \lambda^p\sin(\omega\ln\lambda)((c_{11}c_{21} + c_{12}c_{22})/\Delta), \quad (53)$$

$$a_{12}(\lambda) = -\lambda^p\sin(\omega\ln\lambda)((c_{11}^{\,2} + c_{12}^{\,2})/\Delta), \quad (54)$$

$$a_{21}(\lambda) = +\lambda^p\sin(\omega\ln\lambda)((c_{21}^{\,2} + c_{22}^{\,2})/\Delta), \quad (55)$$

$$a_{22}(\lambda) = \lambda^p\cos(\omega\,\ln\lambda) - \\ - \lambda^p\sin(\omega\ln\lambda)((c_{11}c_{21} + c_{12}c_{22})/\Delta). \quad (56)$$

Since the functions $a_{11}(\lambda)$, $a_{12}(\lambda)$, $a_{21}(\lambda)$ and $a_{22}(\lambda)$ should not depend on the set of variables x_1, x_2,...,x_n, and the functions

$$c_{11}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{12}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{21}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1),$$

$$c_{22}(x_2/x_1, x_3/x_1,\ldots, x_n/x_1)$$

should not depend on λ, the factors

$$(c_{11}c_{21}+c_{12}c_{22})/\Delta,\ (c_{11}^{\,2}+c_{12}^{\,2})/\Delta,$$

$$(c_{21}^{\,2}+c_{22}^{\,2})/\Delta$$

are constants that do not depend on the given set of variables or on λ.

Substituting

$$c_{11}(x_2/x_1, x_3/x_1,\ldots) = \\ = h_a(x_2/x_1, x_3/x_1,\ldots)\cos h_b(x_2/x_1, x_3/x_1,\ldots),$$

$$c_{12}(x_2/x_1, x_3/x_1,\ldots) = \\ = h_a(x_2/x_1, x_3/x_1,\ldots)\sin h_b(x_2/x_1, x_3/x_1,\ldots),$$

$$c_{21}(x_2/x_1, x_3/x_1,\ldots) = \\ = h_c(x_2/x_1, x_3/x_1,\ldots)\sin h_d(x_2/x_1, x_3/x_1,\ldots),$$

$$c_{22}(x_2/x_1, x_3/x_1,\ldots) = \\ = h_c(x_2/x_1, x_3/x_1,\ldots)\cos h_d(x_2/x_1, x_3/x_1,\ldots),$$

we obtain that the constants should be

$$\mathrm{tg}\,(h_b(x_2/x_1, x_3/x_1,\ldots) + h_d(x_2/x_1, x_3/x_1,\ldots));$$

$$h_a(x_2/x_1, x_3/x_1,\ldots)/h_c(x_2/x_1, x_3/x_1,\ldots).$$

Therefore, after substitutions

$$h_a(x_2/x_1, x_3/x_1,\ldots) = s_a\,h(x_2/x_1, x_3/x_1,\ldots),$$

$$h_c(x_2/x_1, x_3/x_1,\ldots) = s_c\,h(x_2/x_1, x_3/x_1,\ldots),$$

$$h_b(x_2/x_1, x_3/x_1,\ldots) = f(x_2/x_1, x_3/x_1,\ldots) + s_b,$$

$$h_d(x_2/x_1, x_3/x_1,\ldots) = -f(x_2/x_1, x_3/x_1,\ldots) + s_d,$$

where s_a, s_c, s_b, s_d are constants; $h(x_2/x_1, x_3/x_1,\ldots)f(x_2/x_1, x_3/x_1,\ldots)$ are auxiliary functions, and after some additional equivalent transformations, we obtain the formulas

$$c_{11}(x_2/x_1, x_3/x_1,\ldots) =$$
$$+ s_{11}h(x_2/x_1, x_3/x_1,\ldots)\cos f(x_2/x_1, x_3/x_1,\ldots) -$$
$$- s_{12}h(x_2/x_1, x_3/x_1,\ldots)\sin f(x_2/x_1, x_3/x_1,\ldots);$$
$$c_{12}(x_2/x_1, x_3/x_1,\ldots) =$$
$$= + s_{11}h(x_2/x_1, x_3/x_1,\ldots)\sin f(x_2/x_1, x_3/x_1,\ldots) +$$
$$+ s_{12}h(x_2/x_1, x_3/x_1,\ldots)\cos f(x_2/x_1, x_3/x_1,\ldots);$$
$$c_{21}(x_2/x_1, x_3/x_1,\ldots) =$$
$$= - s_{22}h(x_2/x_1, x_3/x_1,\ldots)\sin f(x_2/x_1, x_3/x_1,\ldots) +$$
$$+ s_{21}h(x_2/x_1, x_3/x_1,\ldots)\cos f(x_2/x_1, x_3/x_1,\ldots);$$
$$c_{22}(x_2/x_1, x_3/x_1,\ldots) =$$
$$= + s_{22}h(x_2/x_1, x_3/x_1,\ldots)\cos f(x_2/x_1, x_3/x_1,\ldots) +$$
$$+ s_{21}h(x_2/x_1, x_3/x_1,\ldots)\sin f(x_2/x_1, x_3/x_1,\ldots),$$

where s_{11}, s_{12}, s_{21}, s_{22} are constants that do not depend on the given set of variables or on λ.

Such a choice of parameterization for c_{11}, c_{12}, c_{21}, c_{22} is redundant (obviously) because we can establish, for example, $s_a = 1$ and $s_b = 0$ practically without loss of generality, which means that $s_{11} = 1$ and $s_{12} = 0$.

Besides, it is convenient to substitute

$$h(x_2/x_1, x_3/x_1,\ldots)\cos f(x_2/x_1, x_3/x_1,\ldots)$$

with $h_1(x_2/x_1, x_3/x_1,...)$, and

$$h(x_2/x_1, x_3/x_1,\ldots)\sin f(x_2/x_1, x_3/x_1,\ldots)$$

with $h_2(x_2/x_1, x_3/x_1,...)$.

The final form that the general solution for functional equations (9) and (10) takes is

$$a_{11}(\lambda) = \lambda^p\cos(\omega\ln\lambda) + \lambda^p\sin(\omega\ln\lambda) \times \times ((s_{11}s_{21} + s_{12}s_{22})/\Delta^*), \quad (57)$$

$$a_{12}(\lambda) = -\lambda^p\sin(\omega\ln\lambda)((s_{11}^2 + s_{12}^2)/\Delta^*), \quad (58)$$

$$a_{21}(\lambda) = +\lambda^p\sin(\omega\ln\lambda)((s_{21}^2 + s_{22}^2)/\Delta^*), \quad (59)$$

$$a_{22}(\lambda) = \lambda^p\cos(\omega\ln\lambda) - \lambda^p\sin(\omega\ln\lambda) \times \times ((s_{11}s_{21} + s_{12}s_{22})/\Delta^*), \quad (60)$$

$$f_1(x_1, x_2,\ldots, x_n) = x_1^p(s_{11}\cos(\omega\ln x_1) + + s_{12}x_1^p\sin(\omega\ln x_1))h_1(x_2/x_1, x_3/x_1,\ldots) + + x_1^p(-s_{12}\cos(\omega\ln x_1) + + s_{11}\sin(\omega\ln x_1))h_2(x_2/x_1, x_3/x_1,\ldots), \quad (61)$$

$$f_2(x_1, x_2,\ldots, x_n) = x_1^p(s_{21}\cos(\omega\ln x_1) + + s_{22}x_1^p\sin(\omega\ln x_1))h_1(x_2/x_1, x_3/x_1,\ldots) + + x_1^p(-s_{22}\cos(\omega\ln x_1) + + s_{21}\sin(\omega\ln x_1))h_2(x_2/x_1, x_3/x_1,\ldots), \quad (62)$$

where

$$\Delta^* = s_{11}s_{22} - s_{12}s_{21} \neq 0,$$

and s_{11}, s_{12}, s_{21}, s_{22} are arbitrary constants (partially redundant); $h_1(t_2, t_3,...,t_n)$, $h_2(t_2, t_3,...,t_n)$ are arbitrary functions of $(n - 1)$ variables.

Further steps

We can analyze other systems of functional equations of the form (8) with matrices of finite size by a similar scheme. However, complex formulas with many variant branches appear as a result of analysis; in our opinion, they have no particular practical meaning.

Taking into account the analysis given in this article for differentiable functions, all solutions of functional equations of the form (8) are linear combinations of functions taking the form

$$f_{k,p}(x_1, x_2,\ldots, x_n) = = x_1^p(\ln x_1)^k h(x_2/x_1, x_3/x_1,\ldots, x_n/x_1), \quad (63)$$

which correspond to real eigenvalues p with the multiplicity k, and functions taking the form

$$f^{(c)}_{k,p}(x_1, x_2,\ldots, x_n) = x_1^p(\ln x_1)^k\cos(\omega\ln x_1) \times \times h(x_2/x_1, x_3/x_1,\ldots, x_n/x_1), \quad (64)$$

$$f^{(s)}_{k,p}(x_1, x_2,\ldots, x_n) = x_1^p(\ln x_1)^k\sin(\omega\ln x_1) \times \times h(x_2/x_1, x_3/x_1,\ldots, x_n/x_1), \quad (65)$$

which correspond to complex conjugate eigenvalues of the form $p \pm i\omega$ of multiplicity k, where $h(t_2, t_3,...,t_n)$ are some functions of $(n - 1)$ variables.

Regarding the theory on mutually homogeneous functions, we believe that it is sufficient to analyze systems of functional

relations corresponding to isolated fundamental chains of functions taking the form (63) and (64), (65), instead of analyzing systems of functional relations with a general form.

We plan to carry out further investigations analyzing systems of mutually homogeneous functions with infinite chains of functional equations of the form (8).

The calculations in this paper were carried out using the Wolfram Mathematica software [29].

Acknowledgment

We wish to express our sincere gratitude to Anton Leonidovich Bulyanitsa, Professor of Department of Higher Mathematics of Peter the Great St. Petersburg Polytechnic University, for active participation in discussions on the problem.

This study was partially supported by NIR 0074-2019-0009, part of State Task No. 075-00780-19-02 of the Ministry of Science and Higher Education of the Russian Federation.

REFERENCES

1. **Berdnikov A.S., Gall L.N., Gall N.R., Solovyev K.V.**, Generalization of the Thomson formula for harmonic functions of a general type, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 12 (2) (2019) 32–48.

2. **Berdnikov A.S., Gall L.N., Gall N.R., Solovyev K.V.**, Generalization of the Thomson formula for homogeneous harmonic functions, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 12 (2) (2019) 49–62.

3. **Berdnikov A.S., Gall L.N., Gall N.R., Solovyev K.V.,** Donkin's differential operators for homogeneous harmonic functions, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 12 (3) (2019) 45–62.

4. **Berdnikov A.S., Gall L.N., Gall N.R., Solovyev K.V.,** Basic Donkin's differential operators for homogeneous harmonic functions, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 12 (3) (2019) 26–44.

5. **Golikov Yu.K., Krasnova N.K.,** Teoriya synteza elektrostaticheskikh energoanalizatorov [Theory of designing of electrostatic energy analyzers], Saint-Petersburg Polytechnic University Publishing, Saint-Petersburg, 2010.

6. **Golikov Yu.K., Krasnova N.K.,** Application of electric fields uniform in the Euler sense in electron spectrography, Technical Physics. 56 (2) (2011) 164–170.

7. **Golikov Yu.K., Krasnova N.K.,** Generalized similarity principle of similarity in electron spectrography, Prikladnaya Fizika (Applied Physics). (2) (2007) 5–11.

8. **Averin I.A., Berdnikov A.S., Gall N.R.,** The principle of similarity of trajectories for the motion of charged particles with different masses in electric and magnetic fields that are homogeneous in Euler terms, Technical Physics Letters. 43 (2) (2017) 156–158.

9. **Fikhtengol'ts G.M.,** The fundamentals of mathematical analysis, Vol. 1,Oxford, New York, Pergamon Press, 1965.

10. **Gel'fand I.M., Shapiro Z.Ya.,** Generalized functions and their applications, Uspekhi Mat. Nauk. 10 (3) (1955) 3–70.

11. **Gel'fand I.M., Shilov G.E.,** Generalized Functions, Vol. 1: Properties and Operations, AMS Chelsea Publishing, 1964.

12. **Khursheed A., Dinnis A.R., Smart P.D.,** Micro-extraction fields to improve electron beam test measurements, Microelectronic Engineering. 14 (3–4) (1991) 197–205.

13. **Khursheed A.**, Multi-channel vs. conventional retarding field spectrometers for voltage contrast, Microelectronic Engineering. 16 (1–4) (1992) 43–50.

14. **Khursheed A., Phang J.C., Thong J.T.L.,** A portable scanning electron microscope column design based on the use of permanent magnets, Scanning. 20 (2) (1998) 87–91.

15. **Khursheed A.,** Magnetic axial field measurements on a high resolution miniature scanning electron microscope, Review of Scientific Instruments. 71 (4) (2000) 1712–1715.

16. **Khursheed A.,** A low voltage time of flight electron emission microscope, Optik (Jena). 113 (11) (2002) 505–509.

17. **Khursheed A.,** Aberration characteristics of immersion lenses for LVSEM, Ultramicroscopy. 93 (3-4) (2002) 331–338.

18. **Khursheed A., Karuppiah N., Osterberg M., Thong J.T.L.,** Add-on transmission attachments for the scanning electron microscope, Review of Scientific Instruments. 74 (1) (2003) 134–140.

19. **Khursheed A., Osterberg M.,** A spectroscopic scanning electron microscope design, Scanning. 26 (6) (2004) 296–306.

20. **Osterberg M., Khursheed A.,** Simulation of magnetic sector deflector aberration properties for low-energy electron microscopy, Nuclear Instruments and Methods in Physics Research, Section A. 555 (1–2) (2005) 20–30.

21. **Khursheed A., Osterberg M.,** Developments in the design of a spectroscopic scanning electron microscope, Nuclear Instruments and Methods in Physics Research, Section A. 556 (2) (2006) 437–444.

22. **Luo T., Khursheed A.,** Imaging with surface sensitive backscattered electrons, Journal of Vacuum Science and Technology B. 25 (6) (2007) 2017–2019.

23. **Khursheed, A., Hoang, H.Q.,** A second-order focusing electrostatic toroidal electron spectrometer with 2π radian collection, Ultramicroscopy. 109 (1) (2008) 104–110.

24. **Khursheed A.,** Scanning electron microscope optics and spectrometers, World Scientific, Singapore, 2010.

25. **Hoang H.Q., Khursheed A.,** A radial mirror analyzer for scanning electron/ion microscopes, Nuclear Instruments and Methods in Physics Research, Section A. 635 (1) (2011) 64–68.

26. **Hoang H.Q., Osterberg M., Khursheed A.,** A high signal-to-noise ratio toroidal electron spectrometer for the SEM, Ultramicroscopy. 2011. Vol. 11 (8) (2011) 1093–1100.

27. **Khursheed A., Hoang H.Q., Srinivasan A.,** A wide-range parallel Radial Mirror analyzer for scanning electron/ion microscopes, Journal of Electron Spectroscopy and Related Phenomena. 184 (11–12) (2012) 525–532.

28. **Shao X., Srinivasan A., Ang W.K., Khursheed A.,** A high-brightness large-diameter graphene coated point cathode field emission electron source, Nature Communications. 9 (1) (2018) 1288.

29. Wolfram Mathematica, URL: http://wolfram.com/mathematica/

Received 21.01.2020, accepted 02.03.2020.

THE AUTHORS

BERDNIKOV Alexander S.
Institute for Analytical Instrumentation of the Russian Academy of Sciences
26 Rizhsky Ave., St. Petersburg, 190103, Russian Federation
asberd@yandex.ru

SOLOVYEV Konstantin V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
k-solovyev@mail.ru

KRASNOVA Nadezhda K.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
n.k.krasnova@mail.ru

СПИСОК ЛИТЕРАТУРЫ

1. **Бердников А.С., Галль Л.Н., Галль Р.Н., Соловьев К.В.** Обобщение формулы Томсона для гармонических функций общего вида // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2019. Т. 2 № .12. С. 48–32.

2. **Бердников А.С., Галль Л.Н., Галль Р.Н., Соловьев К.В.** Обобщение формулы Томсона для гармонических однородных функций // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2019. Т. № .12 2. С. 62–49.

3. **Бердников А.С., Галль Л.Н., Галль Н.Р., Соловьев К.В.** Дифференциальные операторы Донкина для однородных гармонических функций // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2019. Т. 3 № .12. С. 62–45.

4. **Бердников А.С., Галль Л.Н., Галль Н.Р., Соловьев К.В.** Базисные дифференциальные операторы Донкина для однородных гармонических функций // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2019. Т. 3 № .12. С. 44–26.

5. **Голиков Ю.К., Краснова Н.К.** Теория синтеза электростатических энергоанализаторов. СПб.: Изд-во Политехнического ун-та, 409 .2010 с.

6. **Голиков Ю.К., Краснова Н.К.** Электрические поля, однородные по Эйлеру, для электронной спектрографии // Журнал технической физики. 2011. Т. 2 № .81. С. 15–9.

7. **Голиков Ю.К., Краснова Н.К.** Обобщенный принцип подобия и его применение в электронной спектрографии // Прикладная физика. 2 № .2007. С. 11–5.

8. **Аверин И.А., Бердников А.С., Галль Н.Р.** Принцип подобия траекторий при движении заряженных частиц с разными массами в однородных по Эйлеру электрических и магнитных полях // Письма в Журнал технической физики. 2017. Т. 3 № .43. С. 43–39.

Фихтенгольц Г.М. Курс дифференциального и интегрального исчисления. Т. 1. М.: Физматлит, 616 .2001 с.

10. **Гельфанд И.М., Шапиро З.Я.** Однородные функции и их приложения // Успехи математических наук. 1955. Т. 10. Вып. 3. С. 70–3.

11. **Гельфанд И.М., Шилов Г.Е.** Обобщенные функции и действия над ними. Серия «Обобщенные функции». Вып. 1. 2-е изд. М.: Физматгиз, 1959. 470 с.

12. **Khursheed A., Dinnis A.R., Smart P.D.** Micro-extraction fields to improve electron beam test measurements // Microelectronic Engineering. 1991. Vol. 14. No. 3–4. Pp. 197–205.

13. **Khursheed A.** Multi-channel vs. conventional retarding field spectrometers for voltage contrast // Microelectronic Engineering. 1992. Vol. 16. No. 1–4. Pp. 43–50.

14. **Khursheed A., Phang J.C., Thong J.T.L.** A portable scanning electron microscope column design based on the use of permanent magnets // Scanning. 1998. Vol. 20. No. 2. Pp. 87–91.

15. **Khursheed A.** Magnetic axial field measurements on a high resolution miniature scanning electron microscope // Review of Scientific Instruments. 2000. Vol. 71. No. 4. Pp. 1712 –1715.

16. **Khursheed A.** A low voltage time of flight electron emission microscope // Optik (Jena). 2002. Vol. 113. No. 11. Pp. 505–509.

17. **Khursheed A.** Aberration characteristics of immersion lenses for LVSEM // Ultramicroscopy. 2002. Vol. 93. No. 3–4. Pp. 331–338.

18. **Khursheed A., Karuppiah N., Osterberg M., Thong J.T.L.** Add-on transmission attachments for the scanning electron microscope // Review of Scientific Instruments. 2003. Vol. 74. No. 1. Pp. 134–140.

19. **Khursheed A., Osterberg M.** A spectroscopic scanning electron microscope design // Scanning. 2004. Vol. 26. No. 6. Pp. 296–306.

20. **Osterberg M., Khursheed A.** Simulation of magnetic sector deflector aberration properties for low-energy electron microscopy // Nuclear Instruments and Methods in Physics Research, Section A. 2005. Vol. 555. No. 1–2. Pp. 20–30.

21. **Khursheed A., Osterberg M.** Developments in the design of a spectroscopic scanning electron microscope // Nuclear Instruments and Methods in Physics Research. Section A. 2006. Vol. 556. No. 2. Pp. 437–444.

22. **Luo T., Khursheed A.** Imaging with surface sensitive backscattered electrons // Journal of Vacuum Science and Technology. B. 2007. Vol. 25. No. 6. Pp. 2017–2019.

23. **Khursheed, A., Hoang, H.Q.** A second-order focusing electrostatic toroidal electron spectrometer with 2π radian collection // Ultramicroscopy. 2008. Vol. 109. No. 1. Pp. 104–110.

24. **Khursheed A.** Scanning electron microscope optics and spectrometers. Singapore: World Scientific, 2010. 403 p.

25. **Hoang H.Q., Khursheed A.** A radial mirror analyzer for scanning electron/ion microscopes // Nuclear Instruments and Methods in Physics Research. Section A. 2011. Vol. 635. No. 1. Pp. 64–68.

26. **Hoang H.Q., Osterberg M., Khursheed A.** A high signal-to-noise ratio toroidal electron spectrometer for the SEM // Ultramicroscopy. 2011. Vol. 111. No. 8. Pp. 1093–1100.

27. **Khursheed A., Hoang H.Q., Srinivasan A.** A wide-range parallel radial mirror analyzer for scanning electron/ion microscopes // Journal of Electron Spectroscopy and Related Phenomena. 2012. Vol. 184. No. 11–12. Pp. 525 –532.

28. **Shao X., Srinivasan A., Ang W.K., Khursheed A.** A high-brightness large-diameter graphene coated point cathode field emission electron source // Nature Communications. 2018. Vol. 9. No. 1. P. 1288.

29. Wolfram Mathematica // URL: http://wolfram.com/mathematica/

Статья поступила в редакцию 21.01.2020, принята к публикации 02.03.2020.

СВЕДЕНИЯ ОБ АВТОРАХ

БЕРДНИКОВ Александр Сергеевич – *доктор физико-математических наук, ведущий научный сотрудник Института аналитического приборостроения Российской академии наук.*
190103, Российская Федерация, г. Санкт-Петербург, Рижский пр., 26
asberd@yandex.ru

СОЛОВЬЕВ Константин Вячеславович – *кандидат физико- математических наук, доцент Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого, младший научный сотрудник Института аналитического приборостроения РАН.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
k-solovyev@mail.ru

КРАСНОВА Надежда Константиновна – *доктор физико-математических наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
n.k.krasnova@mail.ru

DOI: 10.18721/JPM
UDC 621.391:681.142

A TRIANGULATION SENSOR FOR MEASURING THE DISPLACEMENTS AND HIGH-PRECISION MONITORING OF THE PRODUCTION PERFORMANCE

V.A. Stepanov, E.N. Moos, M.V. Shadrin,
V.N. Savin, A.V. Umnyashkin, N.V. Umnyashkin

Ryazan State University named for S. Yesenin, Ryazan, Russian Federation

Using the method of laser triangulation as the base, a mobile high-precision sensor has been created for measuring displacements and monitoring of the geometric parameters of workpieces in production. Both the process of signal processing and the operation of the triangulation sensor were accelerated many times owing to the architecture of processes, which was based on a reduced set of commands using simple and effective instructions of the stm32f407vet6 microcontroller. The measurement procedure was carried out by searching for a laser spot, calculating the center of the spot using the center of mass method, converting the centroid into the metric and applying calibration tables. Sensor scan speed amounted to $(3 - 5)\cdot 10^3$ s^{-1}.

Keywords: triangulation sensor, microprocessor, laser diode, spot center, interface, control module

Citation: Stepanov V.A., Moos E.N., Shadrin M.V., Savin V.N., Umnyashkin A.V., Umnyashkin N.V., A triangulation sensor for measuring the displacements and high-precision monitoring of the production performance, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (1) (2020) 47–58. DOI: 10.18721/JPM.13105

ТРИАНГУЛЯЦИОННЫЙ ДАТЧИК ДЛЯ ИЗМЕРЕНИЯ ПЕРЕМЕЩЕНИЙ И ВЫСОКОТОЧНОГО КОНТРОЛЯ ПАРАМЕТРОВ ИЗДЕЛИЯ НА ПРОИЗВОДСТВЕ

В.А. Степанов, Е.Н. Моос, М.В. Шадрин,
В.Н. Савин, А.В. Умняшкин, Н.В. Умняшкин

Рязанский государственный университет имени С.А. Есенина,

г. Рязань, Российская Федерация

На основе метода лазерной триангуляции создан мобильный высокоточный датчик для измерения перемещений и контроля геометрических параметров изделий на производстве. Архитектура процессов, построенная на сокращенном наборе команд, использующих простые и эффективные инструкции микроконтроллера stm32f407vet6, обеспечивает многократное ускорение не только процесса обработки сигнала, но и работы триангуляционного датчика. Процедура измерения осуществляется путем поиска лазерного пятна, расчета расположения центра пятна методом центра масс, перевода центроиды в метрику и применения калибровочных таблиц. Достигнутая скорость сканирования датчика составляет $(3 - 5)\cdot 10^3$ измерений в секунду.

Ключевые слова: триангуляционный датчик, микропроцессор, лазерный диод, центр пятна, интерфейс, модуль управления

Ссылка при цитировании: Степанов В.А., Моос Е.Н., Шадрин М.В., Савин В.Н., Умняшкин А.В., Умняшкин А.В., Умняшкин Н.В. Триангуляционный датчик для измерения перемещений и высокоточного контроля параметров изделия на производстве // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2020. Т. 13. № 1. С. 54–65. DOI: 10.18721/JPM.13105

Introduction

The significance of this study stems from the need for high-precision monitoring of the movements of working bodies in machining centers in production and development of workpieces in such industries as engineering, aerospace industry, military production, as well as in other branches where precise control of geometric parameters or positions of the objects is required.

The triangulation method for measuring the displacements, geometric dimensions and roughness of workpieces with complex surfaces [1 – 6] was used to develop a sensor for measuring displacements and monitoring the geometric parameters of workpieces in dynamics, for example, on a conveyor, without slowing down production. The sensor must provide high accuracy for operation (control) in various industries and a high scanning speed; be compact and have a degree of protection appropriate to the needs of enterprises. The software can process signals from the measuring channel of the sensor transferring the main aspect in the development of applications from hardware to software.

The principle of triangulation distance measurement

The triangulation method of control is based on calculating the required distance in terms of the ratios in the triangle using known system parameters. This allows to measure both the relative change in the distance from the sensor to the controlled object and its absolute value.

The triangulation scheme (Fig. 1) can conditionally be divided into three parts: a radiating (or lighting) channel, a controlled surface, and a receiving channel.

The first part of the meter is a channel consisting of a radiation source and an objective lens, where a probe beam is formed on a controlled surface. A laser diode with a Gaussian distribution is typically used as such a source. The objective lens consists of one or more optical lenses. Its relative position and the position of the laser diode determine the setting of the emitting channel. To set up the laser module (for getting the maximum intensity value), it is necessary to set the constriction to the center of the measuring range and to center the probe beam. The result of good adjustment is a centered beam, whose width and intensity vary symmetrically relative to the center of the measurement range.

The second integral part of the triangulation meter is the controlled surface. Any surface reflects or scatters incident radiation. Scattering of radiation by the surface of a controlled object is used in triangulation as the basis for obtaining information about the distance to this surface.

The triangulation sensor is intended for measuring the distance from the selected point on the probe beam axis to the physical point on the surface with high accuracy. Any controlled surface is characterized by roughness, corresponding to the degree to which the surface is smooth or uneven. The required measurement accuracy is usually inversely proportional to the roughness of the surface monitored. In particular, the surface roughness of microelectronic crystals and

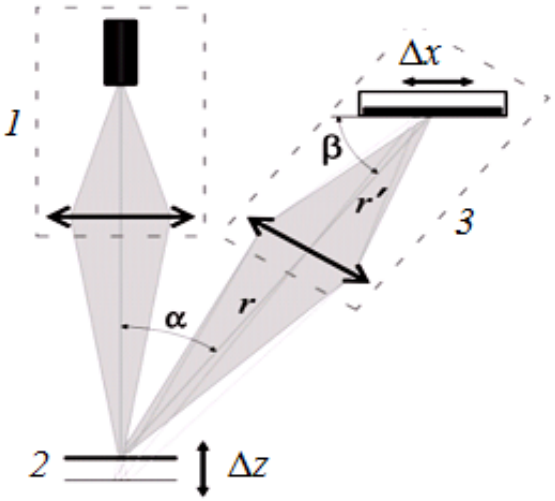

Fig. 1. Schematic diagram of a triangulation meter:
the radiating and receiving channels *1* and *3*, respectively; the controlled surface (*2*); the displacements of the controlled surface (Δz) and a laser spot (Δx); the distances from the controlled surface to the projection lens (r) and from the lens to the photo detector (r')

therefore the measured distance to them have a scale of several micrometers. For example, the accuracy required in the surveying industry is of the order of hundreds of meters.

Industrial dimensional control consists in determining the parameters of metal surfaces, the required control accuracy ranging from a few micrometers (in the nuclear industry) to hundreds of micrometers (in the railway industry).

Each surface also has the property of reflecting or scattering incident radiation. Radiation scattering by the surface of a controlled object is used in triangulation as a physical basis for obtaining information about the distance to this surface. Therefore, the controlled surface is an integral part of the triangulation measuring system.

The third part of the triangulation meter is the receiving channel, consisting of a projection lens and a photodetector. The projection lens generates an image of the probe spot in the plane of the photodetector. The larger the lens diameter *D*, the higher its aperture. In other words, the larger the lens diameter *D*, the more intense the spot image, and the higher its quality. Depending on the specific implementation, either a photodiode array or a position-sensitive receiver is used for capturing the generated image.

The triangulation meter shown in Fig. 2 operates as follows. Emitting channel *1* forms an image of the light spot on controlled surface *2* (*6 – 6`*). Next, the light scattered by the controlled surface enters receiving channel *3*. Thus, an image of the illuminated portion of the controlled surface (light spot) is generated in the plane of the photodetector. If the controlled surface is displaced by Δz (see Fig. 1), the light spot in the plane of the photodetector shifts by Δx. Dependence of the displacement Δz of the controlled surface on the displacement Δx of the light spot in the plane of the photodetector has the following form:

$$\Delta z = r \cdot \sin\varphi / \sin(\alpha - \varphi)$$

where

$$\varphi = \operatorname{arctg}\left(A \cdot \Delta x / (1 + B \cdot \Delta x)\right),$$

$$A = \sin\beta / r', B = -\cos\beta / r',$$

r and r' are the distance from controlled surface *2* to the projecting lens of receiving channel *3*, and that from the lens to the photodetector, despite the fact that the controlled surface is in the center of the range of displacement measurements, respectively.

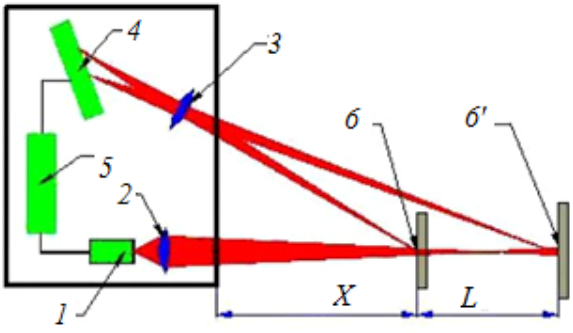

Fig. 2. Principle of operation of the triangulation meter:

the laser (*1*); the lenses of the radiating (*2*) and the receiving (*3*) channels; the dividing plate (*4*); the supply system (*5*); the images of a light spot on the controlled surface (*6 – 6`*)

Software architecture of microcontrollers with ARM core

The ARM core is based on RISC (Reduced Instruction Set Computing) architecture, a processor architecture built on the basis of a reduced set of instructions), using simple and efficient processor instructions that can be executed in a single cycle. The basic concepts of RISC involve shifting the main emphasis in developing applications from hardware to software, since it is much easier to increase application performance by software methods rather than using complex hardware solutions. As a result, programming of RISC processors imposes more stringent requirements for compiler performance compared to CISC architecture (Complete Instruction Set Computing, processor architecture with a wide range of different machine instructions of variable length and different execution times). Microprocessors with CISC architecture (for example, Intel x86) are not so demanding on development software: here the main emphasis is on hardware performance. Fig. 3 shows these differences.

Considering RISC architecture in more detail, we can see that it is based on the following basic principles.

The system of instructions (commands) of the processor. A standard RISC processor has a limited set of instruction types, with each of these instructions executing in a single processor cycle. Individual software algorithms, for example, division, are compiled entirely via software development tools (compilers) or by actual developers. Each processor instruction

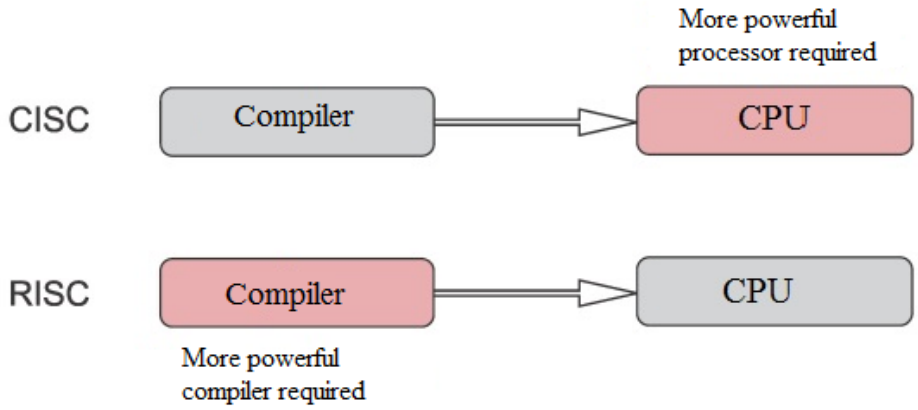

Fig. 3. Architecture of RISC and CISC processes

has a fixed length, which allows to successfully use the pipeline principle to select the next instruction while the previous one is decoded. In contrast, CISC processors have different lengths and can be executed in several machine cycles. Comparing two code snippets for CISC and RISC processors, we can see that the RISC processor needs more instructions. On the other hand, a large number of registers in ARM processors allows computing operations with several variables to be performed very efficiently, since intermediate results of calculations can be placed in registers. In addition, ARM processors can use multiple access instructions for multiple memory locations, which improves the performance of data read/write operations. Additional opportunities for increasing the performance of ARM microprocessors are achieved by using conditional execution instructions, when the next instruction is executed only if the previous instruction set certain flags in the program status register.

Instruction pipeline. Each processor instruction is processed in several stages, which are performed simultaneously. Ideally, in order to achieve maximum performance, the instruction pipeline should move one step in each machine cycle, and decoding of the instruction can be carried out in one step of the pipeline. This approach is different from that adopted for CISC architectures where special microprograms have to be executed to decode instructions.

Using processor registers. RISC processors have many common registers. Each of the registers may contain data or a data address in memory; therefore, registers are local data storages during all operations in the processor. For comparison: CISC processors have a limited set of registers, each with a separate functional purpose, which is why many CISC instructions use a memory cell as one of the operands. For example, the ADD instruction where one of the operands was a memory cell was used for adding two numbers for CISC Intel × 86 processors. Such an instruction requires a lot of machine cycles, reducing the performance of the application, especially if such instructions are used in cyclic calculations. Despite significant differences in architectures, the RISC and CISC architectures are gradually becoming similar. For example, CISC microprocessors are implemented according to RISC principles at the microprogram level. That increases the speed of microinstructions.

The microcontroller stm32f407VET6 and the electrical circuit of the receiving channel

Microcontroller stm32f407VET6. General characteristics of the microcontroller used for operations with the ARM7 core are presented below.

ARM 32-bit Cortex-M4 CPU;

Clock frequency 168 MHz, 210 DMIPS / 1.25 DMIPS / MHz (Dhrystone 2.1);

Support for DSP instructions;

New high-performance AHB-matrix tires;

Up to 1 MB of Flash memory;

Up to 192 + 4 kB SRAM-memory;

Supply voltage of 1.8 − 3.6 V (POR, PDR, PVD and BOR);

Internal RC-generators at 16 MHz and 32 kHz (for RTC);

External clock source of 4−26 MHz and of

32.768 kHz for RTC;

SWD/JTAG debugging modules, ETM module;

Three 12-bit ADCs on 24 input channels (speed up to 7.2 megasamples, temperature sensor);

Two 12-bit DAC;

DMA controller for 16 streams with support for packet transmission;

17 timers (16 and 32 categories);

Two watchdog timers (WDG and IWDG);

Communication interfaces: I2C, USART (ISO 7816, LIN, IrDA), SPI, I2S;

CAN (2.0 B Active);

USB 2.0 FS/HS OTG;

10/100 Ethernet MAC (IEEE 1588v2, MII / RMII);

SDIO controller (SD, SDIO, MMC, CE-ATA cards);

Digital camera interface (8/10/12/14-bit modes);

FSMC-controller (Compact Flash, SRAM, PSRAM, NOR, NAND and LCD 8080/6800); hardware random number generator;

Hardware CRC calculation, 96-bit unique ID;

AES 128, 192, 256, Triple DES, HASH (MD5, SHA-1), HMAC encryption module;

Extended temperature range of 40 – 105 °C.

This controller is selected based on the following parameters:

(i) High performance. High performance is necessary for fast operation of the algorithm for calculating the position of an object.

(ii) High frequency of the core and periphery. The operating frequency of the microcontroller is 168 MHz, which makes it possible to interrogate a linear image sensor with a high frequency, and also allows peripheral data transmission modules to operate at high speeds. The core and peripheral frequencies can be flexibly tuned in this family of controllers, which affects the power consumption of the controller.

(iii) Large number of peripherals. We are primarily concerned with the periphery for data transmission in this study. At the same time, the controller has a large set of different data transfer interfaces, which makes the sensor universal.

Ethernet. The unit is made strictly according to the IEEE802.3 standard. It is possible to transfer data at a speed of 10/100 Mbit/s. Clock synchronization is available: the IEEE1588 v2 protocol is implemented in hardware for this purpose. A fiber optic or copper line requires a third-party transceiver. The PHY transceiver connects directly to the MII or RMII port.

USB (Universal Serial Bus). The system has two separate USB blocks. The first one is USB OTG full-speed, which is fully hardware-implemented and compatible with USB 2.0 standards, as well as OTG 1.0. The USB operates at speeds up to 12 Mbit/s. It is supported in Host/Device/OTG mode. Session Request Protocol (SRP) and Host Negotiation Protocol (HNP) are included.

The second block, USB OTG high-speed, operates in Host/Device/OTG mode with a high speed of 480 Mbps; a transceiver unit operating at high speed through a special ULPI interface is required for this purpose.

USART (Universal Synchronous Asynchronous Receiver Transmitter). Four USART units and two UART (Universal Asynchronous Receiver Transmitter) are integrated in the microcontroller. USART1 and USART6 units allow high-speed data exchange at speeds up to 10.5 Mbit/s. Others support a speed of no more than 5.25 Mbit/s. Thanks to USART, standards such as RS323 and RS485 can be used in the sensor.

Linear Image Sensor ELIS-1024. The ELIS-1024 linear image sensor from Dynamax imaging is used as a photosensitive sensor in the system. This sensor is selected due to its high speed, sufficient resolution with simple controls and low cost.

Its main features are low cost; programmable resolutions of 1024, 512, 256, 128 pixels; high sensitivity; low noise level; clock frequency from 1 kHz to 30 MHz; low dark current; fully customizable clock frequency and frame rate; electrical characteristics: supply voltage of 2.80 – 5.50 V; current consumption of 25 mA; supply voltage of the digital part of 5 V; minimum voltage of the upper logic level of 0.6 V; maximum clock frequency of 30 MHz; saturation output voltage of 4.8 V; output dark voltage of 2.1 V.

This sensor has a resolution control function. The resolution of the sensor can be at 1024, 512, 256, 128 pixels. This is convenient when a high measurement speed is needed.

Eight digital signal lines were created from the microcontroller to operate the sensor:

M1 and M2, controlling the resolution of the sensor and the frame rate and controlled by the table;

Pin PCO-PC5 for operating the image sensor ELIS-1024;

RST (Reset) resets the pixel values and

translates the origin of the pixels to the zero pixel;

SHT (Shutter) is a shutter used to control the exposure of the image sensor;

DATA triggers the image sensor to issue pixels;

RM-pin controls the operation mode of the image sensor, it is directly connected to the total equipment.

Pin CLK are clocking the image sensor; all processes, including laser modulation, occur in the sensor by the clock cycles of this pin; analog signal is used to synchronize with the AD9203 ADC and the microcontroller.

Analog-to-digital converter ADC (AD9203). Data transfer interface (RS485) and receive channel circuit. ADC is an important part of the sensor, the accuracy of measurement directly depends on its conversion accuracy [7, 8]; moreover, it must be fast enough to be able to convert all the signals. We have selected the AD9203 of ADC architecture from Analog Devices. This ADC has a sampling rate of 40 megabytes per second and an accuracy of 10 bits. The ADC transmits data through a parallel interface, which provides greater speed and ease of communication with the microcontroller.

The sensor has several data transfer interfaces for communication with other devices. One of them is UART with the RS485 standard of logical levels. RS485 is a half-duplex multipoint serial data interface. Data transfer is carried out via one pair of conductors using differential signals. The voltage difference between the conductors of one polarity means a logical unit; the difference of the other polarity is zero. RS485 is implemented using the MAX485 chip.

A block diagram for the connection (Fig. 4) and a printed circuit board (Fig. 5) were developed to combine all MCs and other elements of the receiving channel, including the ADC (AD9203) and the data transfer interface (RS485).

Figs. 4 and 5 show the functional and electrical connections between the main blocks and nodes of the electrical circuit of the receiving channel of the triangulation displacement sensor. Evidently, all the electrical connections between the units, including the laser control modules and the ELIS-1024 linear image sensor, are subordinated to the stm32f407VET6 microcontroller.

The electrical circuit was developed in P-CAD (software for Computer Aided Design of electronics by Personal CAD Systems Inc) intended for design of multilayer printed circuit boards, computing and electronic devices. P-CAD includes two main modules: P-CAD Schematic (graphic editor of circuit diagrams) and P-CAD PCB (graphic editor of printed circuit boards), as well as a number of auxiliary programs.

Software STM32Cube TX program.

The STM32Cube TX program is used to configure and further operate the MCs (microcontrollers) and initialize the code of different elements in the circuit board circuit of the receiving channel of the triangulation displacement meter. The program allows to select the necessary MC, specify the clock sources of different buses, initialize the pins, including timers, configure the interrupt. All this is done in graphical mode. The STM32Cube TX program does not allow for errors for operations with hardware. The programmer needs to concentrate directly on solving applied problems: measuring displacements.

The stm32f407vet6 microcontroller is used in the sensor, so ARM Cortex M4 is selected in the Core tab, stm32f4 is in the Series one, stm32f407/417 is in the Line one, LQFP100 is in the Package one. The clock system is configured in the Clock-Configuration tab. The clock of the microcontroller is fully configurable in the STM32Cube TX program. The source of clock signals in this study is an external quartz (8 MHz) generator. At the same time, the system clock frequency (SYSLK) is set to 168 MHz, which is the maximum for our microcontroller.

Asynchronous operation mode is selected to connect the USART serial interface (Universal Asynchronous Receiver-Transmitter) in the USART1 tab; data transfer at a speed of 11500 Bit/s and a data library are connected via USB Full Speed; the integrity of bit parity data is automatically controlled and the parity control is different. When the sum of the number of unit bits in a packet is an even number, and when this sum is odd, USB interruptions are cleared automatically.

Operation of the linear image sensor ELIS-1024. The ELIS-1024 linear image sensor operates in frame-by-frame synchronization mode, when new exposure is set for each new frame.

Data is written to an 8-bit data array with a dimension of 1024 mass-elements [9].

One important point should be noted. The microcontroller operates on 3.3 V, respectively, and the logical unit (high level) will be 3.3 V, and the ELIS-1024 image sensor needs 5 V logic

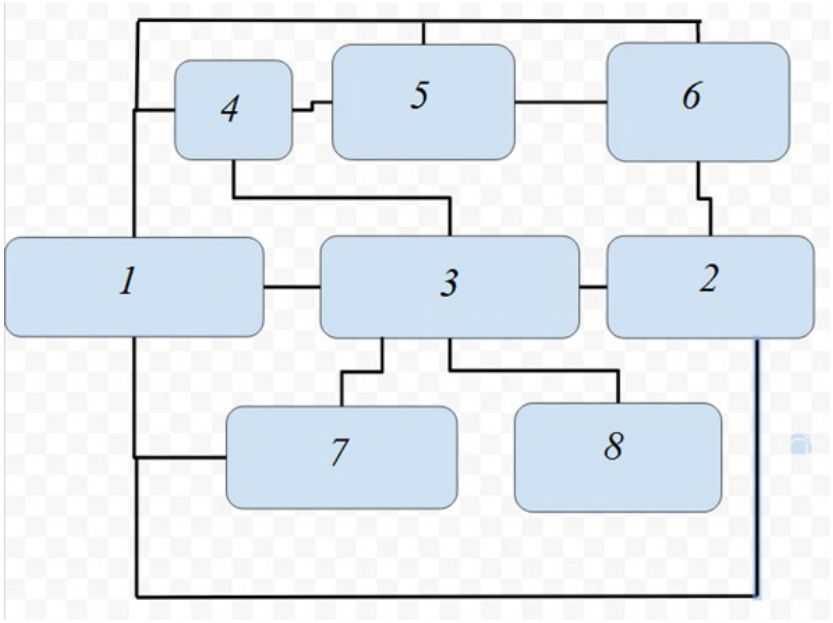

Fig. 4. The electrical block diagram of the receiving channel:
1 is the supply system; *2* is the analog-to-digital converter; *3* is the microcontroller stm32f407vet6; *4* is the interface logic levels RS485 (the main elements of one); *5* is the operational amplifier; *6* is the linear image sensor; *7* is the laser control module; *8* is the data transfer interface

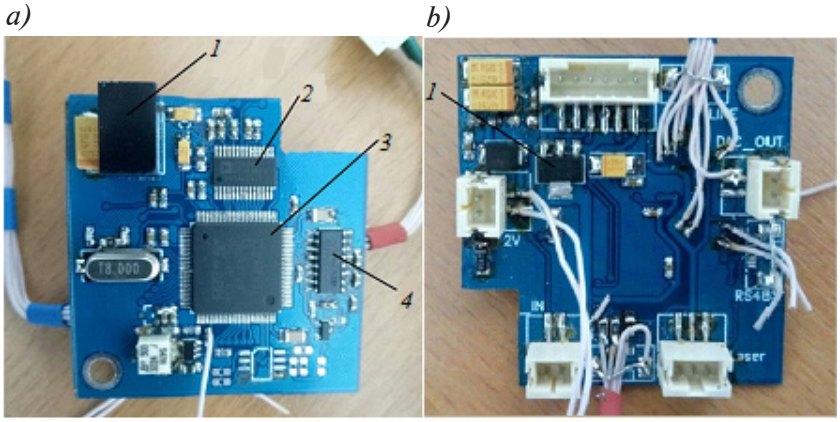

Fig. 5. Appearance of the receiving channel (the face (*a*) and the back (*b*) of the finished printed circuit board with elements);
1 to *4* positions correspond to Fig. 4

levels for normal, fast operation. Therefore, a new level converter should be used. The 74VHCT04AMTCX chip from Fairchild Semiconductor was chosen. In fact, this is not just a converter but also an inverter of levels (6 inverters). This microcircuit was used in the study because it can operate with the required frequencies and voltages. Since the applied microcircuit is an inverter, this is taken into account in the image sensor control function: all signals are inverted. While unity was first supplied earlier for the DATA signal, followed by zero, first zero and then unity is supplied after inversion from the microcontroller.

Processing data from a linear image sensor and calculating distance. After a data array is obtained, it should be processed. According to the triangulation method for finding distance, the beam reflected from the surface of the object passes through the optical system to a linear image sensor. The position of the laser point on the image sensor depends on the distance from the object to the sensor, which is how distance to the object is determined. Therefore, the dis-

tance to the object can be calculated after obtaining data from the image sensor. A graph constructed of data from the linear image sensor clearly shows the laser spot, provided that there is an object in the range visible by the sensor (Fig. 6). The spot has the form of a normal (Gaussian) distribution.

Distribution of the spot can be analyzed to find its exact position on the matrix, which is then translated into the distance to the object. There are several ways to do this.

We used the method for finding the center of mass for a system of material points to calculate the center of the spot in this study. This method allows to calculate the center of the spot with great accuracy that is higher than the resolution of the matrix pixels. The formula for the calculation is as follows:

$$x_C = \frac{\sum m_i \cdot x_i}{\sum m_i}$$

where x_C is the center of the laser spot on the linear image sensor, m_i is the value of the i-th pixel of the spot, x_i is the coordinate of the i-th pixel of the spot.

The center value is calculated only for a region of the spot on the data array with a linear image matrix. The numerator in the function calculating the centroid (the center of the laser spot) is multiplied by 10 because the microcontroller takes longer to perform calculations with fractional numbers.

Application of such a solution allows to perform calculations with larger numbers, it does not matter for the microcontroller how big the number is, allowing to maintain the accuracy of the calculations as a result.

The next steps are to convert the centroid into the metric and convert the position of the laser spot in pixels on the linear matrix into the distance in mm to the sensor. The dependence of the laser spot position (pixels) on the spot – object distance (mm) (it was found from the general optical scheme (see Fig. 7). Calibration tables are introduced to convert distances from subpixels to millimeters.

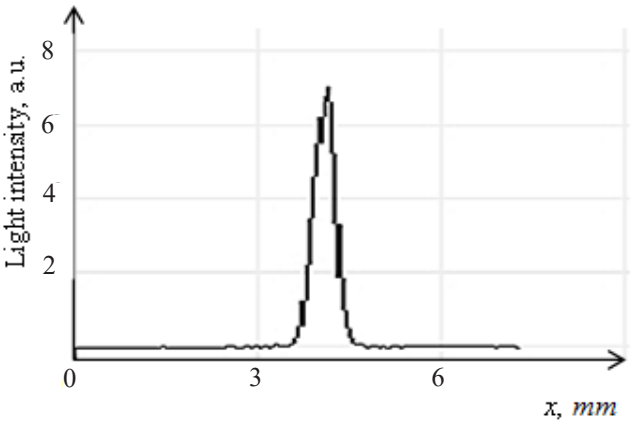

Fig. 6. Laser spot graph (plots of the light intensity versus the x coordinate)

The accuracy of conversion in millimeters depends on the number of elements in the calibration table. The number of table elements in this study was 32.

The bisection method (half division method) was used to search for the range. The half division method allows to exclude half the interval after the each iteration. It is assumed within this method that the function is continuous and has a different sign at the ends of the interval. After the value of the function is calculated in the middle of the interval, one part of the interval is discarded so that the function has a different sign at the ends of the remaining part. Iterations of the bisection method stop if the interval becomes sufficiently small. Since there are 32 values in the calibration table, the method allows to find the interval in just 5 iterations, regardless of the centroid value.

The code of this function consists of about 250 lines, so only part of the code is given. Despite the large code size, the program runs very quickly.

The optical scheme of the displacement sensor

The optical scheme [10] is shown in Fig. 8.

According to Fig. 8, the distance D from the laser (L) to the subject of inquiry (SI) can be found as follows:

$$D = \frac{h}{\tan\theta},$$

where h is the distance between the image sensor (LIS) and the laser (L), θ is the angle between the laser beam and the laser point.

The angle between the laser beam and the returned laser point can be calculated from this formula:

$$\theta = pfc \cdot rpc + ro,$$

where pfc is the number of pixels from the center of the focal plane, rpc is the radian per pixel pitch, ro is the radius offset.

A high reading speed allows to track the position of moving objects, and the resulting accuracy can reach an error of one thousandth of the distance.

An interference filter of 650 nm is installed

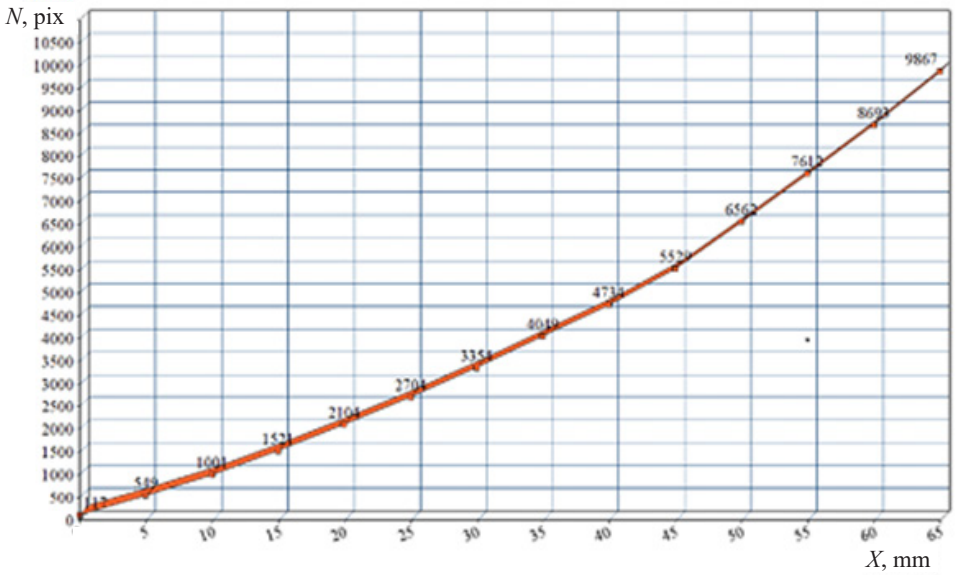

Fig. 7. Plots of the spot position versus the distance from the laser to the measured object (calibration chart)

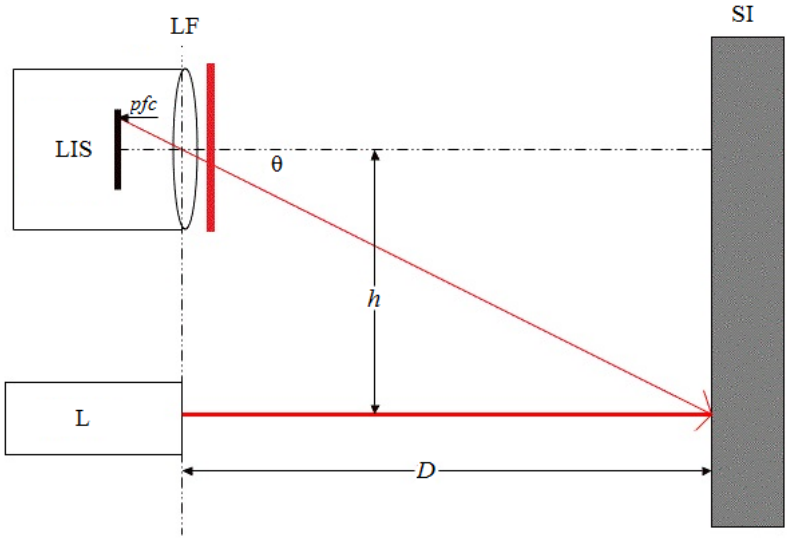

Fig. 8. Optical design of the displacement sensor:
SI – subject of inquiry, L– laser, LF – light filter, LIS – linear image sensor.
Geometric parameters are shown

in front of the lens, which does not transmit light of unnecessary wavelengths, to reduce the influence of external factors and increase accuracy.

Conclusion

The result of the study is a compact triangulation displacement measurement sensor that we have created.

The accuracy of the sensor is 15 – 20 μm, this accuracy is achieved only through correct selection of optics, high-quality filter, accurate calibration on precision equipment, such as CNC machines with digital linear encoders.

The scanning speed is from 3 to 5 thousand measurements per second, which meets most of the needs in the industry. Scanning can be carried out directly on the conveyor without slowing down production.

REFERENCES

1. **Demkin V.N., Stepanov V.A.** Laser methods and means of monitoring the geometric dimensions of products, Measuring Equipment. 69 (2) (2008) 32–35.

2. **Demkin V.N., Stepanov V.A.,** Possibilities of a triangulation laser method for measuring the surface of a complex relief, Metrology. (8) (2007) 32–35.

3. **Demkin V.N., Stepanov V.A.,** Measurement of the roughness profile of materials by the triangulation method, Metrology. (6) (2008) 60–65.

4. **Hausler G.,** Three-dimensional sensors – potentials and limitations, Handbook of computer vision and application, Vol. 1, Academic Press, San Diego, (1999) 485–506.

5. **Crags G., Meuret Y., Danckaert J., Verschaffelt G.,** Characterization of a low-speckle laser line generator, Applied Optics. 51 (20) (2012) 4818–4826.

6. **Tusting R.F., Davis D.L.,** Laser systems and structured illumination for quantitative undersea imaging, Marine Technology Society Journal. 26 (4) (1992) 5–12.

7. **Voisin S., Foufou S., Truchetet F., et al.,** Study of ambient light influence for three-dimensional scanners based on structured light, Optical Engineering. 46 (3) (2007) 030502.

8. **Voegtle T., Schwab I., Landes T.,** Influences of different materials on the measurements of a terrestrial laserscanner (TLS), Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Science. 37, Part B5 (2008) 1061–1066.

9. **Skvortsov A.V., Mirza N.S.,** Algoritmy postroyeniya i analiza triangulatsiy [Algorithms for constructing and analyzing triangulation], Tomsk State University, Tomsk, 2006.

10. **Demkin V.N., Demkin A.V., Shadrin M.V.,** Device for laser scanning, Patent of Russian Federation, IPC G 01 B 11/24, 2012110279/28; declared 03/16/2012; publ. 11/20/2012. Bull. No. 32, 2012.

Received 16.01.2020, accepted 27.01.2020.

THE AUTHORS

STEPANOV Vladimir A.
Ryazan State University named for S. Ysenin
46 Svobody St., Ryazan, 390000, Russian Federation
vl.stepanov@365.rsu.edu.ru

MOOS Evgueniy N.
Ryazan State University named for S. Ysenin
46 Svobody St., Ryazan, 390000, Russian Federation
e.moos@rsu.edu.ru

SHADRIN Maksim V.
Ryazan State University named for S. Ysenin
46 Svobody St., Ryazan, 390000, Russian Federation
addressworken
m.shadrin@russia.ru

SAVIN Vladislav N.
Ryazan State University named for S. Ysenin
46 Svobody St., Ryazan, 390000, Russian Federation
savin-vladislav@mail.ru

UMNYASHKIN Andrew V.
Ryazan State University named for S. Ysenin
46 Svobody St., Ryazan, 390000, Russian Federation
a.umniashkin@kvantron.com

UMNYASHKIN Nicholas V.
Ryazan State University named for S. Ysenin
46 Svobody St., Ryazan, 390000, Russian Federation
n.umniashkin@kvantron.com

СПИСОК ЛИТЕРАТУРЫ

1. **Demkin V.N., Stepanov V.A.** Laser methods and means of monitoring the geometric dimensions of products // Measuring Equipment. 2008. Vol. 69. No. 2. Pp. 32–35.

2**. Демкин В.Н., Степанов В.А.** Возможности триангуляционного лазерного метода измерения поверхности сложного рельефа // Метрология. 2007. № 8. С. 32–36.

3. **Демкин В.Н., Степанов В.А.** Измерение профиля шероховатости поверхности триангуляционным способом // Метрология. 2008. № 6. С. 60–65.

4. **Hausler G.** Three-dimensional sensors – potentials and limitations. Handbook of computer vision and application. Vol. 1. San Diego: Academic Press, 1999. Pp. 485–506.

5. **Crags G., Meuret Y., Danckaert J., Verschaffelt G.** Characterization of a low-speckle laser line generator // Applied Optics. 2012. Vol. 51. No. 20. Pp. 4818–4826.

6. **Tusting R.F., Davis D.L.** Laser systems and structured illumination for quantitative undersea imaging // Marine Technology Society Journal. 1992. Vol. 26. No. 4. Pp. 5–12.

7. **Voisin S., Foufou S., Truchetet F., Page D., Abidi M.** Study of ambient light influence for three-dimensional scanners based on structured light // Optical Engineering. 2007. Vol. 46. No. 3. P. 030502.

8. **Voegtle T., Schwab I., Landes T.** Influences of different materials on the measurements of a terrestrial laser scanner (TLS) // Int. Archives of the Photogrammetry. Remote Sensing and Spatial Information Science. 2008. Vol. 37. Part B5. Pp. 1061–1066.

9. **Скворцов А.В., Мирза Н.С.** Алгоритмы построения и анализа триангуляции. Томск: Изд-во Томского университета, 2006. 160 с.

10. **Демкин В.Н., Демкин А.В., Шадрин М.В.** Устройство для лазерного сканирования. Патент Российской Федерации. МПК G 01 B 11/24. 2012110279/28; заявлено 03/16/2012; опубликовано 11/20/2012. Бюлл. № 32. 2012.

Статья поступила в редакцию 16.01.2020, принята к публикации 27.01.2020.

СВЕДЕНИЯ ОБ АВТОРАХ

СТЕПАНОВ Владимир Анатольевич – *доктор физико-математических наук, профессор кафедры общей и теоретической физики и методики преподавания физики Рязанского государственного университета имени С.А. Есенина, г. Рязань, Российская Федерация.*
390000*, Российская Федерация,* г. Рязань, ул. Свободы, 46
vl.stepanov@365.rsu.edu.ru

МООС Евгений Николаевич – *доктор технических наук профессор кафедры общей и теоретической физики и методики преподавания физики Рязанского государственного университета имени С.А. Есенина, г. Рязань, Российская Федерация.*
390000*, Российская Федерация, г.* Рязань, ул. Свободы, 46
e.moos@365.rsu.edu.ru

ШАДРИН Максим Владимирович – *аспирант кафедры общей и теоретической физики и методики преподавания физики Рязанского государственного университета имени С.А. Есенина, г. Рязань, Российская Федерация.*
390000*, Российская Федерация,* г. Рязань, ул. Свободы, 46
m.shadrin@russia.ru

САВИН Владислав Николаевич – *инженер кафедры общей и теоретической физики и методики преподавания физики Рязанского государственного университета имени С.А. Есенина, г. Рязань, Российская Федерация.*
390000*, Российская Федерация,* г. Рязань, ул. Свободы, 46
savin-vladislav@mail.ru

УМНЯШКИН Андрей Владимирович – *аспирант кафедры общей и теоретической физики и методики преподавания физики Рязанского государственного университета имени С.А. Есенина, г. Рязань, Российская Федерация.*
390000*, Российская Федерация,* г. Рязань, ул. Свободы, 46
a.umniashkin@kvantron.com

УМНЯШКИН Николай Владимирович – *аспирант кафедры общей и теоретической физики и методики преподавания физики Рязанского государственного университета имени С.А. Есенина, г. Рязань, Российская Федерация.*
390000*, Российская Федерация,* г. Рязань, ул. Свободы, 46
n.umniashkin@kvantron.com

DOI: 10.18721/JPM.13106
УДК 532.526.4, 533.6.08

AN EXPERIMENTAL STUDY OF THE FLOW IN THE AREA OF INFLUENCE OF A CYLINDER IMMERSED IN THE FREE CONVECTIVE BOUNDARY LAYER ON A VERTICAL SURFACE

Yu. S. Chumakov, A.M. Levchenya, E.F. Khrapunov

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

New experimental data that quantitatively characterize fields of the mean velocity and temperature, the intensity of temperature and velocity pulsations, and also velocity-temperature correlations in the near zone of a circular cylinder placed on the vertical heated surface at the height corresponding to the fully turbulent flow regime have been presented. Systematic measurements in the middle vertical plane (the plane that contains the cylinder axis) were performed using constant temperature anemometer and resistance temperature detectors. The experimental data was compared with numerical simulation one obtained through solving the RANS equations. The overall data were in good agreement and indicated the cardinal restructuring of the flows both before the cylinder (where the horseshoe-shaped vortex formed) and behind the obstacle (in the near separated area and the recovery one of the natural convective near-wall layer).

Keywords: circular cylinder, free-convective heat exchange, hot wire anemometry, area of influence

Citation: Chumakov Yu.S., Levchenya A.M., Khrapunov E.F., An experimental study of the flow in the area of influence of a cylinder immersed in the free convective boundary layer on a vertical surface, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (1) (2020) 59–69. DOI: 10.18721/JPM.13106

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ТЕЧЕНИЯ В ЗОНЕ ВЛИЯНИЯ ЦИЛИНДРА, ПОГРУЖЕННОГО В СВОБОДНОКОНВЕКТИВНЫЙ ПОГРАНИЧНЫЙ СЛОЙ НА ВЕРТИКАЛЬНОЙ ПОВЕРХНОСТИ

Ю.С. Чумаков, А.М. Левченя, Е.Ф. Храпунов

Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Российская Федерация

Представлены новые экспериментальные данные, количественно характеризующие поля осредненной по времени скорости, осредненной температуры, интенсивности пульсаций скорости и температуры, а также корреляции пульсаций скорости и температуры в окрестности круглого цилиндра, установленного на вертикальной нагретой поверхности, на высоте, соответствующей развитому турбулентному режиму течения. Систематические измерения в средней вертикальной (проходящей через ось цилиндра) плоскости выполнены методами термоанемометрии и термометра сопротивления. Проведено сравнение измеренных профилей осредненной скорости и температуры с результатами численного моделирования на основе уравнений Рейнольдса. Достигнуто хорошее соответствие опытных и расчетных данных, которые в целом указывают на кардинальную перестройку течения как перед цилиндром в области формирования подковообразных вихревых структур, так и за препятствием, в ближней отрывной зоне и зоне восстановления свободноконвективного пристенного течения.

Ключевые слова: круглый цилиндр, свободноконвективный теплообмен, пограничный слой, термоанемометрия, зона влияния

Ссылка при цитировании: Чумаков Ю.С., Левченя А.М., Храпунов Е.Ф. Экспериментальное исследование течения в зоне влияния цилиндра, погруженного в свободноконвективный пограничный слой на вертикальной поверхности // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2020. Т. 1 № .13. С. 66–77. DOI: 10.18721/JPM.13106

Introduction

Characteristic features of heat transfer in case of turbulent natural convection developing on vertical heated surfaces are very important for different applications, such as cooling large surfaces of heat exchangers, construction of high-rise buildings and structures, fire safety, energy industry, safety of nuclear reactors, etc. Many studies consider the problem of a turbulent natural-convective boundary layer developing along a vertical heated plate as a simplified model of such flows.

Most experimental studies on the boundary layer with natural convection analyzed air flow at relatively small differences in absolute temperature (not exceeding 30% of the ambient temperature). The main parameters of the flow and heat transfer, such as profiles of mean temperature and mean velocity distributions, shear stress on the wall were measured by Warner and Arpaci [1], Cheesewright [2], Pirovano et al. [3], Smith [4], Tsuji and Nagano [5], Chumakov [6, 7]. The experiments measuring different turbulence characteristics were carried out by Smith [4], Cheesewright and Doan [8], Miyamoto et al. [9], Cheesewright and Ierokipiotis [10], Tsuji and Nagano [5, 11], Nikolskaya and Chumakov [12], Kuzmitskiy et al. [13]. The accumulated experimental results can be generalized, contributing to a deeper understanding of the basic properties of the flow and the specifics of the turbulent regime for this general case.

The natural-convective boundary layer formed on a heated vertical surface can be substantially disturbed by a single obstacle or a combination of several obstacles in many practical applications. Industrial structures or residential buildings (large-sized containers for storing spent nuclear fuel, buildings with solar panels, etc.) can act as such macro-roughnesses. In some cases, obstacles are deliberately introduced into the natural-convective boundary layer in order to control or control its behavior and thus intensify heat transfer. Refs. [14–16, 17] are examples of such studies. A system of vertical fins is typically installed on the heated surface to improve heat transfer in the natural convective regime. A recent focus has been on *V*-shaped fins capable of susbstantially enhancing heat transfer [18–20]. There is much interest towards problems of flow control and heat transfer enhancement in boundary layers evolving on extended vertical heated plates under turbulent natural convection. Some issues related to using fins of different heights for enhancing heat transfer were studied experimentally in [21, 22]. Ref. [23] compared the effectiveness of enhancers in the form of a long plate and a series of short plates located at the same distance from each other and mounted across the boundary layer flow. The experiments carried out confirmed that using a transverse row of short plates can help achieve greater disturbance of the flow in the wake of these obstacles, significantly enhancing heat transfer.

Flow structure and heat transfer in front of a semi-infinite cylinder piercing the turbulent natural convective boundary layer formed on a vertical heated surface were considered in [24, 25]. Particular attention was paid to using RANS-based three-dimensional simulations to study the effects from horseshoe-shaped vortex structures. Ref. [26] reported on an extended RANS-based study of for the case of a cylinder of finite height disturbing the boundary layer. The authors analyzed the influence of cylinder height and thermal conditions on the cylinder surface, transforming the flow structure and heat transfer in the front and rear parts of this cylinder.

In this study, we measured the mean and fluctuation characteristics of the velocity and temperature fields near a circular cylinder mounted on a vertical heated surface at a height corresponding to developed turbulent flow, analyzing the obtained data.

Experimental setup

Experimental studies were carried out with the testbed set up in the Laboratory of Hydroaerodynamics of the Institute of Applied Mathematics and Mechanics at Peter the Great St. Petersburg Polytechnic University in the 1990s [6, 7], modernized in the last three years (Fig. 1). Natural-convective airflow is generated by vertical aluminum plate *4* 90 cm wide and 495 cm high. 25 heaters (not shown in Fig. 1) are mounted on the back side of the plate; they are controlled by an electronic system capable of maintaining the thermal regime set for a long time.

Setting a specific heating mode for each of the 25 sections, we can simulate different laws for heating the surface along its height, in particular, giving a constant surface temperature. Because the plate is very high, all three regimes of airflow, i.e., laminar, transitional, and fully developed turbulent can be simulated up to the Grashof number of $4.5 \cdot 10^{11}$.

A coordinate device is used to move measuring sensor *13* in the evolving flow; it provides accuracy of about 0.5 mm for motion along the vertical coordinate X, and of about 0.001 mm along the coordinate Y normal to the surface (across the boundary layer); the sensor moves along the normal coordinate using stepper motor *14* by a preset algorithm. Flow parameters are measured fully automatically in one section of the boundary layer.

Notably, sensor *13* can move across the boundary layer with such high accuracy only in one direction, for example, away from the surface. The algorithm moving the sensor to the first point near the surface uses a reversible form of motion, greatly reducing the accuracy with which the coordinates of this point are determined. The accuracy with which the normal coordinate of the first measurement point was determined in this study was not worse than 0.1 mm. The sensor subsequently moves in one direction and the accuracy of movement corresponds to that given above (0.001 mm).

The same procedure is used for measuring velocity and temperature at a given point in space. It consists in the following. The analog signal corresponding to temperature (or velocity) is digitized using an analog-to-digital converter by the parameters set: sample number (N) and sampling rate (Hr). N = 2000, and Hr = 100 Hz in our study; thus, the signal processing time at a given point is 20 s. Next, the mean and root-mean-square (RMS) fluctuation of the given quantity are found.

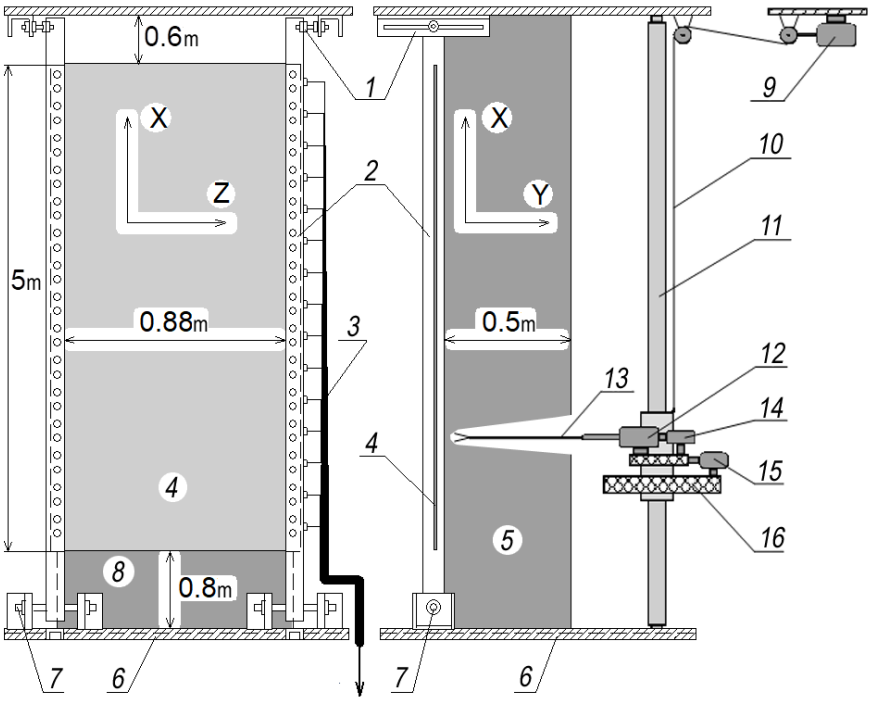

Fig. 1. Schematic of test bench with heated vertical plate and coordinate device:
upper mount *1*; vertical supports *2*; cables *3* of temperature sensors; heated plate *4*; side curtains *5*; foundation *6*; lower hinge mount *7*; rear curtain *8*; electric motors *9* and *15*; cable *10*; guide posts *11*; fixation system *12* for sensor holder; probe *13*; stepper motor *14*; movable carriage *16*

Measuring section and measurement procedure

We considered the region of interaction of a fully developed turbulent layer with a three-dimensional obstacle in the form of a poorly conducting (adiabatic) cylinder 40 mm in diameter, with the same height (Fig. 2, *a*). The cylinder was mounted at a height of 1800 mm, measured from the leading edge of the plate, which corresponds to a Grashof number (determined by the standard technique) of approximately $2 \cdot 10^{10}$ with the plate heated to 60 °C and the ambient air temperature of about 26 °C. The layer flowing onto the cylinder is turbulent with the given Grashof number, and its thickness is approximately four times the height of the cylinder.

We used a thermal anemometer (TA) and a resistance thermometer for systematic combined measurements of mean velocity and temperature fields in the vertical midplane (passing through the cylinder axis), the intensity of velocity and temperature fluctuations and their correlation.

If velocity is measured by the TA in nonisothermal flow, the anemometer readings should be interpreted taking into account the effect of temperature. The given flow is characterized by low mean velocities and a high level of fluctuations, so the current velocities measured by the common method of thermal compensation by mean temperature can be largely inaccurate. The original method of thermal compensation, described in [27], was used for measurements in our study. According to this method, the TA reading corresponding to the current velocity at a given point in space is interpreted taking into account the current temperature at the same point. We used a special calibration setup [27] with uniform motion of the sensor along unevenly heated still air. The setup allows to calibrate the sensors at velocities from 1 to 50 cm/s with air temperatures ranging from 20 to 80 °C. Calibration results are represented as the voltages from the TA depending on flow velocity, and the coefficients in this dependence are functions of temperature.

Thus, to measure velocity in nonisothermal flow, the probe must consist of at least two sensors. One sensor (cold wire) is used to measure temperature, and the other (hot wire) is used to measure voltage, depending on the velocity and temperature of the flow. The measured temperatures are used to determine the calibration coefficients and calculate the velocity at a given point in the flow.

Fig. 2, *b* shows a photograph of a two-wire probe used in this study to measure the current values of temperature and velocity. Tungsten wires 5 μm in diameter and 3.5 mm long serve as sensitive elements. The probe is oriented so that the cold wire is upstream of the hot wire: this reduces the effects from the thermal "microflow" from the hot wire. Both wires are located horizontally parallel to the plate surface, so the probe can be brought very close to it. It should be borne in mind that the given location of the velocity sensor (hot wire) corresponds to the measured magnitude of the current velocity vector lying in the vertical midplane.

Measurement results and discussion

Flow parameters measured in several normal sections in front of the cylinder and behind it are shown in Figs. 3–5. Notably, the "temperature" and "velocity" wires of the sensor are spaced 2 mm apart; as noted above, the velocity wire is located above the temperature one. This explains the small shifts in the vertical coordinate X on the distributions of different measured quantities. The distributions here and below are marked by the distance dX from the corresponding measuring wire to the nearest (leading or trailing) edge of the cylinder. The distance to the plate along the normal to it is normalized to the height of the cylinder h= 40 mm.

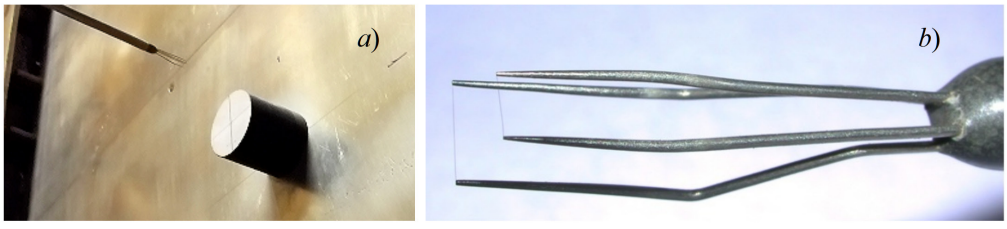

Fig. 2. Fragments of experimental setup: plate *a* with cylinder mounted on it (the measuring probe can be seen nearby); two-wire probe *b* for simultaneously measuring the current values of flow velocity and air temperature.

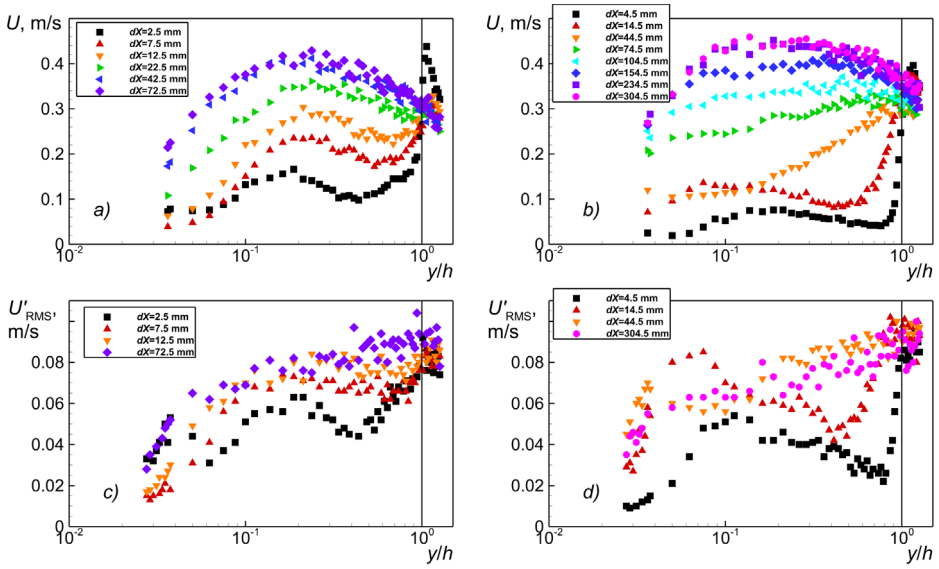

Fig. 3. Measured flow velocity field: mean velocity in front of the cylinder (*a*) and behind it (*b*); velocity fluctuations in front of the cylinder (*c*) and behind it (*d*);
dX are the distances from the corresponding measuring wire to the nearest edge of the cylinder; vertical lines indicate the position of the cylinder end

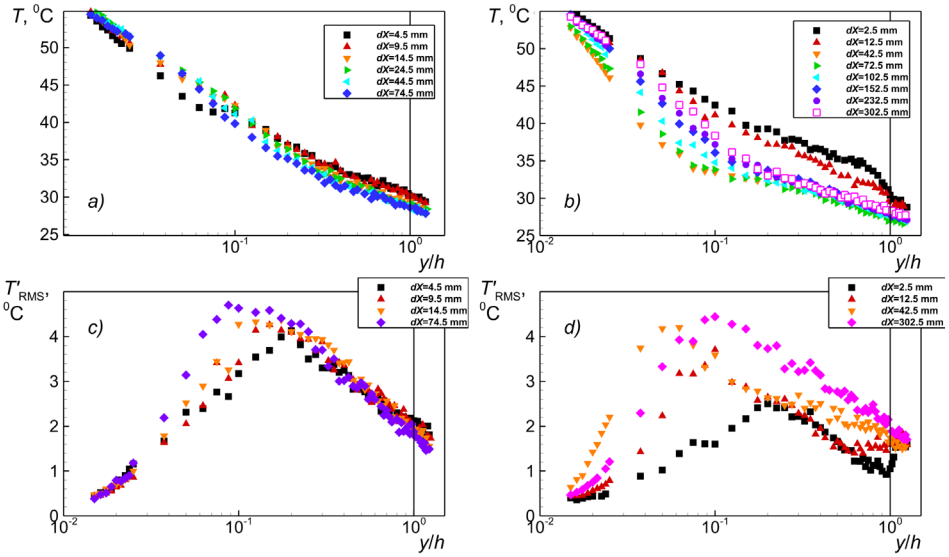

Fig. 4. Measured temperature fields: mean air temperature in front of the cylinder (*a*) and behind it (*b*); temperature fluctuations in front of the cylinder (*c*) and behind it (*d*);
dX are the distances from the corresponding measuring wire to the nearest edge of the cylinder; vertical lines indicate the position of the cylinder end

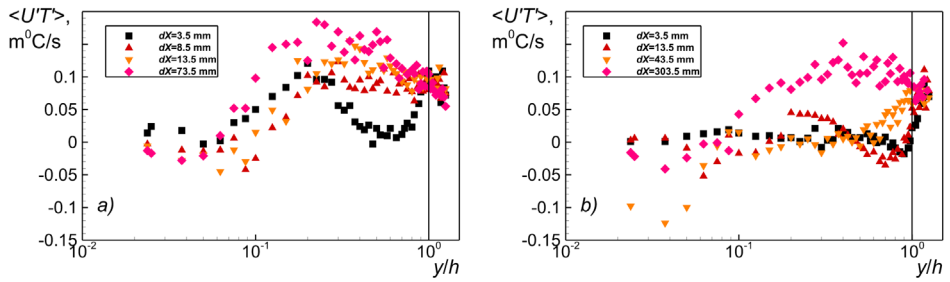

Fig. 5. Measured correlation distributions $\langle U'T' \rangle$ in front of the cylinder (*a*) and behind it (*b*);
dX are the distances from the corresponding measuring wire to the nearest edge of the cylinder; vertical lines indicate the position of the cylinder end

Fig. 3 shows the distributions of mean velocity and its fluctuations. The vertical line marks the position of the cylinder end in this and the following figures with experimental data. The statistically two-dimensional natural-convective boundary layer developed on the plate, with the maximum flow velocity of about 0.4 m/s, slows down as it approaches the leading edge of the cylinder (Fig. 3, *a*), while an increase in velocity magnitudes is observed in the region $y/h > 1$ in the three sections nearest to the leading edge, which corresponds to the zone where the flow through the end of the cylinder accelerates. The region of accelerated flow above the end persists in the first sections after the trailing edge in the near wake behind the cylinder (Fig. 3, *b*), and a substantial decrease in the velocity magnitude is observed near the surface, in the stagnation region of the cylinder, as well as in front of it, especially in recirculation zone. The natural-convective boundary layer is gradually restored downstream. Fig. 3, *c*, *d* shows individual measured distributions of RMS velocity fluctuations. Unfortunately, the TA method does not allow to reliably measure the velocities in the immediate vicinity of the highly conductive wall. In our case, the thickness of the "forbidden zone" is about 2 mm, which corresponds to 5% of the obstacle height.

Similar to Fig. 3, Fig. 4 shows the distributions of mean temperature and its fluctuations in the sections in front of the cylinder and behind it. The temperature distributions in most sections in front of the cylinder and far from it are very similar. The temperature distributions appear to be somewhat less monotonic near the leading edge of the obstacle, and a significant local decrease in temperature is observed at a distance of several mm in front of the edge (in the region less than 10% from the cylinder height). The results of numerical simulation given in [26] indicate that this decrease corresponds to the region where a horseshoe-shaped vortex structure forms, with relatively cold flow from the outer part of the boundary layer attaching to the surface of the plate under the action of a horseshoe-shaped vortex. Greater stratification between the distributions in different sections is observed in the wake of the cylinder (Fig. 4, *b*). The flow is well-mixed in the recirculation area close to the trailing edge of the obstacle, warmed by heat from the hot surface of the plate. The temperatures observed downstream, behind the point where the boundary layer that separated in front of the obstacle at a relatively small distance from the plate (for example, at a normal distance of about 10% of the cylinder height) reattaches, are significantly (10°) lower than those characteristic for an undisturbed

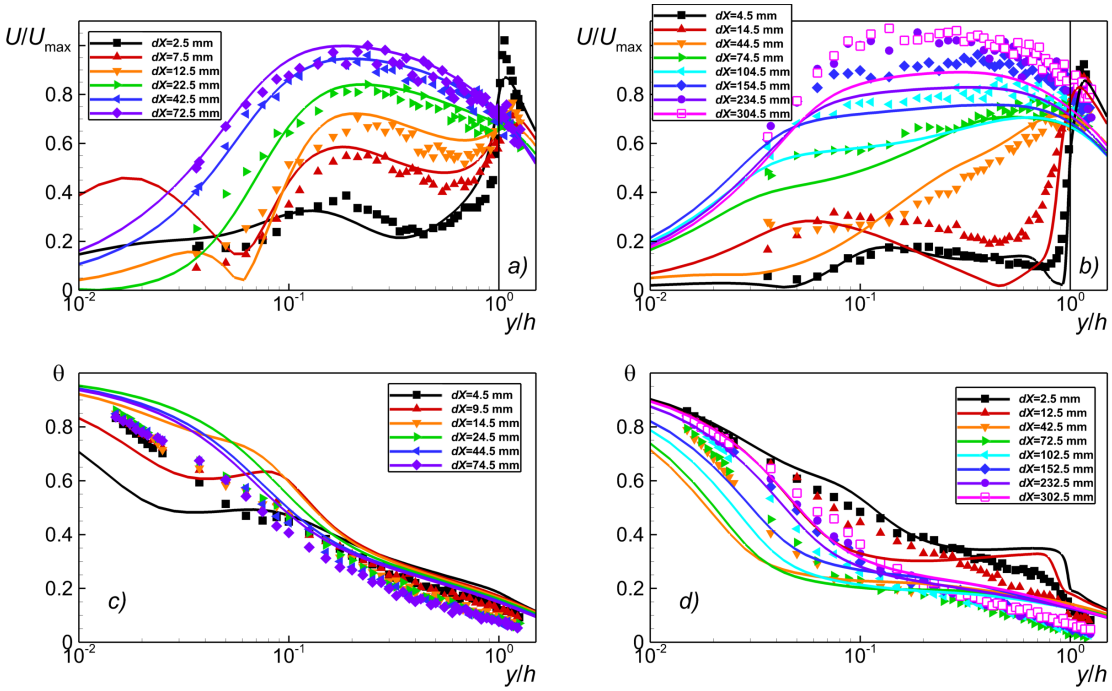

Fig. 6. Comparison of experimental data (symbols) with RANS-based numerical simulation (solid lines): the mean normalized flow velocity in front of the cylinder (*a*) and behind it (*b*) are shown, as well as the mean dimensionless air temperature in front of the cylinder (*c*) and behind it (*d*); *dX* are the distances from the corresponding measuring wire to the nearest cylinder edge

boundary layer at this distance from the plate. The temperature distribution is gradually restored further downstream, corresponding to the case of an undisturbed boundary layer.

The distributions of RMS temperature fluctuations shown in Fig. 4, *c* and *d*, as well as the distributions shown in Fig. 5 for the normalized correlated fluctuations of velocity and temperature make it possible to compare the positions of the fluctuation peaks in front of the cylinder and behind it, and also to estimate the general magnitude of these quantities.

Comparison of experimental results with numerical simulation data

It is of interest to compare the obtained experimental data with the recently published results on numerical simulation of the flow under similar conditions [26]. The numerical study in [26] reports on the structure of three-dimensional flow and heat transfer in the vicinity of a circular cylinder disturbing a turbulent natural-convective boundary layer. The computations were performed using the Reynolds averaged Navier–Stokes equations (RANS) according to Menter's SST turbulence model. The geometric configuration and the conditions adopted in the computations for one of the cases (cylinder sizes, thermal conditions on its surface, parameters of the incident boundary layer) are close to the conditions of the experiments described above. In fact, the computations described in [26] acted as auxiliary for the experiments in our study, making it possible to predict some characteristics of the real flow developing in the vicinity of the given cylinder. Fig. 6 compares the measured mean values with the data obtained by numerical simulation. The velocity U is normalized to its maximum value $U_{\max}$ 72.5 mm away from the cylinder, and the dimensionless temperature θ is determined by the standard formula for such problems

$$\theta = \frac{T - T_a}{T_w - T_a},$$

where T_w, T_a are the temperatures of the heated surface and external space, respectively.

Notably, the computed components of the mean velocity vector were recalculated to obtain the "effective" values of U obtained in measurements with a sensor with one "velocity" wire, which is not sensitive to the direction of the velocity vector but only responds to the current magnitude of the velocity transverse to the wire. Recalculation is based on simple relations using the computational data on the local direction of the averaged velocity vector at the measurement point. Analyzing the data in Fig. 6, we concluded that fairly satisfactory or excellent (for individual distributions) agreement was obtained between the experimental and compational data, generally pointing to major restructuring of the flow both in front of the cylinder, where horseshoe-shaped vortex structures are formed, and behind the cylinder, including the near separated region and the region where natural-convective near-wall flow is restored. Notably, however, there is a pronounced difference between the results of experiments and RANS computations for the latter region. It can be seen from Fig. 6, *b* that the natural-convective region is restored more slowly in the computational model.

Conclusion

We have obtained new experimental data for a fully developed turbulent natural-convective boundary layer interacting with a circular poorly conducting cylinder immersed in it, quantitatively characterizing the fields of time-averaged airflow velocity, mean air temperature, velocity and temperature fluctuation rates, as well as the correlation of velocity and temperature fluctuations.

The family of measured distributions of averaged velocity and temperature was used to compare the results obtained experimentally and by numerical simulation based on RANS approximation. We have obtained fairly satisfactory or excellent (for individual distributions) agreement between the experimental and computational data, generally pointing to major restructuring of the flow both in front of the cylinder, where horseshoe-shaped vortex structures are formed, and behind the obstacle, including the near separated region and the region where natural-convective flow is restored. At the same time, the results of RANS computations indicate that the natural-convective boundary layer is restored with a delay in the far wake of the cylinder, compared with the experimental data.

Measurements have been performed this far only for the vertical midplane passing through the axis of the cylinder; we plan to continue studies on the three-dimensional structure of the flow, analyzing other sections of the working area.

The study financially supported by a Russian Science Foundation grant (project 18-19-00082). RANS computations were carried out using the resources of the Supercomputer Center at Peter the Great St. Petersburg Polytechnic University (www.scc.spbstu.ru).

REFERENCES

1. **Warner C.Y., Arpaci V.S.,** An experimental investigation of turbulent natural convection in air at low pressure along a vertical heated flat plate, International Journal of Heat and Mass Transfer. 11 (3) (1968) 397–406.

2. **Cheesewright R.,** Turbulent natural convection from a vertical plane surface, Journal of Heat Transfer. 90 (1) (1968) 1–8.

3. **Pirovano A., Viannay S., Jannot M.,** Convection naturelle en rйgime turbulent le long d'une plaque plane verticale, Proceedings of the 9th International Heat Transfer Conference, Elsevier, Paris, Amsterdam. 4 (1.8) (1970) 1–12.

4. **Smith R.R.,** Characteristics of turbulence in free convection flow past a vertical plate, Ph.D. Thesis, University of London, 1972.

5. **Tsuji T., Nagano Y.,** Characteristics of a turbulent natural convection boundary layer along a vertical flat plate, International Journal of Heat and Mass Transfer. 31 (8) (1988) 1723–1734.

6. **Chumakov Yu.S., Kuzmitsky V.A.,** Surface shear stress and heat flux measurements at a vertical heated plate under free convection heat transfer, Russian Journal of Engineering Thermophysics. 8 (1–4) (1998) 1–15.

7. **Chumakov Yu.S.,** Temperature and velocity distributions in a free-convection boundary layer on a vertical isothermal surface, High Temperature. 37 (5) (1999) 714–719.

8. **Cheesewright R., Doan K.S.,** Space-time correlation measurements in a turbulent natural convection boundary layer, International Journal of Heat and Mass Transfer. 21(91) (1978) 911–921.

9. **Miyamoto M., Kajino H., Kurima J., Takanami I.,** Development of turbulence characteristics in a vertical free convection boundary layer, Proceedings of the 7th International Heat Transfer Conference, Munich, FRG. 2 (1982) 323–328.

10. **Cheesewright R., Ierokipiotis E.,** Velocity measurements in a turbulent natural convection boundary layer, Proceedings of the 7th International Heat Transfer Conference, Munich, FRG. 2 (1982) 305–309.

11. **Tsuji T., Nagano Y.,** Turbulence measurements in a natural convection boundary layer along a vertical flat plate, International Journal of Heat and Mass Transfer. 31 (10) (1988) 2101–2111.

12. **Nikolskaya S.B., Chumakov Yu.S.,** Experimental investigation of pulsation motion in a free-convection boundary layer, High Temperature. 38 (2) (2000) 231–237.

13. **O.A. Kuzmitskiy, Nikolskaya S.B., Chumakov Yu.S.,** Spectral and correlation characteristics of velocity and temperature fluctuations in a free-convection boundary layer, Heat Transfer Research. 33 (3–4) (2002) 144–147.

14. **Bhavnani S.H., Bergles A.E.,** Effect of surface geometry and orientation on laminar natural convection heat transfer from a vertical flat plate with transverse roughness elements, International Journal of Heat and Mass Transfer. 13 (5) (1990) 965–981.

15. **Burak V.S., Volkov S.V., Martynenko O.G., et al.,** Experimental study of free-convection flow on a vertical plate with constant heat flux in the presence of one or more steps, International Journal of Heat and Mass Transfer. 38 (1) (1995) 147–154.

16. **Aydin M.,** Dependence of the natural convection over a vertical flat plate in the presence of the ribs, International Communications in Heat and Mass Transfer. 24 (4) (1997) 521–531.

17. **Polidori G., Padet J.,** Transient free convection flow on a vertical surface with an array of large scale roughness elements, Experimental Thermal and Fluid Science. 27 (3) (2003) 251–260.

18. **Misumi T., Kitamura K.,** Enhancement techniques for natural convection heat transfer from vertical finned plate, Heat transfer – Japanese Research. 23 (16) (1994) 513–524.

19. **Fujii M.,** Enhancement of natural convection heat transfer from a vertical heated plate using inclined fins, Heat Transfer – Asian Research. 36 (6) (2007) 334–344.

20. **Naserian M., Fahiminia M., Goshayeshi H.R.,** Experimental and numerical analysis of natural convection heat transfer coefficient of V-type fin configurations, Journal of Mechanical Science and Technoljgy. 27 (7) (2013) 2191–2197.

21. **Misumi T., Kitamura K.,** Enhancement of natural convective heat transfer from tall vertical heated plates, JSME B (in Japanese), 65 (640) (1999) 4041–4048.

22. **Komori K., Inagaki T., Kito S., Mizoguchi N.,** Natural convection heat transfer along a vertical flat plate with a projection in the turbulent region, Heat Transfer – Asian Research. 30 (3) (2001) 222–233.

23. **Tsuji T., Kajitani T., Nishino T.,** Heat transfer enhancement in a turbulent natural convection boundary layer along a vertical flat plate, International Journal of Heat and Fluid Flow. 28 (6) (2007) 1472–1483.

24. **Levchenya A.M., Smirnov E.M., Zhukovskaya V.D., Ivanov N.G.,** Numerical study of 3D turbulent flow and local heat transfer near a cylinder introduced into the free-convection boundary layer on a vertical plate, Proceedings of the 16th International Heat Transfer Conference, IHTC-16, August 10–15, 2018, Beijing, China, Paper IHTC16-22916. DOI: 10.1615/IHTC16.hte.022916, (2018) 5493–5500.

25. **Levchenya A.M., Smirnov E.M., Zhukovskaya V.D.,** Numerical study of 3D flow structure near a cylinder piercing turbulent free-convection boundary layer on a vertical plate, AIP Conference Proceedings. 1959 (2018) 050017.

26. **Smirnov E.M., Levchenya A.M., Zhukovskaya V.D.,** RANS-based numerical simulation of the turbulent free convection vertical-plate boundary layer disturbed by a normal-to-plate circular cylinder, International Journal of Heat and Mass Transfer. 144 (December) (2019) 118573.

27. **Kuzmitskii V.A., Chumakov Yu.S.,** Facility for static calibration of a hot-wire anemometer at low velocities in a nonisothermal air medium, High Temperature. 33 (1) (1995) 109–113.

Received 17.01.2020, accepted 31.01.2020.

THE AUTHORS

CHUMAKOV Yuriy S.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
chymakov@yahoo.com

LEVCHENYA Alexander M.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
levchenya_am@spbstu.ru

KHRAPUNOV Evgeniy F.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
hrapunov.evgenii@yandex.ru

СПИСОК ЛИТЕРАТУРЫ

1. **Warner C.Y., Arpaci V.S.** An experimental investigation of turbulent natural convection in air at low pressure along avertical heated flat plate // International Journal of Heat and Mass Transfer. 1968. Vol. 11. No. 3. Pp. 397–406.

2. **Cheesewright R.** Turbulent natural convection from avertical plane surface // Journal of Heat Transfer. 1968. Vol. 90. No. 1. Pp. 1–8.

3. **Pirovano A., Viannay S., Jannot M.** Convection naturelle en rйgime turbulent le long d'une plaque plane vertical // Proceedings of the 9th International Heat Transfer Conference, Paris, Amsterdam: Elsevier, 1970. Vol. 4.1.8. Pp. 1–12.

4. **Smith R.R.** Characteristics of turbulence in free convection flow past a vertical plate. Ph.D. Thesis, University of London, 1972.

5. **Tsuji T., Nagano Y.** Characteristics of a turbulent natural convection boundary layer along a vertical flat plate // International Journal of Heat and Mass Transfer. 1988. Vol. 31. No. 8. Pp. 1723–1734.

6. **Chumakov Yu.S., Kuzmitsky V.A.** Surface shear stress and heat flux measurements at a vertical heated plate under free convection heat transfer // Russian Journal of Engineering Thermophysics. 1998. Vol. 8. No. 1–4. Pp. 1–15.

7. **Чумаков Ю.С.** Распределение температуры и скорости в свободноконвективном пограничном слое на вертикальной изотермической поверхности // Теплофизика высоких температур. 1999. Т. 37. № 5. С. 744–749.

8. **Cheesewright R., Doan K.S.** Space-time correlation measurements in a turbulent natural convection boundary layer // International Journal of Heat and Mass Transfer. 1978. Vol. 21. No. 7. Pp. 911–921.

9. **Miyamoto M., Kajino H., Kurima J., Takanami I.** Development of turbulence characteristics in a vertical free convection boundary layer // Proceedings of the 7th International Heat Transfer Conference. Munich, FRG. Vol. 2. 1982. Pp. 323–328.

10. **Cheesewright R., Ierokipiotis E.** Velocity measurements in a turbulent natural convection boundary layer // Proceedings of the 7th International Heat Transfer Conference. Munich, FRG. Vol. 2. 1982. Pp. 305–309.

11. **Tsuji T., Nagano Y.** Turbulence measurements in a natural convection boundary layer along a vertical flat plate // International Journal of Heat and Mass Transfer. 1988. Vol. 31. No. 10. Pp. 2101–2111.

12. **Никольская С.Б., Чумаков Ю.С.** Экспериментальное исследование пульсационного движения в свободноконвективном пограничном слое // Теплофизика высоких температур. 2000. Т. 38. № 2. С. 249–256.

13. **Kuzmitskiy O.A., Nikolskaya S.B., Chumakov Yu.S.** Spectral and correlation characteristics of velocity and temperature fluctuations in a free-convection boundary layer // Heat Transfer Research. 2002. Vol. 33. No. 3–4. Pp. 144–147.

14. **Bhavnani S.H., Bergles A.E.** Effect of surface geometry and orientation on laminar natural convection heat transfer from a vertical flat plate with transverse roughness elements // International Journal of Heat and Mass Transfer. 1990. Vol. 13. No. 5. Pp. 965–981.

15. **Burak V.S., Volkov S.V., Martynenko O.G., Khramtsov P.P., Shikh I.A.** Experimental study of free-convection flow on a vertical plate with constant heat flux in the presence of one or more steps // International Journal of Heat and Mass Transfer. 1995. Vol. 38. No. 1. Pp. 147–154.

16. **Aydin M.** Dependence of the natural convection over a vertical flat plate in the presence of the ribs // International Communications in Heat and Mass Transfer. 1997. Vol. 24. No. 4. Pp. 521–531.

17. **Polidori G., Padet J.** Transient free convection flow on a vertical surface with an array of large scale roughness elements // Experimental Thermal and Fluid Science. 2003. Vol. 27. No. 3. Pp. 251–260.

18. **Misumi T., Kitamura K.** Enhancement techniques for natural convection heat transfer from vertical finned plate // Heat transfer – Japanese Research. 1994. Vol. 23. No. 16. Pp. 513–524.

19. **Fujii M.** Enhancement of natural convection heat transfer from a vertical heated plate using inclined fins // Heat Transfer – Asian Research. 2007. Vol. 36. No. 6. Pp. 334–344.

20. **Naserian M., Fahiminia M., Goshayeshi H.R.** Experimental and numerical analysis of natural convection heat transfer coefficient of V-type fin configurations // Journal of Mechanical Science and Technology. 2013. Vol. 27. No. 7. Pp. 2191–2197.

21. **Misumi T., Kitamura K.** Enhancement of natural convective heat transfer from tall vertical heated plates // JSME B. (in Japanese). 1999. Vol. 65. No. 640. Pp. 4041–4048.

22. **Komori K., Inagaki T., Kito S., Mizoguchi N.** Natural convection heat transfer along a vertical flat plate with a projection in the turbulent region // Heat Transfer – Asian Research. 2001. Vol. 30. No. 3. Pp. 222–233.

23. **Tsuji T., Kajitani T., Nishino T.** Heat transfer enhancement in a turbulent natural convection boundary layer along a vertical flat plate // International Journal of Heat and Fluid Flow. 2007. Vol. 28. No. 6. Pp. 1472–1483.

24. **Levchenya A.M., Smirnov E.M., Zhukovskaya V.D., Ivanov N.G.** Numerical study of 3D turbulent flow and local heat transfer near a cylinder introduced into the free-convection boundary layer on a vertical plate // Proceedings of the 16th International Heat Transfer Conference (IHTC-16), August 10–15, 2018. Beijing, China. Paper IHTC16-22916. DOI: 10.1615/IHTC16.hte.022916 (2018) 5493–5500.

25. **Levchenya A.M., Smirnov E.M., Zhukovskaya V.D.** Numerical study of 3D flow structure near a cylinder piercing turbulent free-convection boundary layer on a vertical plate // AIP Conf. Proc. 2018. Vol. 1959. P. 050017.

26. **Smirnov E.M., Levchenya A.M., Zhukovskaya V.D.** RANS-based numerical simulation of the turbulent free convection vertical-plate boundary layer disturbed by a normal-to-plate circular cylinder // International Journal of Heat and Mass Transfer. 2019. Vol. 144. December. P. 118573.

27. **Кузьмицкий В.А., Чумаков Ю.С.** Установка для статической калибровки термоанемометра при малых скоростях в неизотермической воздушной среде // Теплофизика высоких температур. 1995. Т. 33. № 1. С. 120–116.

Статья поступила в редакцию 17.01.2020, принята к публикации 31.01.2020.

СВЕДЕНИЯ ОБ АВТОРАХ

ЧУМАКОВ Юрий Сергеевич — *доктор физико-математических наук, профессор Высшей школы прикладной математики и вычислительной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
chymakov@yahoo.com

ЛЕВЧЕНЯ Александр Михайлович — *кандидат физико-математических наук, доцент Высшей школы прикладной математики и вычислительной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
Levchenya-am@spbstu.ru

ХРАПУНОВ Евгений Федорович — *аспирант Высшей школы прикладной математики и вычислительной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
hrapunov.evgenii@yandex.ru

DOI: 10.18721/JPM.13107
УДК 621.455.4; 621.455.34

THE CONTACT IONIZATION ION ACCELERATOR FOR THE ELECTRICALLY POWERED SPACECRAFT PROPULSION: A COMPUTER MODEL

D.B. Dyubo, O.Yu. Tsybin

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

An analytical electrodynamic algorithm has been developed in order to study physical processes and calculate mechanical forces in an ion accelerator. This algorithm is combined with computer simulation of the electromagnetic field and charged particles' trajectories. Computer models of ion accelerators with surface or contact ionization in the injection region were considered as an example. Ultimately, the created theoretical apparatus makes it possible to evaluate the proposed engineering solutions and diagnostic parameters of electric spacecraft propulsions as well as to compare the applicability of various working agents inside.

Keywords: computer modeling, ionization, ion beam, ion accelerator, electrically powered spacecraft propulsion

Citation: Dyubo D.B., Tsybin O.Y., The contact ionization ion accelerator for the electrically powered spacecraft propulsion: a computer model, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (1) (2020) 70–81. DOI: 10.18721/JPM.13107

КОМПЬЮТЕРНАЯ МОДЕЛЬ УСКОРИТЕЛЯ ИОНОВ С КОНТАКТНОЙ ИОНИЗАЦИЕЙ ДЛЯ ЭЛЕКТРОРАКЕТНЫХ ДВИГАТЕЛЕЙ КОСМИЧЕСКИХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

Д.Б. Дюбо, О.Ю. Цыбин

Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Российская Федерация

С целью исследования физических процессов и расчета механических сил в ускорителе ионов, построен аналитический электродинамический алгоритм, объединенный с компьютерным моделированием электромагнитного поля и траекторий заряженных частиц. В качестве примера рассмотрены компьютерные модели ускорителей ионов с поверхностной или контактной ионизацией в области инжекции. Созданный теоретический аппарат позволит оценивать предполагаемые конструкторские решения и параметры разрабатываемых электроракетных двигателей космических летательных аппаратов, сравнивать применение в них различных рабочих веществ.

Ключевые слова: компьютерное моделирование, ионизация, ионный поток, ионный ускоритель, электроракетный двигатель

Ссылка при цитировании: Дюбо Д.Б., Цыбин О.Ю. Компьютерная модель ускорителя ионов с контактной ионизацией для электроракетных двигателей космических летательных аппаратов // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2020. Т. 13. № 1. С. 78–91. DOI: 10.18721/JPM.13107

Introduction

Accelerated ion fluxes in vacuum are widely used in research, medicine, materials science, microelectronics, as well as in technologies for thin film deposition, surface cleaning, etc. [1–5]. Ion accelerators have gained great importance for spacecraft technologies, particularly for electric propulsion spacecraft [6–9]. Fundamental research in this field [10–14] yielded significant results, laying the foundation for theoretical studies [15–19] and computer simulation methods [17–19].

Typical electric propulsion systems are vacuum ion and plasma ion devices converting electromagnetic energy generated by accelerating ions of the propellant into mechanical energy accelerating the spacecraft. Two types of electric propulsion accelerators are commonly used: ion and plasma ion [6, 10, 11]. Electric propulsion encompasses a wide range of complex phenomena, including conversion of the propellant into the vapor-gas phase, ionization, characteristics of plasma in the ionization chamber, injection of plasma and ions into the accelerator and their acceleration, mechanical forces arising and their parameters, beam neutralization, charge exchange, extraction and positioning. Fast ions leaving the accelerator should be neutralized to prevent charging of the spacecraft body, the associated drop in thrust, and secondary discharge phenomena. The beam of neutralized particles is ejected outwards at high velocities, up to tens of km/s, which is much higher than in chemical jet engines. Modern electric propulsion systems mainly use compressed gases, particularly xenon, as propellants. However, xenon is a rare and expensive gas, so efforts are underway to find alternatives for large-scale projects of electric propulsion engines. Electric propulsion engines based on novel methods and technologies should combine simple, reliable and durable design with affordable costs, running on alternative types of propellants, effectively generating the required thrust with reduced consumption of the reactant mass [8, 12–14]. Perfecting the theory and methods for analysis of ionic and mechanical processes in accelerators, as well as conducting flight and ground experiments can establish an effective framework for constructing new electric propulsion systems. Bench tests are run in ground laboratories, typically with large and expensive vacuum systems, taking longer time to operate and consuming considerable amounts of propellants.

Analytical methods are mainly used in theory of electric propulsion for the most simple ion optical and electrodynamic problems [10–15]. Effective solutions for plasma states are obtained by numerically simulating the dynamics of enlarged particles by time steps (the Monte Carlo method) [17–19]. Differential equations in finite-difference form, grid methods, and Fourier transforms are used for simulations of electromagnetic fields and particle trajectories in these fields. However, this classic method is rather complicated, requiring a lot of computer time. The CST Studio Suite for three-dimensional modeling developed by Computer Simulation Technology is an effective tool for trajectory analysis of different electron and ion devices taking into account the intrinsic electric field of the space charge [20, 21]. However, no known studies have reported on applying CST codes to analysis or design of ion or plasma ion electric propulsion systems. While microscopic power characteristics are crucially important, insufficient attention has been given to electrodynamic models and simulations of mechanical processes in accelerators, hindering the progress in new devices. The dependences of thrust on coordinates, shape of electrodes, operating modes of the reactor, and other parameters are not described in known literature. It is usually accepted for simplicity, regardless of structural elements, that the thrust can be expressed in terms of beam parameters

$$F_f = v_{in}(dm_w/dt),$$

where dm_w/dt is the mass flux m_w of propellant in the neutralized beam, t is the time; v_{in} is the velocity of a neutral particle in the beam.

At the same time, the values of dm_w/dt are not measured. The resulting thrust F_f of electrostatic ion and plasma ion engines is formulated in terms of current I and voltage U_d for ions with the mass μ and charge q in the accelerating gap (see, for example, [10, 11]):

$$F_f = I[2U_d(\mu/q)]^{1/2}.$$

It is assumed that the ion field at the output of the accelerator is equal to zero, flux and velocity of particles in the beam are equal to the same values in the accelerator and do not change in the neutralizer; neutralization has 100% efficiency, there are no oppositely charged particles in the acceleration chamber, ions have no transverse velocity components, there are no losses, collisions and charge exchange of ions, etc.

This simplification produces inaccuracies in accounting for the ion-optical properties of the injector, accelerator, and neutralizer, as well as for the microscopic processes associated with generation of mechanical forces. Essentially, the accuracy of the above formulas requires further assessment.

Like the Monte Carlo method and some other similar modeling packages, the CST package is inapplicable to detecting the relationships between mechanical forces in ion accelerators and internal microscopic processes.

The goal of this study consists in constructing an electrodynamic algorithm for determining mechanical parameters, combined with computer simulation in the CST Suite to establish the above-mentioned relationships. Furthermore, our practical task was to apply the developed tools to studying physical processes in ion and plasma ion electric propulsion systems.

Electrodynamic algorithm for determining the mechanical parameters of ion and plasma ion accelerators

The CST model combined with the electrodynamic algorithm was primarily used to analyze ion trajectories in inhomogeneous fields generated by single-stage and multi-stage DC accelerators, taking into account polarization charges and the space charge formed by the charged particle flux. As a result of analysis, we proposed an electrodynamic algorithm accounting for the relationship of mechanical properties with structural elements and modes, characteristics of the main processes and propellants. The algorithm consists of interconnected modules of varying complexity, allowing to simulate both simple elements of the accelerator and their combinations.

The algorithm for determining the mechanical parameters of electrostatic acceleration of ions is based on the following principles.

In combination, the accelerating voltage applied and space charges in the accelerator induce certain polarization charges. These are surface charges bound on electrodes and distributed over their surfaces; the distribution depends on the geometric shape, size and arrangement of the electrodes.

The momentum for engine thrust is generated in the ion acceleration chamber. The thrust applied to the electrodes arises due to Coulomb attraction of surface polarization charges to accelerated ions. The main condition and process for generating mechanical thrust is acceleration of ions by an electric field in the accelerator, their subsequent neutralization and ejection of a beam of neutral particles into space. Accelerated ions should be neutralized because the ions escaping generate an excess charge of the opposite sign. Such a charge inhibits the escaping ions, reducing the thrust, and also causes destructive discharge phenomena in the structural elements of the device. The action of the force F_{TM} produces mechanical stress in the mount of the device fixed on the test bench; the action of this force provides acceleration of the flying spacecraft.

A simplified diagram of a typical single-stage DC ion accelerator for an electric propulsion engine of spacecraft is shown in Fig. 1.

Fig. 1 shows space charge *3* of internal ions with the total mass m and velocity $v(z)$ in the accelerator; beam *5* of neutralized particles is ejected outwards at velocity v_f. The force F_{Ti} acting on the ion flux from bound surface charges *2* is equal to the inertia of the ions with mass m:

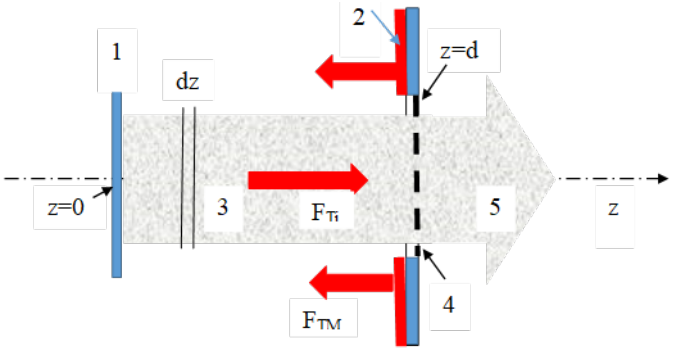

Fig. 1. Simplified diagram of single-stage ion accelerator:
input electrode *1* in ion injection plane ($z = 0$); bound charges *2* on output electrode in ion extraction plane ($z = d$); space charge *Q 3*; ion neutralization region *4*; beam *5* of neutralized particles ejected outwards; F_{TM}, F_{Ti} are the forces acting on the output electrode from the ion flux and on the ion flux from the bound charges, respectively

$$F_{Ti} = m(dv/dt).$$

The force F_{TM} acting on the output electrode from ion flux *3* is expressed as

$$F_{TM} = -v(dm/dt).$$

Ion acceleration is generated by attraction to surface charges, which is calculated as the Lorentz force acting on charged particles from the electric field in the accelerator, taking into account the polarization charges of electrodes. Notably, the parameters of ion flux in the above two formulas refer to the particles in the accelerator rather than the beam.

Ion drift in the accelerating gap is directed along the z axis from injection plane *1* with $z = 0$ to neutralization plane *4*. Instantaneous internal charge and mass in the accelerating gap are equal to Q and m, gap width is equal to d.

Modern plasma ion or Hall-effect thrusters generate propellant plasma directly in the accelerating gap combined with the ionization chamber, while ion thrusters generate propellant plasma in the volume chamber to the left of the injection plane (see Fig. 1), subsequently extracting the ions into the accelerating gap by the electric field.

Processes in the accelerating gap. Acceleration of ions by the force F_{Ti} along the axis z balances the inertia of the accelerated spacecraft in the center of mass and ends with beam ejection. Ions attract the electrode with induced charges bound to it it with the force F_{TM}, resulting in reverse acceleration of spacecraft's center of mass in the direction opposite to the z axis (see Fig. 1). Accordingly, the thrust F_{TM} accelerating the spacecraft is generated only by the ions in the gap and should become zero when the ions pass through the plane of the exit gap and the neutralization region, if the latter coincides with plane *4* (see Fig. 1).

Using the kinetic (mechanical) approach, we obtain the force F_{Ti} by integrating the inertia along the ion acceleration path:

$$F_{Ti}(z) = \int_0^z \frac{dv(z)}{dt} dm(z), \qquad (1)$$

where $dm(z)$ is the mass of each layer dz, $v(z)$ is the velocity of this layer.

Electrodynamically, this force can be calculated as a set of Lorentz forces acting from the electric field E on the charge $dQ(z)$ in each layer dz:

$$F_{Ti}(z) = \int_0^z E(z)\, dQ(z). \qquad (2)$$

Naturally, the forces calculated by Eqs. (1) and (2) should coincide for a given value of the coordinate z, including with $z = d$.

To tentatively test the algorithm and the computer program, we determined the values of the physical quantities included in Eqs. (1) and (2) based on simple models, using, for example, a plane-parallel gap with solid electrodes. The ion current in such an accelerating gap is determined analytically by the Child–Langmuir formula in saturation mode, i.e., with space-charge limited current (SCLC). Accordingly, ions are injected into into the accelerator at zero electric field and with a zero initial velocity. Since the field is zero, the charge on the input surface is zero, and the polarization charge Q_p, equal in magnitude to the internal drift charge Q in magnitude but opposite in sign, is concentrated on the inner surface of the solid output electrode. Well-known formulas include the dependence of current I on voltage U_d according to the three halves power law, distributions of potential $U(z)$, ion velocity $v(z)$, electric field $E(z)$, as well as the total thrust, charge and mass of ions in the gap [1, 2, 10, 15, 16]:

$$I = \frac{4}{9}\sqrt{2\frac{q}{\mu}}\varepsilon_0 S U_d^{\frac{3}{2}} d^{-2},$$

$$E(z) = \frac{4}{3} U_d d^{-\frac{4}{3}} z^{\frac{1}{3}},$$

$$E_0 = \frac{U_d}{d}, E_d = \frac{4}{3} E_0,$$

$$F_{\dot{I}} = \frac{1}{2}\varepsilon_0 S E_d^2 = \frac{8}{9}\varepsilon_0 S E_0^2,$$

$$Q = \frac{4}{3}\varepsilon_0 S E_0, m = \frac{\mu}{q} Q.$$

The following notations are used here: S is the cross-sectional area of the gap; μ is the ion mass; q is the ion charge; ε_0 is the electric constant ($\varepsilon_0 \approx 8.85 \cdot 10^{-12}$ F/m).

Developing this well-known model, we obtain new distributions along the z coordinate for thrust, charge and mass of a thin layer dz inside the accelerating gap:

$$dF_{Ti}(z) = \frac{16}{27}\varepsilon_0 S E_0^2 d^{-\frac{2}{3}} z^{-\frac{1}{3}} dz,$$
$$dQ(z) = \frac{4}{9}\varepsilon_0 S U_d d^{-\frac{4}{3}} z^{-\frac{2}{3}} dz. \qquad (3)$$

Relations (3) describe the mechanical field of velocities and the distribution of propellant masses in the accelerator. The values of the physical quantities satisfy the system of Poisson's equations, equations of motion and continuity (for the given boundary and initial conditions) in the accelerator in quasistatic, nonrelativistic, non-diamagnetic collisionless hydrodynamic approximations. Descriptions of Lorentz and Coulomb forces in the accelerator are related by the Gauss law. For example, the distribution of charge (see Eq. (3)) and electric field $E(z)$ is consistent with the Gauss theorem for each thin layer dz and for the entire gap as a whole. Eqs. (2) and (3) imply the dependence of the force acting on the ions in SCLC mode from zero to the z coordinate in the gap:

$$F_{Ti}(z) = \frac{8}{9}\varepsilon_0 S E_0^2 \left(\frac{z}{d}\right)^{\frac{2}{3}}. \qquad (4)$$

Accordingly, the kinetic power W_i of ion flux (in the spacecraft's own reference frame) is equal to the electric power supplied to the accelerator (excluding heat losses):

$$W_i = \int_0^d v(z) E(z) dQ(z) = \\ = \frac{4}{9}\varepsilon_0 S E_0^2 v(d) = IU_d. \qquad (5)$$

Therefore, the ratio of thrust to input power in the model of a solid output electrode reaches its maximum value and is determined by the formula

$$F_{TM}/W_i = \beta/v(d);\ \beta = 2. \qquad (6)$$

However, the electric field in the gap decreases depending on the shape of the partially open output electrode (grid, aperture, notch, etc.), and the coefficient $\beta < 2$ as a result. According to Eqs. (1)–(6), the mechanical parameters of a complex multi-stage DC accelerator can be calculated by obtaining the spatial distributions of electric field $E(z)$, charge $Q(z)$ and velocity $v(z)$ of ions, taking into account polarization charges and space charge, which can be determined by computer simulation. Computer monitors collecting data on microscopic processes were installed in the region of ion and electron fluxes for these purposes.

Microscopic parameters of trajectories and characteristics of physical processes were obtained for the first time, including distributions depending on the coordinate z along the spacecraft axis: particle velocities, fields $E(z)$ and potentials, charges $dQ(z)$ and currents, taking into account the intrinsic electric field of space charge of ions and electrons.

A more complicated physical process was considered at the next stage; we analyzed the neutralization of the ion flux by injecting an electron beam into this flux.

Testing of two, three and multi-electrode ion accelerators

To verify the calculations, we first simulated and tested a simple planar two-electrode model whose sizes and modes were tailored to closely match the one-dimensional model in SCLC mode. The technique was verified by comparing it with analytical calculations using Eqs. (1)–(6). We additionally tested well-known designs based on Pierce gun algorithms [5, 16].

The three, four, five and six-electrode models were implemented at the second stage. We used a partitioned set of nonplanar electrodes to impose inhomogeneous boundary conditions compensating for the difference between Laplace and Poisson fields at the boundaries. The surfaces for injection into the accelerating gap were given as both smooth planar and smooth parabolic. We simulated emission of ions with different specific charges in SCLC mode in this configuration. In practice, this meant that surface-contact ionization [22–25] was simulated; its prospects for use in ion and plasma ion accelerators are estimated to be better than those of surface-volume ionization [26, 27].

Ion trajectories were constructed for all models considered (partially shown below as examples); microscopic parameters were collected using computer monitors for these trajectories. The surfaces of the boundary parallelepiped surrounding the entire structure were sufficiently far from the accelerating gaps, with a fixed zero potential maintained.

Example calculations Typical dependences of electronic, physical and mechanical parameters on coordinates and operating modes for models of ion accelerators using xenon ions are shown in Figs. 2–6.

The following parameters of two-electrode models were chosen for the examples:

voltage U_d between the plates up to 2 kV,
ion flux diameter 20 mm,
distance between the plates $d = 4$ mm,
emission area $S = 314$ mm^2.

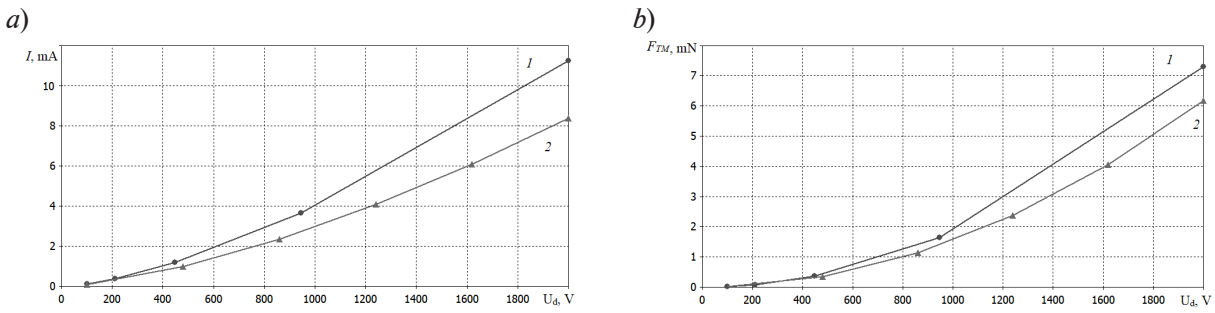

Fig. 2. Numerical (*1*) and analytical (*2*) calculations of current (*a*) and resultant mechanical force (*b*) as functions of voltage U_d between the plates. A two-electrode model A with xenon ions (d = 4 mm) was used

The calculations were carried out for three models:

Model A with two solid electrodes;

model G (grid) with one solid and one grid electrode;

models D5 and D10 with one solid and one grid electrodes, with apertures 5 mm (D5) and 10 mm (D10) in diameter.

The figures show the characteristics obtained by analytical (for SCLC mode) and numerical calculations.

The numerical results are close to theoretical analytical curves with the largest deviation of about 10%. This discrepancy can be explained, firstly, by edge effects due to finite sizes and open boundaries of the accelerating gap, and secondly, by errors in calculating the field and particle velocity near the boundary $z = 0$ in the computer model. Introducing corrections reduced the discrepancy between analytical and computer parameters to 1% or less. The results obtained allowed to consider more complex multi-electrode structures of ion thrusters.

The simple three-electrode 2D models we tested were fundamentally close to real electric propulsion systems constructed for spacecraft. Typical electric propulsion systems use volume ionization in a chamber limited by a grid with multiple apertures [6, 10, 11]. A grid or a solid electrode with an aperture which was an electrostatic focusing lens was placed between

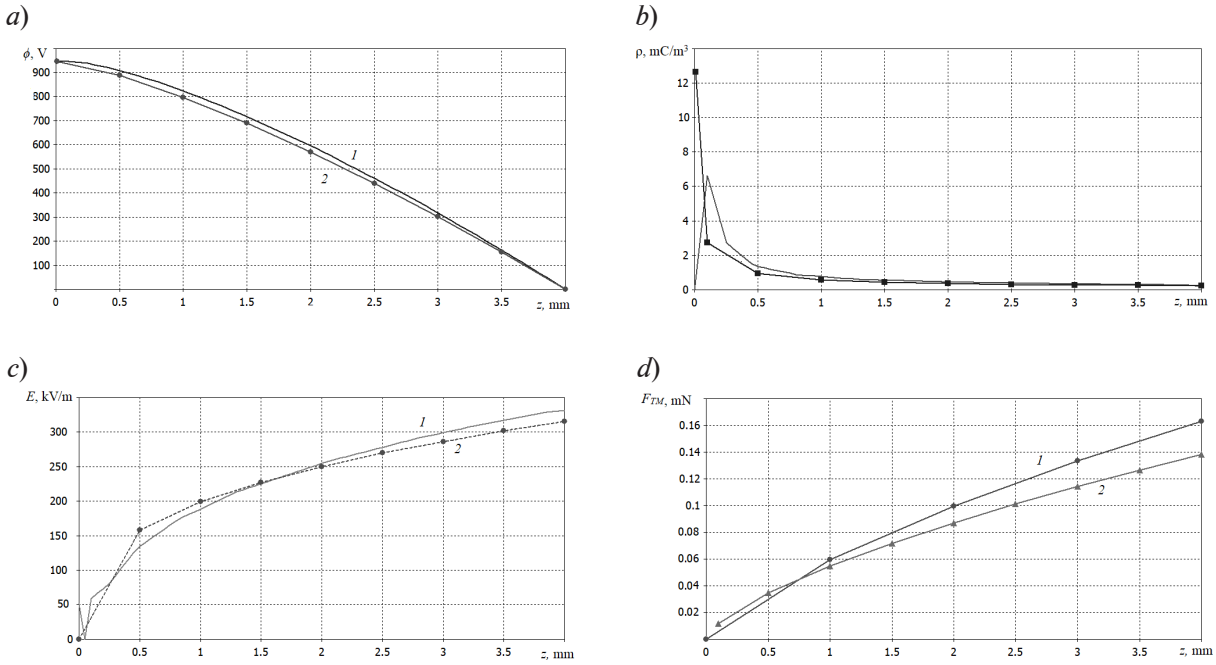

Fig. 3. Numerical (*1*) and analytical (*2*) calculations for distributions of potential (*a*), space charge density (*b*), electric field (*c*) and resultant mechanical force F_{TM} (*d*) along coordinate z along the central axis of ion accelerator. A two-electrode model A with xenon ions (d = 4 mm) was used; U_d = 945.74 V

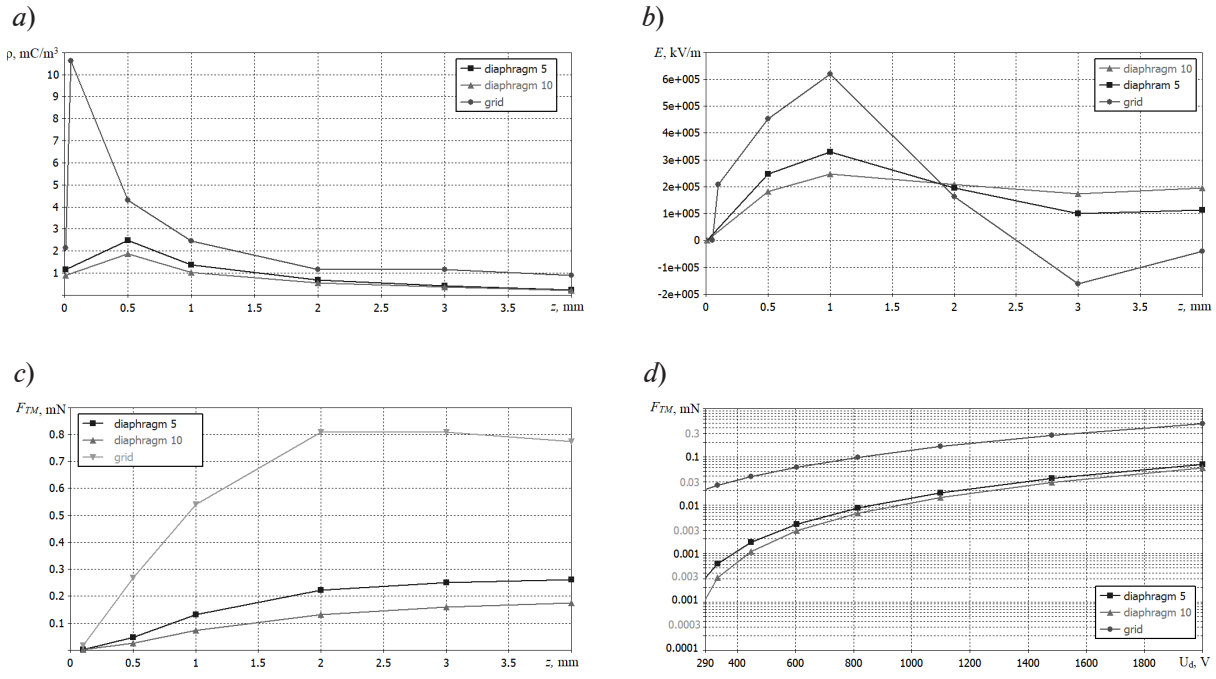

Fig. 4. Numerical calculations for distributions of space charge density (*a*), electric field (*b*) and resultant mechanical force F_{TM} depending on z и U(*d*) along the coordinate z along the central axis of ion accelerator.
Dependences were obtained for G (*1*), D5 (*2*), and D10 (*3*) models.
The radius of the emission spot is 10 mm; potentials of the first (φ_1 with $z = 0$), second (φ_2) and third (φ_3) electrodes φ_i, kV: +1.0; −0.2; +0.2, respectively

two solid electrodes in three-electrode models. Fig. 4 shows typical dependences of electronic, physical and mechanical parameters on the coordinates and operating modes in a three-electrode model with a central electrode mounted in the middle. The values of field and charge density along the z axis were calculated for the model with solid electrodes (see Figs. 2 and 3); the mean field and charge density in the transverse plane were calculated for models with the grid and the aperture (see Fig. 4). Figs. 4 and 5 show the functions $E(z)$, $\rho(z)$, and F_{TM} calculated on the axis z without cross-section averaging (this did not produce significant errors).

Comparing the data in Figs. 2, 3, and 4, we can see that the two-stage scheme with an intermediate grid, similar to practical electric propulsion systems, is optimal in forming the ion flux and achieving the greatest mechanical thrust. These electric propulsion systems use volume ionization in a chamber limited by a grid with multiple apertures [6, 10, 11].

Based on the obtained data, we performed computer simulation of three-dimensional multi-electrode sectioned ion accelerators, intended for Hall-effect electric propulsion systems [6, 10, 11]. A preselected multielectrode scheme of an ion accelerator typical for Hall thrusters, was given in the CST software package. In contrast to the two and three-electrode schemes, where zero potential was imposed for all surfaces of the external boundary parallelepiped, an open upper boundary was used here (Fig. 5). The model parameters are given in the caption to the figure. The model is based on a sectional ion accelerator with optimized thrust generation and neutralization. Similar to modern thrusters, neutralization in the given devices was carried out by an electron beam. The shape and size of the electrodes, the potentials applied to them, the size of the emission region, and the magnitude of charged current were selected based on the data obtained (see Figs. 2 and 3).

Similar constructions were calculated for different sizes and voltages of external sources, specific ion masses, current modes. We determined the characteristics of ion and electron trajectories combined in the accelerator and neutralizer. We assessed whether it was possible to combine the given functions, simulating a range of ion and electron

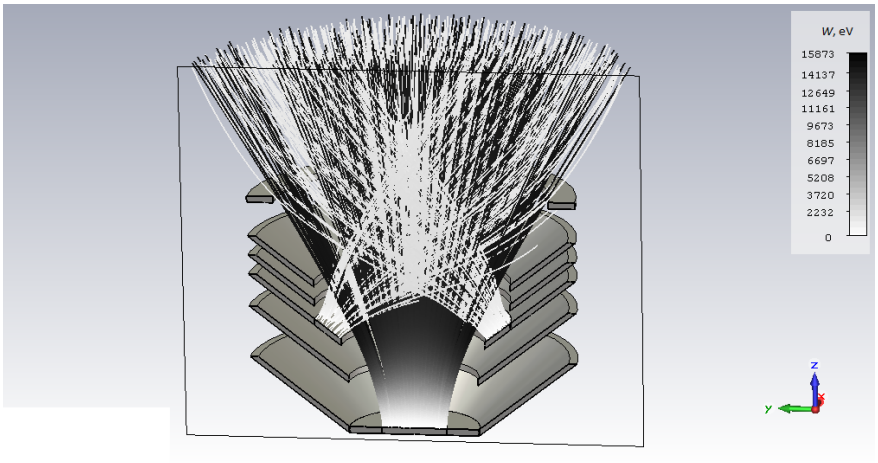

Fig. 5. Schematic model of multi-electrode structure with combined ion and electron trajectories. The ion source is located in the lower part of the model, the electron source is in the middle. The shades of grey correspond to particle energies. Beams intersect near and above the fourth electrode. The potentials of electrodes (from bottom to top electrode) φ_i, kV: 5.0; -0.5; -1.0; 0.5; 0.0; 0.0. ion and electron currents were equal to 17.5 mA

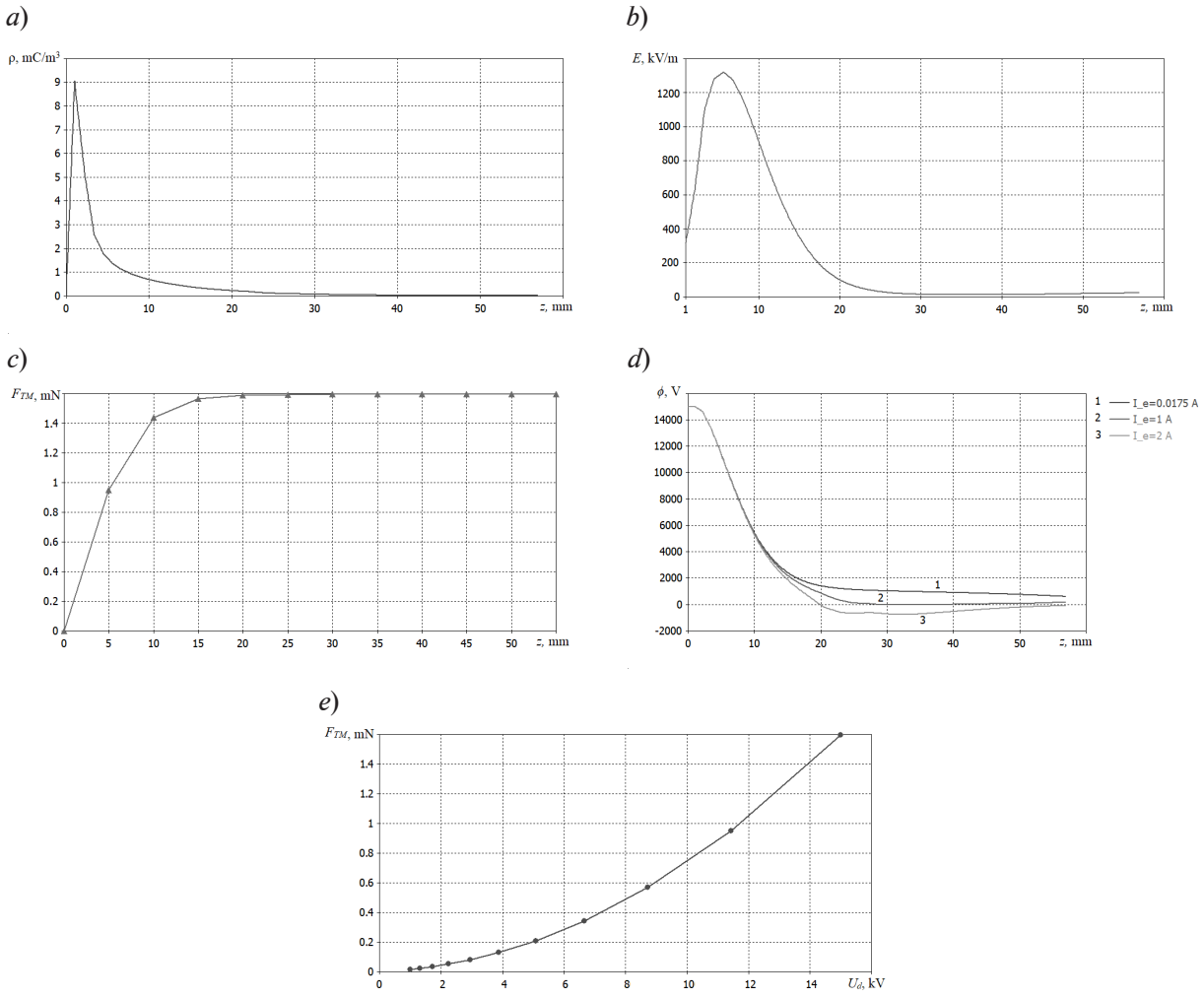

Fig. 6. Numerical calculations for distributions of space charge density (*a*), electric field (*b*), resultant mechanical force F_{TM} (*c*) and potential (*d*) along the coordinate *z* along the central axis of ion accelerator; dependences of resultant mechanical force F_{TM} on the potenial of bottom electrode (*e*); *d* corresponds to electron current I_e, A: 0.0175 (*1*), 1.0 (*2*), 2.0 (*3*); *a, b, c, e* correspond to ion current I_i = 17.5 mA

processes as a single process and creating a virtual electronic cathode.

Fig. 5 shows one of the multi-electrode schemes considered with combined ion and electron trajectories. The electron neutralizer of accelerated cations in this example was based on the principle of embedded beams and a virtual cathode, which is fundamentally similar to an plasma ion Hall thrusters but does not require magnetic sources. This should allow to reduce mass, dimensions and energy consumption.

Fig. 6 shows typical dependences of electron, physical and mechanical parameters on the coordinates and operation modes in the multi-electrode model shown in Fig. 5.

Changing the geometric parameters and potentials of electrodes of the model shown in Fig. 5 in a small range, we optimized the ion flux, electron, physical and power characteristics. In particular, we determined the conditions when the electron beam generates an effective potential well for acceleration and capture of ions. Varying the electron current changed the depth of the potential well trapping the ions. Fig. 6, *d* shows acceleration and neutralizion of the ion flux by space charge of electrons. Varying the electron current over a wide range (17.5 mA, 1.0, and 2.0 A) changed the potential in the region $z > 13$ mm, improving focusing and neutralization of the ion flux.

Conclusion

The CST package was used to construct an analytical electrodynamic algorithm combined with computer simulation of the electromagnetic field and trajectories of charged particles. The algorithm operates by generating mechanical thrust in ion accelerators due to Coulomb interaction of ions moving in vacuum with bound charges on polarized surfaces of the structural elements of the accelerator chamber.

We have considered several models of electrostatic ion accelerators by analytical calculations and computer simulation (using the given package). We have obtained the coordinate dependences of numerical parameters, including the distribution of mechanical forces, potential and field, charge density and current, particle velocity in the accelerating gap. Analysis of the data confirmed that the approaches proposed had satisfactory accuracy and consistency.

The package developed provides a wide range of tools for studying physical phenomena and processes in ionic and plasma ion sources, accelerators and neutralizers for electric propulsion, allowing to assess the designs and parameters of novel devices, comparing different propellants.

The approach presented and the algorithm we have constructed based on this approach may offer potential methods for optimizing newly constructed devices, in particular, by comparing the physical processes for different propellants.

REFERENCES

1. **Forrester T.A.,** Large ion beams: fundamentals of generation and propagation, Wiley-VCH, Weinheim, Germany, 1988.

2. **Aston G.,** High efficiency ion beam accelerator system, Review of Scientific Instruments. 52 (9) (1981) 1325 − 1327.

3. **Lebedev A.H., Shalnov A.V.,** Osnovy fiziki i tekhniki uskoriteley [Basic foundations of accelerators, Physics and techniques], Energoatomizdat Publisher, Moscow, 1991 (in Russian).

4. **Chao A.W., Tigner M.,** Handbook of accelerator physics and engineering, World Scientific Publishing, London, 1999.

5. **Morozov A.I.,** Plazmennyye uskoriteli i ionnyye inzhektory [Plasma accelerators and ion injectors], Nauka, Moscow, 1984 (in Russian).

6. **Gorshkov O.A., Muravlev V.A., Shagayda A.A.,** Khollovskiye i ionnyye plazmennyye dvigateli dlya kosmicheskikh apparatov [Hall and plasma ion thrusters for spacecrafts], Ed. by Koroteyev A.S., Mashinostroyeniye, Moscow, 2008 (in Russian).

7. **Gusev Yu.G., Pilnikov A.V.,** The electric propulsion role and place within the Russian Space Program, Trudy MAI (Network scientific periodic publication) (60) (2012) 1−20. Access Mode: www.mai.ru/science/trudy/.

8. **Mazouffre S.,** Electric propulsion for satellites and spacecraft: established technologies and novel approaches, Plasma Sources Sci. Technol. 25 (3) (2016) 033002.

9. **Levchenko I., Xu S., Teel G., et al.,** Recent progress and perspectives of space electric propulsion systems based on smart nanomaterials, Nature Communications. 9 (4) (2018) 879.

10. **Goebel D.M., Katz I.,** Fundamentals of electric propulsion ion and Hall thrusters, John Wiley & Sons, Hoboken, New Jersey, USA Ch. 1, 6 and 7 (2008) 1−13, 243−389.

11. **Kaufman H.R.** Technology of electron-bombardment ion thrusters, In the book: Advances in electronics and electron physics. Vol. 36. Ed. by L. Marton, Academic Press, New York (1975) 265–373.

12. **Charles C.,** Plasmas for spacecraft propulsion, J. Phys. D: Applied Phys. 42 (16) (2009) 163001.

13. **Cusson S.E., Dale E.T., Jorns B.A., Gallimore A.D.,** Acceleration region dynamics in a magnetically shielded Hall thruster, Physics of Plasmas, 26 (2) (2019) 023506.

14. **Gopanchuk V.V., Potapenko M.Yu.,** Hall effect thrusters for small-sized spacecrafts, IKBFU's Vestnik. (4) (2012) 60–67.

15. **Favorskiy O.N., Fishgoyt V.V., Yantovskiy Ye.I.,** Osnovy teorii kosmicheskikh elektroreaktivnykh dvigatelnykh ustanovok [Foundations of Hall effect thrusters for spacecrafts], Vysshaya Shkola Publishing, Moscow, 1978.

16. **Hassan A., Elsaftawy A., Zakhary S.G.,** Analytical studies of the plasma extraction electrodes and ion beam formation, Nuclear Instruments and Methods in Physics Research, A. 586 (2) (2008) 148–152.

17. **Kalentev O., Matyash K., Duras J., et al.,** Electrostatic ion thrusters – towards predictive modeling, Contributions to Plasma Physics. 54(2) (2014) 235–248.

18. **Lovtsov A.S., Kravchenko D.A.,** Kinetic simulation of plasma in ion thruster discharge chamber. Comparison with experimental data, Procedia Engineering. 185 (2017) 326–331.

19. **Peng X., Keefert D., Ruytent W.M.,** Plasma particle simulation of electrostatic ion thrusters, Journal of Propulsion and Power. 8 (2) (1992) 361–366.20. **Kurushin A.** Basic course of design of microwave devices using CST Studio Suite, One-Book, Moscow, 2014. 433 p.

21. **Kurushin A.A., Plastikov A.N.,** Proyektirovaniye SVCh ustroystv v srede CST Microwave Studio [Design of microwave devices in CST Microwave Studio], MEI Press, 2011.

22. **Zandberg E.Ya.,** Surface-ionization detection of particles (Review), Technical Physics. 40 (1995) 865–890.

23. **Blashenkov N.M., Lavrent'ev G.Ya.,** Surface-ionization field mass-spectrometry studies of nonequilibrium surface ionization, Phys. Usp. 50 (1) (2007) 53–78.

24. **Tsybin O.Yu., Tsybin Yu.O., Hakansson P.,** Laser or/and electron beam activated desorption of ions: a comparative study, In: Desorption 2004, Papers of 10th International Conference, Saint Petersburg (2004) 61.

25. **Tsybin O.Y., Makarov S.B., Ostapenko O.N.,** Jet engine with electromagnetic field excitation of expendable solid-state material, Acta Astronautica. 129 (December) (2016) 211–213.

Received 20.01.2020, accepted 08.03.2020.

THE AUTHORS

DYUBO Dmitry B.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
doobinator@rambler.ru

TSYBIN Oleg Yu.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
otsybin@rphf.spbstu.ru

СПИСОК ЛИТЕРАТУРЫ

1. **Форрестер Т.А.** Интенсивные ионные пучки. М.: Мир, 1992. 354 с.

2. **Aston G.** High efficiency ion beam accelerator system // Review of Scientific Instruments. 1981. Vol. 52. No. 9. Pp. 1325 –1327.

3. **Лебедев А.Н., Шальнов А.В.** Основы физики и техники ускорителей. 2-е изд., перераб. и доп. М.: Энергоатомиздат, 1991. 528 с.

4. **Chao A.W., Tigner M.** Handbook of accelerator physics and engineering. London: World Scientific Publishing, 1999. 650 p.

5. **Морозов А.И.** Плазменные ускорители и ионные инжекторы. М.: Наука, 1984. 269 с.

6. **Горшков О.А., Муравлев В.А., Шагайда А.А.** Холловские и ионные плазменные двигатели для космических аппаратов.

Под ред. акад. РАН А.С. Коротеева. М.: Машиностроение, 2008. 280 с.

7. **Гусев Ю.Г., Пильников А.В.** Роль и место электроракетных двигателей в Российской космической программе // Труды МАИ (электронный журнал). 2012. Вып. № 60. С. 1–20. Режим доступа: www.mai.ru/science/trudy/.8. **Mazouffre S.** Electric propulsion for satellites and spacecraft: established technologies and novel approaches // Plasma Sources Sci. Technol. 2016. Vol. 25. No. 3. P. 033002.

9. **Levchenko I., Xu S., Teel G., Mariotti D., Walker M.L.R., Keidar M.** Recent progress and perspectives of space electric propulsion systems based on smart nanomaterials // Nature Communications. 2018. Vol. 9. No. 4. P. 879.

10. **Goebel D.M., Katz I.** Fundamentals of electric propulsion ion and Hall thrusters. Hoboken, New Jersey, USA: John Wiley & Sons, 2008. Ch. 1, 6 and 7. Pp. 1–13, 243–389.

11. **Kaufman H.R.** Technology of electron-bombardment ion thrusters // Advances in Electronics and Electron Physics. Vol. 36. Ed. by L. Marton, New York: Academic Press, 1975. Pp. 265–373.

12. **Charles C.** Plasmas for spacecraft propulsion // J. Phys. D: Applied Phys. 2009. Vol. 42. No. 16. P. 163001.

13. **Cusson S.E., Dale E.T., Jorns B.A., Gallimore A.D.** Acceleration region dynamics in a magnetically shielded Hall thruster // Physics of Plasmas. 2019. Vol. 26. No. 2. P. 023506.

14. **Гопанчук В.В., Потапенко** М.Ю. Электрореактивные двигатели для малых космических аппаратов // Вестник Балтийского федерального университета им. И. Канта. 2012. Вып. 4. С. 60– 67.

15. **Фаворский О.Н., Фишгойт В.В.,** Янтовский **Е.И.** Основы теории космических электрореактивных двигательных установок. М.: Высшая школа, 1978. 384 с.

16. **Hassan A., Elsaftawy A., Zakhary S.G.** Analytical studies of the plasma extraction electrodes and ion beam formation // Nuclear Instruments and Methods in Physics Research. A. 2008. Vol. 586. No. 2. Pp. 148–152.

17. **Kalentev O., Matyash K., Duras J., Lüskow K.F., Schneider R., Koch N., Schirra M.** Electrostatic ion thrusters – towards predictive modeling // Contributions to Plasma Physics. 2014. Vol. 54. No. 2. Pp. 235–248.

18. **Lovtsov A.S., Kravchenko D.A.** Kinetic simulation of plasma in ion thruster discharge chamber. Comparison with experimental data // Procedia Engineering. 2017. Vol. 185. Pp. 326– 331.

19. **Peng X., Keefert D., Ruytent W.M.** Plasma particle simulation of electrostatic ion thrusters // Journal of Propulsion and Power. 1992. Vol. 8. No. 2. Pp. 361–366.

20. **Kurushin A.** Basic course of design of microwave devices using CST Studio Suite. Moscow: One-Book, 2014. 433 p.

21. **Курушин А.А., Пластиков А.Н.** Проектирование СВЧ устройств в среде CST Microwave Studio. М.: Изд-во МЭИ, 2011. 155 с.

22. **Зандберг Э.Я.** Поверхностно-ионизационное детектирование частиц (Обзор) // Журнал технической физики. 1995. Т. 65. № 9. С. 1–38.

23. **Блашенков Н.М., Лаврентьев** Г.Я. Исследование неравновесной поверхностной ионизации методом полевой поверхностно-ионизационной масс-спектрометрии // Успехи физических наук. 2007. Т. 177. № 1. С. 59–85.

24. **Tsybin O.Yu., Tsybin Yu.O., Hakansson P.** Laser or/and electron beam activated desorption of ions: a comparative study //Desorption 2004, Papers of 10th International Conference. Saint Petersburg, 2004. P. 61.

25. **Tsybin O.Y., Makarov S.B., Ostapenko O.N.** Jet engine with electromagnetic field excitation of expendable solid-state material // Acta Astronautica. 2016. Vol. 129. December. Pp. 211–213.

Статья поступила в редакцию 20.01.2020, принята к публикации 08.03.2020.

СВЕДЕНИЯ ОБ АВТОРАХ

ДЮБО Дмитрий Борисович – *аспирант Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
doobinator@rambler.ru

ЦЫБИН Олег Юрьевич – *доктор физико-математических наук, профессор Высшей инженерно-физической школы Санкт-Петербургского политехнического университета Петра Великого.*
195251 Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
otsybin@rphf.spbstu.ru

DOI: 10.18721/JPM.13108
УДК 536.7:536.1:544.341.2:661.487.1:519.6

THE INTERACTION PROCESSES OF SILICON TETRAFLUORIDE AND HEXAFLUOROSILICATES WITH HYDROGEN-CONTAINING AND OXYGENATED SUBSTANCES: A THERMODYNAMIC ANALYSIS

A.R. Zimin[1], D.S. Pashkevich[1], A.S. Maslova[1], V.V. Kapustin[1], Yu.I. Alexeev[2]

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

[2] LTD "New chemical products", St. Petersburg, Russian Federation

In the paper, the thermodynamic calculations have shown that at the temperatures above 1300 K, the main silicon-containing substance is silicon dioxide in the Si-F-H-O element system, and the main fluorine-containing one is hydrogen fluoride in the same system. The mentioned temperature was realized during the interaction reactions between silicon tetrafluoride, fluorosilicates and hydrogen-containing, oxygen-containing substances in the combustion mode. The high-temperature treatment of silicon tetrafluoride and fluorosilicates in the combustion mode can become the basis of industrial technology for hydrogen fluoride production.

Keywords: silicon tetrafluoride, hydrogen fluoride, silicon dioxide, thermodynamic equilibrium, Gibbs energy

Citation: Zimin A.R., Pashkevich D.S., Maslova A.S., Kapustin V.V., Alexeev Yu.I., The interaction processes of silicon tetrafluoride and hexafluorosilicates with hydrogen-containing and oxygenated substances: a thermodynamic analysis, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (1) (2020) 82–94 DOI: 10.18721/JPM.13108

ТЕРМОДИНАМИЧЕСКИЙ АНАЛИЗ ПРОЦЕССОВ ВЗАИМОДЕЙСТВИЯ ТЕТРАФТОРИДА КРЕМНИЯ И ГЕКСАФТОРСИЛИКАТОВ С ВОДОРОД- И КИСЛОРОДСОДЕРЖАЩИМИ ВЕЩЕСТВАМИ

А.Р. Зимин[1], Д.С. Пашкевич[1], А.С. Маслова[1], В.В. Капустин[1], Ю.И. Алексеев[2]

[1] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация

[2] ООО «Новые химические продукты», Санкт-Петербург, Российская Федерация

Термодинамическими расчетами показано, что в системе элементов Si-F-H-O при температуре выше 1300 K основным кремнийсодержащим веществом является диоксид кремния, а основным фторсодержащим – фторид водорода. Указанная температура реализуется при проведении реакций взаимодействия тетрафторида кремния и фторсиликатов с водородсодержащими и кислородсодержащими веществами в режиме горения. Высокотемпературная обработка тетрафторида кремния и фторсиликатов в режиме горения может стать основой промышленной технологии производства фторида водорода.

Ключевые слова: тетрафторид кремния, фторид водорода, диоксид кремния, термодинамическое равновесие, энергия Гиббса

Ссылка при цитировании: Зимин А.Р., Пашкевич Д.С., Маслова А.С., Капустин В.В., Алексеев Ю.И. Термодинамический анализ процессов взаимодействия тетрафторида кремния и гексафторсиликатов с водород- и кислородсодержащими веществами // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2020. Т. 13. № 1. С. 92–105. DOI: 10.18721/JPM.13108

Introduction

Hydrogen fluoride (HF) is the main fluorinating agent used in production of uranium fluorides in the nuclear fuel cycle, halocarbons, electron gases, etc. Production of anhydrous HF reaches 1.5 million tons per year [1, 2].

Hydrogen fluoride is obtained from fluorspar (natural CaF_2) by decomposition of sulfuric acid [1]:

$$CaF_{2\,sol} + H_2SO_{4\,liq} \rightarrow CaSO_{4\,sol} + 2HF_{gas}. \quad (1)$$

The subscripts here indicate the aggregate states of substances: solid (*sol*), liquid (*liq*) and gaseous (*gas*).

Only high-quality fluorspar with at least 95% content of the original material and a silicon dioxide admixture of not more than 1.5% is used to produce HF [1, 2].

The annual world production of fluorspar exceeds 4 million tons. The main producers are China (generating over 50% of the total world output), Mexico, Mongolia, and South Africa [1, 2].

The deposits of fluorspar in the Russian Federation are depleted, so fluorspar has to be imported. Therefore, developing alternative methods for producing hydrogen fluoride is an important task.

Production of wet-process phosphoric acid from fluorapatite, using the reaction

$$Ca_5(PO_4)_3F_{sol} + 5H_2SO_{4\,liq} \rightarrow 3H_3PO_{4\,liq} + 5CaSO_{4\,sol} + HF_{gas}, \quad (2)$$

forms a substantial amount of hydrogen fluoride, which in turn forms silicon tetrafluoride SiF_4 reacting with silicon dioxide SiO_2 (an admixture in fluoroapatite):

$$SiO_{2\,sol} + 4HF_{gas} \leftrightarrow SiF_{4\,gas} + 2H_2O_{gas}. \quad (3)$$

Tetrafluoride is then hydrolyzed, and the obtained hexafluorosilicic acid H_2SiF_6 is neutralized, and solid fluorine-containing wastes are disposed to landfills.

The amount of these wastes generated in processing fluorapatite is about 2 million tons per year (in terms of fluorine content). Therefore, these wastes from processing fluorapatite can serve as the main industrial source of fluorine.

There is a great number of chemical compounds in the Si-F-H-O system. Some of these compounds and data on their interaction with water, hydrogen and oxygen are given in Table 1 [3–13]. It follows from the data that the most stable compounds in the Si-F-H-O system are SiO_2, HF, SiF_4, H_2O; in addition, it can be seen that hydrogen fluoride can be obtained by hydrolysis of silicon tetrafluoride.

Determination of the temperature range

In order to determine the preferable temperature range for hydrolysis of silicon fluoride, we calculate the Gibbs energy change depending on temperature for the following process:

$$SiF_{4\,gas} + 2H_2O_{gas} \leftrightarrow SiO_{2\,sol} + 4HF_{gas}. \quad (4)$$

More than ten crystalline modifications of SiO_2 are known. The most stable of them are β-quartz, β-tridimite and β-cristobalite, whose melting points are 1550, 1680 and 1720 °C, respectively [4]. The heat capacities and other thermodynamic parameters of these crystalline modifications differ only slightly (by units of percent), so we performed further calculations for one modification, β-tridymite.

The procedure for calculating the Gibbs energy change is given in [14]. We approximated the temperature dependence of heat capacity by a first-degree polynomial. The calculation results are shown in Fig. 1.

Reaction (4) is reversible (↔); the forward reaction rate is equal to the reverse reaction rate at a temperature of 1170 K (referred to as critical, T_{cr}, for convenience from now on). Therefore, hydrogen fluoride can be obtained by hydrolysis of silicon fluoride at a temperature above T_{cr}, quickly cooling the products at temperatures below T_{cr}.

We propose to treat SiF_4 in flames of hydrogen-containing fuel and an oxygen-containing oxidizing agent for high-temperature hydrolysis:

$$SiF_{4\,gas} + x_1 C_m H_n O_k N_{l\,gas} + x_2 O_{2\,gas} \rightarrow \rightarrow SiO_{2\,sol/liq} + 4HF_{gas} + + y_1 CO_{y\,gas} + y_2 N_{2\,gas} - Q, \quad (5)$$

where $C_m H_n O_k N_l$ is the volatile hydrogen-containing substance, $m \geq 0$, $n > 0$, $k \geq 0$, $l \geq 0$; $nx_1 \geq 4$, $2x_2 > 2$, $y = 1; 2$; the standard enthalpy of formation is $Q \sim 10^2 - 10^3$ kJ; the adiabatic temperature of the reaction products is $T_{ad} > 10^3$ K.

Hydrogen, methane and ammonia can serve as hydrogen-containing fuel, and oxygen and air as an oxygen-containing oxidizing agent.

Table 1

Physico-chemical properties of silicon compounds

Compound	Interaction with water	Characteristic reaction
SiF_4 Si_2F_6	$3SiF_4 + 2H_2O \rightarrow SiO_2 + 2H_2[SiF_6]$ (T = 100 °C), $SiF_4 + 2H_2O \rightarrow SiO_2 + 4HF$ (T > 800 °C)	$Si_2F_6 \rightarrow SiF_2 + SiF_4$ (700 °C)
H_2SiF_6	Exists in aqueous solution only	$H_2SiF_6 \rightarrow SiF_4 + HF$
H_3SiF H_2SiF_2 $HSiF_3$	$H_3SiF + H_2O \rightarrow 2HF + (SiH_3)_2O$	$2H_3SiF \rightarrow SiH_4 + H_2SiF_2$, $4HSiF_3 \rightarrow 3SiF_4 + Si + 2H_2$ (20 °C)
SiH_4, Si_2H_6, Si_3H_8, $(SiH_x)_4$	$SiH_4 + 2H_2O \rightarrow SiO_2 + 4H_2$, $Si_2H_6 + 4H_2O \rightarrow 2SiO_2 + 7H_2$	$SiH_4 + 2O_2 \rightarrow SiO_2 + 2H_2O$, $SiH_4 \rightarrow Si + 2H_2$ (400–1000 °C)
H_2SiO_3 H_4SiO_4	Sparingly soluble	$H_2SiO_3 \rightarrow H_2O + SiO_2$
SiO, SiO_2	$SiO + H_2O \rightarrow SiO_2 + H_2$ (T > 500 °C)	$SiO_2 + 2H_2 \rightarrow Si + 2H_2O$ (800 °C)

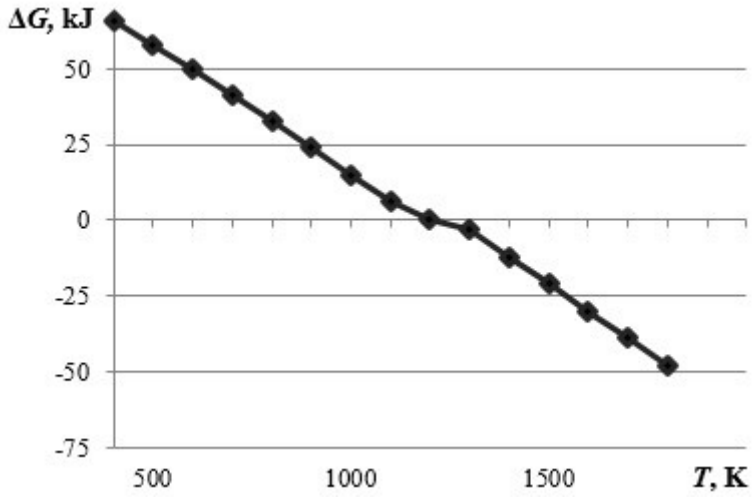

Fig. 1. Temperature dependence of Gibbs energy change for hydrolysis of silicon tetrafluoride (4)

The products of the process are a slightly dust-laden gas flow, i.e.,

$$V_{sol}/V \sim 10^{-5},$$

where V_{sol}, V are the volumes of the solid phase and all products of the process, respectively.

For this reason, the flow should be cooled in a convective heat exchanger of the double pipe type.

Table 2 shows the thermal enthalpies and adiabatic temperatures T_{ad} of the products for process (5). A nonhomogeneous flame loses up to 40% of the released energy due to thermal radiation [15]. In view of this, the temperatures T_{rad} of the reaction products given in Table 2 were calculated taking into account this loss.

Table 3 gives T_{ad} and T_{rad} depending on the initial temperature of the reagents for three crystalline modifications of SiO_2 for the process described by the reaction

$$SiF_{4\,gas} + 2H_{2\,gas} + O_{2\,gas} \rightarrow \rightarrow SiO_{2\,sol/liq} + 4HF_{gas}. \quad (6)$$

It follows from the data given in Table 3 that the values of T_{ad} and T_{rad} are significantly higher than the values of T_{cr} for all cases, so the preferable method for producing hydrogen fluoride from silicon tetrafluoride is by scheme (5). In addition, it is evident that T_{ad} and T_{rad} weakly depend on the structure of crystalline modification of SiO_2.

Fig. 2 shows the Gibbs energy change ΔG depending on temperature in the range of 300–1800 K for reactions (6)–(8) with the ratio of starting components corresponding to production of SiO_2 and HF:

$$3SiF_{4\,gas} + 4NH_{3\,gas} + 3O_{2\,gas} \rightarrow \rightarrow 3SiO_{2\,sol} + 12HF_{gas} + 2N_{2\,gas}, \quad (7)$$

$$SiF_{4\,gas} + CH_{4\,gas} + 2O_{2\,gas} \rightarrow \rightarrow SiO_{2\,sol} + 4HF_{gas} + CO_{2\,gas}. \quad (8)$$

It can be seen from Fig. 2 that the ΔG values are negative for the given processes, which means that processes (6)–(8) are not thermodynamically forbidden for this temperature range.

Table 2

Main thermal parameters for reactions occurring in interaction of silicon tetrafluoride with different compounds (T_0 = 500 K)

Reaction	$-Q$, kJ	T_{ad}	T_{rad}
		K	
$SiF_{4\,gas} + 2H_{2\,gas} + O_{2\,gas} \rightarrow SiO_{2\,sol/liq} + 4HF_{gas}$	384	2491	1843
$3SiF_{4\,gas} + 4NH_{3\,gas} + 3O_{2\,gas} \rightarrow \rightarrow 3SiO_{2\,sol/liq} + 12HF_{gas} + 2N_{2\,gas}$	969	2083	1562
$SiF_{4\,gas} + CH_{4\,gas} + 2O_{2\,gas} \rightarrow \rightarrow SiO_{2\,sol/liq} + 4HF_{gas} + CO_{2\,gas}$	703	3020	2214
$SiF_{4\,gas} + 2H_{2\,gas} + O_{2\,gas} + 4N_{2\,gas} \rightarrow \rightarrow SiO_{2\,sol/liq} + 4HF_{gas} + 4N_{2\,gas}$	384	1836	1407
$3SiF_{4\,gas} + 4NH_{3\,gas} + 3O_{2\,gas} + 12N_{2\,gas} \rightarrow \rightarrow 3SiO_{2\,sol/liq} + 12HF_{gas} + 14N_{2\,gas}$	969	1598	1248
$SiF_{4\,gas} + CH_{4\,gas} + 2O_{2\,gas} + 8N_{2\,gas} \rightarrow \rightarrow SiO_{2\,sol/liq} + 4HF_{gas} + CO_{2\,gas} + 8N_{2\,gas}$	703	1982	1501

Notations: Q is the thermal effect; T_{ad} and T_{rad} are the adiabatic and radiation temperatures; T_0 is the temperature of the starting reagents.

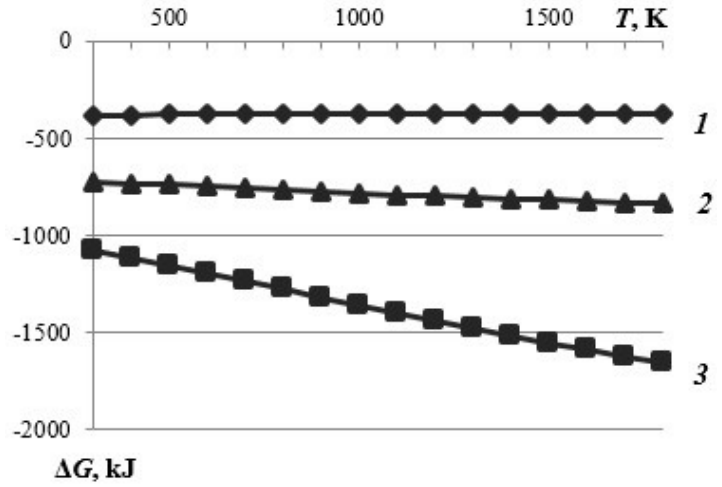

Fig. 2. Temperature dependences of Gibbs energy change for reactions (6) (*1*), (7) (*2*) and (8) (*3*)

Fig. 3 shows the temperature dependences of the Gibbs energy change ΔG for β-quartz, β-tridymite, and β-cristobalite for process (6).

The Gibbs energy change in reaction (8) weakly depends on the structure of crystalline modification of SiO_2: the difference in ΔG does not exceed 5%.

As noted above, the most thermally stable elements in the Si-F-H-O system are SiO_2, SiF_4, H_2O and HF. The thermodynamically equilibrated composition of substances in this system was calculated by minimizing the Gibbs energy for the mixture, varying the concentration of the components with the given atomic ratio [16]:

$$x_1 SiF_{4\ gas} + x_2 H_{2\ gas} + x_3 O_{2\ gas} \rightarrow \rightarrow y_1 SiO_{2\ sol} + y_2 SiF_{4\ gas} + + y_3 HF_{gas} + y_4 H_2 O_{gas}, \quad (9)$$

where x_i, y_i are the stoichiometric coefficients.

The atomic balance equations have the following form:

$$Si: y_1 + y_2 = x_1;\ H: 4y_2 + y_3 = 4x_1;$$
$$F: y_3 + 2y_4 = 2x_2;\ O: 2y_1 + y_4 = 2x_3. \quad (10)$$

If we express y_2, y_3, y_4 in terms of x_1, x_2, x_3 and y_1, i.e.,

$$y_2 = x_1 - y_1;\ y_3 = 4y_1;\ y_4 = x_2 - 4y_1,$$

we obtain:

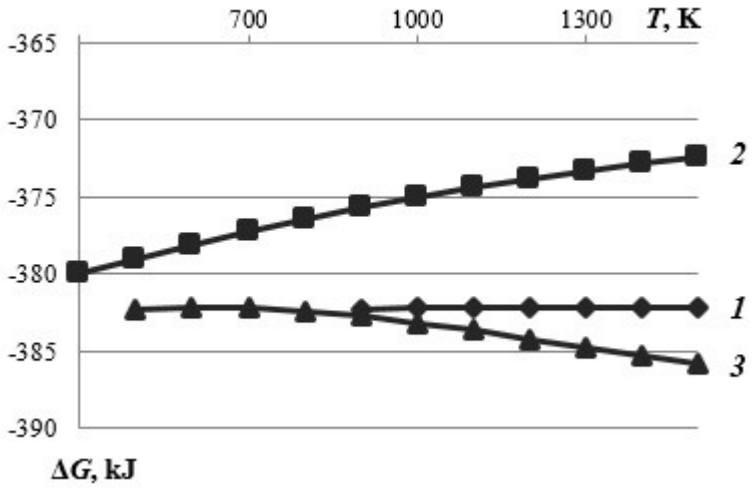

Fig. 3. Temperature dependences of Gibbs energy change for process (6), for three crystalline modifications of silicon dioxide: β-quartz (*1*), β-tridymite (*2*) and β-cristobalite (*3*)

Table 3

Temperature parameters of reaction (6) depending on initial temperature of reagents for different crystalline modifications of silicon dioxide

T_0	T_{ad}	T_{rad}	T_{ad}	T_{rad}	T_{ad}	T_{rad}
	β-quartz		β-tridymite		β-cristobalite	
400	–	–	2544	1760	–	–
600	–	–	2886	2120	2812	2055
800	–	–	3226	2476	3161	2415
1000	3547	2803	3563	2829	3508	2773

Note. All temperatures are given in degrees Kelvin (K).

$$\sum G(x_1, x_2, y_1) = y_1 G(SiO_{2\,sol}) + (x_1 - y_1) G(SiF_{4\,gas}) + 4y_1 G(HF_{gas}) + (x_2 - 4y_1) G(H_2O_{gas}). \quad (11)$$

Given fixed values of x_1, x_2, x_3, we varied y_1 in increments of 0.001, constructing a matrix and then selecting its minimum by comparison.

Fig. 4 shows the temperature dependences for the concentration of products of process (6) in a thermodynamically equilibrated mixture, calculated using the model we formulated [16].

SiO_2 is the main silicon-containing substance at temperatures above 1300 K, and HF is the main fluorine-containing substance; the concentration of SiF_4 does not exceed 3%, and that of H_2O does not exceed 8%.

Calculations in the ASTRA software package

We have tested the model using the ASTRA software package, allowing to calculate the thermodynamically equilibrated composition by entropy maximization [17]. The calculated results for the Si-4F-4H-2O system are given in Table 4.

The results obtained with the ASTRA software package are in qualitative agreement with the calculated data on equilibrium compositions obtained by the method that we have developed.

The ASTRA package was also used to calculate the equilibrium compositions of substances in the Si-4F-C-4H-4O system. The calculated results are given in Table 5.

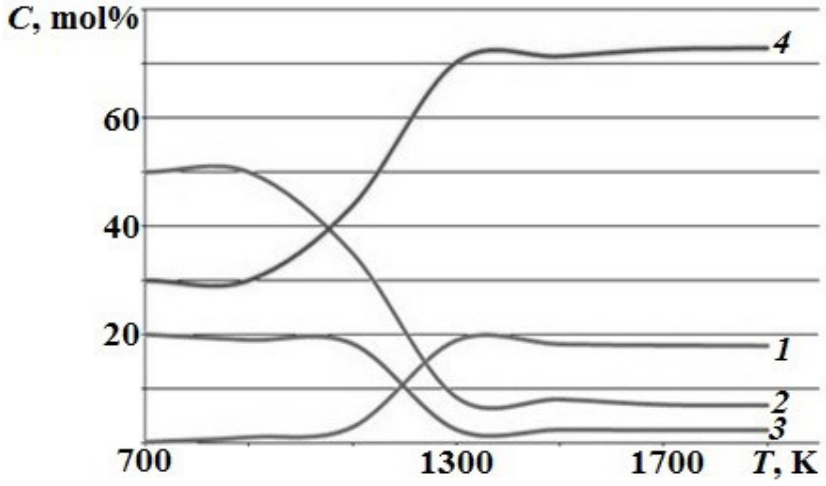

Fig. 4. SiO_2 (*1*), SiF_4 (*2*), H_2O (*3*) and HF (*4*) concentrations depending on temperature for Si-F-H-O system (SiF_4-$2H_2$-O_2 mixture is thermodynamically equilibrated, i.e., $x_1 = x_3 = 1$, $x_2 = 2$)

Table 4

Thermodynamic equilibrium compositions of substances (mol.%) in Si-4F-4H-2O system depending on temperature

T, K	H_2O	HF	SiO_2	SiF_4
500	64.5	2.2	0.5	32.5
700	57.3	10.8	2.7	28.9
900	46.1	24.3	6.1	23.3
1100	35.3	37.3	9.3	17.8
1300	26.9	47.4	11.9	13.6
1500	20.8	54.6	13.7	10.6
1700	16.7	59.6	14.9	8.5
1900	13.7	63.1	15.8	7.1

Notes. 1. The data given were calculated using the ASTRA software package. 2. The O_2 content was less than 0.2 mol.% at all temperatures.

Table 5

Thermodynamic equilibrium compositions of substances (mol.%) in Si-4F-4H-1C-4O system depending on temperature

T, K	H_2O	HF	CO_2	SiO_2	SiF_4
500	48.6	1.7	24.7	0.4	24.4
700	43.3	8.8	23.9	2.2	21.7
900	35.1	19.8	22.5	5.0	17.6
1100	27.0	30.6	21.1	7.6	13.5
1300	20.7	39.0	20.1	9.8	10.4
1500	16.1	45.1	19.3	11.3	8.1
1700	12.9	49.3	18.8	12.3	6.5
1900	10.6	52.3	18.2	13.1	5.4

Note. The data were calculated using the ASTRA software package.

Hydrogen fluoride is the main fluorine-containing substance at temperature above 1100 K in the Si-4F-4H-1C-4O system. Upon reaching 1900 K, the hydrogen fluoride content in the equilibrium mixture amounted to about 50 mol.%, and the silicon tetrafluoride content to about 5 mol.%.

Thus, analyzing the calculated equilibrium compositions of the substances in Si-F-H-O and Si-F-H-C-O systems, we can assume that HF can be the main fluorine-containing substance for SiF_4 processed in flames of hydrogen-containing fuel with an oxygen-containing oxidizing agent at temperatures above 1300 K, and SiO_2 can be the main silicon-containing substance.

Metal and ammonium hexafluorosilicates can be obtained from aqueous solution of H_2SiF_6 and SiF_4 [18, 19]:

$$H_2SiF_{6\,liq} + 2NaCl_{\,liq} \rightarrow \rightarrow Na_2SiF_{6\,sol} + 2HCl_{\,liq}, \quad (12)$$

$$H_2SiF_{6\,liq} + 2NH_4OH_{\,liq} \rightarrow \rightarrow (NH_4)_2SiF_{6\,liq}, \quad (13)$$

$$2NH_4F_{liq} + SiF_{4\,liq} \rightarrow \rightarrow (NH_4)_2SiF_{6\,liq}, \quad (14)$$

$$H_2SiF_{6\,liq} + !\ aCO_{3\,sol} \rightarrow \rightarrow\ !\ 0SiF_{6\,sol} + H_2CO_{3\,liq}. \quad (15)$$

Therefore, we considered whether it was thermodynamically possible to produce hydrogen fluoride from hexafluorosilicates in flames of hydrogen-containing fuel and an oxygen-containing oxidizing agent.

No data are available in literature on thermodynamic functions of the $CaSiF_6$, $(NH_4)_2SiF_6$, Na_2SiF_6 hexafluorosilicates; however, these salts are thermally unstable at temperatures above 370, 250 and 600 °C, respectively [7]:

$$CaSiF_{6\,sol} \rightarrow CaF_{2\,sol} + SiF_{4\,gas}, \quad (16)$$

$$(NH_4)_2SiF_{6\,sol} \rightarrow \rightarrow 2NH_{3\,sol} + SiF_{4\,gas} + 2HF_{gas}, \quad (17)$$

$$Na_2SiF_{6\,sol} \rightarrow 2NaF_{sol} + SiF_{4\,gas}. \quad (18)$$

For this reason, we performed further calculations for their decomposition products.The equations for hydrolysis of sodium and calcium fluorides have the form

$$NaF_{sol} + H_2O_{\,gas} \rightarrow \rightarrow NaOH_{\,sol/liq} + HF_{gas}, \quad (19)$$

$$CaF_{2\,sol} + H_2O_{\,gas} \rightarrow \rightarrow CaO_{\,sol} + 2HF_{gas}. \quad (20)$$

We calculated the Gibbs energy change as function of temperature for these reactions. We found (Fig. 5) that the Gibbs energy change for this reaction follows the inequality $\Delta G > 0$ in the entire temperature range considered. Consequently, reactions (19) and (20) are thermodynamically forbidden in the temperature range $T = 300-2000$ K.

The local maximum on curve *2* (Fig. 5) is due to the fact that the crystal lattice of calcium fluoride changes at a 1424 K, and this compound melts at 1691 K [5].

Table 6 gives the calculated standard enthalpies of formation, temperatures T_{ad} and T_{rad} for the interaction of products of thermal decomposition of hexafluorosilicates in flames of hydrogen-containing fuel and oxygen, with the ratio of the starting components corresponding to production of SiO_2 and HF at the initial temperature $T_0 = 500$ K. Hydrogen-containing fuel is contained in the molecule of hexafluorosilicate in case of ammonium hexafluorosilicate.

The values of T_{rad} in Table 6 significantly exceed the value of T_{cr} obtained for SiF_4 (see Fig. 1), and, therefore, hydrolysis of SiF_4 is thermodynamically possible in the given processes (see Table 6). Notably, sodium and calcium fluorides are not hydrolyzed in the temperature range $T = 300-2000$ K, so only 67% fluorine regeneration is possible from hexafluorosilicates of these elements.

Fig. 6 shows the Gibbs energy change ΔG depending on temperature in the range of 300–1800 K for interaction of thermal decomposition products of hexafluorosilicates with hydrogen-containing substances and oxygen (all reactions in Table 6).

The Gibbs energy changes are negative in the given temperature range, therefore, all the reactions in Table 6 are not thermodynamically forbidden

As the dust-laden gas flow formed by combining the reagents

$$SiO_{2\,sol} + 4HF_{gas}, \qquad (21)$$

is cooled at temperatures below 1170 K, fluorination of SiO_2 occurs:

$$SiO_{2\,sol} + 4HF_{gas} \rightarrow \rightarrow SiF_{4\,gas} + 2H_2O_{gas}. \qquad (22)$$

Because of this, flow (21) should be cooled at the highest rate possible.

Calculation of heat transfer parameters

Kinetic models for reaction (22) are not described in literature. It is thus impossible to give a quantitative assessment for the necessary cooling rate.

We made estimates for the characteristic cooling time of the dust-laden gas flow (21) and the parameters of the convective heat exchanger of the double pipe type with a thermostatically controlled wall for SiF_4 flow rates corresponding to the pilot and industrial-scale setups, based on the data from [20].

The Nusselt number for the dust-laden gas flow was calculated by the following relations:

$$Nu_i = 0.023\,\mathrm{Re}_i^{0.8}\,\mathrm{Pr}_i^{0.4} = Nu(1+\mu)^{0.8}(1-\beta)^{1.12}\left(\frac{1+C_{sol}\mu C_{gas}^{-1}}{1+\mu}\right)^{0.4}, \qquad (23)$$

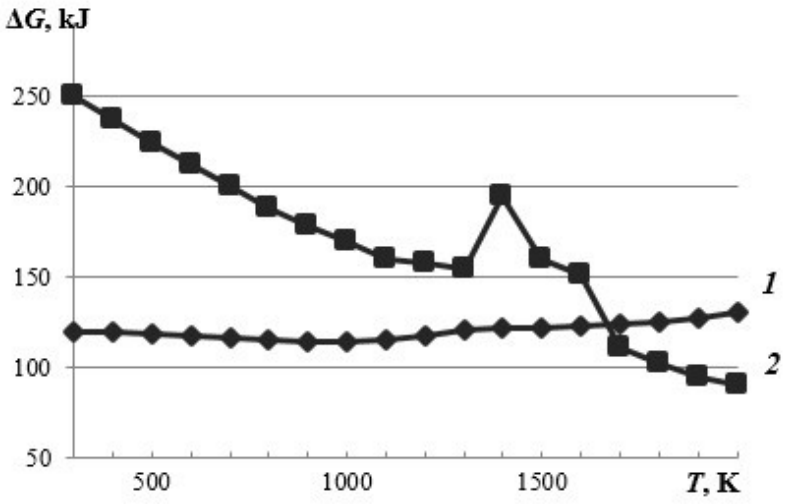

Fig. 5. Temperature dependences of Gibbs energy change for hydrolysis of sodium (*1*) and calcium (*2*) fluorides

Table 6

Main thermal parameters for interaction of thermal decomposition products of hexafluorosilicates with different substances (T_0 = 500 K)

Reaction	-Q, kJ	T_{ad}	T_{rad}
		K	
$SiF_{4\,gas} + 2NH_{3\,gas} + 2HF_{gas} + 1{,}5O_{2\,gas} \rightarrow \rightarrow N_{2\,gas} + SiO_{2\,sol/liq} + 6HF_{gas} + H_2O_{gas}$	539	2211	1643
$2NaF_{sol} + SiF_{4\,gas} + CH_{4\,gas} + 2O_{2\,gas} \rightarrow \rightarrow 2NaF_{sol} + SiO_{2\,sol/liq} + 4HF_{gas} + CO_{2\,gas}$	708	2354	1763
$CaF_{2\,sol} + SiF_{4\,gas} + CH_{4\,gas} + 2O_{2\,gas} \rightarrow \rightarrow CaF_{2\,sol} + SiO_{2\,sol/liq} + 4HF_{gas} + CO_{2\,gas}$	708	2687	1978

The notations for the parameters are given in the caption to Table 2

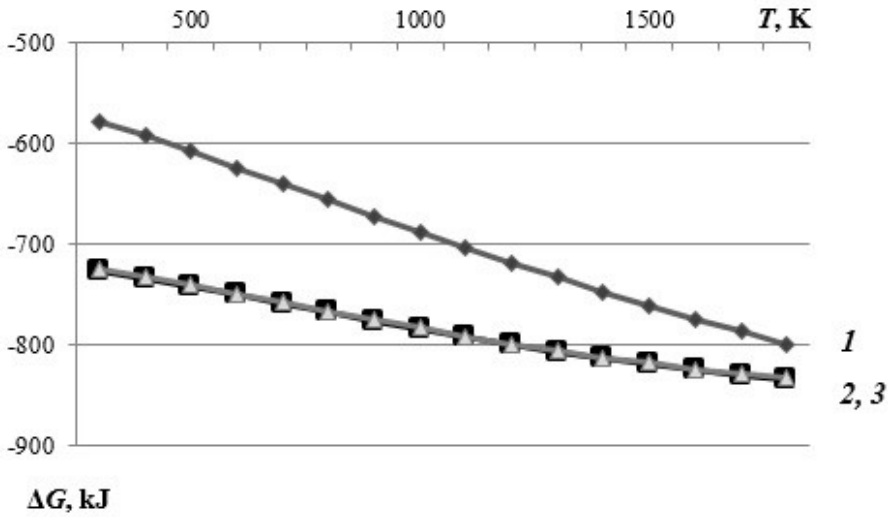

Fig. 6. Temperature dependences of Gibbs energy change for interaction of thermal decomposition products of hexafluorosilicates with hydrogen-containing substances and oxygen
The numbers of the curves correspond to the numbers of reactions in Table 6

$$\mu = \frac{G_{sol}}{G_{gas}} = \frac{\beta}{1-\beta}\frac{\rho_{sol}V_{sol}}{\rho_{gas}V_{gas}}, \qquad \beta = \frac{F_{sol}}{F_s} = \frac{F_{sol}}{F_{sol}+F_{gas}} = \frac{\mu}{\left(\rho_{sol}V_{sol}\Big/\rho_{gas}V_{gas}\right)+\mu}, \tag{24}$$

where G_{sol}, V_{sol} are the flow rate and velocity of the powder; G_{gas}, V_{gas} are the flow rate and velocity of gas; C_{sol}, C_{gas} are the heat capacities of the solid component and gas; ρ_{sol}, ρ_{gas} are the densities of the solid phase and gas, respectively; μ is the flow rate concentration; β is the volumetric concentration of the solid component; F_{sol}, F_{gas} are the volumes of the solid component and gas, respectively; F_s is the system volume.

Table 7 gives the cooling characteristics for slightly dusty (β = $2\cdot10^{-5}$ at T = 1100 K) flow (21) at temperatures from 1100 to 500 K and for the heat exchanger depending on the SiF_4 flow rate and the diameter of the cylindrical heat exchanger.

It follows from the results given in Table 7 that the characteristic cooling time of the flow from 1170 to 500 K is of the order of 10^{-2} s provided that the diameter of the heat exchanger is of the order of tens of millimeters,

Table 7

Characteristics of heat exchanger and slightly dust-laden flow (21) depending on heat exchanger diameter

D, mm	**α**, $W \cdot m^{-2} \cdot K^{-1}$	*L*, m	*u*, m/s	ΔP, kPa	*t*, s	Re	Re_f
20	382	0.68	81	1.47	0.02	16280	24635
30	184	0.94	36	0.29	0.05	10847	16413

Notations: *D*, *L* are the diameter and length of the heat exchanger, α is the heat transfer coefficient; *u* is the dust-laden gas flow velocity; ΔP, *t* are the pressure difference in the heat exchanger and the cooling time of the dust-laden gas flow with a decrease in temperature from 1170 to 500 K; Re, Re_f are the Reynolds numbers for dust-laden gas flows.
Note. The table gives the calculated data for silicon fluoride with the flow rate of 10 g/s (industrial value is 300 tons per year); the wall temperature of the heat exchanger T_{wall} = 100 °C.

its length is of the order of units of meters, and the pressure difference in the heat exchanger is 1.5 kPa. The cooling time is 0.02 seconds for a heat exchanger diameter of 20 mm, so this diameter is considered to be optimal.

Main results and conclusions

Considering the regeneration of fluorine from fluorine-containing materials, we analyzed the existing methods for producing hydrogen fluoride. We have carried out thermodynamic calculations of adiabatic temperature, Gibbs energy change and the equilibrium composition of the reaction products of interaction of silicon tetrafluoride with hydrogen and oxygen-containing substances.

Analysis of the obtained simulation data allowed us to draw the following conclusions.

It is preferable to carry out hydrolysis of SiF_4 aimed at producing SiO_2 and HF at a temperature above 1170 K, followed by rapid cooling of the reaction products.

When SiF_4 is processed in flames of hydrogen-containing fuel (H_2, CH_4, NH_3) and an oxygen-containing oxidizing agent (oxygen, air), the temperature of the reaction products, taking into account thermal radiation of non-homogeneous flame, is significantly higher than 1170 K. The crystalline form of SiO_2 practically does not affect the adiabatic (T_{ad}) and radiation (T_{rad}) temperatures exceeding 1500 and 1200 K, respectively.

SiO_2 is the main silicon-containing substance at temperatures above 1300 K in an equilibrium mixture of substances of the Si-4F-4H-2O system, and HF is the main fluorine-containing substance. The content of SiF_4 is not more than 3%, of water not more than 8% (estimate by the Gibbs energy minimization method).

The results calculated using the ASTRA software package indicate that SiF_4 and H_2O contents in the equilibrium mixture at a temperature of 1900 K are 7% and 14%, respectively. Calculations for the Si-4F-1C-4H-4O system confirm that hydrogen fluoride is the main fluorine-containing substance at 1900 K, the SiF_4 content in an equilibrium mixture is about 5%, carbon dioxide content is 18% and water content is 10%.

HF and SiO_2 can be obtained in flames of hydrogen-containing fuel and an oxygen-containing oxidizing agent using $CaSiF_6$, Na_2SiF_6, $(NH_4)_2SiF_6$, etc., as starting materials: these processes are not forbidden thermodynamically, and the temperature of their products $T_{rad} > 1170$ K. Moreover, hydrolysis of calcium and sodium fluorides is thermodynamically forbidden in the temperature range T = 300--2000 K. Therefore, only 67% regeneration of fluorine is possible from sodium and calcium hexafluorosilicates.

The (SiO_{sol} + $4HF_{gas}$) flow is classified as slightly dusty, so it should be cooled in a convective heat exchanger of the double pipe type. The cooling time can amount to about 10^{-2} s in the temperature range from 1170 to 500 K.

Processing SiF_4 or fluorosilicates in flames of hydrogen-containing fuel and an oxygen-containing oxidizing agent may serve as the basis for producing hydrogen fluoride.

Technology for producing hydrogen fluoride from fluorine-containing by-products and waste products of phosphate fertilizers processed in methane and oxygen flames developed using experimental setup No. 05.608.21.0277. ID RFMEFI60819X0277

REFERENCES

1. Hydrogen fluoride. The Essential Chemical Industry – online, University of York, UK, URL: http://www.essentialchemicalindustry.org/chemicals/hydrogen-fluoride.html.

2. Obzor rynka flyuorita (plavikovogo shpata) v SNG [A review of the fluorite (fluorspar) market in the CIS], 8th edition, Informain Research Team, Moscow, 2016.

3. **Glushko V.P.,** Termodinamicheskiye svoystva individualnykh veshchestv. Spravochnoye izdaniye [Thermodynamic properties of the individual substances. Reference book], the 3d Ed,, Vols. 1–4, Nauka, Moscow, 1979.

4. **Lidin R.A., Molochko V.A., Andreyeva L.L**., Khimicheskiye svoystva neorganicheskikh veshchestv [Chemical properties of the inorganic substances], 3d ed., Chemistry Publishing, Moscow, 2000.

5. Khimicheskaya entsiklopediya [Chemical encyclopedia], In 5 Vols., Vol. 5, Knunyants I.L. is an editor-in-chief, Sovetskaya Entsiklopediya, Moscow, 1990.

6. **Hofmann U., Rudorff V., Haas A., et al.,** Handbuch der Prдparativen Anorganischen Chemie, Herausg. von G. Brauer, Band 3, Stuttgart, Ferdinand Enke Verlag, 1978.

7. **Nekrasov B.V.,** Osnovy obshchey khimii [Fundamentals of general chemistry], Vol. 1., 3d Ed., "Chemistry" Publishing, Moscow, 1973.

8. **Simons J.H. (Ed.),** Fluorene chemistry, 5 Vols., Vol. 1, Academic Press, New York, 1950.

9. **Ryss I.G.**, Khimiya ftora i yego neorganicheskikh soyedineniy [Chemistry of fluorine and its inorganic compounds], Scientific and Technical State Publishing House for Books on Chemistry, Moscow, 1956.

10. **Remy H.,** Treatise on inorganic chemistry: introduction and main groups of the periodic table, Vol. 1, Elsevier, 1959.

11. Khimicheskaya entsiklopediya v 5 t [Encyclopedia on chemistry, 5 Vols.], Vol. 4, Ed. Zefirov N.S., Bolshaya Rossiyskaya Entsiklopedia, Moscow, 1995.

12. **Eseev M.K., Goshev A.A., Horodek P., et al.,** Diagnostic methods for silica-reinforced carbon nanotube based nanocomposites, Nanosystems: Physics, Chemistry, Mathematics. 7 (1) (2016) 180–184.

13. Nitride ceramics: combustion synthesis, properties and applications, Eds A.A. Gromov, L.N. Chukhlomina, Weinheim: Wiley-VCH Verlag, Germany, 2015.

14. **Maslova A.S., Pashkevich D.S.**, Thermodynamic analysis of the process of producing hydrogen fluoride from silicon tetrafluoride in a flame of a hydrogen-containing fuel and an oxygen-containing oxidizing agent, SPbPU Science Week, Materials of Scientific Conference with International Participation, Institute of Applied Mathematics and Mechanics, Polytechnical Institute Publishing, St. Petersburg (2017) 245–247.

15. **Souil J.M., Joulain P., Gengembre E.,** Experimental and theoretical study of thermal radiation from turbulent diffusion flames to vertical target surfaces, Combustion Science and Technology. 41 (1–2) (1984) 69–81.

16. **Zimin A.R., Pashkevich D.S.,** Thermodynamically equilibrium composition of substances in the system of elements U-F-O-H, SPbPU Science Week, Materials of Scientific Conference with International Participation, Institute of Applied Mathematics and Mechanics, Polytechnical Institute Publishing, St. Petersburg (2017) 222–224.

17. **Trusov B.G.,** Code system for simulation of phase and chemical equilibriums at higher temperatures, Engineering Journal: Science and Innovation. Electronic Science and Engineering Publication. (1(1)) (2012) DOI: 10.18698/2308-6033-2012-1-31.

18. **Galkin N.P., Zaytsev V.A., Seregin M.B.,** Ulavlivaniye i pererabotka ftorsoderzhashchikh gazov [Capture and processing of fluorinated gases], Atomizdat, Moscow, 1975.

19. **Zaytsev V.A.,** Proizvodstvo ftoristykh soyedineniy pri pererabotke fosfatnogo syrya [Production of fluoride compounds in phosphate processing], Chemistry Publishing, Moscow, 1982.

20. **Gorbis Z.R.,** Teploobmen i gidromekhanika dispersnykh skvoznykh potokov [Heat exchange and hydromechanics of the dispersed through flows], Energiya, Moscow, 1970.

Received 06.12.2019, accepted 20.12.2019.

THE AUTHORS

ZIMIN Arseniy R.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
fz1min@yandex.ru

PASHKEVICH Dmitriy S.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
pashkevich-ds@yandex.ru

MASLOVA Anastasia S.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
maslovanastya95@gmail.com

KAPUSTIN Valentin V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
Valentin.Kapustin.2014@yandex.ru

ALEXEEV Yuriy I.
LTD "New Chemical Products"
11 Moika River Emb., St. Petersburg, 191186, Russian Federation
alexeev-588@yandex.ru

СПИСОК ЛИТЕРАТУРЫ

1. Hydrogen fluoride. The essential chemical industry – online. University of York, UK. URL: http://www.essentialchemicalindustry.org/chemicals/hydrogen-fluoride.html.

2. Обзор рынка флюорита (плавикового шпата) в СНГ. М.: Исследовательская группа Информайн. Изд. 8-е, 2016. 87 с.

3. **Глушко В.П.** Термодинамические свойства индивидуальных веществ. Справочное издание (3-е изд.). В 4-х тт. М.: Наука, 1979.

4. **Лидин Р.А., Молочко В.А., Андреева Л.Л.** Химические свойства неорганических веществ. 3-е изд., испр. Под ред. Р.А. Лидина. М.: Химия, 2000. 480 с.

5. Химическая энциклопедия. В 5 тт. Т. 2: Д. – М. Редколлегия: Кнунянц И. Л. (гл. ред.) и др. М.: Советская энциклопедия, 1990. 671 с.

6. **Гофман У., Рюдорф В., Хаас А., Шенк П.В., Губер Ф., Шмайсер М., Баудлер М., Бехер Х.-Й., Дёнгес Э., Шмидбаур Х., Эрлих П., Зайферт Х.И.** Руководство по неорганическому синтезу. Пер. с нем. / Под. ред. Г. Брауэра. В 6 тт. Т. 3. М.: Мир, 1985. 392 с.

7. **Некрасов Б.В.** Основы общей химии. Изд. 3-е., испр. и доп. В 2 тт. Т. 1. М.: Химия, 1973. 656 с.

8. Фтор и его соединения. Пер. с англ. Под ред. Д. Саймонса. В 2 тт. Т. 1. М.: Изд-во иностр. лит-ры, 1953. 510 с.

9. **Рысс И.Г.** Химия фтора и его неорганических соединений. М.: Гос. науч.-техн. изд– во хим. лит-ры, 1956. 718с.

10. **Реми Г.** Курс неорганической химии. Пер. с нем. Т. 1. – М.: ИИЛ, 1963. – 922 с., ил.

11. Химическая энциклопедия В 5 тт. Т. 4. П. – Т. Редколлегия: Зефиров Н.С. (Гл. ред.) и др. М.: Большая Российская энциклопедия, 1995. 639 с.

12. **Eseev M.K., Goshev A.A., Horodek P., Kapustin S.N., Kobets A.G., Osokin C.S.** Diagnostic methods for silica-reinforced carbon nanotube based nanocomposites // Nanosystems: Physics, Chemistry, Mathematics. 2016. Vol. 7. No. 1. Pp. 180–184.

13. Nitride ceramics: combustion synthesis, properties and applications. Ed. by A.A. Gromov, L.N. Chukhlomina. Weinheim: Wiley-VCH Verlag, Germany, 2015. 331 p.

14. **Маслова А.С., Пашкевич Д.С.** Термодинамический анализ процесса получения фторида водорода из тетрафторида кремния в пламени водородсодержащего

топлива и кислородсодержащего окислителя // Неделя науки СПбПУ. Материалы научной конференции с международным участием. Институт прикладной математики и механики. СПб.: Изд-во Политехн. ун-та, 2017. С. 245–247.

15. **Souil J.M., Joulain P., Gengembre E.** Experimental and theoretical study of thermal radiation from turbulent diffusion flames to vertical target surfaces// Combustion Science and Technology. 1984. Vol. 41. No. 1–2. Pp. 69–81.

16. **Зимин А.Р., Пашкевич Д.С.** Термодинамически равновесный состав веществ в системе элементов U-F-O-H // Неделя науки СПбПУ. Материалы научной конференции с международным участием. Институт прикладной математики и механики. СПб.: Изд-во Политехн. ун-та, 2017. С. 222–224.

17. **Трусов Б.Г.** Программная система моделирования фазовых и химических равновесий при высоких температурах // Инженерный журнал: наука и инновации. Электронное научно-техническое издание. 2012. № 1 (1). 21 с. DOI: 10.18698/2308-6033-2012-1-31.

18. **Галкин Н.П., Зайцев В.А., Серегин М.Б.** Улавливание и переработка фторсодержащих газов. М.: Атомиздат, 1975. 240 с.

19. **Зайцев В.А.** Производство фтористых соединений при переработке фосфатного сырья. М.: Химия, 1982. 248 с.

20. **Горбис З.Р.** Теплообмен и гидромеханика дисперсных сквозных потоков. М.: Энергия, 1970. 424 с.

Статья поступила в редакцию 06.12.2019, принята к публикации 20.12.2019.

СВЕДЕНИЯ ОБ АВТОРАХ

ЗИМИН Арсений Романович – *аспирант кафедры гидродинамики, горения и теплообмена Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
fz1min@yandex.ru

ПАШКЕВИЧ Дмитрий Станиславович – *доктор технических наук, профессор кафедры гидродинамики, горения и теплообмена Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
pashkevich-ds@yandex.ru

МАСЛОВА Анастасия Сергеевна – *студентка Высшей школы прикладной математики и вычислительной физики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
maslovanastya95@gmail.com

КАПУСТИН Валентин Валерьевич – *аспирант кафедры гидродинамики, горения и теплообмена Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
Valentin.Kapustin.2014@yandex.ru

АЛЕКСЕЕВ Юрий Иванович – *кандидат технических наук, главный конструктор ООО «Новые химические продукты».*
191186, Российская Федерация, г. Санкт-Петербург, наб. р. Мойки, 11.
alexeev-588@yandex.ru

DOI: 10.18721/JPM.13109
УДК 537.531, 621.371, 539.234

NANOSTRUCTURED CARBON AND ORGANIC FILMS: SPECTRAL MICROWAVE AND OPTICAL CHARACTERISTICS

V.V. Starostenko, A.S. Mazinov, A.S. Tyutyunik, I.Sh. Fitaev, V.S. Gurchenko

V.I. Vernadsky Crimean Federal University,
Simferopol, Republic of Crimea, Russian Federation

Microwave and optical transmission and reflection spectra of thin films prepared by casting the aqueous and dichloromethane solutions of fullerene, as well as casting the chloroform solution of 4-methylphenylhydrazone N-isoamylisatin have been recorded in the 2.5 – 4.0, 8.2 – 12.0 GHz and 19 – 110, 330 – 740 THz ranges. The carbon samples precipitated from dichloromethane were established to be the most sensitive to the microwaves. There were 3.4 and 9.1 GHz absorption peaks in their spectrum. The 20 – 50 and 78 – 108 THz IR intervals were chosen for investigation as the most pronounced. The fullerene-containing films, having a linear optical spectrum, exhibited the maximal absorption factor. The organic samples, having a sharp increase of optical absorption in the 599.6 – 713.8 THz. high-frequency region, exhibited an absorption edge of 3.05 eV. In this case the surface photomicrographs demonstrated a rather ramified relief with nontrivial 3D forms dependent on the solution nature, notably prominent for fullerene surfaces.

Keywords: electromagnetic microwaves, fullerene, organic film, optical range, photomicrograph

Citation: Starostenko V.V., Mazinov A.S., Tyutyunik A.S., Fitaev I.Sh., Gurchenko V.S., Nanostructed carbon and organic films: spectral microwave and optical characteristics, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (1) (2020) 95–105. DOI: 10.18721/JPM.13109

СПЕКТРАЛЬНЫЕ СВЧ- И ОПТИЧЕСКИЕ ХАРАКТЕРИСТИКИ НАНОСТРУКТУРИРОВАННЫХ УГЛЕРОДНЫХ И ОРГАНИЧЕСКИХ ПЛЕНОК

В.В. Старостенко, А.С. Мазинов, А.С. Тютюник, И.Ш. Фитаев, В.С. Гурченко

Крымский федеральный университет имени В.И. Вернадского,
г. Симферополь, Республика Крым, Российская Федерация

Представлены спектры пропускания и отражения электромагнитного излучения для тонких пленок, полученных методом полива из растворов фуллеренов в воде и дихлорметане, а также из растворов 4-метилфенилгидразона N-изоамилизатина в хлороформе, в СВЧ- (2,5 4,0 – и 12,0 – 8,2 ГГц) и оптических (110 – 19 и 740 – 330 ТГц) диапазонах. Показано, что наиболее чувствительны к СВЧ-волнам углеродные образцы, осажденные из дихлорметана, на спектре которых отмечены пики поглощения 3,4 и 9,1 ГГц. В инфракрасном диапазоне были выделены частотные интервалы 20 – 50 и 78 – 108 ТГц, где наиболее ярко проявилось взаимодействие электромагнитных волн с образцами. В оптическом спектре пленки, полученные из двух видов фуллеренсодержащих суспензий, имея линейный спектр, обладали максимальным коэффициентом поглощения, а органические образцы с резким увеличением поглощения в высокочастотной области 713,8 – 599,6 ТГц имели край полосы поглощения 3,05 эВ. При этом микрофотографии поверхностей показали достаточно разветвленный рельеф (в особенности для поверхностей фуллерена) с нетривиальными 3D-образованиями, на форму которых влиял тип растворителя.

Ключевые слова: СВЧ электромагнитные волны, фуллерен, органическая пленка, оптический диапазон, микрофотография

Ссылка при цитировании: Старостенко В.В., Мазинов А.С., Тютюник А.С., Фитаев И.Ш., Гурченко В.С. Спектральные СВЧ- и оптические характеристики наноструктурированных углеродных и органических пленок // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2020. Т. 13. № 1. С. 109–120. DOI: 10.18721/JPM.13109

Introduction

New frequency ranges are introduced for modern radio transceivers, as efforts are made to reduce the sizes and weights of the devices, accompanied by steadily decreasing costs; as a result, the search continues for new materials that can effectively interact with electromagnetic radiation in different frequency ranges. Fiber-optic channels transmitting the largest amounts of data traffic [1] and microwave cellular stations providing direct communication with customers [2,3] remain the key communications today.

The interest towards nanostructured carbon derivatives (carbon nanotubes, graphenes, fullerenes) grew considerably in the late 1990s and early 2000s. These structures not only possess unique physical properties [4–6] but also exhibit broadband absorption in combination with other materials. [7, 8]. Using nanocomposites to construct elementary active devices [9] should make it easy to integrate organocarbon elements into existing electronic circuits of modern transceivers.

Combined with organic materials, these elements can serve as a basis for novel emitting [10] and diode structures [11, 12], significantly expanding their operating ranges.

However, such devices have certain drawbacks, primarily, photopolymerization (undesirable changes in properties induced by exposure to light), photostimulated and ordinary oxidation [13, 14] leading to rapid degradation of organic layers used.

Despite wide interest in organocarbon materials, their frequency properties are mainly used in the visible range, while their characteristics in the medium-wave infrared (IR) and microwave ranges are poorly studied.

In this study, we considered the effects of electromagnetic waves of microwave and optical ranges on nanostructured films of C_{60} fullerene and N-isoamylisatin 4-methylphenylhydrazone (IMPH) organic precursor, serving as the main working layers of the corresponding heterojunctions [15].

Measurement procedure and experimental samples

Since the initial studies focused on barrier structures [15], we considered the effect of electromagnetic radiation on thin films, i.e., on the type of matter from which these heterojunctions were made [12]. Examining C_{60} and IMPH samples, we focused on measurements and analysis of reflection and transmission spectra of electromagnetic radiation in the microwave and optical ranges.

The microwave region was represented by two ranges: 2.5–4.0 and 8.2–12.0 GHz. Thin square substrates of two sizes, 18 × 18 and 6 × 6 mm, were prepared from the given materials for measurements on waveguides with cross sections of 72 × 34 and 23 × 10 mm, respectively.

Optical measurements were carried out in two frequency ranges: mid-wave infrared 19–110 THz (650–3650 cm^{-1}) and visible 330–740 THz (405–909 nm). Samples of the same size, 18 × 18 mm, were made for this purpose.

The C_{60} samples were examined in two phase states. One of them was a fullerene-containing aqueous solution (fullerene water system (FWS)), which was 99.9% pure [16]. Another C_{60} was a solid-phase powder obtained by sputtering graphite [17, 18], 99.5% pure.

The primary FWS suspension was synthesized from crystalline C_{60} (20 mg subsample), dissolved in N-methylpyrrolidone (25 ml) using a magnetic stirrer. The resulting purple-brown solution was mixed with distilled water (12.5 to 100 ml). The resulting clear dark red solution was stirred for 1 h and subjected to exhaustive dialysis against deionized water. The dialysate was passed through a filter (0.45 μm-sized pores), producing clear brownish-yellow solution as a result. It was stored at a temperature of 10 °C, protected from light [16].

The organic precursor was prepared according to a procedure similar to that described in [19]; 3-methyl-1-phenyl-4-formyl-pyrazole-5-one (2 mmol) was dissolved in 96% ethanol (25 ml) by stirring and heating. The corresponding 4-chlorobenzoic acid hydrazide (2 mmol) was added to the resulting solution, which was then stirred and heated for 1–2 h until a precipitate

formed. The precipitate was left in mother liquor for a long time (overnight); then it was filtered off, washed with ethanol and dried in air. Target product (weighing 680 mg) was obtained with a yield of 96% by this procedure.

Solutions were prepared for each of the starting materials (IMPH and C_{60}) as active layers were formed. Chloroform was used as solvent for the IMPH compound, and dichloromethane for powder C_{60} (in concentrations of 0.5 mg/ml). There was no need to use additional solvent to prepare the FWS samples. The final stage of sample preparation started after the obtained suspensions were held at room temperature for at least 48 h. This stage consisted in simultaneously depositing aged suspensions (1 ml each) on substrates intended for measurements in the given frequency ranges.

The following notations were introduced for the film samples:

IMPH (N-isoamylisatin 4-methylphenylhydrazone) refers to the samples precipitated on glass from N-isoamylisatin 4-methylphenylhydrazone solutions in chloroform;

FFWS (fullerene from fullerene water system) to the samples precipitated from aqueous solutions of C_{60};

FDCM (fullerene from dichloromethane) to the samples precipitated from dichloromethane suspensions.

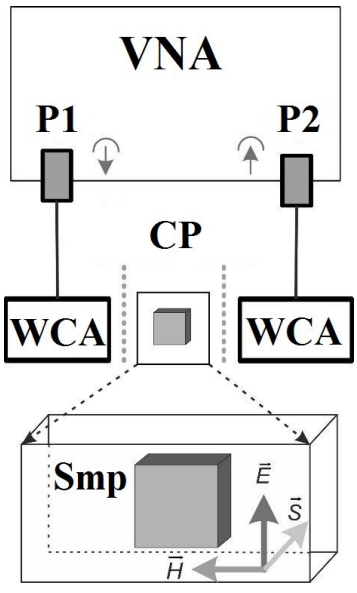

Fig. 1. Block diagram of measurements in waveguide: VNA is the vector circuit analyzer P4226; P1, P2 are the input and output contacts (ports); WCA are the coaxial waveguide transitions: CP is the calibration plane; Smp is the sample in the waveguide (microwave radiation vectors are shown)

Interaction of microwave radiation with fullerene and IMPH films

We previously used the measuring system including the P4226 vector analyzer (Fig. 1) to study the interaction of electromagnetic radiation with thin conducting and semiconducting films [20]. Since the main difficulty in measuring the characteristics of semiconductor fullerene (C_{60}) and organic (IMPH) films was their high ohmic resistance due to small thickness, the measuring system had to be highly sensitive, requiring fine tuning. Measurements were carried out in a closed waveguide in the 2.5–4.0 and 8.2–12.0 GHz ranges to minimize external interference. Through-Reflect-Line calibration was performed to compensate for coaxial waveguide transitions and other interfering factors, using a reflection measure and a quarter-wave line, which yielded fairly accurate results. The effective area of interaction of radiation with the samples was 10% of the cross-sectional area of the waveguide, which helped avoid capacitive and inductive effects from the test sample on the measuring system. The samples were placed in the geometric center of the waveguide cross-section (see Fig. 1) and fixed using a dielectric substrate made from a material that was transparent to microwave radiation. Thus, the sample was at maximum electric field during measurements; since the fundamental mode H_{10} was used, it can be argued that the area of the sample accounted for the largest part of the energy.

The actual interaction of microwave radiation with the samples was determined by the matrix of *S* parameters taking into account the main components S_{21} and S_{11}, corresponding to the radiation directly incident from the first port P1. The initial measurements indicated that the properties of the waveguide with the given structure are close to the properties of a reciprocal two-port network, i.e., the gain is the same in both directions. In view of this, we used the main components S_{21} and S_{11} corresponding to direct incidence from the first port of the VNA.

Recall that the components of *S* parameters are the voltage ratios of the reflected (V_{ref}), incident (V_{inc}) and transmitted (V_{trans}) radiation, i.e.,

$$S_{11} = \frac{V_{ref}}{V_{inc}} \text{ and } S_{21} = \frac{V_{trans}}{V_{inc}};$$

while the powers of the transmitted (P_{trans}) and reflected (P_{ref}) waves are expressed as

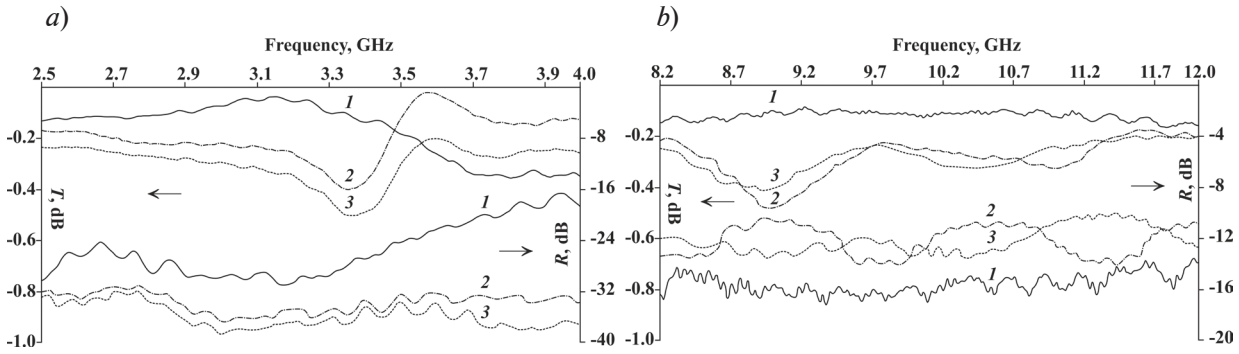

Fig. 2. Frequency spectra of FFWS (*1*), IMPH (*2*) and FDCM (*3*) samples exposed to microwave radiation in 2.5–4.0 (*a*) and 8.2–12.0 (*b*) GHz ranges;
T, *R* are the coefficients of transmitted and reflected power, respectively

$$P_{trans} = \frac{|V_{trans}|^2}{Z_v};$$

$$P_{ref} = \frac{|V_{ref}|^2}{Z_v},$$

where Z_v is the wave impedance.

We first determined the coefficients of transmitted (*T*) and reflected (*R*) power, and then calculated the absorption coefficient *A* (Fig. 2):

$$T = \frac{P_{trans}}{P_{inc}} = \frac{|V_{trans}|^2}{|V_{inc}|^2} = |S_{21}|^2;$$

$$R = \frac{P_{ref}}{P_{inc}} = \frac{|V_{ref}|^2}{|V_{inc}|^2} = |S_{11}|^2;$$

$$A = 1 - |S_{11}|^2 - |S_{21}|^2.$$

Irregular frequency characteristics of the transmission and reflection coefficients confirm our above assumption that the interaction of radiation with thin carbon and organic films has a complex nature. However, the obtained dependences can provide a simplistic explanation for the specific effect of internal structure of the films on the electromagnetic wave. For this reason, we selected the characteristic frequencies ν_1 = 3.4 GHz and ν_2 = 9.1 GHz, at which dips are observed in the frequency dependences of transmittance, for detailed analysis of each of the given spectral ranges (see Fig. 2). In other words, the two materials (IMPH and FDCM) exhibited attenuation of electromagnetic waves at these frequencies. Moreover, the respective curves are similar for both the reflected power and the transmittance at frequencies ν_1 and ν_2. However, maximum transmittance is observed for these structures at frequencies of approximately 3.6 GHz. It is also worth noting that the spectrum is quite uniform and only at these frequencies are anomalies observed, which is obviously due to the specific structure of the material under study. In addition, the reflection and transmission coefficients do not behave anti-symmetrically (curves *2* and *3* in Fig. 2, *a*), suggesting that microwave radiation is absorbed.In contrast to the behavior of IMPH and FDCM samples exposed to microwave radiation, FFWS samples did not possess any pronounced characteristics. However, the samples exhibited an inverse trend to the behavior of other materials at frequencies of 2.5–4.0 GHz: namely, the transmission coefficient decreased with increasing frequency of the incident wave, and the reflection coefficient increased with decreasing transmission coefficient. This suggests that absorption of electromagnetic microwave waves is minimal, and the FFWS material itself has low electrical conductivity, which is, however, higher than that of the other two materials.

Analysis of the general frequency characteristics of the given films led us to conclude that the relationship of the absorbed wave energy with the film volume should be taken into account. The transmission minima at 3.32 and 8.97 GHz were examined more closely. The specific absorbed power *Q* was calculated as the ratio of the power P_{abs} absorbed by the sample to its volume *V*, i.e.,

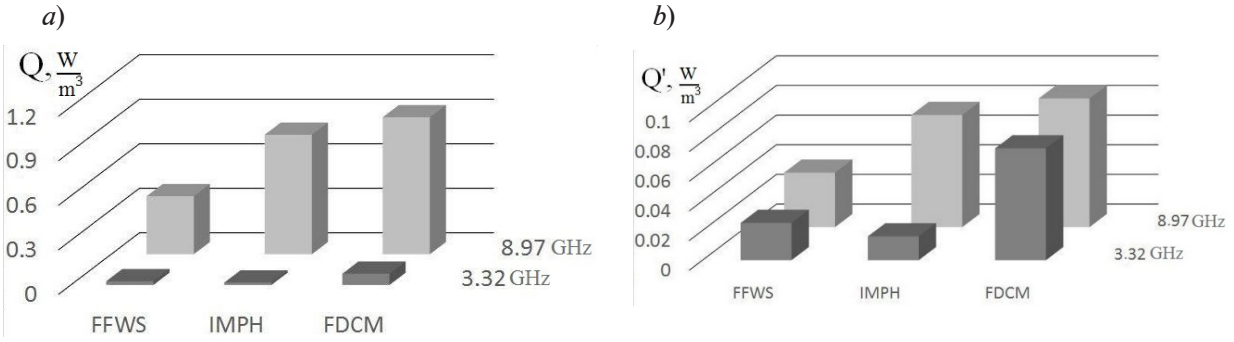

Fig. 3. Graphical representation of specific absorbed microwave power for FFWS, IMPH, and FDCM film samples calculated by Eqs. (1) (*a*) and (2) (*b*)

$$Q = P_{abs}/V, \quad (1)$$

and P_{abs} was calculated as the product of the output power P_{inc} of the VNA generator, equal to −10.00 dBm, multiplied by absorption coefficient A:

$$P_{abs} = P_{inc}\left(1 - |S_{11}|^2 - |S_{21}|^2\right).$$

The volume V was found by averaging the film thicknesses, which we measured using a LOMO MII-4M interference microscope in the most characteristic segments of the samples.

Comparing the specific absorbed power for three samples (Fig. 3, *a*), we found that FDCM films have the highest absorptivity. The lowest absorptivity at 3.32 GHz was observed for IMPH samples, while FFWS films had the lowest absorptivity at 8.97 GHz. Notably, microwave radiation had a constant power at the output of the P4226 generator, so it was impossible to accurately compare the absorptivity of the films at different frequencies. For example, the specific absorbed power was higher at 8.97 GHz than at 3.32 GHz. This effect is not related to the properties of the given materials; it is explained by higher radiation density generated in a waveguide with a smaller cross section. To account for linear calculations, the results were normalized to compare different radiation densities. The normalized specific power (Fig. 3, *b*) obtained follows the expression

$$Q' = Q \cdot \frac{S_{23\times10}}{S_{72\times34}}, \quad (2)$$

where $S_{23\times10}$, $S_{23\times10}$ are the cross-sectional areas of the corresponding waveguide lines.

Thus, the dimensions of the waveguides are taken into account here.

Midwave-IR absorption spectra

The interaction of midwave optical radiation with heterostructure elements was studied with an Agilent Cary 630 FTIR spectrometer in the range of spatial frequencies from 650 to 4000 cm^{-1}, corresponding to direct spectrum of 19.48−119.92 THz, with a resolution of 110 GHz (4 cm^{-1}). The interaction of infrared electromagnetic waves with IMPH, FDCM and FFWS films was particularly pronounced in the frequency ranges of 20−50 and 78−108 THz (667−1667 cm^{-1} and 2601−3602 cm^{-1}).

While the smoothest spectrum for the interaction of microwave radiation with film structures was observed for FFWS samples, the FDCM structures had the smallest number of peaks in the IR range. In particular, a range of relatively narrow absorption bands was observed for the lower frequency range of 20−50 THz (667−1667 cm^{-1}). For example, the peaks observed for FDCM samples at 41.07 and 43.68 THz (1369 and 1457 cm^{-1}) corresponded to the C_{sp3}−H bond, and the peaks at 35.46 and 42.81 THz (1182 and 1427 cm^{-1}) to C_{60}, with the last band coinciding with the band from the alkyl group at 43.68 THz (1456 cm^{-1}) (Fig. 4, *a*). Two characteristic narrow IR absorption bands are clearly seen for FFWS films at 35.43 and 42.81 THz (1181 and 1427 cm^{-1}) (due to C−C bonds) of C_{60}molecules, although they partially overlap with other bands. Absorption bands in the range of 49.46−49.76 THz (1649−1659 cm^{-1}) (due to the C=O bond) for the amide carbonyl group and 29.98−32.97 THz (1000−1099 cm^{-1}) are characteristic for vibrations of the C−O group. In this case, there are no bands characteristic for amino acids (see Fig. 4, *a*). The frequency spectrum of IR absorption by IMPH films is characterized by a significant number of peaks, which is due to numerous chemical bonds in the of N-isoamylisatin 4-methylphenylhydrazone

molecule (see Fig. 4, *a*). Peaks characteristic for vibrations of C=O and C=N atomic groups are found at frequencies of 46.7 and 50.12 THz (1557 and 1671 cm^{-1}). Stretching vibrations of benzene rings play the main role in the frequency range of 40.89–48.26 THz (1363–1609 cm^{-1}). A sequence of absorption maxima is found in the frequency range of 31.59–38.82 THz (1053–1294 cm^{-1}) due to bending and stretching vibrations of C–N, C–C and C–H groups. The main role in the frequency range of 22.30–33.81 THz (743–1127 cm^{-1}) is played by bending vibrations of C–H groups in benzene rings and in the alkyl substituent.

The spectrum is not so diverse at higher frequencies (Fig. 4, *b*), characterized mainly by absorption peaks at frequencies of 75–90 THz (2501–3002 cm^{-1}). In particular, a double peak observed in the range of 83–89 THz (2768–2968 cm^{-1}) for films precipitated from dichloromethane solution, which can be attributed to C_{sp3}–H vibrational modes, appears as a wider single peak for FFWS films (Fig. 4, *b*). However, this peak also has a relatively long absorption band at 90–108 THz (3002–3602 cm^{-1}) with a maximum at 100 THz (3335 cm^{-1}). A series of absorption bands associated with vibrations of the N–H and C–H groups were observed for the IMPH sample in the frequency range of 85.7–101.9 THz (2858–3398 cm^{-1}), (see Fig. 4, *b*).

Microscopy of film surface

Geometry of the surface exposed to such high frequencies of electromagnetic radiation plays an important role, so each of the individual elements and the film as a whole (i.e., the IMPH, FDCM, FFWS compounds) were monitored by reflection and transmission microscopy using a LOMO MII-4M microinterferometer, with enhanced light via a semiconductor laser and with an elongated optical path to a camera with a 1/2 FF 10 MP sensor.

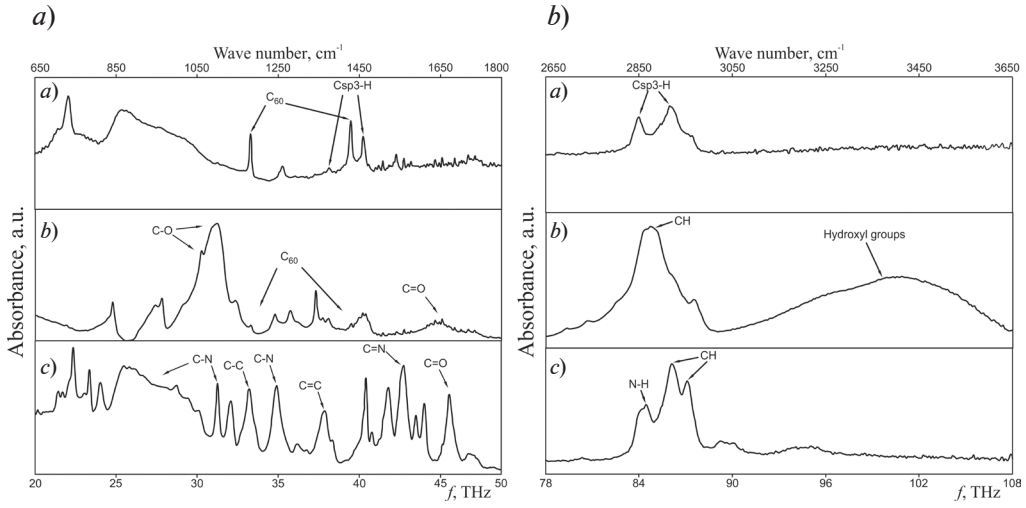

Fig. 4. IR optical absorption spectra of FDCM (*1*), FFWS (*2*) and IMPH (*3*) film samples in 667–1667 cm^{-1} (*a*) and 2601–3602 cm^{-1} (*b*) frequency ranges

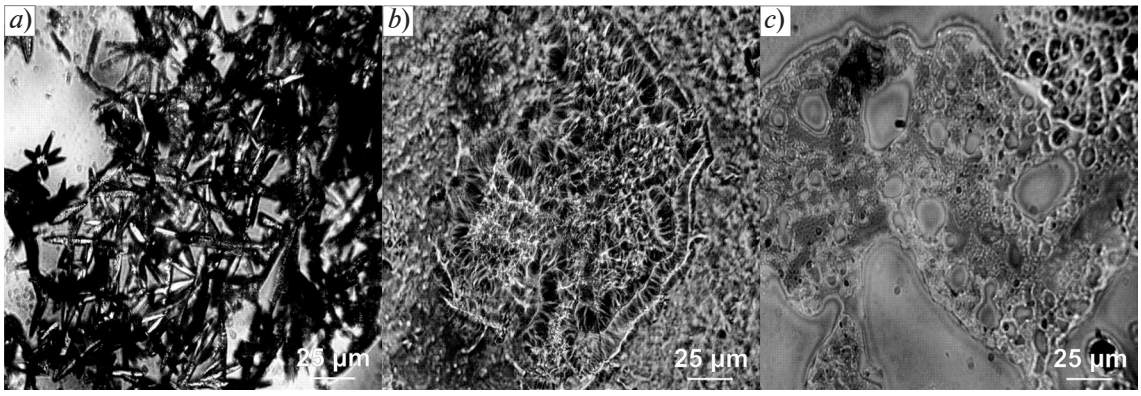

Fig. 5. Micrographs of nanostructured FDCM (*a*), FFWS (*b*) and IMPH (*c*) films

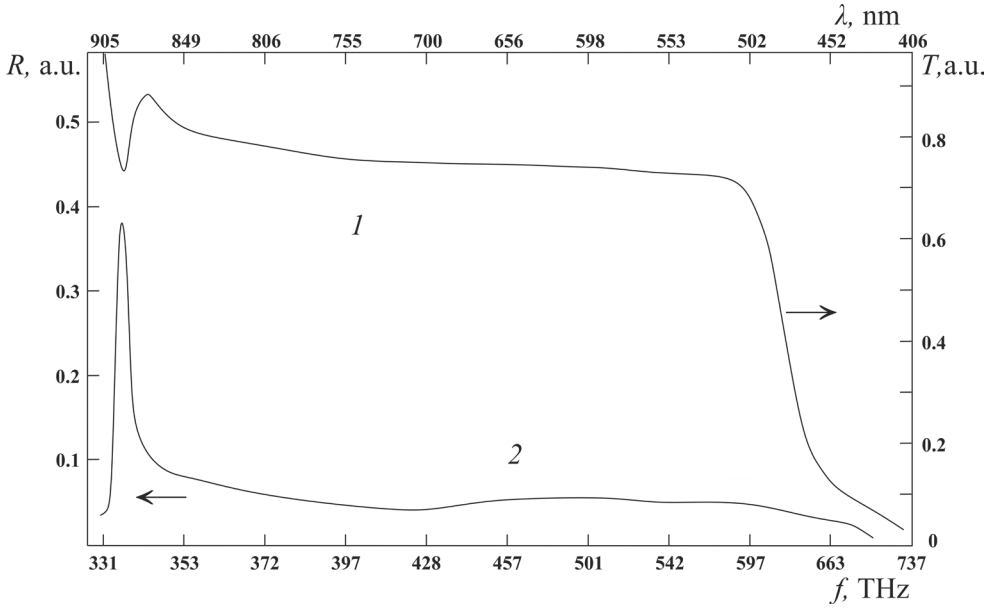

Fig. 6. Optical transmission (*1*) and reflection (*2*) spectra of IMPH thin film in 406–909 nm range

We should note that the surfaces of nanostructured films are irregular, characterized by pronounced separate structures or even regions (Fig. 5). The most characteristic fragments of FDCM, FFWS and IMPH film surfaces are shown.

Distinct microstructures shaped as three-dimensional stars were observed for films precipitated from solution of fullerene in dichloromethane (FDCM), The sizes of individual structures reached 16–20 μm, while film thickness averaged 400–500 nm (see Fig. 5, *a*). FFWS film samples had a fairly uniform surface with localized hexagonal structures. The sizes of individual structures reached 50–80 μm, while film thickness averaged 1.8 μm (see Fig. 5, *b*). The surface of hydrazone films (IMPH) is also relatively uniform, which is explained by considerable length of the 4-methylphenylhydrazone N-isoamylisatin molecule and, in particular, the amyl radical. The film thickness was 1.8–2.0 μm (see Fig. 5, *c*).

Optical transmission and reflection spectra in the visible range

A prism monochromator with an IR filter and a halogen lamp was used for collecting the transmission and reflection spectra of the given films. The spectrometer was calibrated for hydrogen radiation before each series of experiments. A clean substrate was used as a normalizing basis.

FDCM films had the highest absorption: their linear transmission spectrum was at the level of photomultiplier noise and was practically zero. The reflected component was absent for these films. While FFWS samples had similar spectral characteristics, they exhibited a slight dip in the short-wave part of the spectrum.

The optical spectra of light transmission through IMPH films were characterized by sharp minima in the near IR region at 336.85 and 340.68 THz (890 and 880 nm). Accordingly, sharp maxima were observed in the reflection

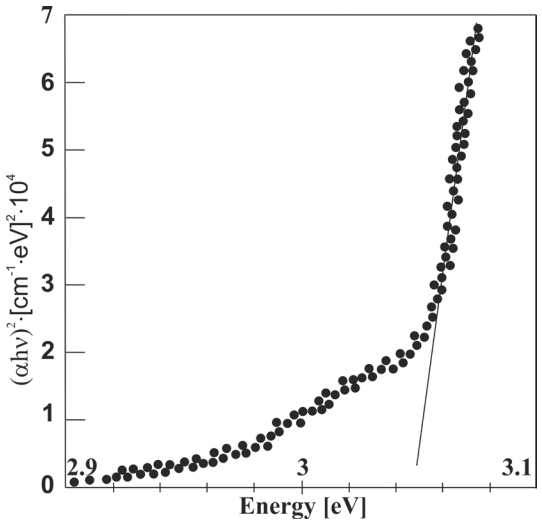

Fig. 7. $\alpha(h\nu)^2$ depending on incident photon energy (energy plot is shown) for IMPH thin film sample

spectra, along with a general decrease in the high-frequency region of 599.6–713.8 THz (500–420 nm) due to absorption in the film (Fig. 6).

We calculated the logarithm of the ratio of transmission coefficient T and reflection coefficient R for the given sample thickness, with subsequent linearization (Fig. 7) with a constant for indirect allowed transitions ($m = 2$) [21]. The formula

$$\alpha h\nu = A\left(h\nu - E_g\right)^m, \qquad (3)$$

was used for the calculations, where α is the absorption coefficient, A is a constant, $h\nu$ is the optical photon energy, E_g is the band gap of the film material.

As a result of the calculations, we obtained the band gap value for the IMPH compound: $E_g = 3.05$ eV.

Conclusion

Almost all film samples of IMPH, FDCM, and FFWS reacted noticeably to electromagnetic radiation in a wide frequency range, i.e., absorption or reflection of incident energy. The infrared region turned out to be the most inhomogeneous the in the range of 20–50 THz (667–1667 cm^{-1}), where a series of narrow-band peaks was observed, with the narrowest bands reaching several hundred gigahertz.

The given structures were less sensitive to microwave radiation. Notably, however, a dip in the transmittance curve was observed at frequencies of 3.4 and 9.1 GHz for the samples precipitated from fullerene suspensions in dichloromethane (FFWS) and from N-isoamylisatin 4-methylphenylhydrazone in chloroform (IMPH).

Sharp minima were observed in the visible absorption spectra at 336.85 and 340.68 THz (890 and 880 nm), accompanied by general decrease in energy in the range of 599.6–713.8 THz (500–420 nm) for IMPH films. Analyzing the obtained experimental data, we have concluded that FDCM films had the highest absorption in all three ranges of electromagnetic radiation considered.

Thus, interaction of electromagnetic radiation with carbon and organocarbon materials can take diverse forms, requiring comprehensive experimental and theoretical studies. We are confident even at this stage that the behavior of microwave, optical absorption and reflection spectra can be controlled by synthesizing complex molecular complexes serving as a basis for heterostructural transitions for experiments in the given frequency ranges.

Acknowledgment

We wish to express our gratitude to the staff of S_{60}Bio (Skolkovo, Moscow) for providing us with a sample of water-soluble fullerene.

The study was financially supported by the Russian Foundation for Basic Research as part of scientific project no. 19-32-90038.

REFERENCES

1. **Kemp S.,** Global digital statshot, URL: https://wearesocial.com/global-digital-report-2019.

2. **Rout S.P.,** 5th generation mobile technology – a new milestone to future wireless communication networks, International Journal of Science and Research. 5 (5) (2016) 529–534.

3. **Kumar A., Gupta M.,** A review on activities of fifth generation mobile communication system, Alexandria Engineering Journal. 57 (2) (2018) 1125–1135.

4. **Baimova J.A., Korznikova E.A., Dmitriev S.V., et al.,** Review on crumpled graphene: unique mechanical properties, Reviews on Advanced Materials Science. 39 (1) (2014) 69–83.

5. **Lebedeva O.S., Lebedev N.G.,** The influence of the stretching and compression deformations on the piezoresistance of the carbon nanotubes and graphene nanoribbons, St. Petersburg State Polytechnical University Journal. Physics and Mathematics (1 (189)) (2014) 26–34.

6. **Eletskii** A.V., Mechanical properties of carbon nanostructures and related materials, Phys. Usp. 50 (3) (2007) 225–261.

7. **Li Y., Liu S., Sun J., et al.,** Effects of the oxygen content of reduced graphene oxide on the mechanical and electromagnetic interference shielding properties of carbon fiber/reduced graphene oxide-epoxy composites. New Carbon Materials. 34 (5) (2019) 489–498.

8. **Wang X., Jiang H.T., Yang K.Y., et al.,** Carbon fiber enhanced mechanical and electromagnetic absorption properties of magnetic graphene-based film, Thin Solid Films. 674 (31) (2019) 97–102.

9. **Chen F.C., Chu C.W., He J., et al.,** Organic thin-film transistors with nanocomposite dielectric gate insulator. Applied Physics Letters. 85 (15) (2004) 3295–3297.10. **Gusev A.N., Kiskin M.A., Braga E.V., et al.,** Novel zinc complex with an ethylenediamine schiff base for high-luminance blue fluorescent OLED applications, The Journal of

Physical Chemistry. 123 (18) (2019) 11850–11859.

11. **Ziminov V.M., Zakharova I.B.,** The rectifying properties of C60 fullerene-based structures, St. Petersburg State Polytechnical University Journal. Physics and Mathematics (2 (146)) (2012) 18–21.

12. **Gusev A.N., Mazinov A.S., Tyutyunik A.S., Gurchenko V.S.** Spectral and conductive properties of film heterostructures based on fullerene-containing material and 4-methylphenylhydrazone N-isoamilisatine; Radio Electronics, Nanophysics and Information Technologies. 11 (3) (2019) 331–336.

13. **Konenkamp R., Priebe G., Pietzak B.,** Carrier mobilities and influence of oxygen in C60 films, Physical Review B. 60 (16) (1999) 11804–11808.

14. **Tapponnier A., Biaggio I., Gunter P.,** Ultrapure C_{60} field-effect transistors and the effects of oxygen exposure, Applied Physics Letters. 86 (11) (2005) 112114.

15. **Gusev A.N., Mazinov A.S., Shevchenko A.I. et al.,** The voltage–current characteristics and photoelectric effect of fullerene C_{60}–N-isoamylisatin 4-methylphenylhydrazone heterostructures, Technical Physics Letters. 45 (10) (2019) 997–1000.

16. **Andreev S.M., Purgina D.D., Bashkatova E.N., et al.,** Facile preparation of aqueous fullerene C_{60} nanodispersions, Nanotechnol. Russia. 9 (7–8) (2014) 369–379.

17 **Mazinov A.S., Gurchenko V.S., Tyutyunik A.S., Shevchenko A.I.,** Influence of structural features of fullerene-containing material on its resistive properties, Ecological Bulletine of the Black Sea Economic Cooperation. 15 (2) (2018) 86–93.

18. **Mazinov A.S., Gurchenko V.S., Tyutyunik A.S., Shevchenko A.I.,** Influence of structural features of fullerene-containing material deposited from solution on its resistive properties, Ecological Bulletine of the Black Sea Economic Cooperation. 15 (4) (2018) 85–92.

19. **Cigan M., Jakusova K., Gaplovsky M., et al.,** Isatin phenylhydrazones: anion enhanced photochromic behavior, Photochemical and Photobiological Sciences. 14 (11) (2015) 2064–2073.

20. **Starostenko V.V., Mazinov A.S., Fitaev I.S., et al.,** Forming surface dynamics of conductive aluminum films deposited on amorphous substrates, Prikladnaya Phyzika. (4) (2019) 60–65.

21. **Al-Saidi I., Sadik F.,** Synthesis and investigation of phenol red dye doped polymer films, Advances in Materials Physics and Chemistry. 6 (5) (2016) 120–128.

Received 18.01.2020, accepted 14.02.2020.

THE AUTHORS

STAROSTENKO Vladimir V.
V.I. Vernadsky Crimean Federal University
4 Vernadskogo Ave., Simferopol, 295007, Republic of Crimea, Russian Federation
starostenkovv@cfuv.ru

MAZINOV Alim S-A.
V.I. Vernadsky Crimean Federal University
4 Vernadskogo Ave., Simferopol, 295007, Republic of Crimea, Russian Federation
mazinovas@cfuv.ru

TYUTYUNIK Andrey S.
V.I. Vernadsky Crimean Federal University
4 Vernadskogo Ave., Simferopol, 295007, Republic of Crimea, Russian Federation
real-warez@mail.ru

FITAEV Ibraim Sh.
V.I. Vernadsky Crimean Federal University
4 Vernadskogo Ave., Simferopol, 295007, Republic of Crimea, Russian Federation
fitaev.i@cfuv.ru

GURCHENKO Vladimir S.
V.I. Vernadsky Crimean Federal University
4 Vernadskogo Ave., Simferopol, 295007, Republic of Crimea, Russian Federation
gurchenko_v@mail.ru

СПИСОК ЛИТЕРАТУРЫ

1. **Kemp S.** Global digital statshot. Режим доступа: https://wearesocial.com/global-digital-report-2019 (дата обращения: 10.01.2020).

2. **Rout S.P.** 5th generation mobile technology – a new milestone to future wireless communication networks // International Journal of Science and Research. 2016. Vol. 5. No. 5. Pp. 529–534.

3. **Kumar A., Gupta M.** A review on activities of fifth generation mobile communication system // Alexandria Engineering Journal. 2018. Vol. 57. No. 2. Pp. 1125–1135.

4. **Baimova J.A., Korznikova E.A., Dmitriev S.V., Liu B., Zhou K.** Review on crumpled graphene: unique mechanical properties// Reviews on Advanced Materials Science. 2014. Vol. 39. No. 1. Pp. 69–83.

5. **Лебедева О.С., Лебедев Н.Г.** Влияние деформаций растяжения и сжатия на пьезорезистивность углеродных нанотрубок и графеновых нанолент // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2014. № 1 (189). С. 26–34.

6. **Елецкий А.В.** Механические свойства углеродных наноструктур и материалов на их основе // Успехи физических наук. 2007. Т. 177. № 3. С. 233–274.

7. **Li Y., Liu S., Sun J., Li S., Chen J., Zhao Y.** Effects of the oxygen content of reduced graphene oxide on the mechanical and electromagnetic interference shielding properties of carbon fiber/reduced graphene oxide-epoxy composites // New Carbon Materials. 2019. Vol. 34. No. 5. Pp. 489 – 498.

8. **Wang X., Jiang H.T., Yang K.Y., Ju A.X., Ma C.Q., Yu X.L.** Carbon fiber enhanced mechanical and electromagnetic absorption properties of magnetic graphene-based film // Thin Solid Films. 2019. Vol. 674. No. 31. Pp. 97–102.

9. **Chen F.C., Chu C.W., He J., Yang Y., Lin J.L.** Organic thin-film transistors with nanocomposite dielectric gate insulator // Applied Physics Letters. 2004. Vol. 85. No. 15. Pp. 3295–3297.

10. **Gusev A.N., Kiskin M.A., Braga E.V., et al.** Novel zinc complex with an ethylenediamine schiff base for high-luminance blue fluorescent OLED applications // The Journal of Physical Chemistry. 2019. Vol. 123. No. 18. Pp. 11850–11859.

11. **Зиминов В.М., Захарова** И.Б. Выпрямляющие свойства структур на основе фуллерена C_{60} // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2012. № 2 (146). С. 18–21.

12. **Gusev A.N., Mazinov A.S., Tyutyunik A.S., Gurchenko V.S.** Spectral and conductive properties of film heterostructures based on fullerene-containing material and 4-methylphenylhydrazone N-isoamilisatine // Radio Electronics, Nanophysics and Information Technologies. 2019. Vol. 11. No. 3. Pp. 331–336.

13. **Konenkamp R., Priebe G., Pietzak B.** Carrier mobilities and influence of oxygen in C_{60} films // Physical Review B. 1999. Vol. 60. No. 16. Pp. 11804–11808.

14. **Tapponnier A., Biaggio I., Gunter P.** Ultrapure C_{60} field-effect transistors and the effects of oxygen exposure // Applied Physics Letters. 2005. Vol. 86. No. 11. P. 112114.

15. **Гусев А.Н., Мазинов А.С., Шевченко А.И., Тютюник А.С., Гурченко В.С., Брага Е.В.** Вольтамперные характеристики и фотоэлектрический эффект гетероструктур фуллерен C_{60} – 4-метилфенилгидразон N-изоамилизатина // Письма в ЖТФ. 2019. Т. 45. № 19. С. 40–43.

16. **Андреев С.М., Пургина Д.Д., Башкатова Е.Н.,** Гаршев **А.В.,** Маерле **А.В., Хаитов М.Р.** Эффективный способ получения водных нанодисперсий фуллерена C_{60} // Российские нанотехнологии. 2014. № 7–8 (9). С. 24–30.

17. **Мазинов А.С., Работягов К.В., Гурченко** В.С., Тютюник **А.С.** Влияние структурных особенностей фуллеренсодержащего материала на его резистивные свойства // Экологический вестник научных центров Черноморского экономического сотрудничества. 2018. Т. 2 № .15. С. 93–86.

18. **Мазинов А.С., Гурченко** В.С., Тютюник **А.С., Шевченко А.И.** Влияние структурных особенностей фуллеренсодержащего материала на его резистивные свойства при осаждении из раствора // Экологический вестник научных центров Черноморского экономического сотрудничества. 2018. Т. 15. № 4. С. 85–92.

19. **Cigan M., Jakusova K., M. Gaplovsky M., Filo J., Donovalova J., Gaplovsky A.** Isatin phenylhydrazones: anion enhanced photochromic behavior// Photochemical and Photobiological Sciences. 2015. Vol. 14. No. 11. Pp. 2064–2073.

20. **Старостенко В.В., Мазинов А.С., Фитаев И.Ш., Таран Е.П.,** Орленсон **В.Б.** Динамика формирования поверхности проводящих пленок алюминия на аморфных подложках // Прикладная физика. 2019. № 4. С. 60–65.

21. **Al-Saidi I., Sadik F.** Synthesis and investigation of phenol red dye doped polymer films // Advances in Materials Physics and Chemistry. 2016. Vol. 6. No. 5. Pp. 120–128.

Статья поступила в редакцию 18.01.2020, принята к публикации 14.02.2020.

СВЕДЕНИЯ ОБ АВТОРАХ

СТАРОСТЕНКО Владимир Викторович – доктор физико-математических наук, заведующий кафедрой радиофизики и электроники Крымского федерального университета имени В.И. Вернадског.
295007, Российская Федерация, Республика Крым, г. Симферополь, пр. Академика Вернадского, 4
starostenkovv@cfuv.ru

МАЗИНОВ Алим Сеит-Аметович – кандидат технических наук, доцент кафедры радиофизики и электроники Крымского федерального университета имени В.И. Вернадского.
295007, Российская Федерация, Республика Крым, г. Симферополь, пр. Академика Вернадского, 4
mazinovas@cfuv.ru

ТЮТЮНИК Андрей Сергеевич – аспирант кафедры радиофизики и электроники Крымского федерального университета имени В.И. Вернадского.
295007, Российская Федерация, Республика Крым, г. Симферополь, пр. Академика Вернадского, 4
real-warez@mail.ru

ФИТАЕВ Ибраим Шевкетович – ведущий специалист кафедры радиофизики и электроники Крымского федерального университета имени В.И. Вернадского.
295007, Российская Федерация, Республика Крым, г. Симферополь, пр. Академика Вернадского, 4
fitaev.i@cfuv.ru

ГУРЧЕНКО Владимир Сергеевич – аспирант кафедры радиофизики и электроники Крымского федерального университета имени В.И. Вернадского.
295007, Российская Федерация, Республика Крым, г. Симферополь, пр. Академика Вернадского, 4
gurchenko_v@mail.ru

DOI: 10.18721/JPM.13110
УДК 519.816

A VECTOR COMPOSED OF MEDICAL PARAMETERS: DETERMINATION OF THE DISTRIBUTION CLASS

V.I. Antonov[1], O.A. Bogomolov[2], V.V. Garbaruk[3], V.N. Fomenko[3]

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

[2]Russian Research Center for Radiology and Surgical Technologies,

St. Petersburg, Russian Federation;

[3]Emperor Alexander I St. Petersburg State Transport University,

St. Petersburg, Russian Federation

In the paper, the authors present a method for determining the distribution class to which a selected random vector with medical parameters as components belongs. The method is based on the statistical significance test. The optimal selection problem for the significance level where the probability of the vector identification error is minimal has been solved. In order to tackle the problem, the authors used the prior information on belonging the vector components to the definite distribution class in which the statistical relationship between the medical parameters was taken into account. The developed mathematical model of patient condition should serve as support of decision-making on further treatment tactics.

Keywords: mathematical simulation, distribution class, significance test, power of test

Citation: Antonov V.I., Bogomolov O.A., Garbaruk V.V., Fomenko V.N., A vector composed of medical parameters: determination of the distribution class, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (1) (2020) 106–113 DOI: 10.18721/JPM.13110

ОПРЕДЕЛЕНИЕ КЛАССА РАСПРЕДЕЛЕНИЯ ВЕКТОРА МЕДИЦИНСКИХ ПОКАЗАТЕЛЕЙ

В.И. Антонов[1], О.А. Богомолов[2], В.В. Гарбарук[3], В.Н. Фоменко[3]

[1]Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Российская Федерация;

[2]Российский научный центр радиологии и хирургических технологий

имени академика А.М. Гранова, Санкт-Петербург, Российская Федерация;

[3] Петербургский государственный университет путей сообщения

Императора Александра I, Санкт-Петербург, Российская Федерация

В статье представлен разработанный авторами метод определения класса распределения, к которому принадлежит выбранный случайный вектор с медицинскими показателями в качестве компонент. Метод основан на статистическом критерии значимости. Решается задача об оптимальном выборе уровня значимости, при котором вероятность ошибки идентификации вектора минимальна. Для этого используется априорная информация о принадлежности компонент вектора к определенному классу распределения, в котором учитывается статистическая зависимость между медицинскими показателями. Разработанная математическая модель состояния пациента должна служить поддержкой принятию решения о выборе дальнейшей тактики лечения.

Ключевые слова: математическое моделирование, класс распределения, критерий значимости, мощность критерия

Ссылка при цитировании: Антонов В.И., Богомолов О.А., Гарбарук В.В., Фоменко В.Н. Определение класса распределения вектора медицинских показателей // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2020. Т. 13. №. С. 121–126. DOI: 10.18721/JPM.13110

Introduction

The goal of this study consisted in constructing a probabilistic model for forecasting medical outcomes of diseases for patients who underwent radical prostatectomy. The model should allow to estimate whether recurrence of the tumor is likely. A database composed of several medical indicators was accumulated for this purpose for groups of patientswho did not suffer recurrence of the tumors, and for those who did. These indicators vary from patient to patient within each group, filling a certain domain in the space of indicators with some density different for the two groups. The system of indicators is combined into a vector, which is regarded as the implementation of a random vector with a distribution law derived from the observed data. This random vector generalizes the experimental data and characterizes the group as a whole. The next step is determining (with a sufficient degree of reliability) whether a vector with the indicators of a particular patient is the implementation of one of the two given random vectors, or, in other words, to which of the two groups the patient most likely belongs to.

We solved this problem using the statistical significance test [1]. One of the two distributions is regarded as the null hypothesis, and the other as the alternative hypothesis. If a random vector falls into the so-called acceptance region, the null hypothesis is accepted. Otherwise, the alternative hypothesis is assumed to hold true. Errors in attributing a vector (classifying it as belonging to a certain probability distribution) by this algorithm can be made in two cases: either the true null hypothesis is erroneously rejected (type I error), or, conversely, the false null hypothesis is erroneously accepted (type II error). Any value (between 0 and 1) can be obtained for the probability of type I errors by choosing an acceptance region. However, changing the probability of type I error also leads to a change in the probability of type II error. Extending the acceptance region obviously reduces the probability of type I errors and increases the probability of type II errors.

Thus, it seems a natural step to choose an acceptance region so as to minimize the probability of type II errors for a given level of significance, that is, the probability of type I error [2, 3].

The problem of choosing an optimal acceptance region in the above sense was solved by introducing the Neyman–Pearson criterion [3]. However, this criterion is used as part of a more general interpretation of the significance test by introducing a certain degree of randomization. As a result, the answer to the question whether the null hypothesis is accepted or rejected is probabilistic.

Practically speaking, the total error of vector attribution by the distribution law is most important. This characteristic consists of two sources: type I and type II errors. If the a priori probabilities of the hypotheses about the distribution law are known, then the probability of the total error can be minimized by choosing an optimal significance level. The above optimization problem is solved in this paper.

The second section of the paper describes a probabilistic model within which we constructed an optimized criterion for attributing a random vector by the distribution law. In the third section, we consider a practical application of this criterion to medical research. Finally, the last section discusses the results obtained and potential options for developing the given method.

Probabilistic model

We consider three-dimensional random vectors with a distribution A or B: $W^{(A)}$ and $W^{(B)}$ in this model. The first two components of the vector are continuous random variables, and the last component takes only the values 0 or 1. The quantities $m_i^{(A)}$, $\sigma_i^{(A)}$, ($I = 1, 2, 3$) are, respectively, the mathematical expectations and standard deviations of the components of the vector $W^{(A)}$. Notations are similar for $W^{(B)}$.

Let $m_i^{(A)}[n]$ be the conditional expectation $W^{(A)}$, ($i = 1, 2$), when $W^{(A)}_3 = n$. We introduce the same notations for conditional standard deviations and covariance of

continuous components. The distribution of the discrete component is given by the quantity $p_n = P\{W_3 = n\}$.

Conditional and unconditional characteristics of continuous components are related by the formulas

$$m_i = \sum_{n=0,1} m_i[n]\, p_n$$

$$\sigma^2{}_i = \sum_{n=0,1} \left(\left(\sigma_i[n]\right)^2 + \left(m_i[n]\right)^2 \right) p_n - m_i{}^2$$

$$\mathrm{Cov}_{12} = \sum_{n=0,1} m_1[n]\, m_2[n]\, p_n - m_1 m_2.$$

The problem solved in this paper is to determine most reliably to which of the distributions (*A* or *B*) the given vector *W* belongs. The significance test is used for this purpose.

Let us call the set

$$\widetilde{D} = \bigcup_{n=0,1} D_n \bigcap \left\{ W_i, i = 1,2,3 \middle| W_3 = n \right\},$$

$$D_n = \left\{ W_i, i = 1,2,3 \middle| x_1^{(n)} \le W_1 \le \right. \qquad (1)$$

$$\left. \le x_2^{(n)} \wedge y_1^{(n)} \le W_2 \le y_2^{(n)} \right\}$$

the acceptance region.

Each of the two values of W_3 has its own range of acceptable values W_1 and W_2. Starting from Eq. (1), we use the symbols $\cup$ and $\cap$ for the operations of union and intersection on sets, the symbol $\wedge$ for conjunction of conditions.

The situation when the vector has the distribution *A* is taken as the null hypothesis H_0. If the vector has the distribution *B*, the alternative hypothesis H_1 is accepted. According to the significance test, if

$$(W_1, W_2) \in \widetilde{D},$$

then hypothesis H_0 is accepted in this and only in this case.

Type I error (erroneously rejecting the null hypothesis) occurs with a probability

$$P_1 = P\left((W_1, W_2) \notin \widetilde{D} \middle| H_0 \right).$$

The probability of type II error (erroneously accepting the null hypothesis) is

$$P_2 = P\left((W_1, W_2) \in \widetilde{D} \middle| H_1 \right).$$

From a practical standpoint, it is preferable to choose the acceptance region so as to obtain the minimum value of P_2 for the given probability P_1, close to zero. Mathematically, the problem is formulated as follows:

$$\min P\left((W_1, W_2, W_3) \in \widetilde{D} \middle| H_1\right) =$$
$$= \min \sum_{n=0,1} p_n{}^{(B)} P\left((W_1, W_2) \in D_n \middle| H_1\right) =$$
$$= \sum_{n=0,1} p_n{}^{(B)} \min P\left((W_1, W_2) \in D_n \middle| H_1\right).$$

Therefore,

$$\left\{x_1{}^{[n]}, x_2{}^{[n]}, y_1{}^{[n]}, y_2{}^{[n]}\right\} = \\ = \underset{x_1{}^{[n]}, x_2{}^{[n]}, y_1{}^{[n]}, y_2{}^{[n]}}{\arg\min} P\left((W_1, W_2) \in D_n \middle| H_1\right), \quad (2)$$

where arg min(f) denotes a function yielding the argument values of $f(x)$ at the minimum point.

We write the expressions for the probabilities of type I and type II errors:

$$\Phi_1{}^{(C)}(n) = F^{(C)}(x_2{}^{(n)}, y_2{}^{(n)})[n] - \\ - F^{(C)}(x_1{}^{(n)}, y_2{}^{(n)})[n],$$

$$\Phi_2{}^{(C)}(n) = F^{(C)}(x_2{}^{(n)}, y_1{}^{(n)})[n] - \\ - F^{(C)}(x_1{}^{(n)}, y_1{}^{(n)})[n],$$

$$P_1 = 1 - \sum_{n=0,1} p_n{}^{(A)} [\Phi_1{}^{(A)}(n) - \Phi_2{}^{(A)}(n)], \\ P_2 = \sum_{n=0,1} p_n{}^{(B)} [\Phi_1{}^{(B)}(n) - \Phi_2{}^{(B)}(n)]. \quad (3)$$

where $F^{(C)}(x,y)[n]$ ($C = A$ or $C = B$) is the conditional function for the distribution of the vector (W_1 and W_2).

Knowing the probability that the vector is attributed erroneously is important for deciding which class, A or B, this random vector belongs to. This probability can be determined if the a priori probability P_A of a vector belonging to class A is known.

Let P_{err} be the probability of erroneous attribution. Then,

$$P_{err} = P_A P_1 + (1 - P_A) P_2. \quad (4)$$

Let $P^{(0)}{}_2(P_1)$ be the probability of type II error, calculated by the optimized algorithm at a significance level P_1. It is natural to set P_1 so that (4) takes the minimum value $P^{(0)}{}_{err}$, i.e.,

$$P_1^{(0)} = \underset{P_1 \in [0,1]}{\arg\min} \left[P_A \cdot P_1 + (1 - P_A) \cdot P_2^{(0)}(P_1) \right]; \\ P_{err}{}^{(0)} = P_A \cdot P_1^{(0)} + (1 - P_A) \cdot P_2^{(0)}(P_1^{(0)}). \quad (5)$$

Table 1

Data set and incidence analysis by patient group

Group of patients	Number of patients		Coefficient of correlation for W_1 and W_2		False attributions (number and total error)		
	$W_3 = 0$	$W_3 = 1$	$W_3 = 0$	$W_3 = 1$	$W_3 = 1$	$W_3 = 1$	Relative error
A (no recurrence)	37	3	0.0058	0,0430	12	0	0.30
B (with recurrence)	33	5	–0.2000	–0.3600	6	1	0.18

Notations: W_1 is the initial PSA level, ng/ml; W_2 is the PSA doubling time, months; W_3 is the surgical margin of the tumor; we assumed that $W_3 = 0$ if there were no abnormal cells, and $W_3 = 1$ otherwise.

Note. The correlation coefficients W_1 and W_2 were found by the formula $R_{XY} = \frac{\mathrm{cov}_{XY}}{\sigma_X \sigma_Y}$.

Table 2

Conditional distributions of continuous components of random vector W_1, W_2

W_3	$p_n^{(A)}$	$p_n^{(B)}$	$m_i^{(A)}$	$m_i^{(B)}$	$\sigma_i^{(A)}$	$\sigma_i^{(A)}$
0	0.925	0.868	12.2; 2200	17.4; 998	10.6; 2410	11.0; 2000
1	0.075	0.132	8.33; 1000	30.9; 265	1.48; 558	20.7; 152
Total value	–	–	11.9; 2110	19.2; 901	10.3; 2350	13.5; 1870

Notations: p_n are the distributions of the discrete component; m_i is the mathematical expectation; σ_i is the standard deviation; the superscripts correspond to the data belonging to patient groups *A* and *B*. Two values correspond to the components W_1 and W_2.

Example application of the attribution algorithm to medical data

The above-described algorithm for attributing random vectors was applied to data for urologic oncology patients who underwent tumor removal surgery. Prostate cancer is considered the most commonly diagnosed cancer in men and the second (according to statistical data) cause of death from cancer [12]. The level of prostate-specific antigen *p* (PSA) in blood serum [5, 6], measured in ng/ml, closely correlates with the volume of the tumor. The tumor's growth rate is characterized by the PSA doubling time [7, 8].

Initially, there were two groups of patients with different outcomes of radical prostatectomy. Each patient was characterized by individual values of preoperative and postoperative factors [9–12]. The array of patients was divided into two groups: tumor recurrence was detected in 33 patients, and no recurrence was observed in 37. Predicting options for further treatment after surgery is an important task, since it affects the final result of radical prostatectomy [9, 13–15].

We selected a total of three factors:

W_1 is the initial PSA, ng/ml;

W_2 is the PSA doubling time, months;

W_3 is the surgical margin of the tumor, i.e., whether any cancer cells are found in the resection line. We assumed that $W_3 = 0$ if these cells were not found, and $W_3 = 1$ otherwise.

Group *A* included patients who did not have recurrences for a certain period of time, and group *B* included patients with recurrences. Table 1 shows the number of patients in groups.

The quantities W_1 and W_2 in group *B* have a noticeable correlation. Table 1 gives the estimates for the correlation coefficients with $W_3 = 0$ and 1.

Table 2 gives the main characteristics of the distributions *A* and *B*.

Let us explain how we constructed the two-dimensional conditional (i.e., with a fixed value of W_3) distribution function of the random vector W_1, W_2 required to calculate the probabilities of type I and type II errors

Let

$$x_k, y_k; (k = \overline{1, N}), \tag{6}$$

be conditional samples of continuous components W_1 and W_2, respectively, arranged in ascending order. Next, let the points

$$(x_{i(j)}, y_{k(j)}); (j = \overline{1, N}) \tag{7}$$

represent the experimental data. Let us

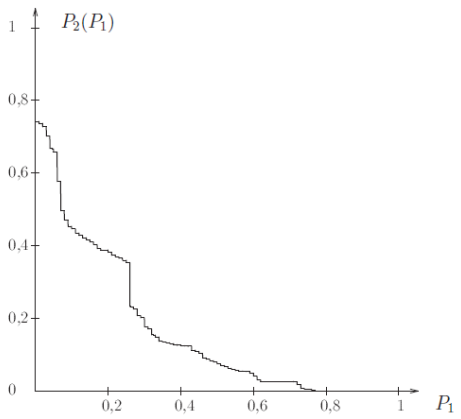

Fig. 1. Probability of type II error as function of probability of type I error

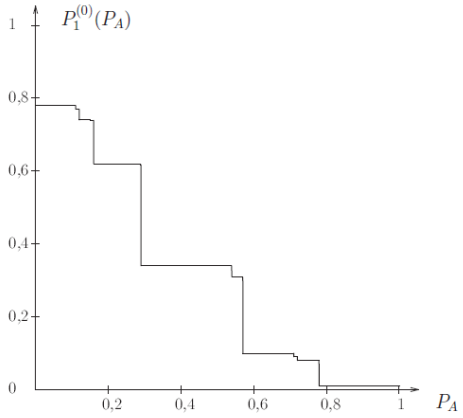

Fig. 2. Optimal significance level as function of a priori probability

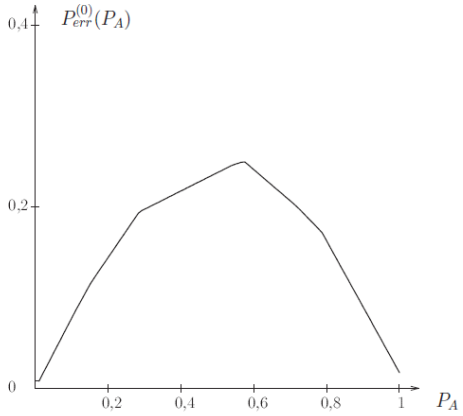

Fig. 3. Minimized attribution error

introduce the notations

$$\begin{aligned}
&\xi_0 = 2x_1 - x_2;\\
&\xi_i = 0.5(x_i + x_{i+1});\ (i = \overline{1, N-1});\\
&\xi_N = 2x_N - x_{N-1};\\
&\eta_0 = 2y_1 - y_2;\\
&\eta_i = 0.5(y_i + y_{i+1}),\ (i = \overline{1, N-1});\\
&\eta_N = 2y_N - y_{N-1}.
\end{aligned} \tag{8}$$

To construct the distribution function, let us divide the rectangle

$$\left[\xi_0, \xi_N; \eta_0, \eta_N\right] \tag{9}$$

into N^2 rectangles of the form

$$\left[\xi_{i-1}, \xi_i; \eta_{k-1}, \eta_k\right];\ \left(i = \overline{1, N}; k = \overline{1, N}\right). \tag{10}$$

Next, let us select from all the rectangles those containing the experimental points (7) and combine them into a set S_e:

$$\left[\xi_{i(j)-1}, \xi_{i(j)}; \eta_{k(j)-1}, \eta_{k(j)}\right];\ (j = \overline{1, N});$$

$$S_e = \bigcap_{i=1}^{N} \left[\xi_{i(j)-1}, \xi_{i(j)}; \eta_{k(j)-1}, \eta_{k(j)}\right].$$

We assume that the random vector W_1, W_2 is evenly distributed inside each of the N rectangles, and the probability of the random vector falling into each of the rectangles is the same and equal to $1/N$. This probability is equal to zero for all other rectangles. The distribution density then has the form

$$\rho(x,y)=\begin{cases}\dfrac{1}{N\cdot\Delta\xi_i\cdot\Delta\eta_k},\\ \text{if } (x,y)\in\\ \in[\xi_{i-1},\xi_i;\eta_{k-1},\eta_k]\subset S_e;\\ 0,\\ \text{if } (x,y)\notin S_e,\end{cases}\quad(12)$$

where $\Delta\xi_i=\xi_i-\xi_{i-1};\quad\Delta\eta_k=\eta_k-\eta_{k-1}.$

In accordance with Eq. (12), the conditional distribution function is an inhomogeneous piecewise bilinear function:

$$F(x,y)=a_{i,k}+b_{i,k}(x-\xi_{i-1})+c_{i,k}(y-\eta_{k-1})+$$
$$+d_{i,k}(x-\xi_{i-1})(y-\eta_{k-1}),\quad(13)$$
$$\text{if } (x,y)\in[\xi_{i-1},\xi_i;\eta_{k-1},\eta_k],$$

where the parameters are obtained from continuity condition $F(x,y)$ and the boundary conditions

$$F(\xi_0,\,y)=F(x,\,\eta_0)=0$$

using recurrence relations

$$b_{i,k}=b_{i,k-1}+d_{i,k-1}\Delta\eta_{k-1},$$
$$c_{i,k}=c_{i-1,k}+d_{i-1,k}\Delta\xi_{i-1},$$
$$a_{i,k}=a_{i-1,k-1}+b_{i-1,k-1}\Delta\xi_{i-1}+$$
$$+c_{i-1,k-1}\Delta\eta_{k-1}+d_{i-1,k-1}\Delta\xi_{i-1}\Delta\eta_{k-1},$$
$$d_{i,k}=\begin{cases}\dfrac{1}{N\cdot\Delta\xi_i\cdot\Delta\eta_k},\\ \text{if } [\xi_{i-1},\,\xi_i;\eta_{k-1},\,\eta_k]\subset S_e;\\ 0,\\ \text{if } [\xi_{i-1},\,\xi_i;\eta_{k-1},\,\eta_k]\notin S_e,\end{cases}\quad(14)$$

with $a_{1,k}=a_{i,1}=b_{1,k}=c_{i,1}=0.$

Fig. 1 shows the dependence of probability of type II error with the optimal acceptance region (1) chosen by Eq. (2). Fig. 2 shows the dependence for the optimal significance level for the given a priori data on whether a patient belongs to group *A*, and Fig. 3 shows the probability for the total error of patient attribution (see Eq. (5).

We applied the attribution algorithm to groups *A* and *B*. Table 1 (right columns) gives the number of errors in determining the group to which the patient belongs. We assumed that a priori probability is $P_A = 0.5$, since the number of patients in both groups is approximately the same. We should also note that the attribution error is close to the maximum value of 0.25 in this case, which can be seen from Fig. 3. The data in Table 1 (false attributions) indicate that the actual total attribution error is close to this estimate.

Conclusion

Example application of the proposed significance test confirms that it can be used in practice, in particular in medicine for predicting complications. Evidently, the probability of error in determining the class to which the given object belongs decreases with increasing number of patients with a known diagnosis.

We should note that the algorithm constructed in this paper is optimal only in the class of significance tests with a connected acceptance region (see Eq. (1)). However, if the distribution has a more complex shape, for example, with a multimodal distribution density, choosing a disconnected acceptance region could produce a more powerful test.

Including a greater number of continuous variables in the test would increase the reliability of the algorithm. However, expanding the number of variables would also make finding the optimal acceptance region more difficult.

REFERENCES

1. **Kendall M.G., Stuart A.,** *Design* and analysis, and time series, Charles Griffin & Co. Ltd., London, 1966.

2. **Brandt S.,** Statistical and computational methods in data analysis, North-Holland Publishing Company, Amsterdam, 1970.

3. **Sevastianov B.A.,** Kurs teorii veroyatnostey i matematicheskoy statistiki [The course of probability theory and mathematical statistics], Nauka, Moscow, 1982. (In Russ.)

4. **Antonov V.I., Blagoveshchenskaya E.A., Bogomolov O.A., et al.,** The exponential model of cell growth: a simulation error, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 11 (3) (2018) 90–98. (In Russ.)

5. **Watt K.W., Lee P.J., M'Timkulu T., et al.,** Human prostate-specific antigen: structural and functional similarity with serine proteases, Proc. Natl. Acad. Sci. USA. 83 (10) (1986) 3166–3170.

6. **Pushkar' D.Yu., Govorov A.V.,** Prostate cancer markers, Experimental and Clinical Urology. (2–3) (2011) 19–21. (In Russ.)

7. **Zharinov G.M., Bogomolov O.A.,** The pretreatment prostate-specific antigen doubling

time: clinical and prognostic values in patients with prostate cancer, Cancer Urology. 10 (1) (2014) 44–48. (In Russ.)

8. **Roberts S.G., Blute M.L., Bergstralh E.J., et al.,** PSA doubling time as a predictor of clinical progression after biochemical failure following radical prostatectomy for prostate cancer, Mayo Clin. Proc. 76 (6) (2001) 576–581.

9. **Bogomolov O.A., Shkolnik M.I., Zharinov G.M.,** The preoperative kinetics of prostate-specific antigen as a predict of relapse-free survival after radical prostatectomy, Cancer Urology. 10 (4) (2014) 47–51. (In Russ.)

10. **Astanin S.A., Kolobov A.V., Lobanov A.I., et al.,** Vliyaniye prostranstvennoy geterogennoy sredy na rost i invaziyu opukholi. Analiz metodami matematicheskogo modelirovaniya [The influence of the spatial heterogeneous medium on a tumor growth and invasion. An analysis by mathematical simulation methods], In: "Medicine in Informatics", Nauka, Moscow (2008) 188–223.

11. **Bezrukov E.A., Lachinov E.L., Martirosyan G.A.,** Factors affecting local reccurence after radical proststectomy, Bashkortostan Medical Journal, Scientific Publication. 10 (3) (2015) 203–205. (In Russ.)

12. **Han M., Partin A.W., Zahurak M., et al.,** Biochemical (prostate specific antigen) recurrence probability following radical prostatectomy for clinically localized prostate cancer, J. Urol. 169 (2) (2003) 517–523.

13. **Bogomolov O.A., Garbaruk V.V., Zhuykov V.N., Tikhomirov S.G**. Pattern recognition in medical diagnostics, Proceedings of the V International Scientific and Methodological Conference. Problems of Mathematical and Natural-Scientific Preparation in Engineering Education, St. Petersburg (2018) 39–41. (In Russ.)

14. **Benzekry S., Lamont C., Beheshti A., et al.,** Classical mathematical models for description and prediction of experimental tumor growth, PLOS Comput. Biol. 10 (8) (2014), e1003800. DOI: 10.1371/journal.pcbi.1003800.

15. **Williams M.J., Werner B., Barnes C.P., et al.,** Identification of neutral tumor evolution across cancer types, Nature Genetics. 48 (3) (2016) 238–244.

Received 21.10.2019, accepted 19.12.2019.

THE AUTHORS

ANTONOV Valeriy I.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
antonovvi@mail.ru

BOGOMOLOV Oleg A.
Russian Research Center for Radiology and Surgical Technologies
70 Leningradskaya St., St. Petersburg, Pesochniy Settl., 197758, Russian Federation
urologbogomolov@gmail.com

GARBARUK Victor V.
Emperor Alexander I St. Petersburg State Transport University
9 Moskovsky Ave., St. Petersburg, 190031, Russian Federation
vigarb@mail.ru

FOMENKO Victor N.
Emperor Alexander I St. Petersburg State Transport University
9 Moskovsky Ave., St. Petersburg, 190031, Russian Federation
vfomenko1943@gmail.com

СПИСОК ЛИТЕРАТУРЫ

1. **Кендалл М., Стюарт А.** Многомерный статистический анализ и временные ряды. Пер. с англ. М.: Наука, 1976. 736 с.

2. **Брандт З.** Статистические методы анализа наблюдений. Пер. с англ. М.: Мир, 312 .1975 с.

3. **Севастьянов Б.А.** Курс теории вероятностей и математической статистики. М.: Наука, 256 .1982 с.

4. **Антонов В.И., Благовещенская Е.А., Богомолов О.А., Гарбарук В.В., Яковлева Ю.Г**. Погрешность экспоненциальной модели роста клеток // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2018. Т. 11. № 3. С. 90–97.

5. **Watt K.W., Lee P.J., M'Timkulu T., Chan W., Loor R.** Human prostate-specific antigen: structural and functional similarity with serine proteases // Proc. Natl. Acad. Sci. USA. 1986. Vol. 83. No. 10. Pp. 3166–3170.

6. **Пушкарь Д.Ю., Говоров А.В.** Маркеры рака предстательной железы // Экспериментальная и клиническая урология. 2 № .2011–3. С. 19–21.

7. **Жаринов Г.М., Богомолов О.А.** Исходное время удвоения простат- специфического антигена: клиническое и прогностическое значение у больных раком предстательной железы // Онкоурология. 2014. Т. 10. № 1. С. 44–48.

8. **Roberts S.G., Blute M.L., Bergstralh E.J., Slezak JM, Zincke H.** PSA doubling time as a predictor of clinical progression after biochemical failure following radical prostatectomy for prostate cancer // Mayo Clin. Proc. 2001. Vol. 76. No. 6. Pp. 576–581.

9. **Богомолов О.А., Школьник** М.И., Жаринов Г.М. Предоперационная кинетика простатспецифического антигена как фактор прогноза безрецидивной выживаемости после радикальной простатэктомии // Онкоурология. 2014. Т. 4 № .10. С. 51–47.

10. **Астанин С.А., Колобов А.В., Лобанов А.И., Пименова Т.П., Полежаев А.А., Соляник Г.И.** Влияние пространственной гетерогенной среды на рост и инвазию опухоли. Анализ методами математического моделирования // Медицина в зеркале информатики. М.: Наука, 2008. С. 223–188.

11. **Безруков Е.А., Лачинов Э.Л., Мартиросян Г.А.** Факторы местного рецидива после радикальной простатэктомии // Медицинский вестник Башкортостана. 2015. Т. 10. № 3. С. 203–205.

12. **Han M., Partin A.W., Zahurak M., Piantadopi S., Epstein J., Walsh P.S.** Biochemical (prostate specific antigen) recurrence probability following radical prostatectomy for clinically localized prostate cancer // J. Urol. 2003. Vol. 169. No. 2. Pp. 517–523.

13. **Богомолов О.А., Гарбарук В.В., Жуйков В.Н., Тихомиров С.Г.** Распознавание образов в медицинской диагностике // Проблемы математической и естественно-научной подготовки в инженерном образовании. Сб. трудов V Международной научно-методической конференции. СПб, 2018. С. 39–41.

14. **Benzekry S., Lamont C., Beheshti A., Tracz**

A., Ebos J.M., Hlatky L., Hahnfeldt P. Classical mathematical models for description and prediction of experimental tumor growth // PLOS Comput. Biol. 2014. Vol. 10. No. 8 (August). e1003800. DOI: 10.1371/journal.pcbi.1003800.

15. **Williams M.J., Werner B., Barnes C.P., Graham T.A., Sottoriva A.** Identification of neutral tumor evolution across cancer types // Nature Genetics. 2016. Vol. 48. No. 3. Pp. 238–244.

Статья поступила в редакцию 21.10.2019, принята к публикации 19.12.2019.

СВЕДЕНИЯ ОБ АВТОРАХ

АНТОНОВ Валерий Иванович – *доктор технических наук, заведующий кафедрой высшей математики Санкт-Петербургского политехнического университета Петра Великого.*
195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
antonovvi@mail.ru

БОГОМОЛОВ Олег Алексеевич – *кандидат медицинских наук, научный сотрудник отделения оперативной онкоурологии Российского научного центра радиологии и хирургических технологий имени академика А.М. Гранова.*
197758, Российская Федерация, г. Санкт-Петербург, пос. Песочный, Ленинградская ул., 70
urologbogomolov@gmail.com

ГАРБАРУК Виктор Владимирович – *кандидат технических наук, профессор кафедры высшей математики Петербургского государственного университета путей сообщения Императора Александра I.*
190031, Российская Федерация, г. Санкт-Петербург, Московский пр., 9
vigarb@mail.ru

ФОМЕНКО Виктор Николаевич – *доктор физико-математических наук, профессор кафедры высшей математики Петербургского государственного университета путей сообщения Императора Александра I.*
190031, Российская Федерация, г. Санкт-Петербург, Московский пр., 9
vfomenko1943@gmail.com