THE MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY

JOURNAL

Physics and Mathematics

VOLUME 13, No.4, 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

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ

САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Физико-математические науки

ТОМ 13, №4

2020

Санкт-Петербургский политехнический
университет Петра Великого
2020

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

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

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

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

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

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

Журнал индексируется в базе данных **Web of Science** (Emerging Sources Citation Index), а также включен в базу данных **«Российский индекс научного цитирования»** (РИНЦ), размещенную на платформе Научной электронной библиотеки на сайте

http://www.elibrary.ru

Адрес редакции и издательства:

Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29.

Тел. редакции (812) 294-22-85.

http://ntv.spbstu.ru/physics

Contents

Condensed matter physics

Simulation of physical processes

Physical electronics

Physical optics

Physical materials technology

Radiophysics

Mathematics

DOI: 10.18721/JPM.13401

AN EXCITON IN THE SUPERSTRONG AND HYPERSTRONG MAGNETIC FIELDS

S.A. Vaganov, A.G. Panfilov, R.P. Seisyan

Ioffe Physical-Technical Institute of the Russian Academy of Sciences,

St. Petersburg, Russian Federation

The transformation of excitonic states in the raising magnetic field has been investigated beginning with the diamagnetic exciton (strong field) through excitonic magnetopolymer (superstrong field) to the QEL exciton (hyperstrong magnetic field). The effects induced by this transformation were considered. The capability of the exciton as a model object to study the hydrogen and positron atoms in superstrong and hyperstrong magnetic fields was also analysed.

Keywords: diamagnetic exciton, excitonic magnetopolymer, superstrong magnetic field, quantum electromagnetic limit, hyperstrong magnetic field

Citation: Vaganov S.A., Panfilov A.G., Seisyan R.P., An exciton in the superstrong and hyperstrong magnetic fields, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 7–20. DOI: 10.18721/JPM.13401

Introduction

The Wannier–Mott exciton (large-radius exciton) is one of the most important objects of semiconductor physics. An immense number of studies and reviews are dedicated to this hydrogen-like formation consisting of an electron and a hole (see, for example, monographs [1, 2] and references therein).

The behavior of excitons in various semiconductor crystals, nano- and heterostructures has been described in detail for magnetic and/or electric fields, under intense laser excitation, etc. [1, 3].

The effect of a weak magnetic field on the Wannier–Mott exciton has been the subject of thorough theoretical and experimental studies. Substantial data have also been accumulated on the Zeeman effect (splitting of exciton energy levels, linear with the field) and the diamagnetic shift (states shifting to higher energies, proportionally to the squared field and radius of the corresponding exciton state).

The transition to strong field effects, such as Landau quantization and the production of diamagnetic excitons (DE) was considered in different semiconductor crystals and low-dimensional structures (see [2] for more details).

Thus, it can be argued that the behavior of the exciton in both weak and moderately strong magnetic fields is fairly well-understood. It would be therefore interesting to carry out experiments answering the following questions:

whether any new magnetic field effects are observed in exciton spectra in stronger fields;

which fields with the magnetic flux density $B > B_{ss}$ can be regarded as superstrong (B_{ss} here and below refers to the magnetic flux density of a superstrong magnetic field), radically changing the behavior of the exciton.

Since the Wannier–Mott excitons are atom-like, it would be natural to look for answers in atomic physics. However, direct comparisons cannot be drawn here, since the 'hydrogen-like' exciton is not completely similar to the hydrogen atom. The importance of this distinction is analyzed in in the first half of this study. The presence of a medium and a small reduced mass shift exciton effects by two to four orders of magnitude on the energy scale and by three to five orders of magnitude on the scale of magnetic fields relative to atomic ones. The role of the spin-orbit interaction turns out to be less significant, while the possibility of exciton production by light (or by electromagnetic field) and its short radiative lifetime in a crystal add important nuances to exciton behavior.

The second half of the study considers the hypothesized exciton effects in super- and

hyperstrong magnetic fields. The present interest towards this problem is due to a number of circumstances.

On the one hand, considerable experimental progress has been achieved over the past 20–30 years: unique laboratory facilities have been created in Russia and in the world for obtaining super-strong magnetic fields [4]; at the same time, band and exciton parameters of many semiconductor materials (including new ones) were established or refined [5].

On the other hand, an obstacle faced by theoretical studies in magnetic-field atomic physics is that direct and even indirect experiments are impossible; for this reason, the fundamentals introduced half a century ago [6–9] saw no further development or conceptual modifications (see the review in [10] and references therein). Using increasingly powerful powerful mathematical methods (for example, second quantization) and exact analytical solutions practically does not alter the results obtained earlier.

Without attempting to consider theoretical models of atomic interactions in the magnetic field (since only experiments can ultimately confirm them), we believe that it is reasonable to apply these concepts to physics of atom-like quasiparticles in solids, i.e., excitons.

This paper is aimed at summarizing the data currently accumulated, making predictions for the new effects that can be expected to be observed in exciton optics and the conditions potentially favoring these effects. In particular, this concerns specific properties of semiconductor materials (existing, and most importantly prospective), as well as the magnitude of superstrong magnetic fields.

Effect of the medium. Scaling of magnetic-field effects

In contrast to an isolated hydrogen atom, the exciton exists in a medium with the dielectric constant $\varepsilon > 1$, reducing the Coulomb interaction by a factor of ε, and the exciton binding energy by a factor of ε^2 times. Screening of the Coulomb interaction also increases the exciton radius by a factor of ε, which determines many key properties of the exciton, for example, size quantization in nanostructures.

Furthermore, since the electron is not a particle with the mass m_e in a crystal, but rather a quasiparticle with an effective mass m^*_e, defined as

$$m^*_e = (2/\hbar^2)\, d^2E/dk^2$$

near the extremum on the dependence $E(k)$ for the corresponding conduction band, and the effective mass m^*_h of the hole, i.e., the 'nucleus' of a hydrogen-like exciton, is much smaller than the mass of a proton, the reduced exciton mass

$$\mu = m^*_e m^*_h/(m^*_e + m^*_h)$$

is much smaller than m_e, in contrast to the reduced mass of a hydrogen atom $\mu_H \cong m_e$.

Consequently, the Bohr radius of the exciton

$$a_{ex} = a_0\varepsilon/(\mu/m_e) \quad (1)$$

is 10–100 times larger than the Bohr radius of the hydrogen atom ($a_0 \cong 0.53$ Å), and the exciton binding energy

$$Ry_{ex} = Ry\,(\mu/m_e)/\varepsilon^2 \quad (2)$$

is 100–1000 times lower than the ionization energy of the hydrogen atom $Ry \cong 13.6$ eV.

Additionally, the mass of the nucleus for hydrogen and muonium atoms (proton p^+, mu meson μ^+) is much larger than the electron mass, and therefore the contribution of the magnetic moment of the nucleus to the energy of interaction of these atoms with the magnetic field can be neglected:

$$E_M(B) = e\hbar B/m_e + e\hbar B/m_{p,\mu^+} \cong e\hbar B/m_e. \quad (3)$$

However, the contribution of the nucleus cannot be neglected for the exciton (as well as for the positronium atom), and the total energy of interaction of the exciton with the magnetic field becomes higher than that of the atom by a factor of m_e/μ:

$$E_M(B) = e\hbar B/m^*_e + e\hbar B/m^*_h = e\hbar B/\mu. \quad (4)$$

Summarizing all these characteristics, we find that the value of the magnetic flux density B_{Ryex} at which the interaction energy $E_M(B)$ between the magnetic moment of the exciton and the field becomes equal to the binding energy Ry_{ex} turns out to be lower by a factor of $\varepsilon^2(\mu/m_e)^{-2}$ for the exciton than for the hydrogen atom:

$$B_{Ryex} = \mu Ry_{ex}/e\hbar = \\ = m_e Ry\,(\mu/m_e)^2/\varepsilon^2 = B_{Ry}\cdot(\mu/m_e)^2/\varepsilon^2, \quad (5)$$

with the difference reaching 3–5 orders of magnitude.

The values of the corresponding masses, binding energies and characteristic magnetic fields B_{Ry} for hydrogen-like atoms and fields B_{Ryex} for excitons in various semiconductor crystals are given in Tables 1 and 2. The values for excitons in Table 2 are approximate rather

Table 1

Values of basic physical quantities for hydrogen and hydrogen-like atoms

Atom	Nucleus	m_{nuc}/m_e	μ/m_e	Ry, eV	B_{Ry}, kT
Hydrogen H	Proton p^+	1836	1.000	13.6	117.5
Muonium Mu	μ meson μ^+	207	0.995	13.5	116.5
Positronium Ps	Positron e^+	1	0.500	6.8	29.0

Notations: m_{nuc}, m_e, μ are the masses of the nucleus, electron and muon, respectively; Ry is the binding energy, B_{Ry} is the corresponding characteristic magnetic field.

Table 2

Values of basic physical quantities for exciton in various crystals

Crystal	m^*_h/m_e	m^*_e/m_e	μ/m_e	a_{ex}, Å	Ry_{ex}, meV	B_{Ryex}, T
ZnS	1.76	0.340	0.285	16.5	49.00	120
ZnO	0.59	0.280	0.190	22.0	42.50	70
ZnSe	0.78	0.160	0.130	29.0	35.00	39.0
GaN	0.80	0.200	0.160	31.0	25.00	35.0
CdS	0.68	0.210	0.160	31.0	25.00	34.0
ZnTe	0.60	0.120	0.100	46.0	18.00	15.5
CdSe	0.45	0.110	0.088	61.0	11.50	8.70
CdTe	0.63	0.096	0.083	65.0	11.00	7.80
InP	0.60	0.079	0.070	95.0	6.00	3.60
GaAs	0.50	0.063	0.056	125	4.40	2.10
GaSb	0.28	0.041	0.036	231	2.00	0.62
InAs	0.41	0.024	0.023	350	1.35	0.27
HgTe	0.32	0.031	0.028	400	0.86	0.21
InSb	0.42	0.014	0.014	673	0.60	0.07

Notations: m^*_h, m^*_e are the effective masses of the hole and the electron, respectively; a_{ex} is the Bohr radius of the exciton; Ry_{ex} is the binding energy of the hole and the electron in the exciton, B_{Ryex} is the corresponding characteristic magnetic field.

than reference. The differences reach 10–15% in original studies; preference is given to self-consistent data: all values in Eq. (2) can be obtained experimentally, and they must be consistent, which is not observed in a number of publications.

The material medium in which the exciton exists 'scales' the magnetic-field effects: the action of the field B on the exciton is equivalent to the action of a much larger field,

$$B^* = \varepsilon^2/\mu^2\, B,$$

on the hydrogen atom. As a result, the magnetic fields available in laboratories, equal to approximately 1 T, allow to simulate the behavior of hydrogen atoms in fields of 10^3–10^5 T in studies of excitons in semiconductor crystals.

There is another option for further reducing the "model" fields. Since optical methods are primarily employed to study diamagnetic excitons, the study of excited exciton states is the most informative. Applying mathematical processing to magnetoabsorption oscillations arising in the spectrum provides information about both the Landau levels and the parameters of the diamagnetic exciton in a specific material at a given field value [2]. The most pronounced oscillations correspond in this case to fieldless exciton states with the principal quantum number n = 3–5. At the same time, since there is vritually no difference in the symmetry and other properties of the states such as, for example, 1S, 2S, 3S, etc., another scaling is introduced: the effect of the

magnetic field $B_{Ryex,n}$ on the exciton state nS is equivalent to the effect of a larger field

$$B_{Ryex,n=1} = n^2 B_{Ryex,n}$$

for state $1S$.

Similar scaling takes place for states nP, nD, and so on.

Summarizing the above, we can assume that the behavior of the exciton state $4S$, for example, in CdSe crystals in a magnetic field of 3–4 T (it is not at all a record) can be used to obtain information on the behavior of the ground state $1S$ of hydrogen atoms in megatesla-level magnetic fields that are superstrong for a hydrogen atom. Fig. 1 illustrates this point for a CdSe crystal (we previously published less detailed diagrams in [11]). It is evident that, for example, for exciton states n= 3, 4, the transition from the behavior characteristic for weak fields (Zeeman splitting, the energy shift of lines is quadratic (diamagnetic) with the field) to the behavior characteristic for strong fields (multiplet splitting, the shift is linear with the field (cyclotron)) occurs in cadmium selenide in fields noticeably smaller than B_{Ryex} = 9 T.

Thus, laboratory studies of the exciton using scaling of the magnetic field make it possible to simulate the behavior of the hydrogen atom in fields unattainable on Earth: fields of hundreds of thousands of Tesla can arise only in the vicinity of exotic objects of the Universe, such as neutron stars; constant magnetic fields up to 40 T (and pulsed fields up to 1500 T) can be generated in special laboratories.

In 1999, the State Prize of Russia in Science and Technology was awarded to physicists from the All-Russian Research Institute of Experimental Physics (Sarov, Russia) for their work on explosive generation of superstrong magnetic fields and solid state studies in magnetic fields of the 10-MG range [4]. A record field of 2800 T was reached, and the maximum field reached by the setup for optical experiments was 800 T.

There is another notable circumstance. The criterion for a strong field in atomic physics is the ratio between the interaction energy of the magnetic moment of an electron with an external magnetic field $E_M(B)$ and the energy of the spin-orbit interaction Δ_{SO}. The transition from a weak magnetic field to a strong one is accompanied by a transition from the Zeeman effect to the Paschen–Back effect in experiments. The criterion for a superstrong magnetic field B_{Ry} for a hydrogen atom is the ratio of the same interaction energy of the electron

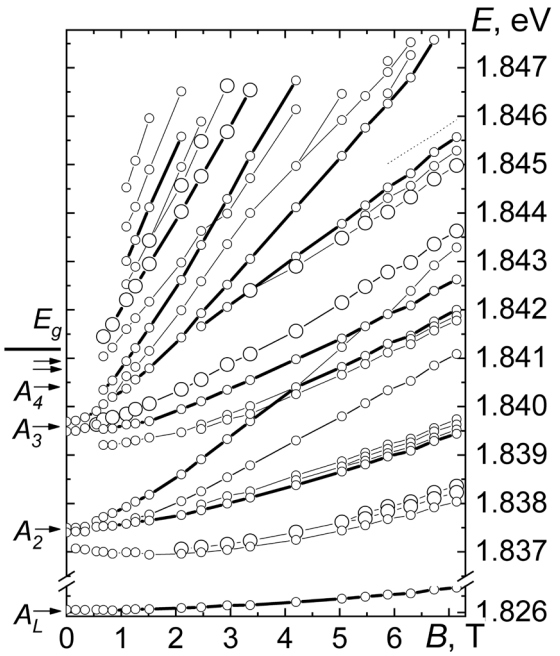


Fig. 1. Fan diagram of exciton states (series A) in a CdSe crystal in an external magnetic field $H \perp c$, $H \perp k$. Polarization $\boldsymbol{E} \parallel \boldsymbol{c}$.
The thicknesses of the lines and the radii of the circles reflect the intensities and widths of the absorption spectral lines, respectively

magnetic moment with an external field $E_M(B)$, and the binding energy of an electron with a proton in a hydrogen atom, Ry (the total potential energy of the hydrogen atom in the ground state, 1 Hartree = $2Ry$, is often considered instead of Ry). Because the ionization energy of the hydrogen atom in the ground state is four orders of magnitude higher than the energy of the spin-orbit interaction, fields over 100,000 T are already superstrong for the hydrogen atom. Table 1 shows the values of B_{Ry} for hydrogen, for the case $E_M(B_{Ry}) = Ry$ and for hydrogen-like atoms with the corresponding values of the binding energy.

As a rule, $Ry_{ex} << \Delta_{SO}$ for an exciton, i.e., an exciton in a 'superstrong magnetic field' (from an 'atomic' standpoint) is essentially the same as in a 'strong magnetic field' (a diamagnetic exciton). The behavior of the DE is already sufficiently well understood, at least for magnetic fields that are transitional from weak to strong [2].

Since the most interesting features of the behavior of hydrogen-like atoms in superstrong magnetic fields (see below) are manifested in the range B = 10–1000B_{Ry}, in terms of

our methodology, it is convenient to use the term 'superstrong field' for hydrogen-like muonium, positronium and exciton provided that $B > B_{ss} \approx 10\ B_{Ry}(B_{Ryex})$.

Hydrogen and hydrogen-like atoms in a superstrong field

Only theoretical representations exist for the behavior and transformations of hydrogen atoms in magnetic fields above 10^6 T ('super-strong field') [6–8]. A consistent point of different models is that there is a geometrical rearrangement of the atom: it is initially spherical, subsequently contracting along the axes perpendicular to the direction of the magnetic field and turning into an ellipsoid (spindle or cigar); see Fig. 4. The ionization potential of a hydrogen atom presumably increases to thousands of electron volts in gigatesla fields, i.e., grows by several orders of magnitude [6, 8].

It is worth repeating that the dependences obtained in [6] and [8] are almost identical, even though more than 20 years have passed between the publication of these studies.

Fig. 2 shows how the binding energy of excitons Ry_{ex} would change in magnetic fields lower than the atomic ones by many orders of magnitude. The given dependences are not the result of exact calculations: here we only assume that the dependence $Ry_{ex}(B)$ and the theoretically predicted dependence $Ry(B)$ of atomic physics considering magnetic fields [6] are similar (i.e., have the form $\sim \ln^2(B/B_{Ry})$). For convenience,, only a few crystals are shown: from wide-gap semiconductors to narrow-gap ones.

It is even less important to refine the dependences $Ry_{ex}(B)$ than in atomic physics: although the changes shown in Fig. 2 are radical (the exciton binding energy should change by several times in experimentally achievable magnetic fields), it is still difficult to observe them experimentally. As as a matter of fact, these changes occur as the electronic spectrum is rearranged, with Landau levels forming for carriers of both types. In this case, the dependence of the energy of the Landau levels on the magnetic field is linear:

$$E_{L,n}(B) = E_M(B)\cdot(n_L + 1/2), \qquad (6)$$

and it is much stronger than the logarithmic dependence of the binding energy of an exciton (in this case, already diamagnetic) on the field, so the dependence $Ry_{ex}(B)$ is actually imperceptible against $E_{L,n}(B)$.

Considering the behavior of several levels at once will hardly help explain the changes

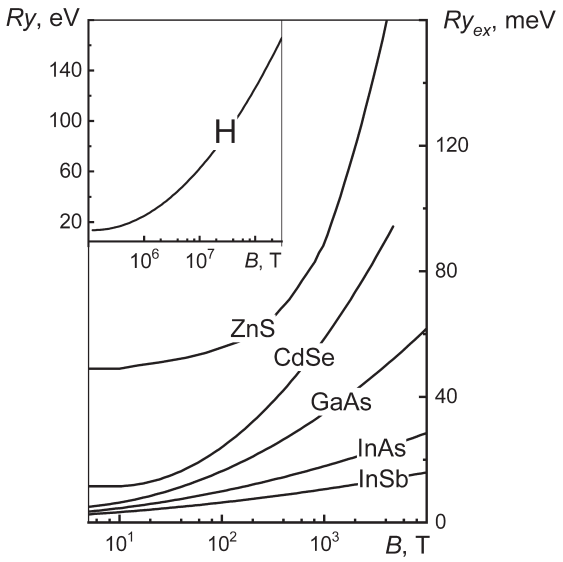


Fig. 2. Dependences of exciton binding energy on the magnetic field in various crystals (the inset shows a similar dependence for a hydrogen atom)

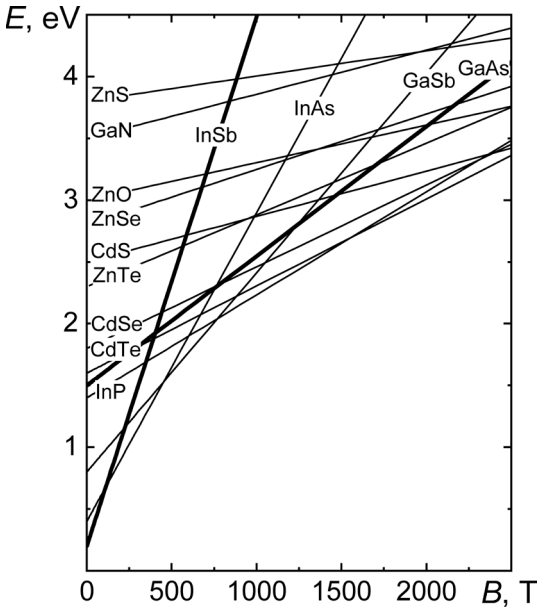


Fig. 3. Dependences of transition energy between the ground ('zero') Landau levels of an electron and a hole on the magnetic field (several dependences are highlighted for clarity)

occurring: the Landau levels are equidistant only in the first approximation. Furthermore, the binding energy of the exciton attached to each specific Landau level depends on the value of n_L [2]. Consider, for example, indium antimonide (InSb), one of the first materials studied (the effect of oscillations of magnetoabsorption in InSb was first analyzed

in detail back in the 1950s–1960s): as the superstrong magnetic field B_{ss} amounts to several Tesla due to the extremely low binding energy of the exciton, the magnetic field dependences in InSb characterizing the energy of the diamagnetic exciton levels in fields up to 8 T (i.e., about 100 B_{Ryex}) were described without relying on the concept of increasing the exciton binding energy in such fields [2].

Interestingly, if the quantity E_M from Eq. (4) is used in Eq. (6), it is evident (Fig. 3) that for the value of the transition energy between the ground Landau levels of an electron and a hole ($n_L = 0$) for most semiconductors in an ultrastrong magnetic field is shifted to the visible part of the spectrum, convenient for experimental observation.

However, the most intriguing effect of a superstrong magnetic field is possibly the theoretically predicted modification of interatomic interaction which in the case of the exciton should manifest itself primarily in exciton luminescence rather than absorption spectra; the study of these spectra has received far less attention.

Interatomic (interexciton) interaction in a superstrong field

Neutral atoms strongly elongated along the direction of the field will have a large electric quadrupole moment, so they will begin to strongly attract each other along the magnetic field lines, weakly repelling in directions perpendicular to the field. The so-called magnetopolymers may form as a result, which are chains of spindle-like atoms modified by the field, arranged along the lines of the field (Fig. 4). Individual atoms disappear in such a polymer with a further increase in the field, while electrons are collectivized into a single elongated cloud, with protons that are atomic nuclei located along its axis [9].

This theoretical picture may have already been confirmed experimentally by the detected thermal emission spectra of neutron stars. A part of this radiation cannot penetrate the atmosphere of a neutron star, and absorption bands are observed in the X-ray region of the spectrum. The main band is typically located in the region $E_a \approx 200$ –700 eV [12], and several less pronounced bands are sometimes observed at energies $2E_a$, $3E_a$ etc. Initial attempts to relate these bands to the cyclotron frequencies of electrons or protons, or to oxygen in the stellar atmosphere were unsuccessful.

However, this absorption can be explained

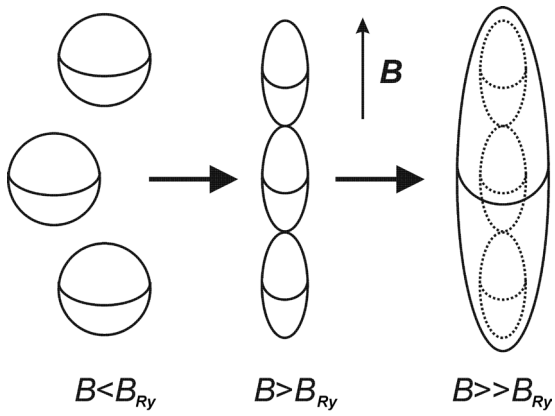


Fig. 4. Formation of atomic/excitonic magnetopolymers in superstrong magnetic fields

by ionization of hydrogen atoms in a superstrong magnetic field: an increase in the ionization potential to 200–700 eV ($15 - 50Ry$) is possible in fields of the order of $10^8 - 10^9$ T, i.e., specifically in the fields characteristic for old neutron stars (much smaller fields, namely, 10^6–10^7 T, are typical for young neutron stars (supernova remnants)).

An X-ray quantum can simultaneously ionize several atoms making up the magnetopolymer; moreover, collectivization of electrons occurs in such fields (as pointed out above). As a result, bands should be observed at energies E_a, $2E_a$, $3E_a$, etc., up to a relatively low binding energy of atoms into a polymer. This can explain the presence of additional absorption bands in the spectrum.

Interestingly, even though the surface temperature of neutron stars approaches a million degrees ($kT \approx 100$ eV), hydrogen atoms in their atmosphere should remain predominantly unionized, since $kT < E_a$. In other words, the atmosphere of an old neutron star hypothetically contains atoms (albeit unusual), molecules (as the atmosphere of the Sun), and magnetopolymers.

Going back from atomic physics to the physics of excitons, we should note that biexcitons have been studied only in several types of crystals and in fact in a relatively weak magnetic field, while triexcitons, etc., are merely a theoretical idea. It has already been discovered that, in contrast to an exciton, a biexciton does not transform into radiation in a one-photon process

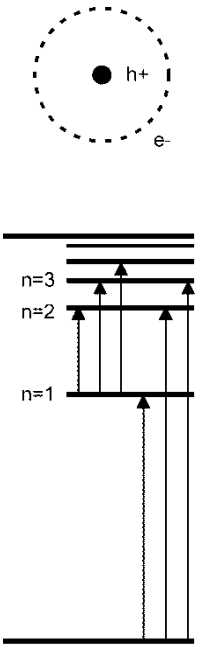
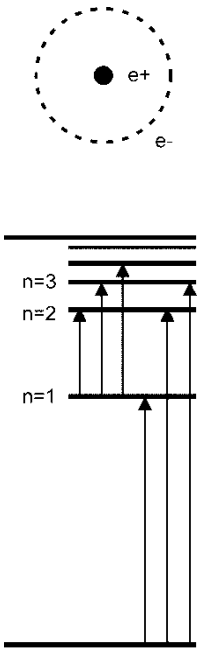
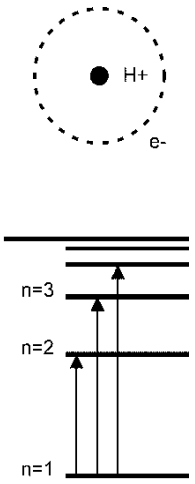
$$X_2 \rightarrow h\nu,$$

where X is an exciton, but rather does so in two stages [13]:

$$X_2 \rightarrow X_{n=2} + h\nu_1 \rightarrow h\nu_1 + h\nu_2.$$

Table 3

Comparison of exciton with hydrogen and positronium atoms

Characteristic	Exciton (X)	Positronium (Ps)	Hydrogen (H)
Structure	Hole + electron ($h^+ + e^-$)	Positron + electron ($e^+ + e^-$)	Proton + electron ($p^+ + e^-$)
Reduced mass	$0.01 – 0.30 m_e$	$0.5 m_e$	$\approx m_e$
Size, Å	10–1000	≈ 2	≈ 1
Binding energy, MeV	1–100	≈ 6800	≈ 13600
Dimerization energy, eV	Biexciton X_2: 0.0001–0.0010	Dipositronium Ps_2: 0.4	H_2 molecule: 4.5
Annihilation (Energy)	1, 2,... photons; $h\nu \approx E_g \approx 0.1 – 10$ eV)	2, 3,... photons; $h\nu_\Sigma = 2m_e c^2 \approx 1$ MeV)	impossible; $H^+ + e^- = n^0 + \nu_e$ ($m_H c^2 \approx 1$ GeV)
Lifetime, s	10^{-5}–10^{-7}	Parapositronium $1.25 \cdot 10^{-10}$ Orthopositronium $1.43 \cdot 10^{-7}$	Stable
Optical transitions			

Furthermore, annihilation of a positronium atom (and even more so a molecule) is also a multiphoton process (see Table 3).

Since the biexciton theory is poorly developed even for weak fields, not to mention strong and superstrong ones, studies of high-density excitons (which is a necessary condition for observing bi- and polyexcitons in weak magnetic fields) in superstrong magnetic fields can yield new, unexpected and interesting results.

Production of an exciton by light and its short lifetime

An extremely important difference between an exciton and a hydrogen atom is that the exciton can be produced by light.

In this aspect, exciton is much more similar not to the hydrogen atom but to the exotic positronium atom (Table 3). Table 3 also shows one more difference of an exciton (and a positronium atom) from a hydrogen atom, its (their) instability. Increasing the binding energy of the exciton, the magnetic field also increases the simultaneous lifetime of the electron and hole. However, it is unclear whether this trend will persist in extremely strong magnetic fields; if it does persist, it may contribute to producing a high-density exciton gas. In this case, the gas may possibly be obtained cold, i.e., the conditions for Bose–Einstein condensation of excitons can be generated [14].

A relative 'overheating' of the electronic subsystem usually takes place in the absence of a field or in magnetic fields that are not very strong. The reason for this is that the exciton gas excited by a short light pulse does not always have time to reach equilibrium, and the temperature of the exciton subsystem turns out to be uncoordinated with the temperature of the crystal lattice.

Aside from the exciton (electron-hole pair), many quasiparticle excitations can be generated in a crystal. They include, for example, phonons corresponding to the collective vibrations of the atoms in the crystal lattice. Since the energy of acoustic phonons has a lower limit of zero, these excitations can be produced in any magnetic field. Figuratively speaking, an electron and a hole moving along Larmor orbits emit acoustic phonons around them. The stronger the field, the more phonons are produced and the more the lattice overheats.

The larger the magnetic field that is to be generated, the less time there is for this task. The time for the field to reach its maximum value in the aforementioned record experiments was 5–20 μs; after that, a thermal explosion occurred, followed by destruction of both the magnet and the sample.

This means that the situation is different in hyperstrong magnetic fields. The exciton system may not have time to heat up before the sample is destroyed and will remain relatively cold. A high-density cold excitonic gas is an ideal situation for Bose-Einstein condensation [14].

Hyperstrong magnetic field

Theory [6–9] predicts for magnetic fields above $B_{hs} \approx 10^{10}$ T (B_{hs} here and below refers to the flux density of a hyperstrong magnetic field) that the hydrogen atom should be no longer spindle-shaped but rather needle-shaped, i.e., the ratio of the longitudinal to transverse linear dimensions of the ellipsoid should reach 200.

However, the proposed model is apparently no longer applicable in such a strong field. Evidently, the energy of interaction of the electronic magnetic moment of the hydrogen atom with the external magnetic field $E_M(B)$ exceeds the level of 10^6 eV in the given hyperstrong field, i.e., the energy becomes sufficient for the production of electron-positron pairs from vacuum, and the so-called quantum electrodynamic Schwinger limit (QEL, $B_{cr} = m_e^2c^2/e\hbar \cong 4.4\cdot10^9$ T) is reached, above which absolutely unpredictable effects can be observed (see below).

To produce an electron-positron pair due to the interaction of a magnetic field with a positronium atom rather than a hydrogen atom taking into account the Landau quantization (electrons and positrons are produced in a strong magnetic field only in states corresponding to the ground Landau level), the magnetic field B must become such that the energy of interaction with it is

$$2[E_{\mathrm{L},0e^+}(B) + E_{\mathrm{L},0e^-}(B)]$$

(compare Eqs. (3) and (6)) is higher than simply

$$m_e c^2 + m_{e^+} c^2 = 2m_e c^2,$$

namely, it must follow the expression

$$E = 2m_e c^2 + E_{\mathrm{L},0e^+}(B) +$$

$$+ E_{\mathrm{L},0e^-}(B) = 2m_e c^2 + E_M(B)/2. \quad (7)$$

In other words, the required interaction energy $E_M(B) = E$ must be higher than $4m_ec^2$, and the corresponding hyperstrong magnetic field must be

$$B_{hs} = B_{cr} \cong 4.4\cdot10^9 \text{ T}.$$

Since the production/annihilation of an electron-positron pair is a fundamentally multiphoton process (see Table 3), the required energy can either be 'collected' from two electrons of one atom, or obtained by exciting already existing bound electron-positron pairs (positronium atoms). In the latter case, cascade pair generation occurs. We should also note that a multiphoton process implies that

'subbarrier' production of an electron-positron pair is possible (similar to absorption of light below the absorption edge in semiconductors).

An energy that is a million times less than the energy for the production of an electron-positron pair in vacuum is required for the production of an electron-hole pair in a crystal. It is equal to the energy E_g that is the band gap, and varies from several electron volts for wide-gap semiconductors to tenths of electron volts for narrow-gap semiconductors. Similarly to Eq. (7), we have the following criterion for an exciton in the magnetic field:

$$E_M(B) > 2E_g,$$

$$B > B_{hs} = 2\mu E_g/e\hbar.$$

The corresponding dependence of the interaction energy on the magnetic field in the crystal, taking the form

$$E_M(B) = E_g + E_{\mathrm{L},0h}(B) + E_{\mathrm{L},0e}(B),$$

was already shown in Fig. 3 for different semiconductors.

The exciton in such a hyperstrong magnetic field should begin to emit not only phonons (see above) but also new electron-hole pairs.

However, before we consider this possibility, it should be borne in mind that there are also optical phonons, whose energy is comparable to the binding energy of an exciton (in contrast to acoustic phonons with their low energy). The interaction of excitons with optical phonons was analyzed in detail theoretically and experimentally as early as the 1950s–1970s (see, for example, [15]). The dependences of exciton-phonon interaction on the degree of ionicity of semiconductor crystals, their defect structure, temperature, etc. were described.

In fact, the actual motion of the Wannier–Mott exciton in the crystal was proved in a study of 'phonon repetitions', i.e., spectral lines caused by the emission of phonons by moving excitons (energy relaxation of excitons). A characteristic 'Maxwellian' shape of these lines corresponds to the energy

$$E_X - nE_{\mathrm{LO}},$$

where E_X is the energy of the ground exciton state ($E_X = E_g - Ry_{ex}$); E_{LO} is the optical phonon energy; n = 1, 2.

The given shape of the lines and their temperature dependence were described up to n= 4 [16].

However, the modification of exciton-phonon interaction in a magnetic field still remains virtually unexplored, especially for the case of a strong field, when the theoretical parameter $E_M(B)/Ry_{ex}$ grows large.

Thus, new exciton-phonon effects can be expected in superstrong and more so in hyperstrong magnetic fields. At the very least, numerous exciton lines of 'phonon repetitions' should appear in the luminescence spectra.

Going back to potential production of not only phonons but also new electron-hole pairs ('cloning' of excitons, so to speak) in a hyperstrong magnetic field, by analogy with optical phonons, '(e-h)- repetitions' of the ground exciton state bound to the zero Landau levels of the electron and hole can be expected to appear. However, the theory for this is completely untenable, leaving space for interpretation.

The energy $E_M = 2E_g$, for example, in an InSb crystal ($E_g \approx 0.2$ eV), corresponds to the magnetic field $B_{hs} \approx 50$ T, which is achievable in modern laboratories. Table 4 shows the values of B_{hs} for several other materials.

Even more surprising exciton effects are also possible. An attempt was made in [17] to hypothesize what happens to a substance at magnetic fields above the QEL. The vacuum supposedly becomes polarized, and the light entering the region with such a field changes its speed and wavelength. Consequently, birefringence is observed, similar to, for example, anisotropic crystals. Moreover, high-energy photons begin to spontaneously split into several quanta or merge into one.

Notably, such splitting/merging in a hyperstrong magnetic field can be quite simply described in terms of exciton physics as follows: the quanta $h\nu \approx 2E_g$ produce excitons (e + h pairs), these pairs/excitons form a magnetic molecule (magnetopolymer, see Fig. 4) whose extended 'hypernucleus' is surrounded by collectivized electrons. This magnetopolymer of N excitons annihilates with a certain probability, i.e., a quantum of light an energy $Nh\nu$ is emitted. Or, vice versa, a quantum $Nh\nu$ generates a magnetopolymer of N excitons; each of these excitons annihilates, emitting a quantum of light with the energy $h\nu \approx 2E_g$.

Thus, the emission spectrum of semiconductors in a hyperstrong magnetic field excited by light with an energy of about N_1E_g should exhibit several bright emission lines separated by an energy of about N_2E_g, due to both the above-described 'splitting/merging' of light quanta, and by cascade generation of electron-hole pairs. The spectrum of the

radiation appearing will be complex, with line shifts due to the binding energies of the biexciton, the magnetopolymer, reflecting transitions to the non-ground state of the exciton, etc.

It is evident that the QEL imposes a constraint on the increase in the binding energy of an electron in a hydrogen atom. A similar situation apparently happens in the case of an exciton. Unfortunately, the existing theories are not applicable near the QEL, and the quantitative results of the calculations are still unreliable.

Furthermore, it is notable that the ratio of superstrong and hyperstrong fields is hundreds of thousands for the positronium and hydrogen atoms, while for excitons it does not exceed 100 (see Table 4, where superstrong fields $B_{ss} > 10\ B_{Ry}$). Correspondingly, the radius of the hydrogen or positronium atom transverse to the field can decrease by two orders of magnitude upon reaching hyperstrong fields, and the exciton radius near $B \approx B_{hs}$ by only several times (by less than five times for narrow-gap materials and slightly more for wide-gap materials, according to the most optimistic estimates). However, we should note that the corresponding magnetic fields are far beyond the current experimental capabilities This somewhat weakens the research potential of the exciton as a model object for analyzing the behavior of hydrogen-like atoms in superstrong magnetic fields, although the effect of exciton magnetopolymerization should still be observed.

Energy density of a hyperstrong field

Another radical difference in the behavior of excitons in a hyperstrong magnetic field from the behavior of hydrogen and positronium atoms under the same conditions is in their characteristic energy densities of the given field.

The magnetic field is a special type of matter through which moving charged particles or bodies with a magnetic moment interact. Interacting with the magnetic moment of the electron, this field increases the electron's energy, generating by itself a certain energy density.

The record energy density currently achieved in a constant magnetic field is 6 kJ/cm^3, and 3 MJ/cm^3 in a pulsed one [4]. The energy density of an ultrastrong field with respect to the positronium atom is $W_{hs} \cong 8 \cdot 10^{18}$ J/cm^3,

Table 4

Parameters of energy and magnetic field for some crystals and the positronium atom

Crystal	E_g, eV	B_{hs}, T	Ry_{ex}, meV	B_{ss}, T	B_{hs}/B_{ss}
Exciton					
InSb	0.2	47	0.60	0.7	67
InAs	0.4	160	1.35	2.7	59
GaSb	0.8	500	2.00	6.2	81
GaAs	1.5	1 400	4.40	21	67
InP	1.4	1 700	6.00	36	47
CdTe	1.6	2 300	11.00	78	30
CdSe	1.8	2 700	11.50	87	31
ZnTe	2.3	4 000	18.00	155	26
GaN	3.5	9 800	25.00	350	28
CdS	2.5	6 800	25.00	340	20
ZnSe	2.8	6 200	35.00	390	16
ZnO	3.0	9 900	42.50	700	14
ZnS	3.8	19 000	49.00	1200	16
Positronium Ps					
–	$\Sigma(h\nu)$, eV	B_{cr}, T	Ry_{Ps}, meV	B_{ss}, T	B_{cr}/B_{ss}
–	10^6	$9 \cdot 10^6$	$7 \cdot 10^3$	$3 \cdot 10^5$	$3 \cdot 10^4$

Notations: E_g, Ry_{ex} are the band gap and binding energy of the exciton; B_{hs}, B_{ss} are the corresponding characteristic magnetic fields; $\Sigma(h\nu)$, B_{cr}, Ry_{Ps}, B_{ss} are the respective quantities for the positronium atom.

or $1.3 \cdot 10^8$ MeV/Å^3 for $B_{hs} = B_{cr} \cong 4.4 \cdot 10^9$ T. At the same time, if we imagine a space that is densely (or completely) filled with positronium atoms in the absence of a field, the corresponding energy density will be $W_{ps} \approx 0.1$ MeV/Å^3 ($\Sigma(h\nu) \approx 1$ MeV per positronium volume $V_{ps} \approx 10$ Å^3), i.e., a billion times less. Taking into account the doubling of the positronium energy in a hyperstrong field and a simultaneous hundred-fold decrease in its transverse radius increases W_{ps} by four orders of magnitude, i.e., does not change the ratio $W_{hs} >> W_{ps}$. It is this circumstance that leads to the so-called boiling of vacuum, i.e., production of a large number of virtual particles in magnetic fields B_{Wps}, much weaker than the magnetic field B_{cr} that is a hyper-strong field of the quantum electrodynamic limit (subbarrier production, see previous section), materialization of these particles and their subsequent annihilation (de-excitation). Such light emission by production and subsequent annihilation of pairs can actually be interpreted as transformation of energy of a constant magnetic field into electromagnetic radiation. In any case, the main effect of a hyper-strong field that is the production of pairs should be already observed in characteristic fields an order of magnitude smaller than the hyper-strong ones.

The role of a vacuum in semiconductor crystals is played by a medium with a dielectric constant different from unity; instead of virtual electron-positron pairs, all kinds of virtual quasiparticles with different energies are produced, including virtual electron-hole pairs. Following the above 'atomic/positronium' concept, the characteristic values of B_W can be given depending on the band gap and the binding energy of the exciton (its radius) (Table 5). For simplicity, Table 5 shows the parameters of only three abstract semiconductors: wide-gap (CdS-ZnTe type), narrow-gap (InSb type) and 'intermediate' (GaSb type). Aside from the hyperstrong magnetic field density W_{hs}, the 'exciton energy' density W_1 is given for each material (the density is expressed as the ratio E_g values to the exciton volume V_{ex}, i.e., it is assumed that the given crystal is densely filled with excitons). The factor accounting for the change in the radius and energy of the electron-hole pair does not exceed 100 in a superstrong field. The energy W_2 was calculated assuming a superdense exciton gas with one exciton per crystal cell (or electron-hole liquid, see below). The magnetic fields whose energy density determines the values of W_1 and W_2 are also given.

We should note that a magnetic field equal to only 5 T corresponds to a field energy density of 10 J/cm^3, or 1.6 10^4eV/Å^3 in the case of a narrow-gap material, which corresponds to the exciton density $\rho \cong 8 \cdot 10^{20}$cm^{-3}, or one exciton per six cells. In this regard, it should be mentioned that the Wannier–Mott exciton (an exciton of large radius) is a quasiparticle, which

Table 5

Dependence of energy densities and corresponding characteristic magnetic fields on the parameters of semiconductor crystals

Quantity	Value for semiconductor		
	wide-gap	intermediate	narrow-gap
E_g, eV	2.4	0.8	0.2
V_{ex}, 10^6 Å^3	0.08	10	500
B_{hs}, T	5000	500	50
W_{hs}, J/cm^3	10^7	10^5	10^3
W_1, J/cm^3	2	$5 \cdot 10^{-3}$	$2 \cdot 10^{-5}$
B_1, T	2	0.1	0.007
W_2, J/cm^3	1000	250	60
B_2, T	50	25	12

Notations: E_g, V_{ex} are the band gap and the volume of the exciton; B_{hs} is its characteristic magnetic field; W_{hs} is the energy density corresponding to B_{hs}; W_1, W_2 are the energy densities of superstrong and hyperstrong fields; B_1, B_2 are the magnetic fields corresponding to W_1 and W_2.

is not bound by the condition for 'dense filling' of space; the wave function of such an exciton is distributed over the entire crystal (in contrast to the Frenkel exciton localized on the cell at each moment of time); therefore, the exciton density can hypothetically increase infinitely. However, as the density increases, on the one hand, the effects of mutual screening of Coulomb interaction by excitons begin within each of them, and as a result, the excitonic gas transforms not into an excitonic liquid but into an electron-hole one; on the other hand, an exciton is a particle described in the one-electron approximation, and as the exciton density increases, this description becomes less and less valid

It is clear that not all of the energy of the magnetic field is transferred to exciton excitations. Nevertheless, if electron-hole pairs are generated at a sufficiently high rate, then all the known effects of a high-density excitonic gas ($\rho > 10^{18}$ cm^{-3}) will become possible: the formation of biexcitons (triexcitons, etc., up to, again, magnetopolymers) or excitonic/electron liquid, Bose-Einstein condensation. In other words, even given weak laser excitation of the crystal (when the exciton density produced by the laser is extremely low), it is nevertheless possible to obtain a high exciton density in extreme magnetic fields.

Materials with potential applications in magneto-optical studies on superstrong and hyperstrong magnetic fields

It can be concluded that the lower the binding energy of an electron and a hole Ry_{ex} (the larger the radius of the Wannier–Mott exciton a_{ex}), the more pronounced the discussed effects are. Since Ry_{ex} is typically the lower, the smaller the semiconductor band gap E_g (see, for example, Table 4), we can assume that a narrow-gap material with the band gap $E_g <$ eV has potential for applications.

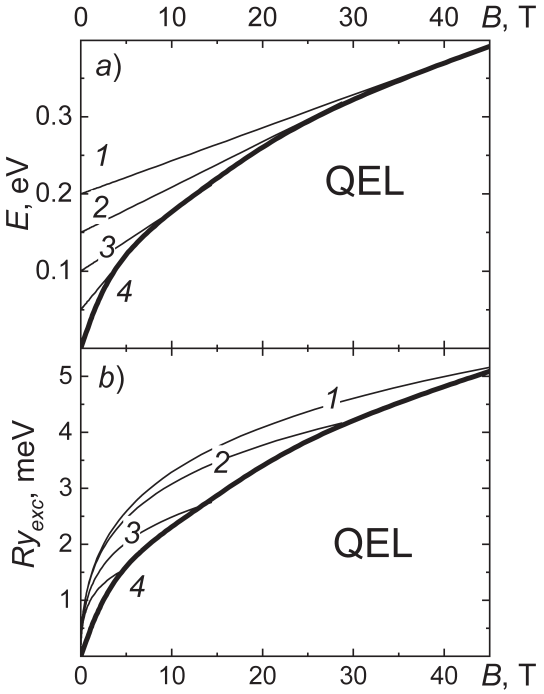


Fig. 5. Calculated dependences of ground Landau level energy (*a*) and binding energy of diamagnetic exciton (*b*) on the magnetic field in four semiconductors (numbers of the curves correspond to numbers in Table 6)

It seems reasonable to assess (if the current trends continue) the estimated parameters of such a hypothetical material (Table 6). Accordingly, it is possible to calculate the dependences of the energy of the ground ('zero') Landau level and the binding energy of a diamagnetic exciton on the magnetic field (compare Figs. 2 and 3). The calculated data obtained for three types of materials given in Table 6 (in addition to the well-known type *1*) are shown in Fig. 5.

Table 6

Parameters of extremely narrow-gap semiconductors

Number	E_g, meV	μ/m_e	ε	Ry_{ex}, meV	B_{ss}, T	B_{hs}, T
1	200	0.014	17.3	0.60	0.70	47.0
2	150	0.010	17.5	0.47	0.40	25.0
3	100	0.007	17.8	0.31	0.20	13.0
4	50	0.004	18.0	0.17	0.06	3.5

Notes. 1. The notations for the quantities correspond to those given in Tables 2 and 4, ε is the dielectric constant. 2. Variant *1* refers to, for example, InSb crystals.

Analyzing the data in Table and Fig. 5, we can see that even in the case of a material with E_g = 0.05 eV, in magnetic fields up to 10 T that are not at all record, the exciton line (or optical transition between the ground Landau levels of an electron and a hole) shifts from the infrared to the visible region of the spectrum, as in the vast majority of other semiconductors. Furthermore, increasing the exciton binding energy expands the temperature range where the exciton lifetime is sufficiently long: by 10 K (in a rough approximation), with Ry_{ex} increasing by 1 meV.

Unfortunately, the most narrow-gap compounds among homoatomic and binary semiconductors, for example, crystals such as InSb and mercury-based compounds with an inverted band structure have the values of $E_g \approx$ 0.2–0.3 eV: HgS, HgSe, HgTe.

Several extremely narrow-gap ternary (tertiary) compounds ($E_g \leq$ 0.1 eV) based on InBi are already known: $InAs_{1-x}Bi_x$, $InSb_{1-x}Bi_x$, as well as mercury chalgogenides but in the form of solid solutions, i.e., with a large number of structural defects that interfere with the existence of excitons (for example, the main problem in the case of the InBi compound is the formation of bismuth nanocrystals). We hope that further technological progress will nevertheless allow to obtain crystals of sufficient quality with the values of $E_g \leq$ 0.1 eV.

Conclusion

Let us now summarize the main effects expected for excitons in extremely strong magnetic fields.

The most interesting modification of the interexciton interaction in a superstrong field, with $B > B_{ss} \approx 10\ B_{Ryex}$, leads to greater stability of biexcitons and the appearance of exciton 'polymers'. In this case, the shift to the visible spectrum of the transition energy between the ground Landau levels of electron and hole in a superstrong field simplifies the experiments for most semiconductors. The rearrangement of the wave function of the exciton ground state from a spherical to a needle-like shape is not as substantial as that of a hydrogen atom in the corresponding superstrong atomic fields; therefore, the effect from increasing the binding energy is likely to be weak. At the same time, an increase in the exciton lifetime, which is still difficult to estimate, can result in generating a high-density cold exciton gas, i.e., produce the conditions for Bose-Einstein condensation of excitons.

In a hyperstrong magnetic field, given the magnetic flux density

$$B >; B_{hs} = 2\mu E_g/e\hbar$$

for an exciton in a solid or

$$B > B_{hs} = 4\mu m_e c^2/e\hbar$$

for hydrogen/positronium in a vacuum, the "cloning" of excitons (or, respectively, electron-positron pairs), the effect of magnetically induced birefringence and the effect of proportional multiplication/division of the frequency of the light incident on the crystal become possible.

Acknowledgments and closing remarks

We wish to express our gratitude to the experts of the International Science and Technology Center for supporting Project No. 3986 on Simulation of the low-dimensional state of the matter in experiments on excitons in high and ultra-high magnetic fields [18] that unfortunately received no funding from the Center. Experiments with superstrong and hyperstrong magnetic fields are extremely expensive. Nevertheless, we hope that the groundwork we have laid will prove useful in the nearest future; we believe that magnetic field experiments with the endless possibilities offered by the exciton will yield invaluable data not only on the properties of solids and magnetic atmospheres of neutron stars but possibly even the matter itself.

REFERENCES

1. Excitons: selected chapters (Modern problems in condensed matter sciences), Abridged edition by E.I. Rashba (author), M.D. Sturge (editor), Elsevier Science Ltd; Abridged Edition (May 1, 1987).

2. **Seysyan R.P.,** Spektroskopiya diamagnitnykh eksitonov [Spectroscopy of diamagnetic exitons]. Nauka, Moscow, 1984 (in Russian).

3. **Combescot M.,** Semiconductors in strong laser fields, Phys. Reports. 221 (4) (1992) 167–247.

4. **Boriskov G.V., Bykov A.I., Dolotenko M.I., et al.,** Research in ultrahigh magnetic field physics, Phys. Usp. 54 (4) (2011) 421–427.

5. Database SpringerMaterials – properties of materials, URL: http://materials.springer.com.

6. **Kadomtsev B.B., Kudryavtsev V.S.,** Matter in a superstrong magnetic field, Soviet Physics JETP. 35 (1) (1972) 76–80.

7. **Liberman M.A., Johansson B.,** Properties of matter in ultrahigh magnetic fields and the structure of the surface of neutron stars, Phys. Usp. 38 (2) (1995) 117–136.

8. **Korolev A.V., Liberman M.A.,** Bose condensation and superfluidity of excitons in a high magnetic field, Phys. Rev. B. 50 (19) (1994) 14077–14089.

9. **Kadomtsev B.B., Kudryavtsev V.S.,** Molecules in a superstrong magnetic field, Soviet Physics JETP Letters. 13 (9) (1971) 42–54.

10. **Moskalenko S.A.**, **Dumanov E.V.**, **Podlesny I.V., et al.,** Coherence of two-dimensional electron-hole systems: spontaneous breaking of continuous symmetries: a review, Physics of the Solid State. 55 (8) 1563–1595.

11. **Gelmont B.L., Mikhaylov G.V., Panfilov A.G., et al.,** Diamagnitnyye eksitony v geksagonalnykh kristallakh A_2B_6, Soviet Physics. Solid State. 26 (6) (1987) 1730–1739.

12. **Potekhin A.Y.,** Cyclotron harmonics in opacities of isolated neutron star atmospheres, Astronomy & Astrophysics. 518 (2010) A24, 1–12.

13. **Razbirin B.S., Uraltsev I.N., Mikhaylov G.V.**, Biexcitons in A_2B_6 crystals, Soviet Physics JETP Letters. 25 (4) (1977) 174–177.

14. **Kornell E.A., Wieman K.E., Ketterle W.,** Nobelevskiye lektsii po fizike – 2001 [Nobel lectures on physics – 2001], Phys. Usp. 173 (12) (2003) 1319 (in Russian).

15. **Gross E.F., Permogorov S.A., Razbirin B.S.,** Annihilation of excitons and exciton-phonon interaction, Sov. Phys. Usp. 14 (3) (1971) 104–112.

16. **Klochikhin A.A., Permogorov S.A., Reznitskii A.N.,** Multiphonon processes in resonant scattering and exciton luminescence of crystals, Soviet Physics JETP. 44 (6) (1976) 2230–2251.

17. **Kouveliotou C., Duncan R.C., Thompson C.,** Magnetars, Scientific American. 288 (2, February). (2003) 35–41.

18. Simulation of the low-dimensional state of the matter in experiments on excitons in high and ultra-high magnetic fields, URL: http://www. istc. int/en/

Received 04.08.2020, accepted 03.09.2020.

THE AUTHORS

VAGANOV Sergei A.
Ioffe Physical-Technical Institute of the Russian Academy of Sciences
26 Politekhnicheskaya, St. Petersburg, 194021, Russian Federation
SV.Exciton@mail.ioffe.ru

PANFILOV Andrei G.
Ioffe Physical-Technical Institute of the Russian Academy of Sciences
26 Politekhnicheskaya, St. Petersburg, 194021, Russian Federation
a.panfilov@mail.ioffe.ru

SEISYAN Ruben P.
Ioffe Physical-Technical Institute of the Russian Academy of Sciences
26 Politekhnicheskaya, St. Petersburg, 194021, Russian Federation
rseis@ffm.ioffe.ru

DOI: 10.18721/JPM.13402

TEMPERATURE EVOLUTION OF DIFFUSE SCATTERING IN THE STRONTIUM-BARIUM NIOBATE SBN-60 SINGLE CRYSTAL IN THE RELAXOR STATE

P.Yu. Vanina[1], N.S. Silin[1], A.A. Bosak[2], D.Yu. Chernyshov[1], A.A. Naberezhnov[3]

[1]Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation;
[2] European Synchrotron Radiation Facility, Grenoble, France;

[3] Ioffe Physical Technical Institute of the Russian Academy of Sciences, St. Petersburg, Russian Federation

Temperature dependences of synchrotron radiation diffuse scattering on the model relaxor single crystal $Sr_{0.6}Ba_{0.4}Nb_2O_6$ (SBN-60) have been studied in the temperature interval from 90 to 290 K in the vicinities of lattice points (332), (412) and (002) in the [001], [010] and [1$\bar{1}$0] directions. This diffuse scattering (DS) has been shown to be a strongly anisotropic with intensity $I_{DS}(T, q)$ proportional $q^{-\alpha}$ where q is a reduce wavevector. The temperature dependences of the parameters α have been obtained, and it is established, that the microscopic modification of structure in the SBN-60 continues on cooling below the temperature of transition to the relaxor state.

Keywords: diffuse scattering, synchrotron radiation, relaxor, strontium-barium niobate, crystal structure

Citation: Vanina P.Y., Silin N.S., Bosak A.A., Chernyshov D.Y., Naberezhnov A.A., Temperature evolution of diffuse scattering in the strontium-barium niobate SBN-60 single crystal in the relaxor state, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 13 (4) (2020) 21–28. DOI: 10.18721/JPM.13402

Introduction

Solid solutions of strontium barium niobate (SBN) are model uniaxial ferroelectrics of the tungsten bronze family with a partially disordered structure. If the strontium content is $50\% < x < 75\%$, these $Sr_xBa_{1-x}Nb_2O_6$ solutions (SBN-x) exhibit distinct relaxor properties, which become more pronounced with increasing strontium concentration [1].

An SBN-x structure is based on two types of NbO_6 octahedra, linked by oxygen atoms into a 3D network. This network includes three types of structural channels directed along the polar axis c (along the [001] direction). Threefold channels are always empty in SBN-x, the medium-sized fourfold channels (A_1) are filled only with strontium atoms, the wide fivefold channels (A_2) are filled with barium and strontium atoms [2]. A characteristic peculiarity of the crystal structure of such compounds is that there are five divalent cations and one vacancy for six possible sites in A_1 and A_2 channels, which determines the statistical nature of the cation distribution [2].

Much attention is paid to solutions of the SBN-x type, especially at $x \approx 60\%$, due to their relaxor, pyroelectric, dielectric, electro-optical and nonlinear optical properties [3–6]. Furthermore, it was found [7] that there is a critical strontium concentration in the range of 50–60%, above which normal ferroelectric domains do not exist: instead of them needle-like nanopolar domains are observed.

The SBN-60 crystal has a tetragonal structure with the following lattice parameters at room temperature:

$$a = b = 12.4566(9)\ \text{Å},\ c = 7.8698(6)\ \text{Å}.$$

We should also note that the SBN-60 structure is modulated, with the modulation vectors

$$\mathbf{q}_1 = 0.3075(6)\ (\mathbf{a}^* + \mathbf{b}^*),$$

$$\mathbf{q}_2 = 0.3075(6)\ (\mathbf{a}^* - \mathbf{b}^*),$$

where $\mathbf{a}^*$ and $\mathbf{b}^*$ are reciprocal lattice vectors.

The most adequate solution for the crystal structure of an SBN-61 single crystal (with a close concentration) is offered in [8], where it is described as an incommensurate structure in five-dimensional space within the superspace group $P4bm$ ($pp1/2$, $p-p1/2$).

Modulation of the structure was first observed in [9], and it was confirmed in [8] that it could be associated with the collective distortion of oxygen octahedra and the random occupation of 4*c* sites in the SBN structure with strontium and barium ions. Similar critical dynamics was also considered for an SBN-60 single crystal (similar in its structure and properties to SBN-61 doped with cerium). It has been established in [10–12] that transverse diffuse scattering (DS) by ferroelectric fluctuations is observed near the reciprocal lattice points of the (00L) type upon cooling from the high-temperature paraelectric phase, as the temperature of the transition to the relaxor state is approached; the DS intensity increases with decreasing temperatures, while the correlation length r_c increases. DS consists of two components in the transition region; the first of these components is well described by the Lorentzian

$$G_1 \sim 1/(\kappa^2 + q^2),$$

and the second by the squared Lorentzian

$$G_2 \sim 1/(\kappa^2 + q^2)^2.$$

Here κ is the inverse correlation length ($r_c \sim 1/\kappa$), q is the reduced wave vector.

Notably, the intensity of the second component increases sharply upon reaching the transition temperature. The dependence of the correlation length $r_c(T)$ on temperature has been obtained, and it has been found that this parameter 'freezes' below 340 K and remains practically constant up to 290 K [10]. Applying an external electric field suppresses both components of diffuse scattering, primarily, the component described by the function G_2 [10, 13].

On the other hand, the peculiarities of the crystal structure of SBN-60 (and its temperature evolution) on cooling in the relaxor phase are rather poorly understood at present: the only results available point to the existence of two different contributions to diffuse scattering [14]. Both types of scattering are considerably anisotropic: the first type is observed in the vicinity of superstructure reflections and is likely due to scattering by displacement waves in the system of oxygen octahedra, and the second is associated with the presence of local regions of correlated displacements of strontium and barium ions in fivefold SBN-60 channels.

This paper presents the first results on the temperature evolution of DS upon cooling from room temperature to 90 K.

Experimental study

The measurements were carried out on the ID29 beamline at the European Synchrotron Radiation Facility (ESRF). The wavelength of incident radiation was $\lambda = 0.7749$ E, $\Delta\lambda/\lambda \approx 2 \cdot 10^{-4}$. The samples were needle-like crystals approximately 100 μm thick.

Two-dimensional distributions of the X-ray scattering intensity were obtained using a PILATUS 6M pixel detector [15]. Adjustment of the orientation matrix and preliminary reconstruction of the reciprocal space were carried out via the CrysAlis software package [16]. The measurements were taken on cooling in the temperature range from 290 to 90 K, temperature stability was maintained to within 1.5 K.

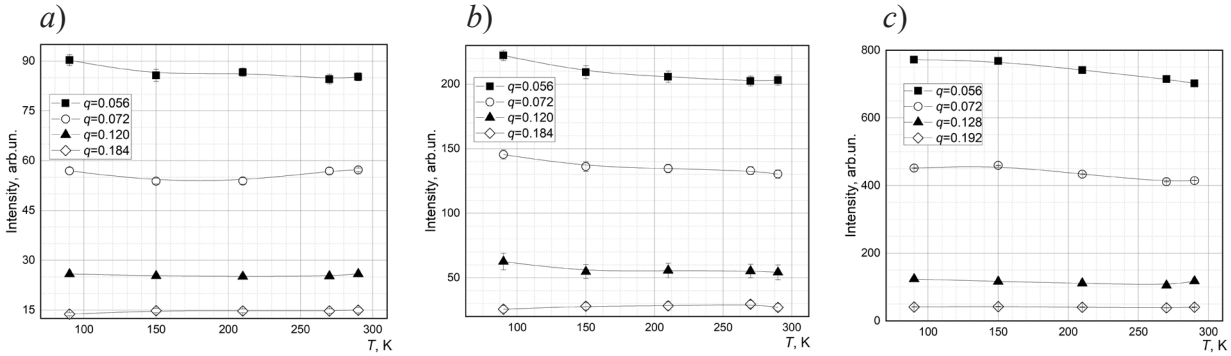


Fig. 1. Diffuse scattering intensities $I_{DS}(q,T)$ of synchrotron radiation as function of temperature in SBN-60 at different values of q in the [001] direction in the vicinity of points (332) (*a*), (412) (*b*), (002) (*c*).
The curves in the figures are the result of a smoothing procedure

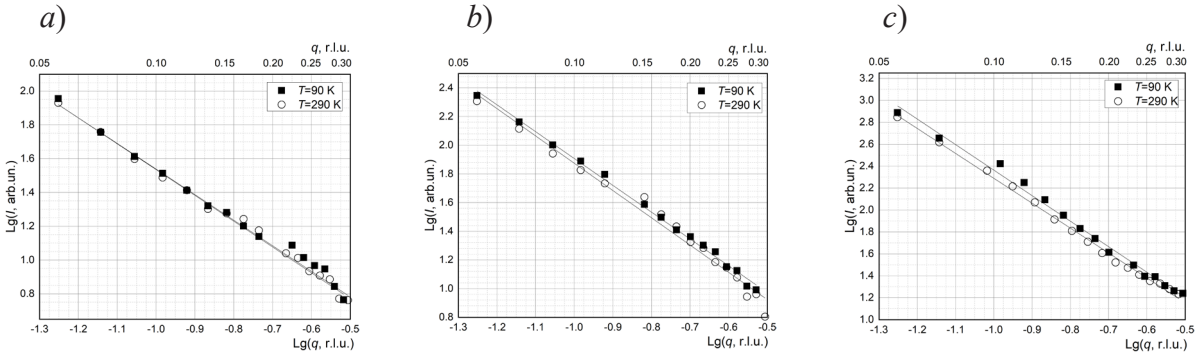


Fig. 2. DS intensities depending on the reduced wave vector (on a doubly logarithmic scale) in SBN-60 at 90 and 290 K in the [001] direction for points (332) (*a*), (412) (*b*), (002) (*c*). The values of q are given in r.l.u. (reduced lattice units). The lines in the figures are the result of fitting

Experimental results and discussion

The temperature evolution of DS has been studied in the vicinity of reciprocal lattice points (002), (412) and (332) in three directions: [001], [010] and [1$\bar{1}$0] to obtain information on the potential anisotropy of diffuse scattering.

The dependence of the intensity I from the reduced wave vector q has been fitted by the following expression:

$$I(q) = I_{DS}(q) + I_{\mathrm{Br}}(q) + I_{bkg}(q). \qquad (1)$$

The first term corresponds to diffuse scattering DS, which can be fitted ~~to~~ with the dependence $I_{DS}(q) \sim q^{-\alpha}$ at sufficiently large q ($q > \kappa$). The second term $I_{\mathrm{Br}}(q)$ describes the Bragg contribution, and the third, $I_{bkg}(q)$, is the linear contribution of the background to the total scattering.

The elastic peak was described by a Gaussian; the exact position of the q_c maximum was determined from the fit. The background was approximated by a linear function, the parameters were determined from averaging over 10–14 points, which were chosen to the left and right of q_c at large q, when the contributions from both the elastic peak and the DS became negligible. With $|q| < 0.09$, the strong Bragg peak did not allow to reliably detect the weaker DS in all cases, while diffuse scattering at $|q| > 0.35$ could not be differentiated from the background.

Let us consider the behavior of the temperature dependences of DS intensity in the SBN-60 single crystal in the direction [001]. It is clear from Fig. 1 that the DS intensity is practically independent of temperature in this direction for all values of q.

Next, the dependences of the DS intensities on q have been plotted for this direction on a doubly logarithmic scale in the entire investigated temperature range (Fig. 2). It is evident from the graphs that the relation $I_{DS}(q) \sim q^{-\alpha}$ is satisfied for all dependences $I_{DS}(q)$. The coefficients α corresponding to the slopes of the obtained straight lines are given in Table.

Table

Values of parameter α for straight lines $\lg(I_{DS}) = -\alpha\lg(q) + b$, obtained at different temperatures for different directions and lattice points in SBN-60 single crystal (see Fig. 2)

Reciprocal lattice point	Direction in crystal	Value of parameter α				
		90 K	150 K	210 K	270 K	290 K
(332)	[001]	1.53(11)	1.49(2)	1.52(5)	1.51(4)	1.51(3)
	[010]	3.52(10)	3.47(9)	3.45(7)	3.18(4)	2.89(10)
	[1$\bar{1}$0]	3.48(6)	3.49(4)	3.50(5)	3.36(5)	2.87(5)
(412)	[001]	1.87(3)	1.81(4)	1.86(6)	1.81(5)	1.85(6)
	[010]	3.32(5)	3.23(6)	3.22(6)	3.04(4)	2.80(5)
	[1$\bar{1}$0]	3.18(5)	3.25(5)	3.27(5)	3.24(6)	2.83(5)
(002)	[001]	2.33(5)	2.33(4)	2.31(4)	2.27(4)	2.27(3)
	[010]	3.20(5)	3.15(5)	3.11(5)	2.87(5)	2.57(3)
	[1$\bar{1}$0]	2.76(1)	2.68(2)	2.57(2)	2.24(4)	1.92(4)

We should note here that these parameters remain virtually unchanged over the entire temperature range considered for all reciprocal lattice points. In other words, diffuse scattering does not change upon cooling in the direction [001].

The temperature dependences of DS intensities of X-ray radiation in SBN-60, $I_{DS}(q,T)$, have been obtained by a similar procedure in the directions [010] and [1$\bar{1}$0] for values of the reduced wave vector in the range of $0.112 < |q| < 0.288$ in the temperature range from 90 to 290 K (Fig. 3). It can be seen from the data in Fig. 3 that the diffuse scattering intensity $I_{DS}(q,T)$ increases upon cooling for all the given points at small q.

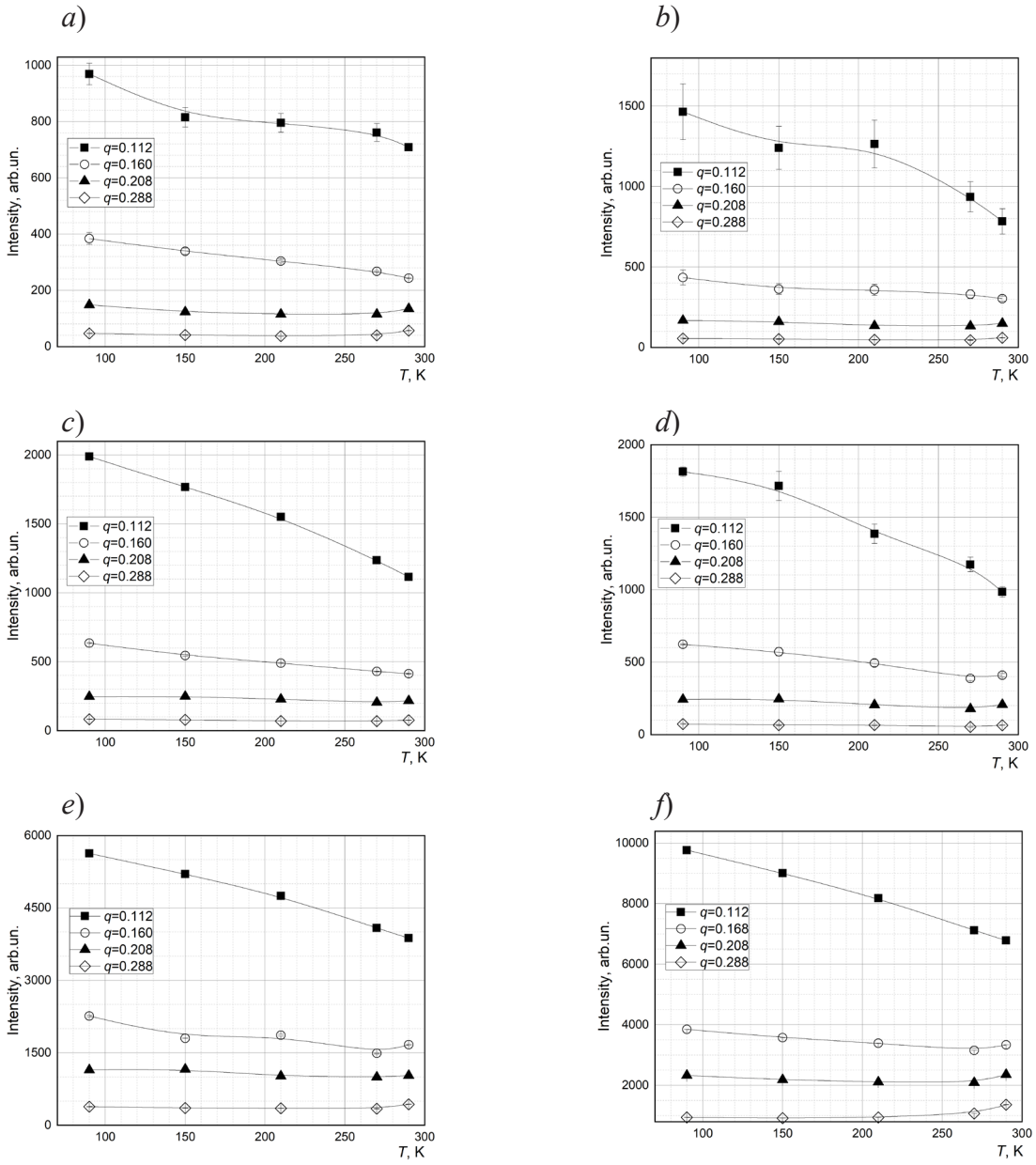


Fig. 3. Temperature dependences of DS intensities $I_{DS}(q,T)$ in SBN-60 single crystal at different values of q in the vicinity of point (332) in the directions [0$\bar{1}$0] (*a*), [1$\bar{1}$0] (*b*), in the vicinity of point (412) in the directions [010] (*c*), [1$\bar{1}$0] (*d*) and in the vicinity of point (002) in the directions [010] (*e*), [1$\bar{1}$0] (*f*).
The curves in the figures are the result of a smoothing procedure

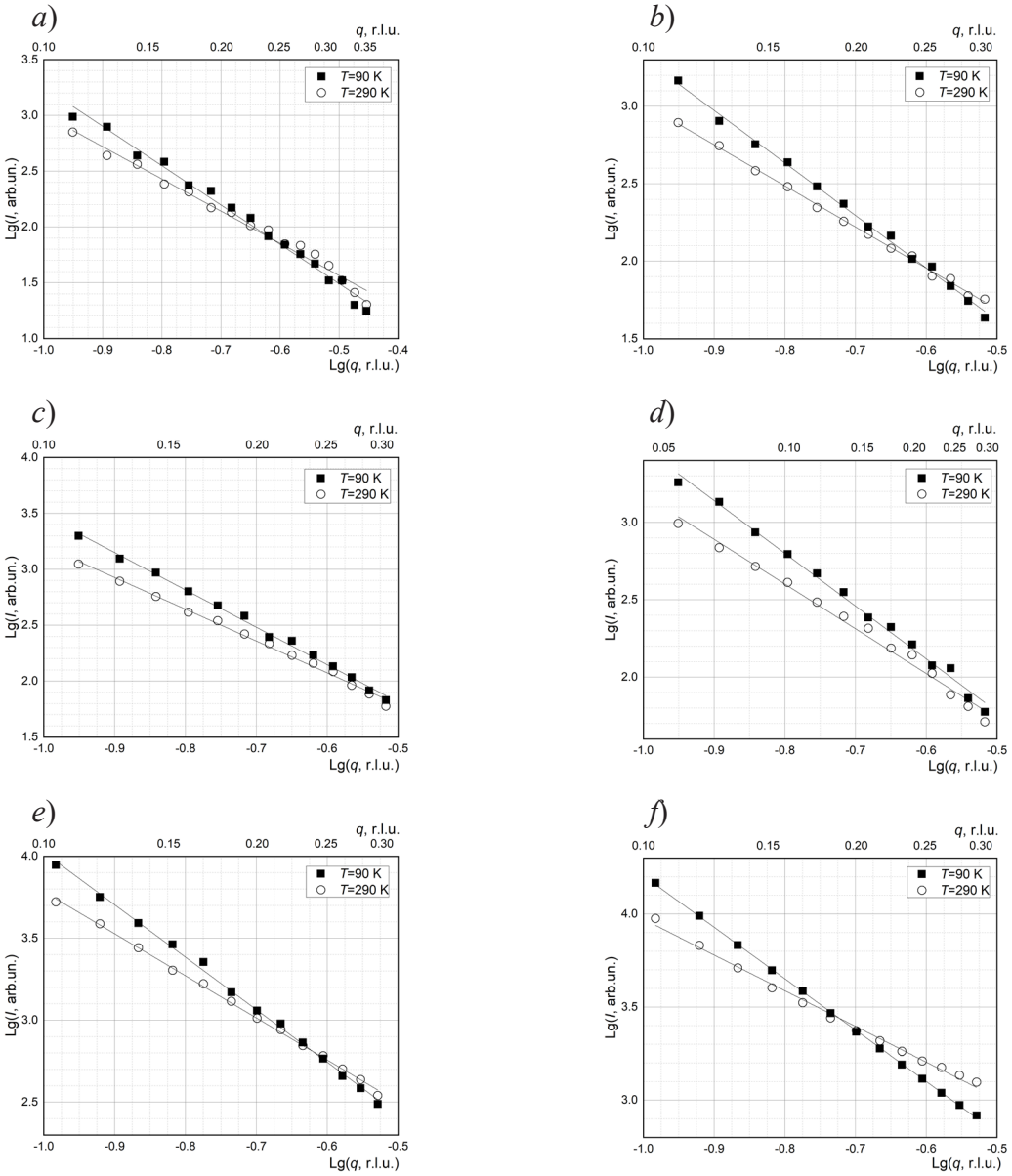


Fig. 4. DS intensities depending on the reduced wave vector (on a doubly logarithmic scale) in SBN-60 at 90 and 290 K in the vicinity of point (332) in the directions [010] (*a*) and [1$\bar{1}$0] (*b*), in the vicinity of point (412) in the directions [010] (*c*) and [1$\bar{1}$0] (*d*), in the vicinity of point (002) in the directions [010] (*e*) &[1$\bar{1}$0] (*f*). The values of q are given in r.l.u. The lines in the figures are the result of fitting

Next, the dependences of DS intensities on the reduced wave vector q have been plotted on a doubly logarithmic scale for the entire temperature range of 90–290 K (see Table). In Fig. 4, the dependences $I_{DS}(q)$ at 290 and 90 K are presented for comparison. Apparently, the linear character of the obtained dependences is clearly observed in these cases, i.e., there is a power-law dependence of the intensity DS in the vicinity of these point: $I_{DS}(q) \sim q^{-\alpha}$. Notably, the parameters α characterizing the slope of the curves increase significantly on cooling, reaching the following values: 3.52 and 3.48 for point (332) in the directions [010] and [1$\bar{1}$0], respectively; 3.32 and 3.18 for point (412), and 3.20 and 2.76 for point (002) at $T = 90$ K (see Table).

Thus, analysis of the data obtained points to the existence of two contributions to diffuse scattering:

DS-1, which increases with decreasing temperature along the directions [010] and [1$\bar{1}$0], with the parameter α increasing as well;

DS-2, which is practically independent of temperature along the direction [001] in the entire range of 90–290 K, the same as the parameter α; however, the values of the parameter α differ in the vicinity of different reciprocal lattice points.

Importantly, the obtained temperature dependences of the DS-1 and DS-2 intensities give reason to believe that neither type of scattering is thermal diffuse one (TDS), for which the intensity should increase on heating [17].

As mentioned above, it has been found in [10] that an additional contribution from the squared Lorentzian G_2 appears upon cooling from the paraelectric phase along with the contribution described by the Lorentzian G_1, which is typical for critical scattering; this contribution increases at transition to the relaxor phase. As a result, the total diffuse scattering is described by the superposition of the functions G_1 and G_2, and at sufficiently large values of q ($q > \kappa$) the experimental value of α in the scattering law $q^{-\alpha}$ exceeds 2. We also observe the same effect in this case for DS-1 at temperature decreasing, and the growth of α upon cooling indicates an increase of G_2 contribution into DS-1. In a sense, similar behavior of the parameter α has been observed in [18] for the lithium doped $KTaO_3$ single crystal (1.6% Li) below T_C (the phase transition temperature is about 35 K).

The function G_1 corresponds in direct space (in the isotropic three-dimensional case) to a correlation function of the form

$$C_1 = \exp(-\kappa r)/r,$$

characterized by the presence of a 'cut-off' factor $1/r$, that is, such a correlator is characteristic of regions of local ordering (correlation) with rather sharp boundaries.

The functions G_2 corresponds in direct space to a correlator of the form

$$C_2 = \exp(-\kappa r),$$

containing no 'cut-off' factor, i.e., such a correlation function is typical for regions with smeared boundaries.

Thus, it can be concluded from the observed growths of the DS-1 intensity and the slope parameter α upon cooling that local rearrangement of the structure continues in SBN-60 as the temperature decreases (at least down to 90 K), with an increase in the contribution from regions with blurred boundaries, which is what contributes to an increase in α. In turn, a growth of the DS-1 intensity upon cooling indicates an increase in the 'density' (i.e., the total number) of such regions.

The approaches for interpreting the data on small-angle neutron and X-ray scattering (SANS and SAXS) can be used to explain the absence of temperature dependences of both the DS-2 intensity and the slope parameters α in this case. It is known that in the case of a mass fractal with the fractal dimension D for parameter α in the scattering law $I(q) \sim q^{-\alpha}$ (I is the intensity, q is the transferred momentum) the relation $D = \alpha$, where $1 < \alpha < 3$, holds true [19–21]. For example, $\alpha = 1$ for a one-dimensional chain of mass spheres, and $1 \le D \le 3$, where D is a non-integer, for aggregates constructed from spheres without overlapping. The parameter D corresponds to the distribution of mass in space. This approach offers a logical explanation for the experimental values obtained for the parameter α, if, for example, the corresponding agglomerates of atomic chains in SBN-60 are regarded as scattering objects.

Conclusion

We have considered the temperature dependences for diffuse scattering intensity of synchrotron radiation, $I_{DS}(T,q)$ (q is the reduced wave vector), in a single crystal of strontium-barium niobate $Sr_{0.6}Ba_{0.4}Nb_2O_6$ in the vicinity of the crystal lattice points (332), (412) and (002), in three fundamental directions: [001], [010], [1$\bar{1}$0].

It has been established that the $I_{DS}(T,q)$ dependences are well described by the power law $q^{-\alpha}$ on a doubly logarithmic scale; the temperature dependences of the parameters α have been obtained.

We have found that diffuse scattering is substantially anisotropic in the vicinity of the given points and consists of two components (DS-1 and DS-2) with fundamentally different behavior both the temperature dependences of the intensities and the parameters α.

The intensity of the DS-1 component and the corresponding parameter α increase at temperature decreasing, which indicates that the evolution of the local structure in SBN-60, associated with the smearing of the boundaries of the regions of local ordering, continues on cooling.

As for the behavior of the DS-2 component observed in the [001] direction, neither the intensity nor the parameter α depend on temperature in the range considered (90–290 K), but the value of α depends on the reciprocal lattice point in whose vicinity the measurements have been taken. It is likely that the nature of DS-2 is associated with scattering by mass fractal structures.

The study was carried out at Peter the Great St. Petersburg Polytechnic University within the framework of a State Task for Basic Research (topic code 0784-2020-0025). D.Yu. Chernyshev expresses his gratitude to the Russian Foundation for Basic Research for partial financial support of the studies on this subject (grant no. 19-29-12023).

The authors declare no conflict of interest.

REFERENCES

1. **Viehland D., Xu Z., Huang W.-H.,** Structure-property relationships in strontium barium niobate. I. Needle-like nanopolar domains and the metastably-locked incommensurate structure, Philos. Mag. A. 1995. Vol. 71 (2) (1995) 205–217.

2. **Chernaya T.S., Maksimov B.A., Verin I.A., et al.,** Crystal structure of $Ba_{0.39}Sr_{0.61}Nb_2O_6$ single crystals, Crystallography Reports. 42 (3) (1997) 375–380.

3. **Glass A.M.,** Investigation of the electrical properties of $Sr_{1-x}Ba_xNb_2O_6$ with special reference to pyroelectric detection, J. Appl. Phys. 40 (12) (1969) 4699–4713.

4. **Ewbank M.D., Neurgaonkar R.R., Cory W.K., Feinberg J.,** Photorefractive properties of strontium-barium niobate, J. Appl. Phys. 1987. Vol. 62 (2) (1987) 374–380.

5. **Romero, J.J., Aragy C., Gonzalo J.A., et al.,** Spectral and thermal properties of quasiphase-matching second-harmonic-generation in Nd^{3+}: $Sr_{0.6}Ba_{0.4}(NbO_3)_2$ multiself-frequency-converter nonlinear crystals, J. Appl. Phys. 93 (5) (2003) 3111–3113.

6. **Simagina L.V., Mishina E.D., Semin S.V., et al.,** Second harmonic generation in microdomain gratings fabricated in strontium-barium niobate crystals with an atomic force microscope, J. Appl. Phys. 110 (5) (2011) 052015–052020

7. **Kuz`minov Yu.S.,** Segnetoelektricheskie kristally dlya upravleniya lazernym izlucheniem [Ferroelectric crystals to control laser emission], Nauka, Moscow, 1982 (in Russian).

8. **Woike T., Petricek V., Dusek M., et al.,** The modulated structure of $Ba_{0.39}Sr_{0.61}Nb_2O_6$. I. Harmonic solution, Acta Crystallographica, Section B. 59 (Part 1, February) (2003) 28–35.

9. **Schneck J., Toledano J.C., Whatmore F, Ainger F.W.,** Incommensurate phases in ferroelectric tetragonal tungsten bronzes, Ferroelectrics. 36 () (1981) 327–330.

10. **Borisov S.A., Okuneva N.M., Vakhrushev S.B., et al.,** Critical neutron scattering in a uniaxial relaxor $Sr_{0,6}Ba_{0,4}Nb_2O_6$, Physics of the Solid State. 55 (2) (2013) 334–341.

11. **Chillal S., Koulialias D., Gvasaliya S.N., et al.,** Phase transition of chemically doped uniaxial relaxor ferroelectric, J. Phys.: Condens. Matter. 27 (43) (2015) 435901 (8p.).

12. **Ondrejkovic P., Kempa M., Kulda J., et al.,** Dynamics of nanoscale polarization fluctuations in a uniaxial relaxor, Phys. Rev. Lett. 113 (16) (2014) 167601.

13. **Ondrejkovic P., Kempa M., Savinov M., et al.,** Electric-field influence on the neutron diffuse scattering near the ferroelectric transition of $Sr_{0.61}Ba_{0.39}Nb_2O_6$, Phase Transit. 89 (7–8) (2016) 808–816.

14. **Vanina P.Yu, Vakhrushev S.B., Naberezhnov A.A., Bosak A.A.,** Multiscale local ordering in the prototypical uniaxial relaxor $Sr_{0.6}Ba_{0.4}Nb_2O_6$ single crystal at room temperature, J. Phys.: Condens. Matter. 31 (No. 17) (2019) 175401.

15. **Brμnnimann C., Eikenberry E.F., Henrich B., et al.,** The PILATUS 1M detector, J. Synchrotron. Radiat. 13 (Part 2, March) (2006) 120–130.

16. Agilent, CrysAlis PRO, Agilent Technologies, Yarnton, Oxfordshire, England, 2011.

17. **Willis B.T.M., Pryor A.W.,** Thermal vibrations in crystallography, Cambridge University Press, Cambridge, 1975.

18. **Andrews S.R.,** X-ray scattering study of the random electric dipole system $KTaO_3$: Li, J. Phys. C. Solid State Phys. 18 (7) (1985) 1357–1376.

19. **Jullien R.,** Fractal aggregates, Comm. Cond. Mat. Phys. (Comm. Mod. Phys. Pt. B). 13 (4) (1987) 177–205.

20. **Avdeev M.V., Aksenov V.L.,** Small-angle neutron scattering in structure research of magnetic fluids, Physics Uspekhi. 53 (10) (2010) 971–993.

21. **Cousin F.,** Small angle neutron scattering, EPJ Web of Conferences. 104 (2015) 01004 (1–49).

Received 05.09.2020, accepted 10.09.2020.

THE AUTHORS

VANINA Polina Yu.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
p.yu.vanina@gmail.com

SILIN Nikita S.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
silin.ns@edu.spbstu.ru

BOSAK Alexey A.
European Synchrotron Radiation Facility
71, Avenue des Martyrs, Grenoble, 38000, France
alexei.bossak@esrf.fr

CHERNYSHOV Dmitry Yu.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
chernyshov65@gmail.com

NABEREZHNOV Alexander A.
Ioffe Physical Technical Institute of the Russian Academy of Sciences
26 Polytekhnicheskaya St., St. Petersburg, 194021, Russian Federation
alex.nabereznov@mail.ioffe.ru

DOI: 10.18721/JPM.13403

RHEOLOGICAL PARAMETERS' EFFECT ON THE ELECTRET PROPERTIES OF POLYVINYLIDENE FLUORIDE

Yu.A. Gorokhovatsky, D.E. Temnov, Yu.I. Sotova

Herzen State Pedagogical University of Russia, St. Petersburg, Russian Federation

It is known that polymer films based on a copolymer of polyvinylidene fluoride and tetrafluoroethylene P(VDF-TFE) have piezoelectric properties, and these properties appear only after preliminary stretching of the films. The polar crystalline β-phase of the copolymer is responsible for the piezoelectric properties in the P(VDF-TFE), the percentage of this phase in the polymer can vary. In this work, we have studied the influence of the orientational stretching rate on the stability of the electret, and, as a consequence, the piezoelectric state. The influence of rheological parameters on the polymer structure and the parameters of electrically active defects responsible for polarization processes is shown.

Keywords: electret state, polyvinylidene fluoride, tetrafluoroethylene, piezoelectric effect, thermostimulated spectroscopy

Citation: Gorokhovatsky Yu.A., Temnov D.E., Sotova, Yu.I., Rheological parameters' effect on the electret properties of polyvinylidene fluoride, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 13 (4) (2020) 29–34. DOI: 10.18721/JPM.13403

Introduction

The piezoelectric properties of polyvinylidene fluoride (PVDF) and copolymers based on it are used in various piezoelectric elements [1 – 3]. Piezoelectric properties of PVDF were observed for the first time in 1969 by H. Kawai [4]. Even then, it became clear to researchers that piezoelectric properties of PVDF films appear only after their preliminary stretching. Since there are certain difficulties in the production of dielectric films with piezoelectric properties from a PVDF homopolymer associated with a high coercive field of about 10^8 V/m (this is two orders of magnitude higher than that of piezoceramics) [5], there are usually used piezoelectric polymers with a lower coercive field based on vinylidene fluoride copolymers with trifluoroethylene (VDF-TrFE) and vinylidene fluoride with tetrafluoroethylene (VDF-TFE), which also have piezoelectric properties after stretching several times in comparison with the original length [6–9].

In this work, we have studied thermally stimulated depolarization currents in P(VDF-TFE) films manufactured at different rates of preliminary stretching and electreted in a negative corona discharge field.

It should be noted that prior to this, no proper attention was paid to the effect of the rate of preliminary stretching on charge relaxation in P(VDF-TFE) films.

Samples and research methods

The subjects of inquiry were copolymers P(VDF-TFE) of the F2ME trademark with different rates of preliminary stretching. The pre-stretching rate (percentage increase in length per minute) was 5, 30, 50 and 200. The samples were examined by IR spectroscopy using the attenuated total reflection (ATR) on the FT-IR spectrometer FSM-1202 to determine the degree of crystallinity.

We also studied the thermally stimulated depolarization currents in P(VDF-TFE) films, in which the corona-electret state was created. The corona-electret state was created as follows: the samples were heated to a temperature of about 70°C, placed in positive or negative corona discharge (field of about 1 MV/cm) and held at an elevated temperature for 10 min, then the samples were cooled down in the corona discharge field to room temperature. After that, the depolarization currents were measured in the films using a Setaram TSC II device; measurements were carried out in the linear heating mode at two heating rates: 6 °C/min and 9 °C/min.

Experimental results and their discussion

IR spectroscopy. Fig. 1 shows the IR transmission spectra obtained using the ATR method for copolymer P(VDF-TFE) with pre-stretching rates (in % per min): of 5, 30, and 200. To

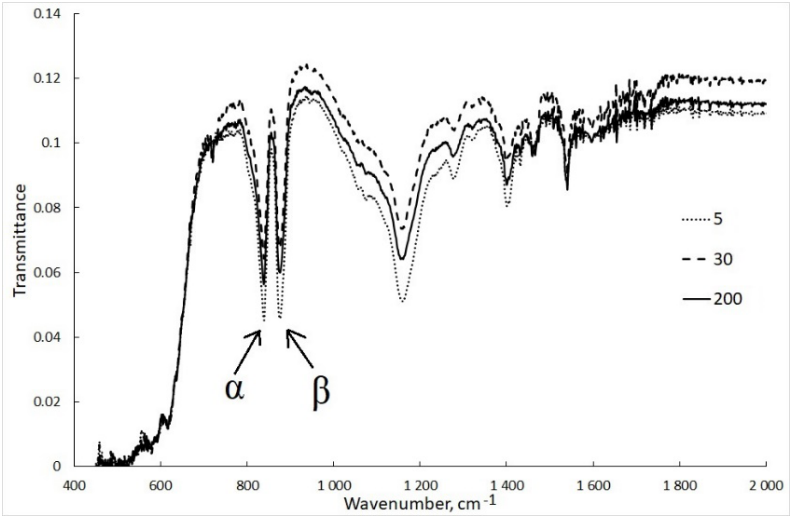


Fig. 1. ATR-IR spectra in copolymer P(VDF-TFE) for different rates of preliminary stretching (in % per min)

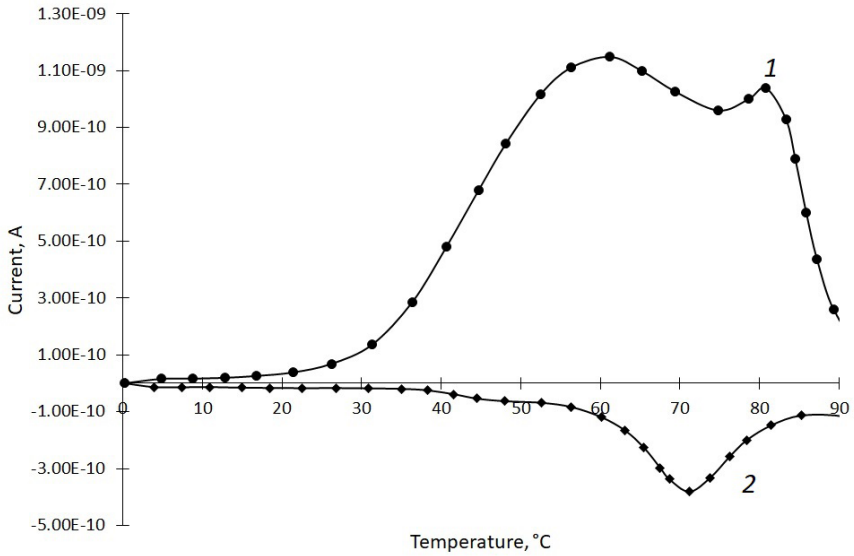


Fig. 2. Thermally stimulated depolarization currents in the P(VDF-TFE) with a preliminary stretching rate of 50 %/min, electreted in the field of the negative (*1*) and positive (*2*) corona discharges

determine the degree of crystallinity of PVDF, we studied the conformationally dependent band at 840 cm^{-1} (crystalline phase, α peak in Fig. 1) and the conformationally independent band at 876 cm^{-1} (amorphous phase, β peak in Fig. 1). Since there are no bands corresponding to the α phase for copolymers of VDF with TFE, the ratio of the intensities of these bands characterizes the degree of crystallinity of the polymer. By calculating the peak area ratio, it was obtained that the degree of crystallinity of the copolymer increased slightly from 49 % to 57 % as the stretching rate increased. Thus, it can be concluded that an increase in the stretching rate leads to a slight increase in the proportion of the crystalline β phase responsible for the piezoelectric properties of P(VDF-TFE). It should be noted that the quality of a piezoelectric element is determined not only by the magnitude of its piezoelectric modulus, but also by its stability.

Thermally stimulated depolarization currents. Fig. 2 shows a comparison of the curves of the depolarization currents P(VDF-TFE) with a preliminary stretching rate of 50 %/min electreted in the field of a positive and negative corona discharge (heating rate was 6°C/min).

The curves show two closely located peaks, the magnitude and temperature position of which strongly depend on the polarity of the

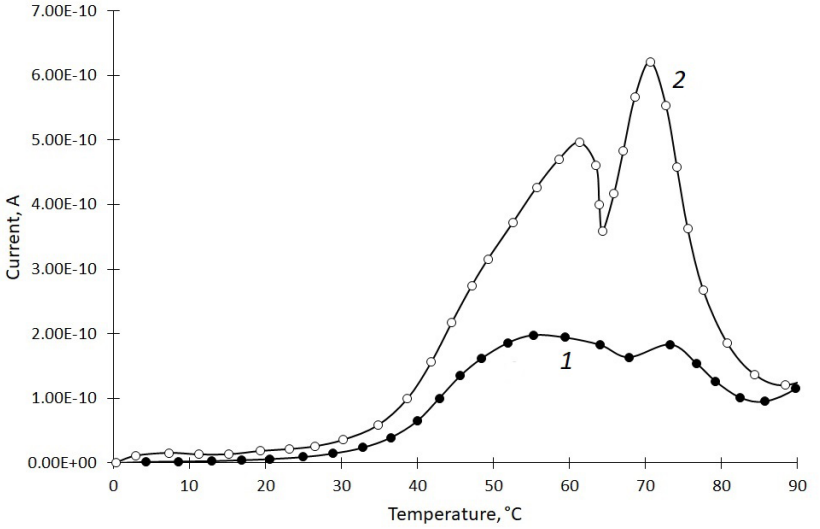


Fig. 3. Depolarization currents in the P(VDF-TFE) electreted in the field of the negative corona discharge at two values of a preliminary stretching rate (%/min): 30 (*1*) and 200 (*2*)

corona electrode. It should be noted that for other rates of preliminary stretching, the result was similar, i.e., the peak on the curve for negative polarity of the corona electrode was higher in magnitude and shifted along the temperature axis position compared with the peak for the positive polarity of the corona electrode. This difference suggests that these peaks are associated with the relaxation of the homocharge, i.e., with the release of charge carriers from deep subsurface traps. Such traps can be caused, for example, by various defects in the structure of polymer films. Apparently, the number and depth of traps for positive and negative charge carriers are different: for negative charge carriers the traps are deeper and their number is greater. This result is consistent with literature data [10].

The dependence of the depolarization currents in the P(VDF-TFE) electreted in the field of the negative corona discharge on the rate of preliminary stretching is shown in Fig. 3 (heating rate was 6°C/min). It can be seen that the high-temperature peak increases in magnitude and shifts to the region of higher temperatures with an increase in the stretching rate. Since we concluded earlier that this peak is associated with the release of charge carriers from traps, which, in turn, are caused by structural defects, it can be assumed that with a faster orientation of polymer films, the number of polymer chain defects increases, which means that the number of traps also increases.

Using the method of varying the samples heating rate, the parameters of the traps (activation energy and frequency factor) for negative charge carriers were calculated. For this purpose, we measured the depolarization currents of the P(VDF-TFE) films with a preliminary stretching rate of 30 %/min, electreted in the field of the negative corona discharge, at two different heating rates (6 °C/min and 9 °C/min). The results are shown in Fig. 4, *a*. It is seen that with an increase in the heating rate, the peak increases in magnitude and shifts to the right along the temperature scale. This result is consistent with the theory of thermally stimulated depolarization currents. Calculations revealed that, in this case, the activation energy of traps for negative charge carriers is 0,89 ± 0,05 eV, and the frequency factor is 10^{12} s^{-1} (with an accuracy of half a decade).

For the P(VDF-TFE) films with a preliminary stretching rate of 200%, electreted in the field of a negative corona discharge, the depolarization currents obtained for different heating rates are shown in Fig. 4,*b*. The chart also shows a shift in the peak towards higher temperatures with an increase in the heating rate from 6°C/min to 9°C/min. In this case, calculations give the following trap parameters: activation energy 0,93 ± 0,05 eV, frequency factor 10^{13} s^{-1} (with an accuracy of half a decade). Thus, it can be concluded that with an increase in the rate of preliminary stretching, the activation energy (depth) of traps for negative charge carriers slightly increases.

The curves in Fig. 3 also show a low-temperature peak (in the region of 50 − 60°C), which is partially superimposed on the high-temperature peak (in the region of 75°C).

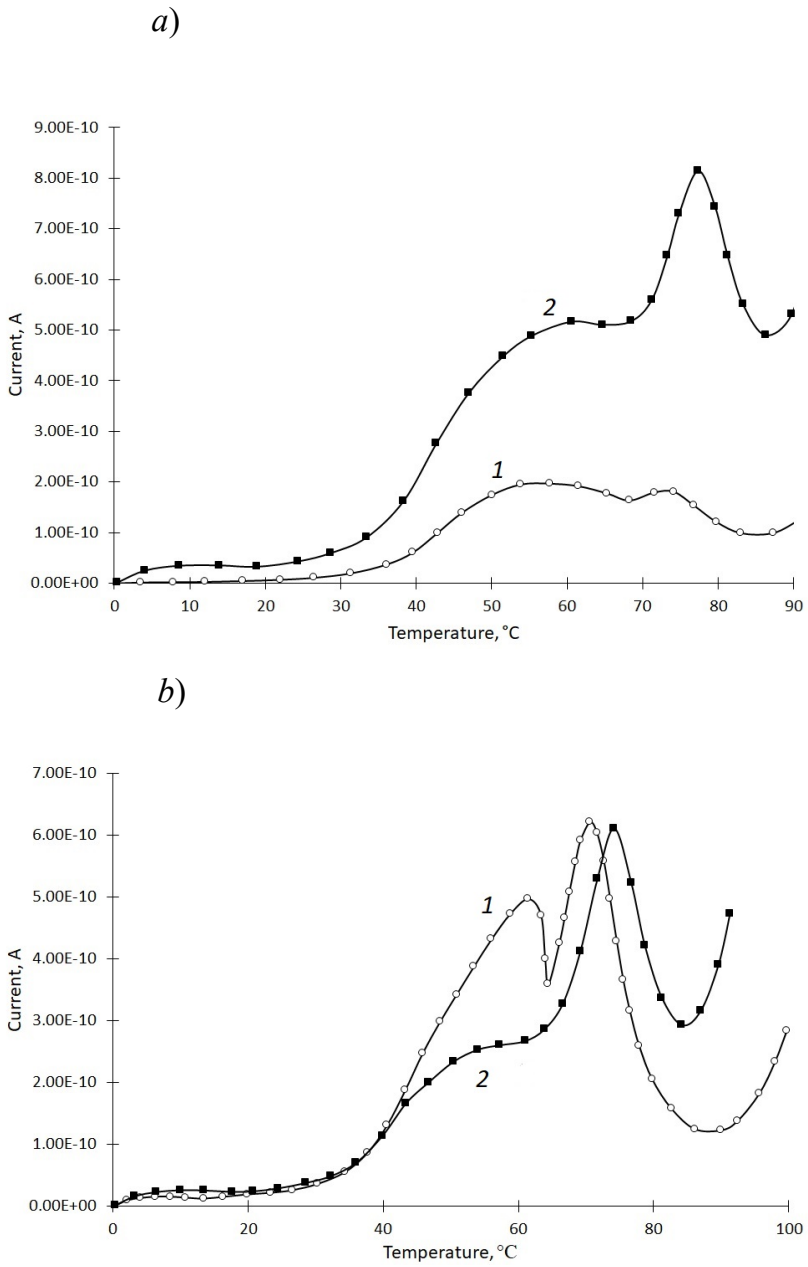


Fig. 4. Thermally stimulated depolarization currents in the P(VDF-TFE) with preliminary stretching rate of 30 %/min (*a*) and 200 %/min (*b*), electreted in the negative corona discharge field, for a heating rate of 6°C/min (*1*) and 9°C/min (*2*)

Since the growth of the low-temperature peak correlates with the growth of the high-temperature peak, we assume that the process in the region of 60°C is associated with the reorientation of polar structures that are present in the amorphous phase. With an increase in the stretching rate, the value of the homocharge, as noted above, increases, which means that the field of this homocharge also increases, and therefore the reorientation of polar structures becomes more and more evident.

It should be noted that it is the presence of the component associated with the reorientation of polar structures present in the amorphous phase of P(VDF-TFE) that determines the stability of the electret state, and, accordingly, the stability of the piezoelectric state. The increase in the tensile rate leads to an increase in this component and, as might be expected, to an improvement in the stability of the electret and piezoelectric properties of the copolymer under study.

Summary

IR spectroscopy data on the subjects of inquiry showed that with an increase in the stretching rate, the proportion of their ferroelectric β-phases increases. On the curves of thermally stimulated depolarization currents, two closely spaced peaks were found: a low-temperature one with a maximum in the region of 60°C and a high-temperature one with a maximum at 75°C. The high-temperature peak is associated with the release of charge carriers from deep near-surface traps, i.e., with relaxation of homocharge, the value of which depends on the rate of preliminary stretching. The activation energy of the traps increases with an increase in the stretching rate from 0,89 eV to 0,93 eV. The low-temperature peak is due to the reorientation of polar structures that are present in the amorphous phase. As the stretching rate increases, i.e., with an increase in homocharge, the proportion of oriented polar structures also increases. The presence of a component associated with the reorientation of polar structures determines the stability of the electret and, consequently, the piezoelectric state.

The reported study was funded by RFBR, project number 19-32-90161 and the Ministry of Education of the Russian Federation as part of a state task (project No. FSZN-2020-0026).

Работа выполнена при финансовой поддержке РФФИ, грант No 19-32-90161, а также Государственного задания Минобрнауки России (проект No FSZN-2020-0026).

REFERENCES

1. **Chen X., Han X., Shen Q.D.,** PVDF-based ferroelectric polymers in modern flexible electronics, Advanced Electronic Materials. 5 (3) (2017) 1600460 (1–18).

2. **Dong W., Xiao L., Hu W., et al.,** Wearable human-machine interface based on PVDF piezoelectric sensor, Transactions of the Institute of Measurement and Control. 39 (4) (2017) 398–403.

3. **Zhukov V.B., Katunin A.A., Seleznev I.A.,** Hydroacoustic tranducers based on perspective piezomaterials, In: Proceedings of the 14th All-Russian Conference «Advanced Technologies of Hydroacoustics and Hydrophysics». (2018) 11–16.

4. **Kawai H.,** The piezoelectricity of poly (vinylidene fluoride), Japanese Journal of Applied Physics. 8 (6) (1969) 975–983.

5. **Lushcheykin G.A.,** New polymer-containing piezoelectric materials, Physics of the Solid State. 48 (6) (2006) 963–964.

6. **Tan S., Hu X., Ding S., et al.,** Significantly improving dielectric and energy storage properties via uniaxially stretching crosslinked P(VDF-co-TrFE) films, Journal of Materials Chemistry A. 1 (35) (2013) 10353–10361.

7. **Ma S., Ye T., Zhang T., et al.,** Highly oriented electrospun P(VDF-TrFE) fibers via mechanical stretching for wearable motion sensing, Advanced Materials Technologies. 3 (7) (2018) 1800033 (1–7).

8. **Kochervinskiy V.V., Bedin S.A., Razumovskaya I.V., et al.,** Molecular mobility and structuring in textured films of the ferroelectric copolymer of vinylidene fluoride with tetrafluoroethylene, Polymer Science. Series A. 3 (58) (2016) 255–267.

9. **Tasaka S., Miyata S.,** Effects of crystal structure on piezoelectric and ferroelectric properties of copoly(vinylidenefluoride□ tetrafluoroethylene), Journal of Applied Physics. 57 (3) (1985) 906–910.

10. **Murayama N., Oikawa T., Katto T., Nakamura K.,** Persistent polarization in poly(vinylidene fluoride). II. Piezoelectricity of poly(vinylidene fluoride) thermoelectrets, Journal of Polymer Science: Polymer Physics Edition. 13 (5) (1975) 1033–1047.

Received 24.09.2020, accepted 13.10.2020.

THE AUTHORS

GOROKHOVATSKY Yuriy A.
Herzen State Pedagogical University of Russia
48 Moyka Emb., St. Petersburg, 191186, Russian Federation
gorokh-yu@yandex.ru

TEMNOV Dmitry E.
Herzen State Pedagogical University of Russia
48 Moyka Emb., St. Petersburg, 191186, Russian Federation
tde@herzen.spb.ry

SOTOVA Yulia I.
Herzen State Pedagogical University of Russia
48 Moyka Emb., St. Petersburg, 191186, Russian Federation
juliasotova1992@mail.ru

DOI: 10.18721/JPM.13404

AN IMPLICIT ECONOMICAL ALGORITHM FOR NUMERICAL INTEGRATION OF THE EQUATION SYSTEM DESCRIBING A MULTIPHASE FLOW STATE WITH COMMON PRESSURE

S.V. Bulovich

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

An economical scheme for numerical integrating the system of differential equations has been proposed for the model of a multiphase medium with a common pressure (in the barotropic approximation) in liquids. The algorithm allows us to consider an arbitrary number of liquids and admits the possibility of degeneration of this parameter to one liquid in the calculation process. The equations of states for liquids have no restriction related to the finite compressibility of the medium, i.e. the liquid can be incompressible. An implicit method for generating a solution is used. The efficiency is ensured by the fact that the algorithm for constructing the inverse matrix is based on the splitting scheme for physical processes and the solvability of equations within scalar runs. As an example, the flow variant for three liquids is considered.

Keywords: multi-fluid model, barotropic approximation, implicit algorithm, numerical simulation, Cauchy problem

Citation: Bulovich S.V., An implicit economical algorithm for numerical integration of the equation system describing a multiphase flow state with common pressure, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 35–45. DOI: 10.18721/JPM.13404

Introduction

The concept of interpenetrating continua, introduced by Rakhmatulin [1] and describing the case when the same region of space contains volume fractions of a substance in different aggregate states, is one of the most productive directions of mathematical modeling of phenomena occurring in heterogeneous media. The theoretical definition of fluids is used to identify the objects considered. A number of mathematical models have been formulated within this framework; these models can be organized into a certain hierarchy in terms of the degree to which they are complete and adequate for characterizing the processes occurring [2]. This makes it possible to differentiate the composition, level and direction of velocity, the temperatures and pressures even for the same aggregate state.

The thermodynamic and gas-dynamic functions characterizing the state of multiphase flow are found using the required number of balance ratios. Their number can be formally reduced if equivalent functions in different fluids are in equilibrium. The model introduced by Bayer and Nunziato [3] is the most complete from this standpoint: it is completely unsteady with respect to temperatures, velocities and pressures in fluids. Because the relaxation time of the pressure field in fluids is considerably less than the relaxation time of temperature and velocity, the process of pressure equalization can be excluded so that the final equilibrium state can be considered. Then, in terms of providing a complete description of the processes, the second-best model is the one assuming equal pressure in fluids. According to the common classification, these are the so-called models with a common pressure.

The mathematical model can be further simplified by considering barotropic processes in the class of flows where thermal processes do not play the governing role. In this case, the multi-fluid model consists of continuity equations, momentum equations, equations relating density and pressure for each of the liquids, and the algebraic balance ratio of the volume fraction of liquids. The system of

equations written in a quasi-one-dimensional statement for a channel of variable cross-section has the form

$$\begin{gathered}\frac{\partial\left(A\alpha_k\rho_k\right)}{\partial t}+\frac{\partial\left(A\alpha_k\rho_k u_k\right)}{\partial x}=0,\\ \frac{\partial\left(A\alpha_k\rho_k u_k\right)}{\partial t}+\frac{\partial\left(A\alpha_k\rho_k u_k^2+A\alpha_k p\right)}{\partial x}\\ -p\frac{\partial\left(A\alpha_k\right)}{\partial x}=0,\\ \sum_{k=1}^{k=K}\alpha_k=1,\ \rho_k=\rho_k\left(p\right).\end{gathered}\qquad(1)$$

where α_k is the volume fraction of kth liquid; ρ_k, kg/m^3, u_k, m/s, are its density and velocity; k = 1, 2, K (K is the total amount of liquids); p, Pa, is the pressure in the channel section, common for all liquids; A, m^2, is the cross-sectional area of the channel; t, s, is the time; x, m, is the coordinate along the channel generatrix.

System of differential equations (1) consists of terms in the form of first derivatives with respect to space and time. The source terms describing mass transfer in liquids and the processes of momentum change due to interaction of phases on the interface surface or to interaction of liquids with the channel wall are not included. Differential relations are typically not used to construct these terms and, therefore, such terms do not affect the type and properties of system of equations (1).

Numerical methods for integrating a system of differential equations are widely apply to obtain a solution to system of equations (1). More and more algorithms and various difference schemes are developed. Not only the completeness and accuracy with which the given processes are described but also the efficiency of the schemes are constantly improved. These factors are closely related.

Even focusing only on the specifics of the problem solved (steady or unsteady statement, presence or absence of discontinuities in the form of shock waves or contact surfaces, intense mass and heat transfer between phases, etc.), it is impossible to make an unequivocal choice of the method for approximating the differential equations: explicit, semi-implicit or completely implicit schemes; centered or with an offset stencil of the initial data.

Let us consider several examples.

Explicit schemes are typically based on the Lax–Friedrichs, Lax–Wendroff methods, or a combination of both, for example the MASTA scheme [4]. The flux-corrected transport method [5], generalized to the multi-fluid model, is still popular.

Semi-implicit methods [6, 7], where part of the terms of the equation are approximated on the upper time layer, allow to partially remove the restriction on the value of the time integration step.

Fully implicit algorithms are based on the Newton–Raphson iterative method, ensuring full agreement of function values on the upper time layer [8, 9]. The OLGA and KATARE [10, 11] industrial codes are implemented on the basis of fully implicit schemes.

Generally, the algorithm and the discretization stencil are chosen based on the characteristic properties of the system of differential equations solved. There are well-proven techniques for solving unsteady problems that are applicable to systems exhibiting, according to the generally accepted classification, hyperbolic or parabolic properties. However, some models for describing the unsteady state of a multiphase medium can be classified as elliptical. This primarily concerns the multi-fluid model with common pressure. It was established within the framework of the two-fluid approach (K = 2) that system of equations (1) is always non-hyperbolic if the velocities in the liquids are not equal in magnitude; the Cauchy problem is incorrect for this system of equations. In other words, formally speaking, such a system of equations cannot describe the state of a heterogeneous medium [12]. However, several techniques are known to offer some sort of solution for this problem.

Notably, if the number of compressible fluids degenerates to one, i.e., upon transition to the one-fluid approximation, hyperbolic properties are restored for the remaining equations of the system. However, if the remaining fluid is incompressible, the system of equations preserves elliptic properties. Therefore, choosing a specific algorithm for solving the system of equations with common pressure only complicates the problem.

Two main approaches have been established for this class of problems, where either hyperbolic or parabolic properties are generated for the system of equations. In both cases, the result is achieved by adding the terms which change the type of the given system of equations to the momentum balance equations. Apparently, each of the methods somewhat perturbs the initial system of equations.

In this paper, we are going to consider the approach in which the system of equations acquires parabolic properties.

Parabolization of the system of equations was chosen for a number of reasons. Firstly, this method for regularizing the solution is equally applicable for any number of liquids. Secondly, no restriction is imposed on the equation linking density and pressure. Thirdly, using iterated derivatives to regularize the computational process formally brings the initial system of equations closer to the system of Navier–Stokes equations, although the value of the transfer coefficient does not follow from the physical statement of the problem. Notably, the so-called pressure at the interface (the zone of contact between liquids) is used for the case when the hyperbolic properties of the system of differential equations are reconstructed. Recommendations on the magnitude of this pressure have been developped for the two-fluid model. It proved impossible to uncover the proof for the effectiveness of this technique within the common pressure model for a system of equations in a three-fluid statement or for more fluids.

Regularization of the Cauchy problem by using higher-order derivatives has been discussed for over 40 years now. It was established in [13] that higher-order derivatives are used to suppress short-wave perturbations. Ref. [14] confirmed that it was possible to stabilize and suppress long-wave perturbations for stratified flows. Higher-order derivatives were used in [15] in the equations of mass and momentum balance, achieving linear stability of for all wavelengths (wave numbers). Completely incompressible flows were considered in [16]. Various viscosity schemes are discussed in [17, 18].

An algorithm proposed in [19] for numerical integration of system of equations (1) is aimed at describing multi-fluid flow within the framework of the barotropic approximation for equilibrium pressure. The drawbacks of the method are also described. If the calculations are based on explicit approximation of finite-difference values, then the method is conditionally stable and has a limitation imposed on the time integration step. The main limitation in this case is associated with the stability of the calculations approximating diffusion-type terms. This can be overcome by using an explicit-implicit or fully implicit algorithm. Using the explicit-implicit approximation and, in particular, diffusion terms by the Allen–Sheng [20] or DuForth–Frankel [21] method allows to preserve the efficiency of calculating the explicit algorithm and eliminate the limitations on the time integration from modeling the diffusion process. However, the limitations due to explicit approximation of convective terms remain. These limitations can be lifted within a fully implicit algorithm. Additionally, the implicit algorithm can be expected to be more consistent with the elliptic properties of the initial system of equations.

Two techniques were used in this study to solve system of equations (1), similar to the approach we used in [19].

Firstly, the type of the given system of equations was changed to parabolic by supplementing the momentum equations with higher-order derivatives. The transfer coefficient necessary for this purpose (similar to the interpretations adopted in computational mathematics, it should be called artificial viscosity) is determined by the parameters of the problem solved and the spatial discretization step.

Secondly, an evolutionary differential equation with respect to pressure was formulated instead of the algebraic balance equation of the volume fraction.

As a result of the modifications introduced, system of equations (1) takes the following form:

$$\begin{gathered}
\frac{\partial\left(A\alpha_k\rho_k\right)}{\partial t}+\frac{\partial\left(A\alpha_k\rho_k u_k\right)}{\partial x}=0,\\
\frac{\partial\left(A\alpha_k\rho_k u_k\right)}{\partial t}+\frac{\partial\left(A\alpha_k\rho_k u_k^2\right)}{\partial x}+\\
+A\alpha_k\frac{\partial p}{\partial x}=\frac{\partial}{\partial x}\left(\nu A\frac{\partial\left(\alpha_k\rho_k u_k\right)}{\partial x}\right),\\
\frac{A}{\rho c^2}\frac{\partial p}{\partial t}+\sum_{k=1}^{K}\left(\frac{1}{\rho_k}\frac{\partial\left(A\alpha_k\rho_k u_k\right)}{\partial x}\right)=0,\\
\rho_k=\rho_k\left(p\right),\ c_k^2=\frac{\partial p}{\partial\rho_k},\ \rho=\sum_{k=1}^{K}\alpha_k\rho_k,\\
\frac{1}{\rho c^2}=\sum_{k=1}^{K}\left(\frac{\alpha_k}{\rho_k c_k^2}\right),\ \nu=\Delta x c,
\end{gathered}\tag{2}$$

where c_k, c, m/s, are, respectively the speed of sound in the kth fluid and that in a heterogeneous medium, found by the Wood formula; ρ is the density of the mixture.

The artificial viscosity coefficient was chosen based on the study conducted in [17].

Numerical method

We use the finite-volume approach for numerical integration of system of equations (2). To do this, we divide the computational domain into cells of equal length. All functions are defined relative to the center of the cell, the fluxes are generated on the edges of the cell. Let us now introduce the variables

$$q_i = \alpha_i \rho_i \text{ and } m_i = \alpha_i \rho_i u_i;$$

and also vectors

$$Q = \begin{bmatrix} \alpha_1 \rho_1 \\ \vdots \\ \alpha_K \rho_K \\ \alpha_1 \rho_1 u_1 \\ \vdots \\ \alpha_K \rho_K u_K \end{bmatrix} = \begin{bmatrix} q_1 \\ \vdots \\ q_K \\ m_1 \\ \vdots \\ m_K \end{bmatrix},$$

$$F = \begin{bmatrix} \alpha_1 \rho_1 u_1 \\ \vdots \\ \alpha_K \rho_K u_K \\ \alpha_1 \rho_1 u_1^2 \\ \vdots \\ \alpha_K \rho_K u_K^2 \end{bmatrix} = \begin{bmatrix} m_1 \\ \vdots \\ m_K \\ q_1 u_1 \\ \vdots \\ q_K u_K \end{bmatrix},$$

$$G = \nu \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \partial m_1 / \partial x \\ \vdots \\ \partial m_K / \partial x \end{bmatrix}, \; H = \frac{\partial p}{\partial x} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \alpha_1 \\ \vdots \\ \alpha_K \end{bmatrix}.$$

Here Q is the vector of conservative variables; F is the flux vector; G is the vector of diffusive transfer; H is the resulting force in the channel of variable section, associated with pressure.

The two-layer finite-volume implicit scheme for system of equations (2), written for an arbitrary ith cell, takes the form

$$\frac{(AQ)_i^{n+1} - (AQ)_i^n}{\Delta t} + \frac{(AF)_{i+1/2}^{n+1} - (AF)_{i-1/2}^{n+1}}{\Delta x} - \frac{(AG)_{i+1/2}^{n+1} - (AG)_{i-1/2}^{n+1}}{\Delta x} + AH_i^{n+1} = 0,$$

$$\left(\frac{A}{\rho c^2}\right)_i^n \frac{p_i^{n+1} - p_i^n}{\Delta t} + \sum_{k=1}^{K} \left(\frac{1}{(\rho_k)_i^n} \frac{(Am_k)_{i+1/2}^{n+1} - (Am_k)_{i-1/2}^{n+1}}{\Delta x} \right) = 0. \quad (3)$$

Let us formulate the problem with respect to the vector of increments

$$\Delta Q_i = Q_i^{n+1} - Q_i^n$$

upon transition from a known time layer to the next time layer:

$$\frac{A\Delta Q_i}{\Delta t} + \frac{(A\Delta F)_{i+1/2} - (A\Delta F)_{i-1/2}}{\Delta x} - \frac{(A\Delta G)_{i+1/2} - (A\Delta G)_{i-1/2}}{\Delta x} + A\Delta H_i = \Delta Q_i^n,$$

$$\left(\frac{A}{\rho c^2}\right)_i^n \frac{\Delta p_i}{\Delta t} + \sum_{k=1}^{K} \left(\frac{1}{(\rho_k)_i^n} \frac{(A\Delta m_k)_{i+1/2} - (A\Delta m_k)_{i-1/2}}{\Delta x} \right) = \Delta p_i^n, \quad (4)$$

where the right-hand sides of Eqs. (4) ΔQ_i^n and Δp_i^n are determined by the formulas

$$\Delta Q_i^n = -\begin{bmatrix} \frac{(AF)_{i+1/2}^n - (AF)_{i-1/2}^n}{\Delta x} - \\ -\frac{(AG)_{i+1/2}^n - (AG)_{i-1/2}^n}{\Delta x} + AH_i^n \end{bmatrix},$$

$$\Delta p_i^n = -\left[\sum_{k=1}^{K} \left(\frac{1}{(\rho_k)_i^n} \frac{(Am_k)_{i+1/2}^n - (Am_k)_{i-1/2}^n}{\Delta x} \right) \right]. \quad (5)$$

The diffusive transfer on the cell face is approximated in Eqs. (5) by standard two-point differences. The concept of characteristic properties of transfer equations is used to generate the function values at the cell face in terms of convective transfer and terms related to pressure. To obtain a sign-definite direction of velocity through the face, the mass flux and the pressure on the right face of the ith cell with the index $(i + 1/2)$ are determined by the expressions

$$\left(u_k\right)_{i+1/2} = 0.5\left(\left(u_k\right)_i + \left(u_k\right)_{i+1}\right) \geq 0;$$

$$m_{i+1/2} = m_i + 0.5\Delta x l_i^m;$$

$$M_k^2 \leq 1, p_{i+1/2} = p_{i+1} - 0.5\Delta x l_{i+1}^p;$$

$$M_k^2 > 1, p_{i+1/2} = p_i + 0.5\Delta x l_i^p;$$

$$\left(u_k\right)_{i+1/2} = 0.5\left(\left(u_k\right)_i + \left(u_k\right)_{i+1}\right) < 0;$$

$$m_{i+1/2} = m_{i+1} - 0.5\Delta x l_{i+1}^m;$$

$$M_k^2 \leq 1, p_{i+1/2} = p_i + 0.5\Delta x k_i^p;$$

$$M_k^2 > 1, p_{i+1/2} = p_{i+1} - 0.5\Delta x k_{i+1}^p,$$

where $M_k^2 = u_k^2/c_k^2$ is the Mach number.

A piecewise linear representation is used for the behavior of a function in a cell is used; the index l denotes the slopes of functions in the cells.

The following procedure was used to select the slope of the function:

$$\min \operatorname{mod}\left(\frac{f_i - f_{i-1}}{\Delta x}, \frac{f_{i+1} - f_i}{\Delta x}\right)$$

selecting the minimum absolute value among the one-sided 'backward' and 'forward' differences. These formulas provide a second-order accuracy on smooth solutions.

The procedure for linearizing the flux vector on the face was carried out using simplified dependences, so that only the most important factors characterizing the impact of the flux on the change in the function in the cell were preserved:

$$\left(u_k\right)_{i+1/2} \geq 0 \qquad \Delta\left(\alpha_k \rho_k u_k^2 + \alpha_k p\right)_{i+1/2} \approx$$
$$\approx \left(u_k^n\right)_{i+1/2}\left(\Delta m_k\right)_i + \left(\alpha_k^n\right)_{i+1/2}\Delta p_{i+1},$$

$$\left(u_k\right)_{i+1/2} < 0 \qquad \Delta\left(\alpha_k \rho_k u_k^2 + \alpha_k p\right)_{i+1/2} \approx$$
$$\approx \left(u_k^n\right)_{i+1/2}\left(\Delta m_k\right)_{i+1} + \left(\alpha_k^n\right)_{i+1/2}\Delta p_i. \qquad (6)$$

This linearization method allows to formulate a system of equations for the increments of functions on the upper time layer, based on the simplest two-point relations between equivalent functions. Since the motion of liquids is multidirectional motion, this yields tridiagonal matrix dependences.

Using the notations

$$\Lambda f_i = \left(f_i - f_{i-1}\right)/\Delta x$$

for the two-point difference operator directed 'backward' and

$$\overline{\Lambda} f_i = \left(f_{i+1} - f_i\right)/\Delta x$$

for the difference operator directed 'forward', conjugate to Λ, we obtain that system of equations (4) for the case $(u_k)^n_i \geq 0$ takes the form

$$\frac{\left(A\Delta q_k\right)_i}{\Delta t} + \Lambda\left(A\Delta m_k\right)_i = \left(\Delta q_k\right)_i^n;$$

$$\frac{\left(A\Delta m_k\right)_i}{\Delta t} + \Lambda\left(Au_k^n \Delta m_k\right)_i -$$
$$-\Lambda\left(A\nu^n \overline{\Lambda}\Delta m_k\right)_i +$$
$$+A\alpha_k^n \overline{\Lambda}\Delta p_i = \left(\Delta m_k\right)_i^n; \qquad (7)$$

$$\left(\frac{A}{\rho c^2}\right)_i^n \frac{\Delta p_i}{\Delta t} +$$
$$+\sum\left(\frac{1}{\rho_k^n}\Lambda\left(A\Delta m_k\right)_i\right) = \Delta p_i^n.$$

Importantly, the operator conjugate to the operator of convective transfer is used in each equation of momentum to approximate the pressure gradient. If it turns out that $(u_k)^n_i < 0$ at this point, then the cyclic replacement of Λ with $\overline{\Lambda}$ is used in Eq. (7).

The solution to system of equations (7) is constructed by splitting the equations of momentum into two steps by physical processes:

$$\left[\begin{array}{l} \dfrac{A}{\Delta t} + \Lambda\left(Au_k^n ..\right) - \\ -\Lambda\left(A\nu^n \overline{\Lambda}\left(..\right)\right) \end{array}\right]\Delta m_k^* = \left(\Delta m_k\right)_i^n;$$

$$\frac{A}{\Delta t}\Delta m_k + A\alpha_k \overline{\Lambda}\Delta p = \frac{A}{\Delta t}\Delta m_k^*.$$

where Δm_k^* is the auxiliary value of the function.

The first step takes into account convective and diffusive transfers in the equations of momentum, the second takes into account the terms related to pressure.

The equations of the first step are solved within standard scalar runs. The number of these equations depends on the amount of liquids K in the multi-fluid model.

An evolutionary equation for pressure is used to solve the system of equations at the second step. Substituting the value Δm_k into the evolutionary equation for pressure, we can obtain an equation of the following form for Δp:

$$\left(\frac{A}{\rho c^2}\right)_i^n \frac{\Delta p}{\Delta t} - \Delta t \sum \frac{1}{\rho_k^n} \Lambda\left(A\alpha_k \bar{\Lambda} \Delta p\right) = = \Delta p_i^n - \sum\left(\frac{1}{\rho_k^n} \Lambda\left(A \Delta m_k^*\right)\right). \quad (8)$$

Notably, depending on the direction of the fluid flow, a sequence of operators, either

$$\Lambda\left(A\alpha_k \bar{\Lambda} \Delta p\right), \text{ or } \bar{\Lambda}\left(A\alpha_k \Lambda \Delta p\right),$$

appears after the substitutions are performed; these operators are one of the possible representations of the higher-order derivative

$$\frac{\partial}{\partial x}\left(A\alpha_k \frac{\partial \Delta p}{\partial x}\right)$$

with second-order accuracy. Both operators are equivalent at the difference level. The standard representation of the higher-order derivative generates a tridiagonal matrix of coefficients in front of unknown Δp. Then the system of linear equations is resolved within scalar sweep.

After the pressure increment is found, the remaining increments of the functions Δm_k and Δq_k for all K liquids are determined by explicit formulas

$$\Delta m_k = \Delta m_k^* - \Delta t \alpha_k \bar{\Lambda} \Delta p,$$

$$\Delta q_k = \frac{\Delta t}{A}\left(\Delta q_k^n - \Lambda\left(A \Delta m_k\right)\right).$$

The formulation of the boundary conditions depends on the problem solved. Boundary conditions can be imposed relatively simply for an impenetrable boundary. If a cell outside the boundary is used, the condition of zero flux through the control surface is guaranteed by assigning functions in this cell according to the following rule: all functions must be even, except for the velocity, which is taken with the opposite sign. Similar conditions are also imposed on the increments of the functions. Permeable boundaries generate a wide variety of options; considering these options falls outside the scope of our study.

Example calculations

To illustrate how the proposed algorithm works, let us take to statements of the problem on the breakdown of an arbitrary discontinuity. The first statement describes the behavior of functions in the three-fluid approximation given asymptotic degeneration of the second and third liquids. The second statement covers a more general case.

It can be expected that the solution of the problem in the multi-fluid statement will be close to the solution of the problem in the single-fluid statement for small volume fractions of the second, third, and subsequent liquids.

For a single compressible fluid, there is an analytical solution to the problem on the breakdown of an arbitrary discontinuity, which is represented by a two-wave configuration combining waves of two types: a rarefaction wave and a shock wave. The solution to this problem has two distinctive characteristics in the barotropic approximation. First, the intensity of the shock wave is described by the Poisson adiabat. Secondly, the resulting solution always has a 'tailored' contact surface, i.e., a marker is needed to identify it, since all gas-dynamic functions on the contact surface coincide. Notice that the the barotropic approximation model describes the characteristics of a rarefaction or a compression wave without distortion. A low-intensity shock wave is also described satisfactorily.

Statement of the problem on breakdown of discontinuity in the barotropic approximation. If the states of such gas-dynamic functions as pressure, density and velocity to the left of the initial breakdown are denoted by the variables p_L, ρ_L, u_L, and the states to the right of the breakdown by p_R, ρ_R, and u_R, respectively, then the value of the functions on the contact surface P, $R=R(P)$ and U is found by solving the system of equations

$$U - u_L + \frac{P - p_L}{a_L} = 0,$$

$$U - u_R - \frac{P - p_R}{a_R} = 0.$$

where a_L, a_R are the mass fluxes through the gas-dynamic singularity in the form of a wave to the left or to the right of the initial discontinuity, respectively.

Either a rarefaction wave or a shock wave is chosen by comparing the current pressure Pon the contact surface with the pressures p_L and p_R. If a shock wave is detected to the left or to the right, the mass flux is determined by the formula

$$a_{L,R} = \sqrt{\frac{P - p_{L,R}}{\frac{1}{\rho_{L,R}} - \frac{1}{R}}}.$$

The mass flux for the left or right rarefaction wave is determined by the expression

$$a_{L,R} = \frac{\gamma - 1}{2\gamma} \rho_{L,R} c_{L,R} \frac{1 - \frac{P}{p_{L,R}}}{1 - \left(\frac{P}{p_{L,R}}\right)^{\frac{\gamma-1}{2\gamma}}},$$

where γ is the adiabatic exponent for the equation of state in the form of a power law of pressure depending on density:

$$\frac{p}{p_0} = \left(\frac{\rho}{\rho_0}\right)^{\gamma},$$

in this case p_0, ρ_0 are the values of the functions that determine the reference level.

Evidently, the values a_L and a_R depend on the solution to the problem and can be determined via an iterative procedure.

The quantitative values of the constants of thermodynamic functions used and their equations of state for fluids are summarized in Table.

The given set of dependences of density on pressure allows to confirm that the algorithm is capable of running calculations in a wide range of compressibilities of the given substances.

Let us consider a cylindrical pipe of constant cross-section, 1 m long and closed at both ends. The volumetric content of liquids at the initial time is

$$\alpha_1 + \alpha_2 + \alpha_3 = $$
$$= 0.999998 + 1.0 \cdot 10^{-6} + 1.0 \cdot 10^{-6},$$

the pressure is kept constant and equal to 7.0 MPa, and the velocity is 100 m/s.

This problem statement is noteworthy in that the given two-wave configuration is constructed in combination with the boundary conditions of the problem, i.e., the condition of impermeability through the end surfaces. A rarefaction wave appears at the left end of the pipe, propagating to the right. A shock wave appears at the right end of the pipe, propagating to the left. The intensity of the waves (pressure change) depends on the velocity of the moving gas.

The calculations in this study were carried out on a sequence of grids containing 1000 (long dashes), 2000 (short dashes), 4000 (solid line) discretization intervals of the computational domain. Fig. 1 shows the dependences of pressure and velocity for liquid no. 1 (see Table) on the longitudinal coordinate at the initial moment (t = 0) and for two subsequent times: t = 0.2 ms (line 1) and 0.4 ms (line 2). The obtained numerical values of the functions correspond to the analytical solution of the problem. It can be confirmed by direct verification for both cases that until the waves

Table

Equations of state containing thermodynamic functions with constant values for three types of liquids

#	Liquid	Equation of state
1	Vapor	$\rho = \left(\frac{p}{1.0e5}\right)^{0.714}$
2	Incompressible	$\rho = 1.0e3$
3	Weakly compressible	$\rho = 1.0e3\left(\frac{p}{1.0e6}\right)^{0.3}$

Notations: ρ, kg/m^3, is the density; p, Pa, is the pressure

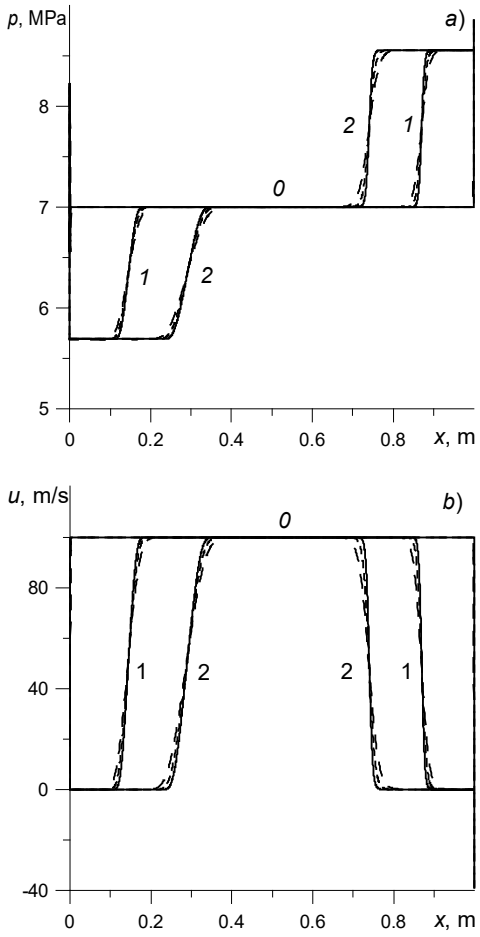


Fig. 1. Calculated dependences of pressure (*a*) and velocity (*b*) for liquid no. 1 (vapor) along the longitudinal coordinate for instants $t = 0.0$, 0.2 and 0.4 ms (curves *0*, *1* and *2*, respectively) The data were obtained on a sequence of grids containing 1,000 (long dashes), 2,000 (short dashes), and 4,000 (solid lines) discretization intervals

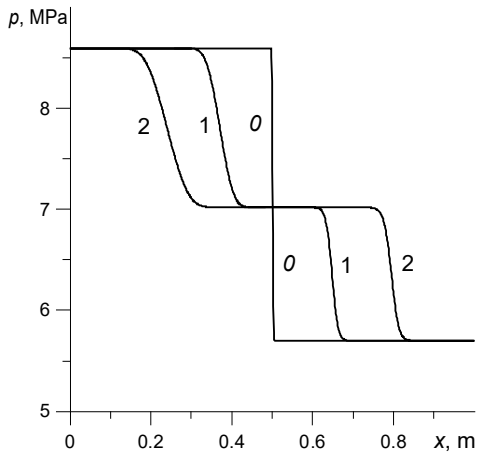


Fig. 2. Calculated pressure dependences along the longitudinal coordinate for instants $t = 0.0$, 0.2 and 0.4 ms (curves *0*, *1* and *2*, respectively)

start interacting, the gas is stationary in the vicinity of the ends of the pipe behind each of the waves.

Next, the flow in the shock tube was calculated in the three-fluid approximation on a grid containing 1000 intervals. The pressure to the left of the aperture was taken equal to 8.6 MPa, and to 5.7 MPa to the right. The volume fractions of liquids were

$$\alpha_1 + \alpha_2 + \alpha_3 = $$
$$= 0.8000 + 0.1999 + 1.0 \cdot 10^{-4}.$$

Similar dependences were obtained for the behavior of gas-dynamic functions on the longitudinal coordinate at the initial time and at times $t = 0.2$ and 0.4 ms (curves *1* and *2*, respectively).

Fig. 2 shows the pressure distribution, and Figs. 3–5 the distributions of velocities, densities and volume fractions. Without the terms accounting for transfer of momentum and mass between liquids or different interactions of liquids with the channel wall, system of equations (1) is purely model and serves only to illustrate the performance of the algorithm. There is however one phenomenon that the given system of equations can illustrate. It is the problem on the breakdown of discontinuity with a jump in the channel's cross-sectional area. In the single-fluid approximation, the area ratio is a parameter that makes the problem non-selfsimilar but allows for parametric study. Another discontinuity appears at a jump in the channel's cross-sectional area. Going back to the problem on the breakdown of discontinuity in a channel of constant cross-section but for the case of a multi-fluid flow, we can regard the current values of the volume fractions in the cross section as elements of the area. In situations where the condition of dynamic compatibility of the equations solved leads to a jump-like change in the volume fraction, the corresponding discontinuities can be observed in the behavior of pressure and fluid flow rates (see Figs. 2 and 3).

Conclusion

We have proposed a model of a multiphase medium with a common pressure in liquids, along with a resource-efficient scheme for numerical integration of a system of differential equations. Let consider the key points of the proposed algorithm.

Firstly, no limitations are imposed on the number of liquids.

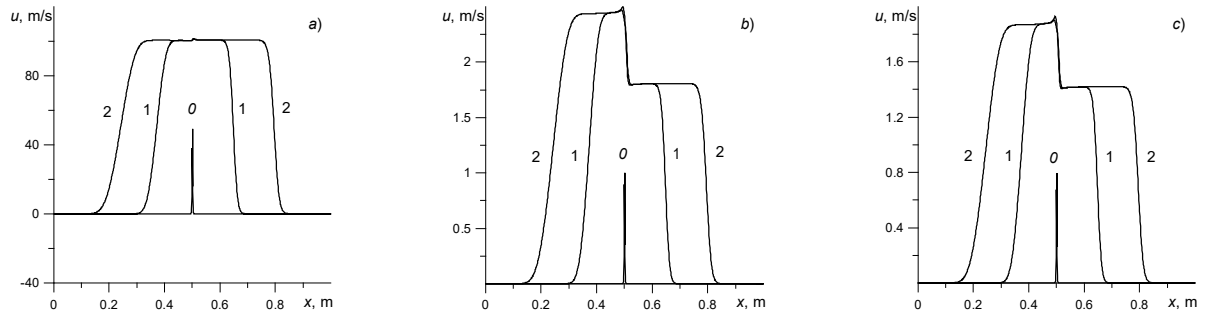


Fig. 3. Distributions of velocities for liquids no. 1 (*a*), no. 2 (*b*), no. 3(*c*)) (see Table) along the longitudinal coordinate for instants $t = 0.0$, 0.2 and 0.4 ms (curves *0*, *1* and *2* , respectively)

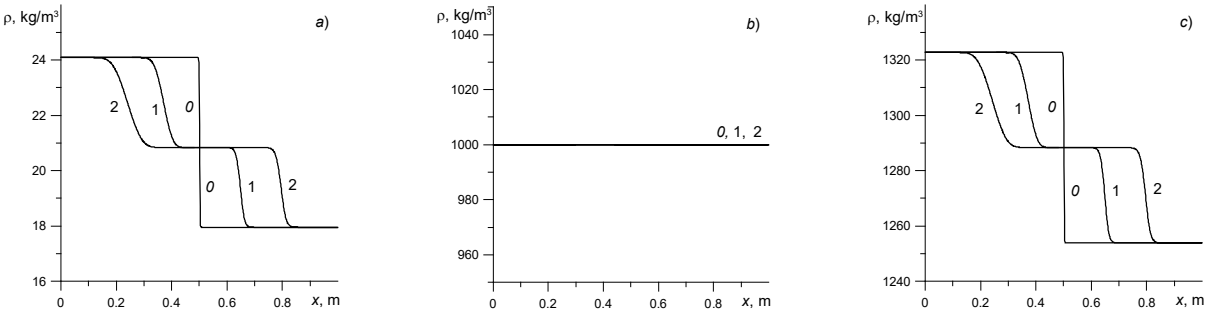


Fig. 4. Distributions of densities for liquids no. 1 (*a*), no. 2 (*b*), no. 3(*c*)) (see Table) along the longitudinal coordinate for instants $t = 0.2$ and 0.4 ms (curves *0*, *1* and *2* , respectively)

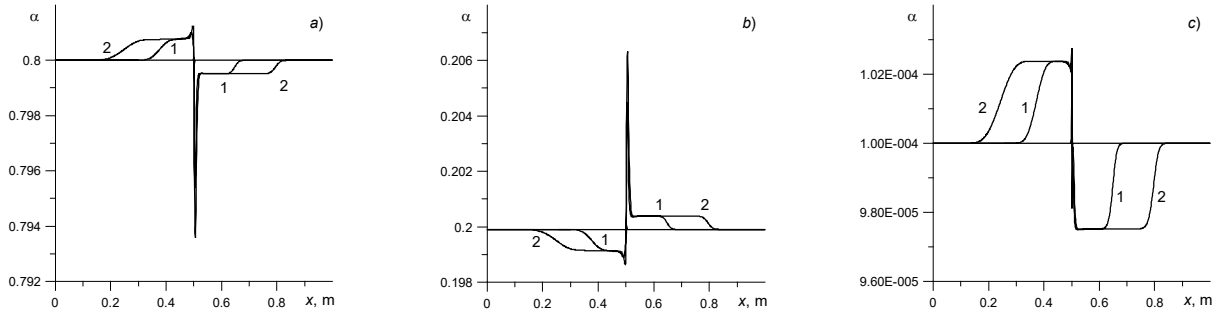


Fig. 5. Distributions of volume fractions for liquids no. 1 (*a*), no. 2 (*b*), no. 3 (*c*) (see Table) along the longitudinal coordinate for instants $t = 0.2$ and 0.4 ms (curves *1* and *2*, respectively)

Secondly, the technique regularizing the Cauchy problem using higher-order derivatives allows to avoid distortions to the propagation velocity of acoustic perturbations.

Thirdly, the proposed algorithm has certain advantages over other implicit methods, including, importantly, the following:

the algorithm does not require iterations and does not require a matrix run with a block dimension equal to $2K + 1$, corresponding to the number of unknowns in the problem;

calculations by the proposed algorithm require performing $K + 1$ scalar sweeps and $2K$ calculations using explicit formulas. Compared with the implicit algorithm for non-iterative computations, this allows to reduce the number of operations per discretization point by $(2K + 1)^2$ times.

REFERENCES

1. **Rakhmatulin Kh.A.**, Osnovy gazodinamiki vzaimopronikayushchikh dvizheniy szhimayemykh sred [Gas-dynamics fundamentals of interpenetrating motions of compressible media], Journal of Applied Mathematics and Mechanics. 20 (2) (1956) 184–195 (in Russian).

2. **Linga G., Fletten T.,** A hierarchy of non-equilibrium two-phase flow models, ESAIM: Proceedings and Surveys. 66 (9) (2019) 109–143.

3. **Baer M., Nunziato J.,** A two-phase mixture theory for the deflagration-to-detonation transition (DDT) in reactive granular materials. International Journal of Multiphase Flow. 12 (6) (1986) 861–889.

4. **Munkejord S.T.,** A numerical study of two-fluid models with pressure and velocity relaxation, Advances in Applied Mathematics and Mechanics. 2 (2) (2010) 131–159.

5. **Figueiredo A.B., Baptista R.M., Rachid F.B.F., Bodstein G.C.R.,** Numerical simulation of stratified-pattern two-phase flow in gas pipelines using a two-fluid model, International Journal of Multiphase Flow. 88 (January) (2017) 30–49.

6. **Evje S., Flatten T.,** On the wave structure of two-phase flow models, SIAM Journal on Applied Mathematics. 67 (2) (2007) 487–511.

7. **Yoon H.Y., Park I.K., Kim Y.I., et al.,** A fast-running semi-implicit numerical scheme for transient two-phase flows on unstructured grids, Numerical Heat Transfer, Part B: Fundamentals: An International Journal of Computation and Methodology. 56 (6) (2010) 432–454.

8. **Frepoli C, Mahaffy J.H., Ohkawa K.,** Notes on the implementation of a fully-implicit numerical scheme for a two-phase three-field flow model, Nuclear Engineering and Design. 225 (12) (2003) 191–217.

9. **Nascimento J.C.S., Santos A., Pires A.P.,** A fully-implicit solution for the single-pressure two-fluid model with sharp discontinuities, Computers and Fluids. 175 (15 October) (2018) 214–229.

10. **Barre F., Bernard M.,** The CATHARE code strategy and assessment, Nuclear Engineering and Design. 124 (3) (1990) 257–284.

11. **Bendiksen K.H., Maines D., Moe R., Nuland S.,** The dynamic two-fluid model OLGA: theory and application, SPE Production Engineering. 6 (2) (1991) 171–180.

12. **Klebanov L.A., Kroshilin A.E., Nigmatulin B.I., Nigmatulin R.I.,** O giperbolichnosti, ustoychivosti i korrektnosti zadachi Koshi dlya sistemy differentsialnykh uravneniy dvukhskorostnogo dvizheniya dvukhfaznykh sred [On hyperbolicity, stability and correctness of the Cauchy problem for a differential equations system of the two-speed motion of two-phase media], Journal of Applied Mathematics and Mechanics. 46 (1) (1982) 83–95 (in Russian).

13. **Arai M.,** Characteristics and stability analyses for two-phase flow equation systems with viscous terms, Nuclear Science and Engineering. 74 (2) (1980) 77–83.

14. **Holmas H., Sira T., Nordsveen M., et al.,** Analysis of a 1D incompressible two-fluid model including artificial diffusion, IMA Journal of Applied Mathematics. 73 (4) (2007) 651–667.

15. **Vreman A.W.,** Stabilization of the Eulerian model for incompressible multiphase flow by artificial diffusion, Journal of Computational Physics. 230 (4) (2011) 1639–1651.

16. **Fullmer W.D., Lee S.Y., Bertodano M.A.L.D.,** An artificial viscosity for the ill-posed one-dimensional incompressible two-fluid model, Nuclear Technology. 185 (3) (2014) 296–308.

17. **Bulovich S.V., Smirnov E.M.,** Experience in using a numerical scheme with artificial viscosity at solving the Riemann problem for a multi-fluid model of multiphase flow, AIP Conference Proceedings. 1959 (1) (2018) 050007.

18. **Smirnov S.I., Bulovich S.V., Smirnov E.M.,** Calculation of unsteady two-phase quasi-onedimensional channel flow based on the two-fluid model and the artificial viscosity numerical scheme, Journal of Physics: Conference Series, 2018. Vol. 1135 (Proceedings of International Conference PhysicA.SPb/2018. Saint-Petersburg, October, 23–25) (2018) 012103.

19. **Bulovich S.V.,** The system of multiphase flow equations in the equilibrium barotropic approximation: a numerical scheme, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 12 (3) (2019) 17–25.

20. **Allen J.S., Sheng S.I.,** Numerical solution of the compressible Navier –Stokes equation for the laminar near wake, Physics of Fluids. 13 (1) (1970) 37–52.

21. **DuFort E.C., Frankel S.P.,** Stability conditions in the numerical treatment of parabolic differential equations, Mathematical Tables and Other Aids to Computation. 7 (43) (1953) 135–152.

Received 15.07.2020, accepted 03.09.2020.

THE AUTHOR

BULOVICH Sergei V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
bulovic@yandex.ru

DOI: 10.18721/JPM

THERMOKINETIC PROCESSES IN THE CORIUM TRAPS OF HIGH-TEMPERATURE MELT DURING THE REACTOR ACCIDENTS AT AN A-PLANT: SIMULATION

A.N. Kovalenko[1,3], A.O. Koptyukhov[2], D.K. Meshcheryakov[3], A.P. Schuklinov[3,4]

[1] Ioffe Physical Technical Institute of the Russian Academy of Sciences, St. Petersburg, Russian Federation;

[2] NRC "Kurchatov Institute" – PNPI, Gatchina of Leningrad region, Russian Federation;

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

[4] JSC "Atomproekt", St. Petersburg, Russian Federation

The two-stage model of changing the thermal and phase state of the active zone melt in the cooled subreactor crucible trap when interacting with the sacrificial material and when crystallizing the diluted corium after the gravitational inversion of its oxide and metal components is presented. The simulation is based on the generalized formulation of Stefan's task. The results of the end-to-end calculation of the COMSOL Multiphysics package show that the thermokinetic processes examined generally reduce the temperature and density of the corium, reduce heat flows on the trap body, minimize the release of hydrogen and radioactive fission products with their retention in a controlled thermal and phase state until full crystallization.

Keywords: melt active zone, subreactor crucible trap, sacrificial material, thermal state

Citation: Kovalenko A.N., Koptyukhov A.O., Meshcheryakov D.K., Schuklinov A.P., Thermokinetic processes in the corium traps of high-temperature melt during the reactor accidents at an a-plant: simulation, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 46–58. DOI: 10.18721/JPM.13405

Introduction

Data on severe accidents at nuclear power plants indicate that the physical causes are related to critical disruption of heat removal from the reactor core, with reactor elements overheating and melting as a result [1–3]. Technical solutions for retaining the corium melt (a mixture of uranium, plutonium, zirconium oxides, fragments of nuclear fuel fission, absorbing materials, metals of support structures such as iron, nickel, chromium, etc.) in the reactor, developed for external vessel cooling in reactors with low and medium power [4], proved to be unreliable for preventing wall melting in high power reactor vessels. As an alternative [5], the concept of ex-vessel melt retention devices (MRD) was proposed for this type of reactors as an additional barrier for the propagating reaction products, confining them in a controlled thermal and phase state until complete crystallization.

For European EPR reactors, this concept is implemented based on a lateral compartment for melt spreading, cooled from above by water [6]. Subreactor crucible traps with filler made of sacrificial material (SM), cooled by water from the outside, are used as heat exchangers in Russian VVER reactors [7] (Fig. 1).

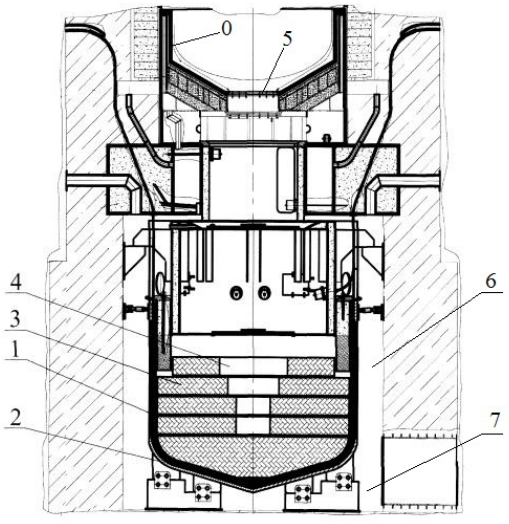


Fig. 1. Schematic of crucible-type melt retention devices for Russian NPPs with VVER reactors
reactor vessel *0*; cooled trap *1*; MRD supports *2*; pellets *3* of sacrificial material (SM); assembly *4* of the container holding the SM; melt plug *5*; lateral heat exchanger *6*; channel supplying cooling water and lower heat exchanger *7* (according to RF patent 2514419 [7])

Combining the functions of a water-cooled crucible for trapping the melt and a collector with the SM in the MRD greatly expands the range of problems considered, especially since the composition of the incoming melt may be unclear and there is insufficient data on its interaction with the sacrificial material [8–10].

Full-scale studies of these processes are impossible; moreover, small-scale setups only offer limited possibilities to reproduce them. For this reason, mathematical modeling is very important for understanding the physical mechanisms underlying the retention of the corium melt captured by the trap in controlled thermal and phase states, as well as for substantiating the technical solutions adopted.

Notably, solving these problems via well-known specialized software codes developed for modeling various scenarios of severe accidents and their individual stages at the in-reactor stage may be considerably difficult for a number of reasons. Practical experience accumulated in computational studies of crucible traps [11–15] with some of these codes (SOCRAT, GEFEST-ULR, KORSAR, ASTEC), as well as benchmarking of these and other codes at the Joint Research Center (JRC) of the Directorate for Energy, Transport and Climate of the European Commission (Petten, Netherlands) [16] indicates that there is a large scatter of data both between different codes and between different users of the same code. This can be largely due to the differences not only in approaches to constructing physical and mathematical models but also in nodalization schemes, computational algorithms and criteria for evaluating the results obtained.

In this regard, the approach accounting for the interrelated factors characterizing thermokinetics and chemical phase, with the possibility of end-to-end computations of conjugated mechanical, thermal, hydrodynamic and chemical phenomena has certain benefits for simulating the processes of interaction of the sacrificial material with the melt in a crucible trap and subsequent changes in its state up to complete crystallization.

Thermokinetic processes in the crucible trap

The processes of interaction between the high-temperature corium melt and the energy-absorbing components of the sacrificial material melting in the corium, consisting of a mixture of light iron oxides Fe_2O_3 and aluminum Al_2O_3, occur in a water-cooled crucible trap. The above processes must meet certain requirements:

provide a reduction in corium temperature and density, as well as a reduction in heat fluxes to the trap body;

preserve the subcriticality of the melt;

minimize the yields of hydrogen and radioactive fission products [9, 10, 17, 18].

The relationships between mass and heat transfer, substance transformations and conversion of various types of energy in these processes are determined by a set of coupled physical phenomena of mechanics, thermal physics, hydrodynamics, and nuclear physics.

A significant role in these relationships is played by chemical reactions of endothermic reduction of oxides from the composition of the sacrificial material and exothermic oxidation of unoxidized elements of the melt, including zirconium by the redox reaction

$$2/n\ \mathrm{MeO}_n + \mathrm{Zr} \rightarrow \mathrm{ZrO}_2 + 2/n\ \mathrm{Me}$$

(Me is the metal, n is its oxidation degree), helping to minimize hydrogen release during vapor-zirconium contact.

The thermal effects of these reactions are comparable in power to the magnitude of residual heat release of fission products and their nuclear transformations [9, 10, 19]. It is of great practical interest to explore the mechanism of interaction between the sacrificial oxide material and the melt containing strong reducing agents in the liquid-phase combustion mode with a total endothermic effect in the reaction zone [20]. It can occur if the products of chemical interaction are in liquid or dispersed in liquid state with both metalized and oxidized component of the melt.

Additional dilution of corium due to the unreacted part of the melted iron and aluminum oxides from the composition of SM leads to a noticeable decrease in the density of the oxide components of the melt initially appearing, heavier compared to the metal ones, and the subsequent gravitational inversion of their initial stratification. This makes it possible to additionally supply cooling water to the melt surface without the risk of steam explosion [21].

In this case, a refractory skull crust forms under non-isothermal conditions, preventing mass and heat transfer both on the surface of the corium and between its immiscible oxide and metal layers. There is also a risk that crystallization zones of individual melt components may form, with secondary centers of increased radioactivity appearing where heavier (relative to the entire melt) oxides of radioactive matter locally accumulate. To

preserve subcriticality, the composition of SM includes immobilizing radionuclide additives of refractory heavy elements, such as gadolinium oxide, which absorb thermal neutrons [18]. Localization of heavy oxides with large residual heat generation can also lead to boiling of uranium dioxide with release of components into the gas phase, with an ablative interaction with the steel wall of the trap body itself, melting with insufficient external cooling [8].

Correct analysis of the processes occurring in the interaction of corium and sacrificial materials is associated with taking into account the phase transformations of the components when changing the composition and thermodynamic parameters of the system, reflected in experimental or calculated phase state diagrams.

Experimentally obtaining the full diagram for this system is very resource-intensive and the proportions of the corium composition are uncertain, due to the large number of parameters of the possible mechanism for its formation in the nodalization scheme adopted (according to the estimates conducted in [22], the number of parameters is about 10^3—10^4), so this stage replaced by construction of computed diagrams characterizing the composition, conditions, and properties based on the chemical thermodynamics of phases of variable composition [9].

Problem statement

We consider the phase equilibrium approximation of multicomponent thermodynamic systems with variable composition for a two-stage mathematical model of thermokinetic processes in the corium melt trap, describing the set of thermal, chemical and hydrodynamic phenomena occurring, as well as correlated thermodynamic factors, taking into account the results of laboratory studies at small-scale facilities [23].

At the first, 'pre-inversion' stage, the mathematical model is formulated as a Stefan problem for a material environment of the corium trap with mobile heat sources of endothermic decomposition of the melting SM, heat of exothermic reactions of chemical interaction with the SM of underoxidized zirconium, chromium, and uranium entering into the trap in the core melt, and also the heat of residual heat release of radioactive elements. In this case, the isotherm corresponding to the melting temperature of the SM components divides the computational domain into subregions, where the condition of continuity for temperature and the condition for changing the heat flux density by a value equal to the product of the enthalpy of phase transition by the normal component of the velocity of the interface are imposed at the boundary. In a generalized form [24], this problem is reduced to a single nonlinear heat conduction and mass transfer equation in the entire melt region, which includes the above nonuniform conjugation conditions for temperature and heat flux at the previously unknown (free) interfaces directly into the equation itself:

$$c_{eff}\rho\left(\frac{\partial T}{\partial \tau}+\mathbf{v}\,\mathrm{grad}\,T\right)=\mathrm{div}\left(\lambda\,\mathrm{grad}\,T\right)+f. \quad (1)$$

Here ρ, kg/m^3, is the density of the medium component; T, K, is the temperature; τ, s, is the time; **v,** m/s, is the local velocity of the medium (vector); λ, W/(m·K), is the heat transfer coefficient; the term (**v** grad T) accounts for the temperature change due to convective transport of the medium; f, W, defines the power of internal heat sources; c_{eff}, J/(kg·K), defines the effective specific heat capacity, accounting for the heat of the phase transition at its boundary in concentrated form.

During melting, the derivative function for the effective heat capacity is as follows:

$$c_{eff}=c_{p,\mathrm{sol}}\cdot(1-\varphi(T))+$$
$$+c_{p,\mathrm{liq}}\cdot\varphi(T)+\delta(T-T_{\mathrm{mel}})L\frac{d\varphi}{dT}, \quad (2)$$

where $c_{p,sol.}$, $c_{p,liq}$, J/(kg·K), are the heat capacities of the solid and liquid phases, respectively; $\varphi(T)$ is the phase transition fraction function determined by the equilibrium phase diagram of the system; L, J/kg, is the specific heat capacity of the phase transition; $\delta(T)$ is the delta function; T_{mel}, K, is the phase transition temperature.

The model takes into account heat fluxes not only in the reaction layer at the mobile boundaries of phase transitions but also in the volume of interacting materials, including those caused by residual heat release from products of nuclear transformations and heat exchange of the corium melt with the SM. The intensity of the heat exchange depends both on the kinetics of the ongoing chemical reactions, which depends in turn on the temperature conditions of activation (primarily, the initial temperature of the incoming melt), and on the mixing rate of the reaction products with the corium melt, which is assumed to be instantaneous.

The heat transfer during natural convection in the melt is based on unsteady Reynolds-averaged Navier–Stokes equations for a viscous incompressible fluid, which determine the effect of convective velocity fluctuations in the form of turbulent kinetic energy of the fluctuations and its dissipation due to internal friction:

$$\rho\frac{\partial \mathbf{v}}{\partial \tau}+\rho(\mathbf{v}\cdot\nabla)\mathbf{v}= \\ =\nabla[-p\mathbf{I}+(\mu+\mu_m)(\nabla\mathbf{v}+(\nabla\mathbf{v})^{\mathrm{T}})- \\ -\frac{2}{3}(\mu+\mu_B)(\nabla\mathbf{v})\mathbf{I}-\frac{2}{3}\rho k\mathbf{I}]+\rho\mathbf{g}. \quad (3)$$

Here ρ, kg/m^3, is the bulk density; **v**, m/s, is the flow velocity (vector); p, Pa, is the pressure; μ, Pa·s, is the dynamic viscosity; μm, Pa·s, is the turbulent viscosity; **g**, m^2/s, is the free fall acceleration (vector); k, J, is the turbulent kinetic energy (turbulent fluctuations, $k=\Sigma_i(\overline{V_i'^2}/2)$; the subscript T denotes the transposed tensor; ∇ is the Hamiltonian operator; **I** is the unit tensor.

The equation for turbulent kinetic energy k and its dissipation rate ε, due to viscous friction

$$\mu_m=\rho c_\mu k^2/\varepsilon,$$

is closed in accordance with the standard (k-ε) turbulence model [25].

In the given conditions, when the intensity of turbulent heat transfer is caused by the thermogravitational mechanism of turbulent energy generation, its value depends on the established regime of thermal stratification of oxide and metal layers of the corium. The stability of thermal stratification depends on the direction of heat supply and removal through the boundaries of the layer relative to the vector of gravity.

Eq. (3) is solved together with the continuity equation:

$$\nabla\cdot(\rho\mathbf{v})=0. \quad (4)$$

For hydrodynamic equations, the usual no-slip conditions are imposed at the contact interface with the body, and the sliding condition along the symmetry axis and the upper free boundary (the normal velocity components are equal to zero). In the case of crust formation, the no-slip condition is also imposed at the upper boundary.

For the temperature problem on the side and bottom faces of the melt pool, contacting the cooled walls of the trap body, third-kind boundary conditions of heat transfer are imposed at a given temperature T_{cool} of the cooling water and the heat transfer coefficient α:

$$q_{cool}=-\lambda(\partial T/\partial h)=\alpha(T-T_{cool}), \quad (5)$$

where h is the normal to the surface, the heat transfer coefficient α was computed for well-developed steady flow in the annular channel of the water cooling jacket of the outer walls of the trap body according to recommendations in [26].

The initial thermal state of the filled trap corresponds to the temperature of the oxide and metal components of the melt of 3000 K and the temperature of the sacrificial material of 400 K.

In the second, 'post-inversion' stage of the melt pool at $L=0$ and $\varphi(T)=1$, the cooling processes of the melt components after the gravitational inversion of its oxide and metal components are simulated, taking into account natural convection, external cooling of the trap body, and radiative heat transfer over the pool surface. The oxide and metal components of the inverted melt layers are assumed to be homogeneous in structure. The given third-kind boundary conditions are preserved for heat transfer on the side and bottom surfaces of the melt pool contacting with the cooled walls of the trap body.

The conditions on the free surface of the pool, bordering the emitted aerosols, are complemented by accounting for heat transfer by radiation

$$\lambda(\partial T/\partial h)+\alpha(T-T_{cool})+ \\ +\acute{\varepsilon}\sigma(T^4-T_{mean}^4)=0, \quad (6)$$

where έ is the radiation coefficient (emissivity) of the melt surface, σ, W·m^{-2}·K^{-4}, is the Stefan–Boltzmann constant, T_{med}, K, is the temperature of the medium above the corium mirror.

The heat transfer coefficient α was taken into account based on the data from [21], when cooling water was supplied to the surface of the oxide melt using the dependence for film boiling of water on a solid upward-facing surface according to the recommendations in [26].

Initial conditions for this simulation stage are imposed according to the results of computations at the previous stage.

Decreasing the temperature of the melt components during cooling to their crystallization temperature at this model stage again leads to the Stefan problem. The differences between this statement and the one presented above are that the chemical heat of the interaction between the melt and the SM is not taken into account, while potential formation of spatial crystallization zones is accounted for, including those with a variable phase transition temperature, depending on

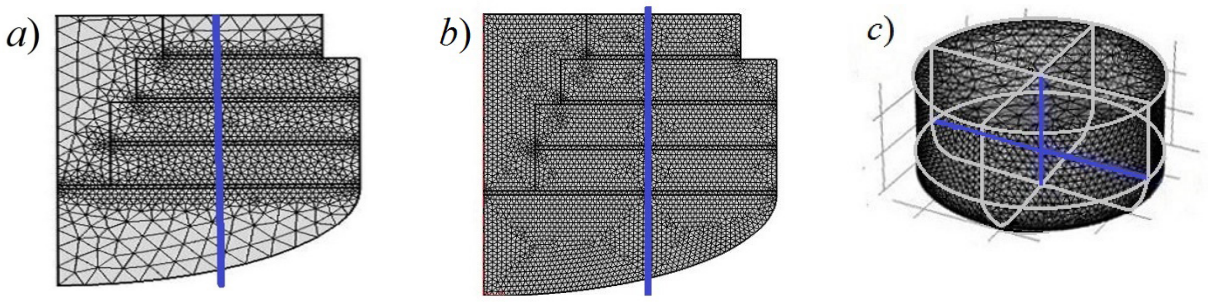


Fig. 2. Generated finite-element meshes for computed trap regions with selected control sections (blue lines) at different stages: at the 'pre-inversion' stage taking into account the initial location of the iron pellets (*a*) and their melting (*b*), and at the 'post-inversion' stage taking into account the location of the lighter oxide layer over the heavier metal layer, in the corium melt pool relative to the marked horizontal interface (*c*)

the location of solidus and liquidus lines in the phase diagram of the corium medium. Methods of thermodynamic analysis based on conditional minimization of Gibbs free energy using a set of different solution models are used to construct them [9].

Computational aspects

Simulation of the computational model was performed numerically using the COMSOL Multiphysics software package developed for modeling related multiphysics mechanical, thermal, electromagnetic, chemical, and hydrodynamic phenomena and systems [27]. This made it possible to carry out end-to-end computation of thermokinetic processes with phase transformations in the cooled high-temperature corium melt during its interaction with the sacrificial material both at the level of an individual structural element [28] and the SM container as a whole. Compatibility of the package with programming languages of Matlab/Simulink, AutoCAD, SolidWorks, Excel systems allows to include additional models for computing boundary conditions, setting material properties, etc.

The special modules available in the program automatically provide a suitable numerical discretization and solver configuration for a given mathematical model (1)–(6). In this case, the set is implemented based on the finite element method using the method of end-to-end counting with smoothing of coefficients without explicit separation of free interphase boundaries in the Stefan problem. After finding the solution, these boundaries were identified as isotherms corresponding to the temperatures of the considered phase transitions.

Initial geometric parameters of the trap, mass, composition and properties of various media and corium melt components, values of residual heat release power, characteristics of the sacrificial material and kinetics of chemical reactions were taken based on MRD designs for 1000–1200 MW VVER reactors.

The geometry of the problem solved was assumed to be two-dimensional axisymmetric. The generated meshes of the computational domain for the considered problems are shown at different simulation stages in Fig. 2.

The boundary layer was resolved using an anisotropic grid whose step decreased at the boundaries of the computational domain. The minimum size of the structural mesh element was 0.003 m and the maximum size was 0.119 m, for a total of 4,076,212 elements and the size of the computational domain of 2.74×2.85 m. The performed computations were compared with the results obtained at another minimum discretization step of 0.016 m, while the practical convergence of the solution remained at the same level.

Computational construction of the phase diagrams involved [9] for the components in the interaction system of the corium melt and sacrificial material was based on the TernAPI program for thermodynamic analysis, freely available from the website of Chemical Department of Moscow State University [29], using the algorithm of convex shells proposed.

Computational results and discussion

The results for computed changes in the thermal and phase states of the corium melt entering the trap during its interaction with the melting sacrificial material at the first, 'pre-inversion', stage of the simulation are shown in Figs. 3 and 4.

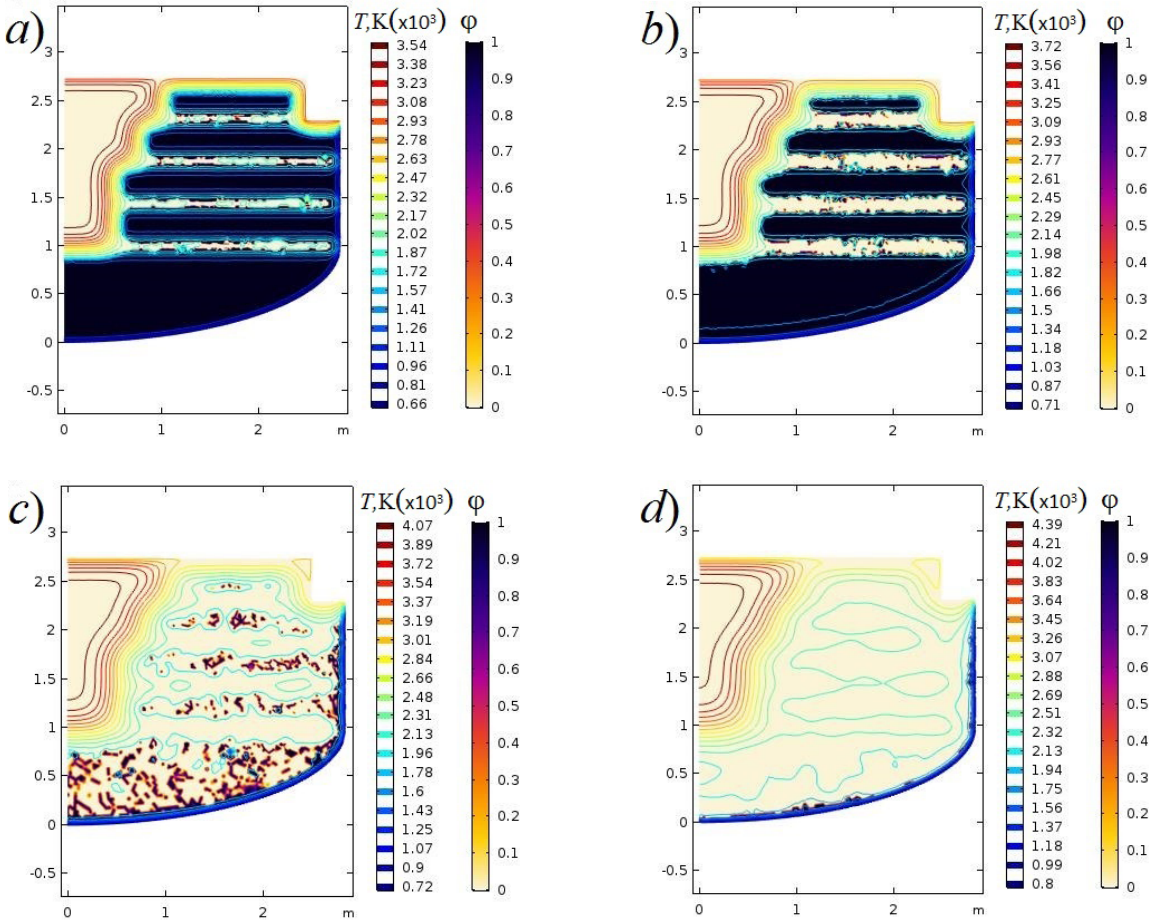


Fig. 3. Changes in temperature fields (*T,* K) and ratios φ of solid and liquid phases (dark and light regions, respectively) for corium and SM melt after it enters the trap and before inversion of oxide and metal components at different time points, min: in 10 (*a*) 30 (*b*), 60 (*c*), 90 (*d*). The ratio φ of solid to liquid phases is reflected by the color saturation in the range from 1 to 0

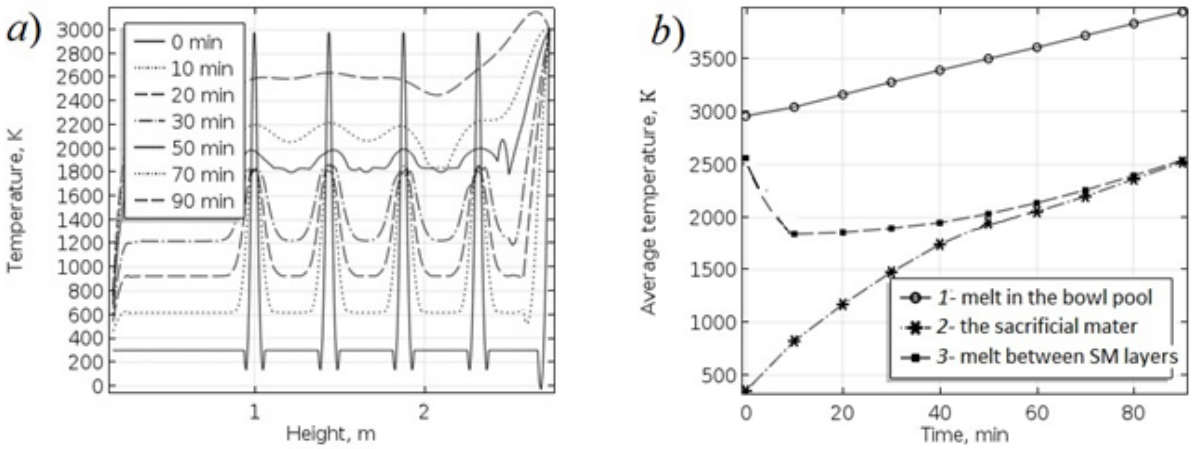


Fig. 4. Temperature distribution profiles in the control section of the trap (see Fig. 2,*b*) for different moments in time (*a*); mean temperatures over the corium (*1*) and SM (*2*) melt at the stages after entering the trap and before inversion of oxide and metal components (*b*). The stages shown are: melt inside the pool bowl *1*, SM *2*, melt between the pellet layers of SM *3*

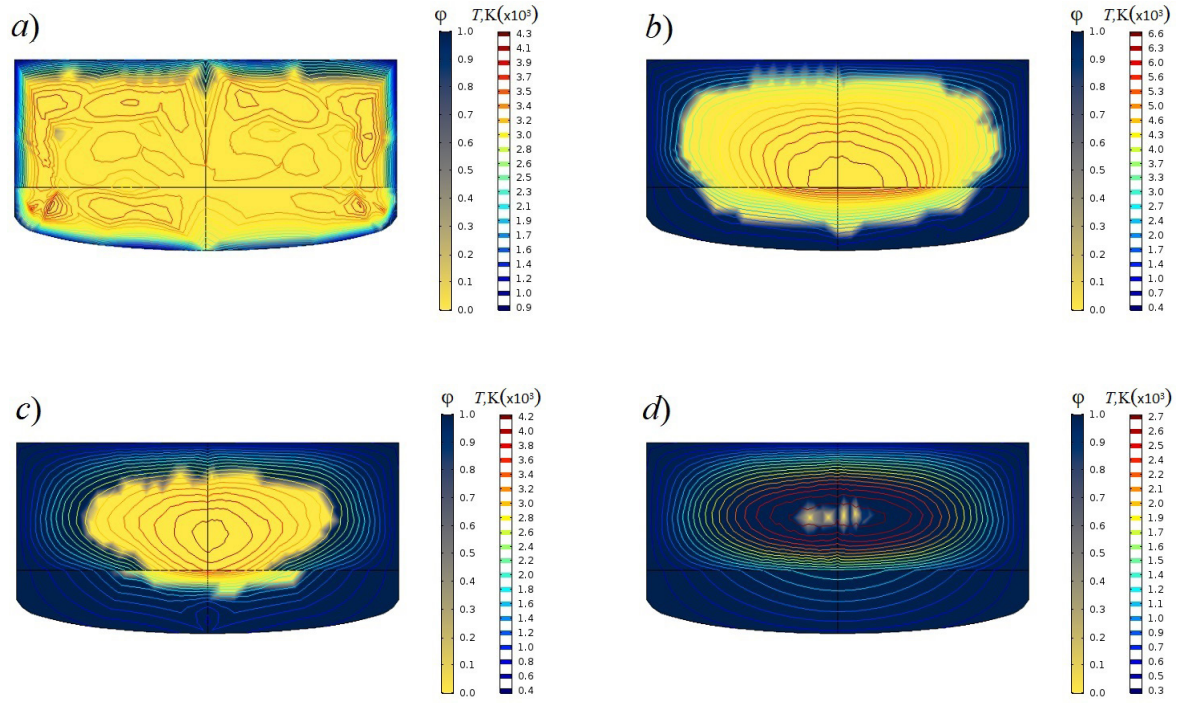


Fig. 5. Evolution of fields of temperature (T, K) and ratios φ for solid and liquid phases (dark and light regions, respectively) of the corium melt after its interaction with SM and inversion of oxide and metal components at different moments of time, h: after 5 (*a*) 40 (*b*), 65 (*c*), 78 (*d*)

Evidently, in the initial period, this interaction is accompanied by a decrease in the corium temperature, despite the residual heat release, due to the heat input for heating and melting of the sacrificial material. Thermokinetic processes are characterized by rather abrupt cooling of the high-temperature melt between the filler modules, up to its short-term solidification, which is replaced by overall melting together with the structural elements of the SM with redox reactions occurring between them.

Next, there is an increase in the temperature of the melt under the influence of residual energy release of radioactive elements. Together with additional dilution of the corium due to the unreacted part of the melted iron and aluminum oxides from the composition of the SM this leads to a noticeable decrease in the density of the oxide components of the melt initially arrived, heavier compared with the metal ones, and their rapid reverse gravitational inversion. Taking into account the estimated high rate of surfacing of the oxidized component and movement to the trap bottom of the metallized component [23], the duration of these processes is approximately 20–25 min at the stage of initial corium cooling and about 90–120 min before the final melting of the SM with inversion of the initial stratification of the metal and oxide layers of the melt.

Figs. 5 and 6 show computed results for the changes in thermal and phase states of the corium melt after chemical interaction with the SM and inversion of immiscible oxide and metal components of different densities, with molten oxides moving to the upper part and molten metals to the lower part of the melt pool.

The results of computations using the equations of motion for natural convection of the corium melt components are shown in Fig. 7.

Apparently, for the given directions of heat supply and removal through the boundaries of the upper oxide layer of the melt relative to the gravity vector, upward motion occurs in its central region and downward motion at the periphery near the cooled walls of the crucible. The motion of flows in the lower metal layer has an opposite character with unsteady thermal stratification under the interface with the oxide layer of the melt. According to the experimental data in [30], generation of turbulent flows of metal and oxide melts in free thermogravitational convection appears with the product of Grashof and Prandtl criteria

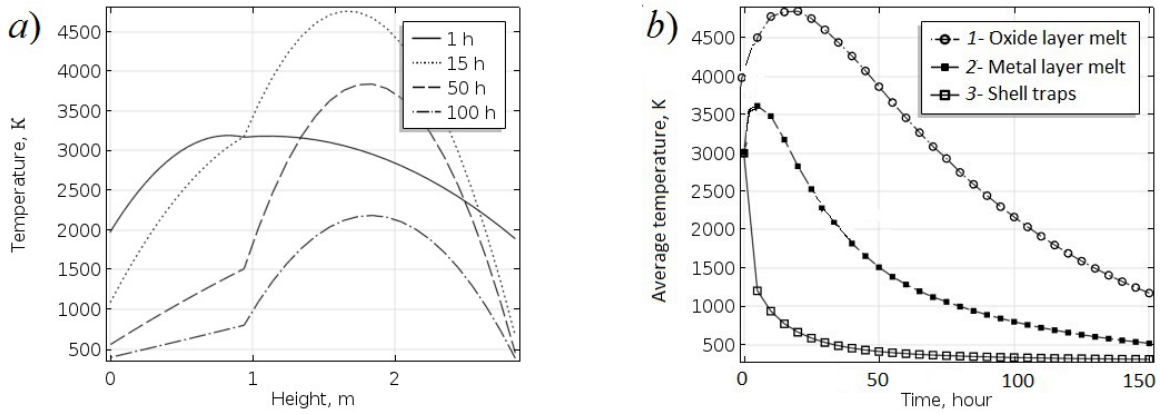


Fig. 6. Temperature distribution profiles in control section of the trap (see Fig. 2,*c*) for different moments of time (*a*) and average temperature values over the corium melt after inversion of oxide and metal components (*b*).
Data for oxide (*1*) and metal (*2*) melt layers, as well as for the trap shell (*3*) are shown

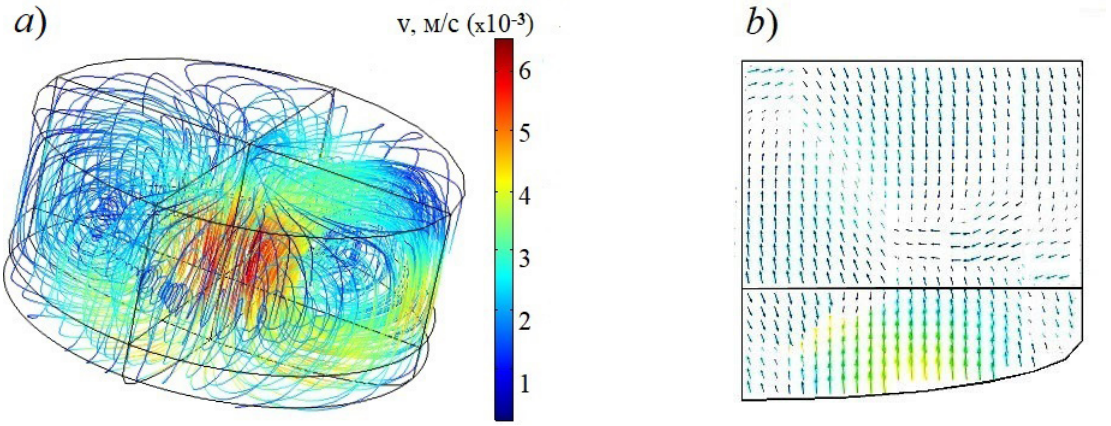


Fig. 7. Streamlines in the melt pool (*a*) and velocity field in its vertical section (*b*)

exceeding 47,000. For the parameters of the problem considered, this condition is satisfied with a large margin. Similar phenomena will be observed near the solidifying phase interface, since the melt is always superheated relative to the crystallization temperature.

After inversion of the melt components, the temperature of its oxide layer, where the sources of residual heat release are mainly concentrated, continues to grow for some time and, after reaching its maximum, begins to gradually decrease, as the trap body's side walls are cooled externally and water is supplied to the upper surface of the oxide layer of the melt, without risk of steam explosion [21].

The temperature of the metal layer changes almost synchronously due to heat transfer in its surface region adjacent to the oxide layer, remaining below the temperature of the latter due to external cooling of the bottom of the trap body. Stabilization of thermal processes occurs after 10–20 h, with further transition to quasi-regular cooling of the melt. A slightly higher cooling rate of the metal layer and more fusible components cause its faster solidification, compared to the refractory components of the oxide layer. A fairly large molten core remains there for quite a long time (40–65 h), and even by 80 h separate pockets of the liquid phase remain in the crystallized zones. Throughout these processes, the heat fluxes from the corium to the trap walls are in the range from 0.2 to 0.6 MW/m^2, which does not exceed the critical heat load [14] on the external water-cooled surface of the crucible body with a sufficient margin.

Conclusion

The presented two-stage model of thermokinetic processes and the results of end-to-end computations of changes in thermal and phase states of the corium melt in the core entering the trap during severe accidents of NPP reactors details the characteristics of the melt's interaction with the sacrificial material (SM) and the patterns of subsequent cooling to crystallization temperature provided in the design of the trap.

From a physical standpoint, the selected model stages are justified by the phase transitions of the aggregate state of the system components including melting of the sacrificial material based on iron and aluminum oxides in the high-temperature corium melt during the initial period of their interaction and subsequent crystallization of the products of this interaction during intense external cooling of the trap body. Mathematical modeling is based on a generalized statement of the Stefan problem with free interfaces taking into account the residual radioactive heat release from the corium, the thermal effects of melting of iron oxides and redox reactions of oxides during oxidation of underoxidized elements in the melt, heat transfer by natural convection in the melt, heat transfer to cooling water through the shell walls and thermal radiation in the inner space.

The phenomenon of rapid gravitational inversion of immiscible oxide and metal components of different densities with molten oxides moving to the upper part of the melt pool and molten metals to the lower part is a natural delimitation of the model stages. Such a change in the pool structure appears due to a significant decrease in the density of the initially heavier oxide components of the melt as compared to the metal ones after chemical interaction of the melt with the SM.

The COMSOL Multiphysics software package used in the computations does not create difficulties in solving the considered problems by end-to-end computation and analysis of thermokinetic processes in crucible-type ex-vessel melt retention devices, used in known specialized software codes developed for simulating various severe accident scenarios and their individual stages at the in-reactor stage.

The results obtained indicate that the thermokinetic processes considered generally provide a decrease in the temperature and density of the corium, a decrease in heat fluxes to the trap body, and a minimization of hydrogen and radioactive fission products. At the same time, the scatter in the estimates for the modes and times of cooling of the corium melt with uncertain composition requires in-depth consideration of the thermokinetics of crystallization processes, extremely important because of the risk of secondary centers with increased radioactivity appearing during local accumulation of heavier (relative to the entire melt) radioactive substance, as well as release of components into the gas phase. Their study at the next stages of the project can consist in refining the specifics of spatial kinetics of the computational model in the 3D statement, verifying and validating the model's applicability, in particular taking into account the problems of assessing the thermal state and the thermal stresses in the trap body.

This work was supported by a grant from the Russian Foundation for Basic Research (Project No. 19-08-01181).

REFERENCES

1. **Walker S.J.,** Three-mile island: A nuclear crisis in historical perspective, University of California Press, Berkeley, 2004.

2. **Abagyan A.A, Arshavskii I.M, Dmitriev V.M., et al.,** Computational analysis of the initial stage of the accident at the Chernobyl' atomic power plant, Atomic Energy. 71 (4) (1991) 785–795.

3. **Gonzalez A.J., Akashi M., Boice J.D., et al.,** Radiological protection issues arising during and after the Fukushima nuclear reactor accident, Journal of Radiological Protection. 33 (3) (2013) 497–571.

4. **Theofanous T.G., Lin C., Addition S., et al.,** In-vessel coolability and retention of a core melt, Nuclear Engineering and Design. 169 (1–3) (1997) 1–48.

5. **Kim H.Y., Bechta S., Foit J., Hong S.W.,** In-vessel and ex-vessel corium stabilization in light water reactor, Science and Technology of Nuclear Installations. 2018 (2018) ID 3918150 (3 p).

6. **Fischer M.,** Main features of the EPR melt retention concept (Rep. FZKA-6475), Forschungszentrum Karlsruhe GmbH Technik und Umwelt (Germany). 32 (14) (2000) 508–517; Proceedings of the OECD Workshop on Ex-Vessel Debris Coolability, Alsmeyer H. (Ed.), Karlsruhe, Germany, 15–18 November, 1999.

7. **Bezlepkin V.V., Sidorov V.G., Kukhtevich V.O., et al.,** Ustroystvo lokalizatsii i okhlazhdeniya koriuma yadernogo reaktora [An arrangement to localize and cool down the nuclear reactor's

corium], Pat. No. 2514419, Russian Federation, MPK G21C 9/016.; SPb., OAO «Golovnoy Institut «VNIPIET» is a declarant and a patentee, No. 2012124161; declar. 20.02.14; publ. 27.04.2014, Bull. No. 23 (II Ch.). 11 p.

8. **Stolyarevskiy A.Ya.,** Spasayet li lovushka? [Does a trap rescue?], Atomnaya Strategiya. (4(89)) (2014) 16–18 (in Russian).

9. **Gusarov V.V., Almjashev V.I., Khabensky V.B., et al.,** Physicochemical modeling and analysis of the interaction between a core melt of the nuclear reactor and a sacrificial material, Glass Physics and Chemistry. 31 (1) (2005) 53–66.

10. **Asmolov V.G., Sulatskii A.A., Beshta S.V., et al.,** The interaction of nuclear reactor core melt with oxide sacrificial material of localization device for a nuclear power plant with water-moderated water-cooled power reactor, High Temperature. 2007. T. 45(1) (2007) 22–31.

11. Integral codes development and application for NPP safety analysis, Proceedings of IBRAE RAS, Ed. by L. A. Bolshov; Nuclear Safety Institute (IBRAE) RAS, Nauka, Moscow, 2011, Iss. 12 (in Russian).

12. **Lityshev A.V., Pantyushin S.I., Aulova O.V., et al.,** Experience in carrying out calculations of severe beyond design basis accidents for VVER RP using code SOKRAT, Proceedings of the 9th International Scientific and Technical Conference "Safety Assurance of NPP with VVER", Podolsk, Russia, May 19–22, SCB "Gidropress", Podolsk (2015) 1–22.

13. **Zvonarev Yu.A., Melnikov I.A., Shmelkov Yu.B.,** Simulation of corium pool behavior on RPV bottom under severe accident for VVER-1000, Proceedings of the 10th International Scientific and Technical Conference "Safety Assurance of NPP with VVER", Podolsk, Russia, May 16–19, SCB "Gidropress", Podolsk (2017) 1–22.

14. **Fiskov A.A., Bezlepkin V.V., Semashko S.Ye., et al.,** Opyt obosnovaniya ustroystva lokalizatsii rasplava LAES-2 [An experience on foundation for the arrangement of the melt localization at Leningrad NPP-2], Proceedings of the Industry Research-to-Practice Conference «The Youth NFC: Science, Production, Ecological Safety», St. Petersburg, «Atomenergoproyekt», 2010. P. 5.

15. **Vasilenko V.A., Migrov Yu.A., Dragunov Yu.G., Bykov M.A.,** Heat-hydraulic design code KORSAR: Development status and experience, Proceedings of the 3rd International Scientific and Technical Conference "Safety Assurance of NPP with VVER", Podolsk, Russia, May 26–30 2003, SCB "Gidropress", Podolsk. 5 (3) (2003) 204–212.

16. **Sangiorgi M., Grah A., Pascal G., et al.,** In-vessel melt retention (IVMR) analysis of a VVER1000- NPP // JRC Technical Reports EUR 27951. JRC Science Hub, European Union, 2016. 250 p.

17. **Udalov Yu.P., Fedorov N.F., Soloveychik Ya., Pavlova E.A.,** Novyye funktsionalnyye oksidnyye materialy dlya yadernogo reaktorostroyeniya [New functional materials for nuclear rocket production], Industry & Chemistry. 80 (12) (2003) 3–10 (in Russian).

18. **Gusarov V.V., Almyashev V.I., Khabenskiy V.B.,** et al., Novyy klass funktsionalnykh materialov dlya ustroystva lokalizatsii rasplava aktivnoy zony yadernogo reaktora [A new class of functional materials for making the arrangement for localization the melt's core region], Rossiyskiy Khimicheskiy Zhurnal. XLIX (4) (2005) 42–53 (in Russian).

19. **Stivens J.,** Posleavariynyy otvod tepla ot oblomkov razrushennoy aktivnoy zony [Post-accident heat removal from the wreckage of the destroyed core], Atomnaya Tekhnika za Rubezhom (12) (1984) 14–22 (in Russian).

20. **Gusarov V.V., Almjashev V.I., Khabensky V.B., et al.,** Physicochemical simulation of the combustion of materials with the total endothermal effect, Glass Physics and Chemistry. 33 (5) (2007) 492–497.

21. **Lopukh D.B., Loginov I.A., Granovskii V.S., et al.,** Experimental investigation of processes arising when flooding a steel melt, Thermal Engineering. 48 (9) (2001) 725–731.

22. **Lagunenko O.S., Krasnov V.O., Dovyd'kov S.A.,** Fuel in room 305/2. Probable formation scenario for nuclear-dangerous zones, Problemi bezpeki atomnikh yelektrostantsiy i Chornobilya [Safety Problems of Nuclear Power Plants and Chernobyl]. (24) (2015) 51–59 (in Russian).

23. **Almyashev V.I., Granovskiy V.S., Khabenskiy V.B., et al.,** Experimental study of corium melt oxidation processes in the reactor vessel, "Tekhnologii Obespecheniya Zhiznennogo Tsikla Yadernykh Energeticheskikh Ustanovok" [Maintenance Technology of the Life Cycle of Nuclear Power Installations]. (4 (10)) (2017) 59–84 (in Russian).

24. **Samarskiy A.A., Vabishchevich P.N.,** Vychislitelnaya teploperedacha [Computational heat transfer], Editorial URSS, Moscow, 2003 (in Russian).

25. **Snegiryov A.Yu.,** Chislennoye modelirovaniye turbulentnykh techeniy [Numerical simulation of turbulent flows], Polytechnical Institute Publishing, St. Petersburg, 2009.

26. **Kutateladze S.S., Leontiev A.I.,** Heat transfer, mass transfer, and friction in turbulent boundary layers, Hemisphere, New York ,1990.

27. COMSOL Multiphysics, Software products COMSOL, Burlington, MA. URL: https://www.comsol.ru.

28. **Golovacheva V.G., Kovalenko A.N., Meshcheryakov D.K., et al.,** Correlation between heat – mass transfer, chemical reactions and phase transformations in corium melt localization devices during severe nuclear power plant accidents, Diversity in Nuclear: International Youth Nuclear Congress (NIYNC2020), 8 –13 March 2020, Sydney, Australia (2020) 176–179.

29. **Voskov A.L., Dzuban A.V., Maksimov A.V.,** TernAPI program for ternary phase diagrams with isolated miscibility gaps calculation by the convex hull method, Fluid Phase Equilibria. 2015. Vol. 388 (25 February) (2015) 50–58.

30. **Konakov P.K., Veryovochkin G.E., Gorenkov L.A., et al.,** Teplo- i massoobmen pri poluchenii monokristallov [Heat and mass exchange when manufacturing monocrystals], Metallurgy Publishing, Moscow, 1971 (in Russian).

Received 20.07.2020, accepted 27.08.2020.

THE AUTHORS

KOVALENKO Anatoliy N.
Ioffe Physical Technical Institute of the Russian Academy of Sciences,
Peter the Great St. Petersburg Polytechnic University
26 Polytekhnicheskaya St., St. Petersburg, 194021, Russian Federation
ras-kan@mail.ru

KOPTYUKHOV Artem O.
NRC «Kurchatov Institute» – PNPI
Mcr. Orlova Roshcha, Gatchina 1, Leningradskaya oblast,188300, Russian Federation
t44h@yandex.ru

MESHCHERYAKOV Dmitry K
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
fess_i@bk.ru

SCHUKLINOV Alexey P.
JSC "Atomproekt";
Peter the Great St. Petersburg Polytechnic University
82A Savushkina St., St. Petersburg, 197183, Russian Federation
mupol@mail.ru

DOI: 10.18721/JPM.13406

INFLUENCE OF THE SLOPE ANGLE OF A LOCALLY HEATED SURFACE ON THE HEAT TRANSFER DURING BIFURCATION OF A PURE THERMAL PLUME

E.F. Khrapunov, Yu.S. Chumakov

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

The results of numerical simulation of the puffing regime of a pure thermal plume, which forms above a slightly inclined (with respect to the horizon) surface have been presented. The flow structure in the near-wall region was investigated and the fields of the velocity components were analyzed. Particular attention was paid to the analysis of the distribution of instantaneous, time- and surface-averaged values of the heat transfer coefficient α. Based on the presented results, it was concluded that there was an optimal slope angle of the surface maintaining the most efficient heat transfer with a developed periodic flow regime.

Keywords: natural convection, numerical simulation, bifurcation, heat transfer, puffing

Citation: Khrapunov E.F., Chumakov Yu.S., Influence of the slope angle of a locally heated surface on the heat transfer during bifurcation of a pure thermal plume, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 59–67. DOI: 10.18721/JPM.13406

Introduction

Analyzing the studies available to us on free-convection flows, we can see that significant progress has been made, first of all, regarding flows formed along vertical heated surfaces, when the vector of the free-fall acceleration **g** is parallel to the heated surface [1, 2]. This is largely due to the possibility to describe such flows within the boundary layer approximation that has appeared in the second half of the 20th century. Tools for numerically solving equations were constructed, allowing for detailed analysis of the characteristics of both laminar and fully developed turbulent boundary layer.

In turn, studies of flows formed over horizontal locally heated surfaces perpendicular to vector **g**, the so-called free-convection plumes, have long presented considerable difficulties [3, 4]. In this case, upward flow is observed over a small part of the surface (near its center), and a wall layer forms over the rest of the surface under the action of a favorable pressure gradient produced by the upward jet. Obviously, the presence of upward flow, as well as the section with a smooth transition from the wall layer to the upward jet does not allow the boundary layer approximation to be used to describe the plume as a whole. Such an approximation proves to be valid over a small section of the surface away both from its edge and from its center, as well as at a considerable distance from the disk surface in the region of fully developed upward flow. The section of the plume between these two regions (the transition region) cannot be fundamentally described within the above approximation. For a long time, the only way to numerically determine the flow characteristics in this region was to 'merge' the solutions obtained in the regions where the boundary layer equations were applicable.

At present, the significantly improved range of experimental and numerical methods allows to describe the flow as a whole, without dividing it into fragments, in particular focusing on analysis of changes in not only the averaged but also the instantaneous characteristics of the plume. For instance, a bifurcation of the solution was discovered in [6–9], as a result of which a stable steady regime is replaced by a stable self-oscillatory one, or puffing. The essence of puffing, which is a particular case of Hopf bifurcation, is periodic destruction of the near-wall layer by the evolving toroidal vortex structures [10]. While some studies [7, 9] considered the changes in some characteristics of the plume during bifurcation, the influence of parameters of the heated surface (for example, the incline relative to the horizontal position) on the flow as a whole is still poorly understood. The question about the mechanism of bifurcation remains unclear. The inverse effect of fluctuations in the plume on the characteristics of

heat transfer with the heated surface is practically unstudied. At the same time, the structure of the flow near an inclined heated surface in the absence a puffing flow regime is described in sufficient detail [11–15].

The first study considering the influence of the slope angle of the heated surface on heat transfer was apparently [11], discussing the flow evolving near a heated rectangular plate tilted relative to the vertical position. It was found that tilting the surface to angles up to 40° had practically no effect on the characteristics of the formed free-convection boundary layer. As a result, it was concluded that the characteristics of heat transfer for an incline can be determined by simply replacing the modulus of vector **g** by the modulus of its component parallel to the surface in the well-known dependences obtained for vertical plates.

It was found in subsequent works that the flow characteristics vary only slightly at large slope angles as well [12]. However, there is a limiting angle, at which a cardinal change in the flow structure is observed: the flow is detached from the surface at a certain distance from the lower edge, and a pure thermal plume is formed. In particular, this change in flow structure does not allow using the dependences obtained earlier to describe the heat transfer at large surface slope angles relative to the vertical position.

One of the first works studying the structure of the flow over a heated surface slightly inclined with respect to the horizon is [14]. The surface temperature varied over a wide range, but since the experimental study was conducted in water, the maximum temperature did not exceed 368 K. The streamlines were visualized using aluminum powder illuminated by a special source. It was found that the transition point of the near-wall layer into an ascending jet remains on the surface at slope angles less than 10° relative to the horizontal position. At larger angles, the transition point coincides with the plate's boundary point. The most recent results (of those known to us) were presented in [15], also indicating that the surface inclination relative to the horizon has the greatest influence on the flow structure at angles up to 10° in the case of a steady pure thermal plume.

In our opinion, the key reason why it is small slope angles that have a major influence on characteristics of the plume is the weakening of upward flow during formation of the flow in the near-wall region. As the slope angle increases, the transition region from the near-wall to the upward flow shifts towards the edge of the surface elevated above the horizon, and relatively stable near-wall flow is formed over most of the surface at an angle of about 10°. The slope angle gradually ceases to affect its characteristics. A further increase in this angle in this case ultimately leads to the near-wall flow becoming predominant.

The goal of this study was to obtain and analyze separate instantaneous pictures of the dynamic and thermal fields of the puffing flow regime of a pure thermal plume formed over a slightly inclined (in the slope range from 0° to 5° with respect to the horizon) surface.

In view of this goal, we consider in detail the fragments of flow in the near-wall region and analyze the fields of velocity component distributions within one puffing period. A particular focus was on analysis of the distribution patterns of instantaneous (within a single oscillation period) and time-averaged values of the heat transfer coefficient α.

Problem statement

We simulated free-convection flow over a heated disk of radius $R = 95$ mm in an unbounded space. The angle γ, which makes the free-fall acceleration vector **g** with the disk surface, varied from 90° (corresponding to the case of a horizontal surface) to 85°. A schematic of the computational domain and the reference system of angle γ are shown in Fig. 1,*a*. The ANSYS Fluent software allowing to solve the Navier–Stokes equations in the Boussinesq approximation for a compressible medium was used for numerical simulation. The height of the computational domain was $H = 1.5$ m, the diameter of the domain was $D = 1$ m.

Based on the results of preliminary experimental studies, partially given in [10], it was concluded that the contribution of turbulent mixing to the processes occurring near the disk surface at moderate heating temperatures was insignificant. For this reason, laminar 3D flow was simulated in the presence of conjugate heat transfer between the heated disk and the ambient. The thermal properties of air in the computational domain remained constant (their values were taken at room temperature T_a), since their variation seems insignificant in the given temperature range.

A quasi-structured mesh including $1.1 \cdot 10^6$ elements was used for the main computations, refined towards the assumed boundaries of the plume and to the lower surface of the computational domain. A fragment of the computational grid is shown in Fig. 1,*b*.

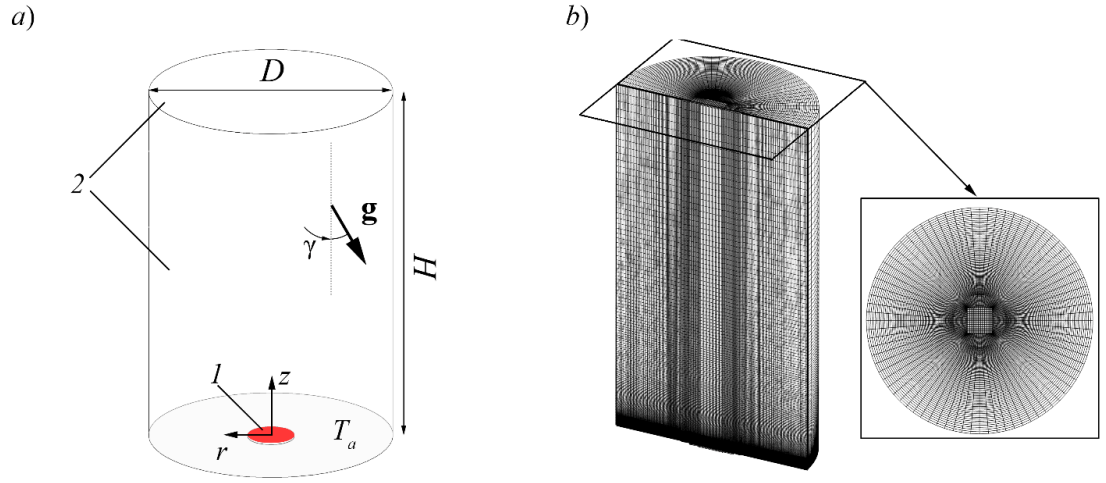


Fig. 1. Schematic of computational domain (*a*) and fragment of computational mesh (*b*): heated disk *1* (conjugate heat transfer conditions); free boundaries *2*

The boundary conditions were as follows (see Fig. 1, *a)*. The temperature of the disk's bottom surface was set equal to T_w = 383 K and remained constant during the simulation, the temperature of the horizontal surface of the computational domain also remained constant, equal to the ambient air temperature (T_a = 293 K). The temperature of the disk's upper surface was found from the condition of conjugate heat transfer with the air adjacent to the surface. Zero overpressure conditions were set at all other boundaries (side and top). Restrictions were imposed on the air flowing into the computational domain through these boundaries: the air temperature remained constant, equal to 293 K, and the velocity vector was always normal to the corresponding flow boundary.

In this study we analyzed a periodic flow regime stable over time. As a criterion that the regime was stabilized, we used data on the time variation of the temperature at the control points of the near-wall region, as well as the time variation of the heat transfer coefficient averaged over the disk surface (as shown, for example, in [10]).

The coordinate system used for representing the results is given in Fig. 1,*a*. The dimensionless components v_z* and v_r* are used to describe the velocity fields. The ratio of the kinematic viscosity coefficient ν to the disk radius *R* is used as the velocity scale.

Traditionally, the Grashof number Gr, determined by the disk radius and the characteristic temperature difference between the surface and the ambient air, is used as the main parameter to describe pure thermal plumes. Since the temperature of the disk's upper surface is determined by conjugate heat transfer conditions and is not known in advance, the temperature of the disk's lower surface T_w, set as boundary conditions within the framework of numerical simulation, is used to construct the Grashof number. The Grashof number is constructed accounting for the slope of the surface with respect to the free-fall acceleration vector **g**:

$$Gr_\gamma = \frac{g \cdot \beta \cdot (T_w - T_a) \cdot R^3}{\nu^2} \cdot cos(\gamma), \qquad (1)$$

where β is the coefficient of thermal expansion of air, taken as $3.6 \cdot 10^{-3}$ K^{-1}.

The given results correspond to the Grashof number Gr = $8.65 \cdot 10^6$, determined from dependence (1) in the case of a zero slope angle of the surface.

Analysis of results

Fairly detailed analysis of the characteristics of steady free-convection flows near inclined surfaces, as well as the fundamentals on the bifurcation of the solution, given in [10] for the case γ = 0°, allows to formulate the proposed scheme for the flow under consideration. The presence of a slope should lead to an acceleration of air in the near-wall region [16] and a shift of the transition region from near-wall to ascending flow toward the corresponding surface boundary. In this case, it is unclear whether the solution bifurcation will persist. At the same time, ascending flow shifted relative to the center of the heated surface may prove to be a noticeable obstacle to accelerated near-wall flow.

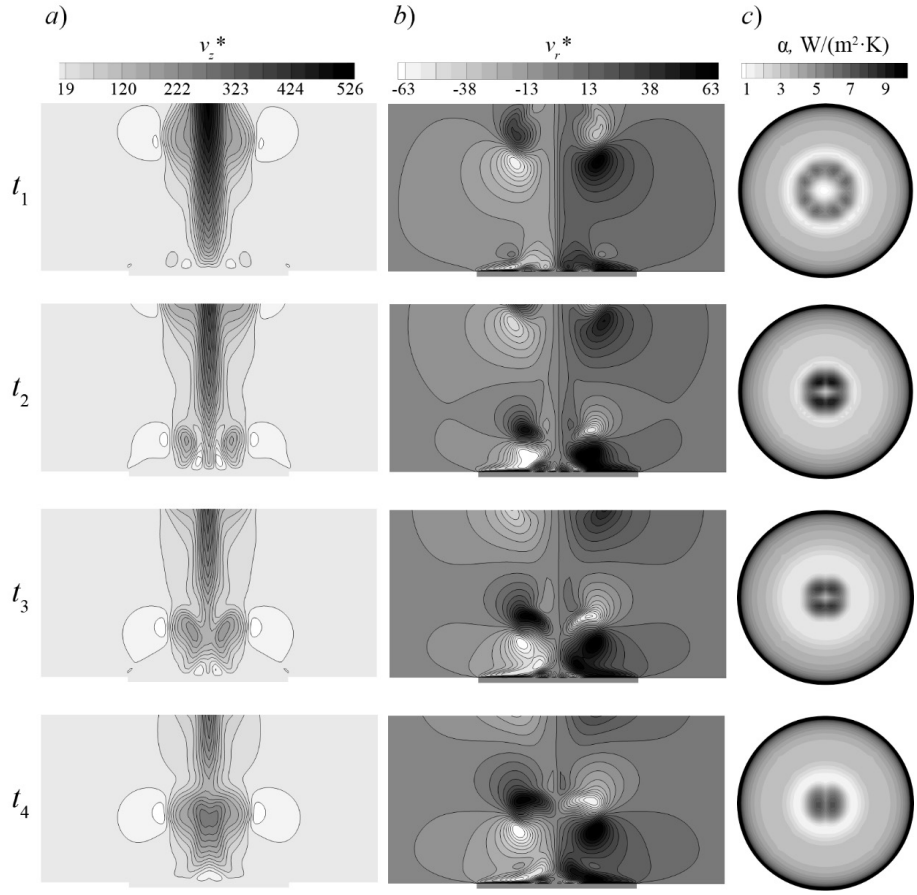


Fig. 2. Changes in axial (*a*) and radial (*b*) velocity components; changes in heat transfer coefficient (*c*) within one oscillation period when $\gamma = 0^o$

To analyze the effect of surface inclination on bifurcation, let us briefly consider the features of the flow for the case where the angle $\gamma = 0^o$ (see Fig. 1).

Fig. 2 shows the change in the flow characteristics in the *Orz* plane for $\gamma = 0^o$ (the case of a horizontal surface) within one plume oscillation period, as well as the distribution of instantaneous values of the heat transfer coefficient over the disk surface. It can be clearly seen that symmetric fields of the distribution of the axial and radial velocity components were obtained near the disk surface. It is also shown in [9, 10] that the flow in the near-surface region is axisymmetric. Axial symmetry is violated downstream (upstream of the disk surface) as a result of the development of instability during the rise of heated vortex structures in the surrounding cold ambient.

The formation of symmetric vortex structures leads to the emergence of regions where the heat transfer coefficient is increased over the disk surface, which can be clearly seen in Fig. 2.*c*. A region of increased heat transfer coefficient emerges because cold (temperature T_a) air is 'trapped' near the heated surface where the toroidal vortex separates. A similar conclusion is confirmed by the results of studies in [6, 8].

Notice that at each given moment of time, it is possible to detect a region on the surface of the disk where the heat transfer coefficient increases up to the values α^t_{max}, the maximum value of the coefficient on the disk surface at a given moment of time. In this case, the maximum value of the coefficient, α^t_{max}, for the full period of oscillations is over twice the α^t_{max} values of the coefficient in the case of a stable plume [17]. Within a single oscillation period, the region of increasing heat transfer coefficient is transformed: the radius of this region decreases, the region shrinks toward the center of the disk.

It should also be noted that an unusual structure that does not possess axial symmetry (see Fig. 2,*c*) appears on the disk surface (near the center) in visualizations of the heat transfer coefficient. This structure appears near the central prismatic block of the mesh (see Fig. 1,*b*), but this may be due to the general topology of

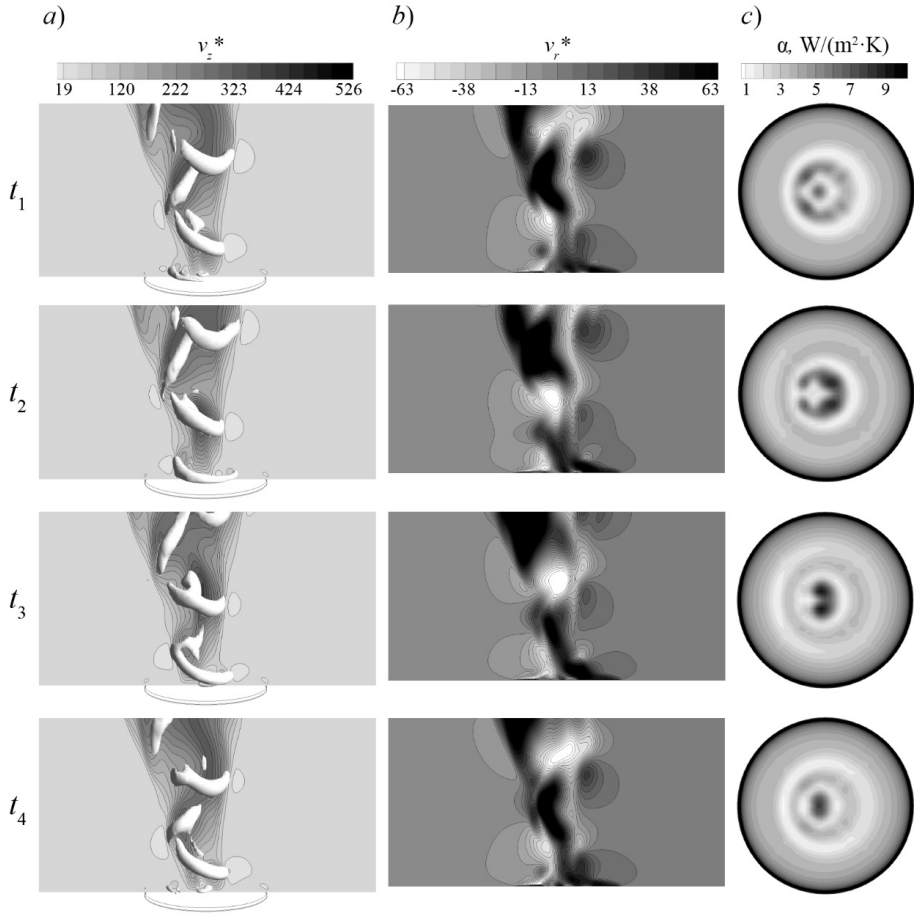


Fig. 3. Variations of axial (*a*) and radial (*b*) velocity components, as well as the heat transfer coefficient (*c*) at $\gamma = 1^{\circ}$ within one oscillation period

the mesh. We did not estimate the influence of mesh topology on the resulting solution as it is beyond the scope of our study.

Fig. 3 shows the change in the flow characteristics in the *Orz* plane for $\gamma = 1^{\circ}$ within one oscillation period. It can be seen that even a small surface slope angle leads to a noticeable change in the flow structure. For visual representation of the vortex structures, *Q*-criterion isosurfaces are shown on the fields of the v_z* component. Evolution of these isosurfaces for the case $\gamma = 0^{\circ}$ is described in [10]. We should emphasize that isosurfaces do not visualize the exact shapes and sizes of the vortex, only the part of the flow that has the highest vortex intensity. If we consider regions with different levels of vortex intensity, it is possible to carry out a general analysis of the changes in the characteristics of the pure thermal plume for an inclined surface.

The flow patterns shown in Fig. 3 allow us to conclude that when the horizontal surface is inclined, instead of an axisymmetric vortex, two areas with increased vortex intensity form, separated by the ascending flow. In this case, a vortex structure is formed at the initial moment of time, located 'under' the ascending flow. The air trapped in this area is heated stronger than the air in the near-wall layer forming on the other side of the ascending plume. As a result, there is a local increase in the Archimedean force, and, as a consequence, an increase in the axial velocity v_z*.

As vortices develop further, a 'vortex track' is formed in the space above the heated surface, where the vortex structures with different slopes alternate. Notably, at all moments of time within one period, the presence of multidirectional velocities v_r* is observed in the plume formation region. In this case, the surface inclination indeed, leads to flow acceleration and to the general dominance of positive radial velocities (see the coordinate and angle reference systems in Fig. 1) in the near-wall region.

Let us turn to the analysis of the peculiarities of the distribution of the heat transfer coefficient over the disk surface inclined by 1°. The size and shape of the region where the coefficient α increases significantly change as a result

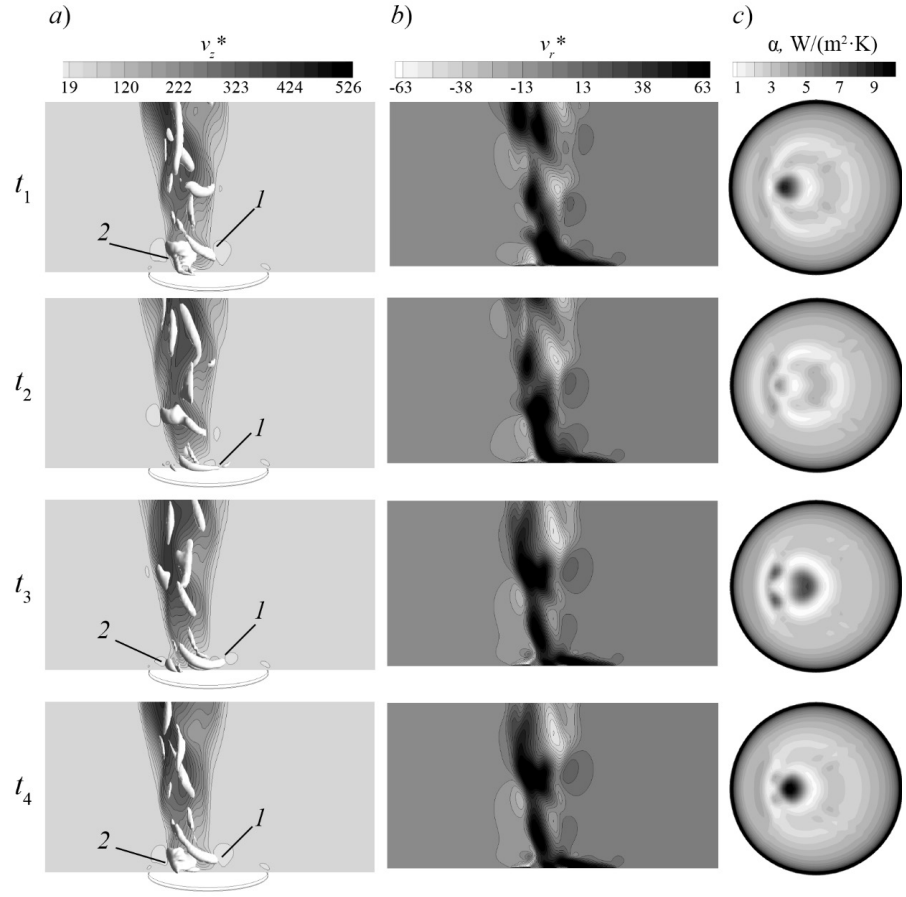


Fig. 4. Changes in the axial (*a*) and radial (*b*) velocity components, as well as the heat transfer coefficient (*c*) within one oscillation period at $\gamma = 3^o$. The vortex structures bending around the plume (*1*) and localized (*2*) are shown

of violating the symmetry of the flow in the near-wall region. Thus, in Fig. 3,*c*, the region with the maximum value of the heat transfer coefficient is observed first on the left and then on the right from the center of the disk.

Fig. 4 shows the change in the flow characteristics in the *Orz* plane for the value of angle $\gamma = 3^o$ within one oscillation period. Compared to the previous case ($\gamma = 1^o$), the delay in the formation of two vortex structures, separated by the ascending flow, becomes more noticeable. An elongated vortex structure can be distinguished in the plume, which bends around the ascending flow and a localized vortex, which appears to be 'trapped' between the inclined plume and the disk surface. The corresponding vortex structures are marked by numbers *1* and *2* in Fig. 4. Note that the transition point of near-wall into ascending flow at $\gamma = 3^o$ is visibly shifted from the center of the disk.

Analysis of the distribution fields of the heat transfer coefficient suggests that an increase in the surface inclination leads to a decrease in the size of regions with elevated values of the coefficient α. Within a single puffing period, it is possible to distinguish moments of time when there is no region of a pronounced increase in the coefficient α on the disk surface, which is not characteristic of smaller surface slope angles. Note also that the region of significant growth of α in this case is relatively small.

The value of the angle $\gamma = 5^o$ between the free-fall acceleration vector **g** and the disk surface (Fig. 5) is interesting because it seems to be a limiting case of a stable periodic flow regime. The transition point from the near-wall to ascending flow shifts practically to the disk boundary. A similar conclusion is drawn in [14]. Complex vortex motion is formed in the near-wall region, which, using the *Q*-criterion, can be represented as separate regions of increasing vortex intensity. As a result, there is reason to assume that a large number of small disconnected vortex structures are formed above the disk, which weakly interact with each other as they ascend in the surrounding air.

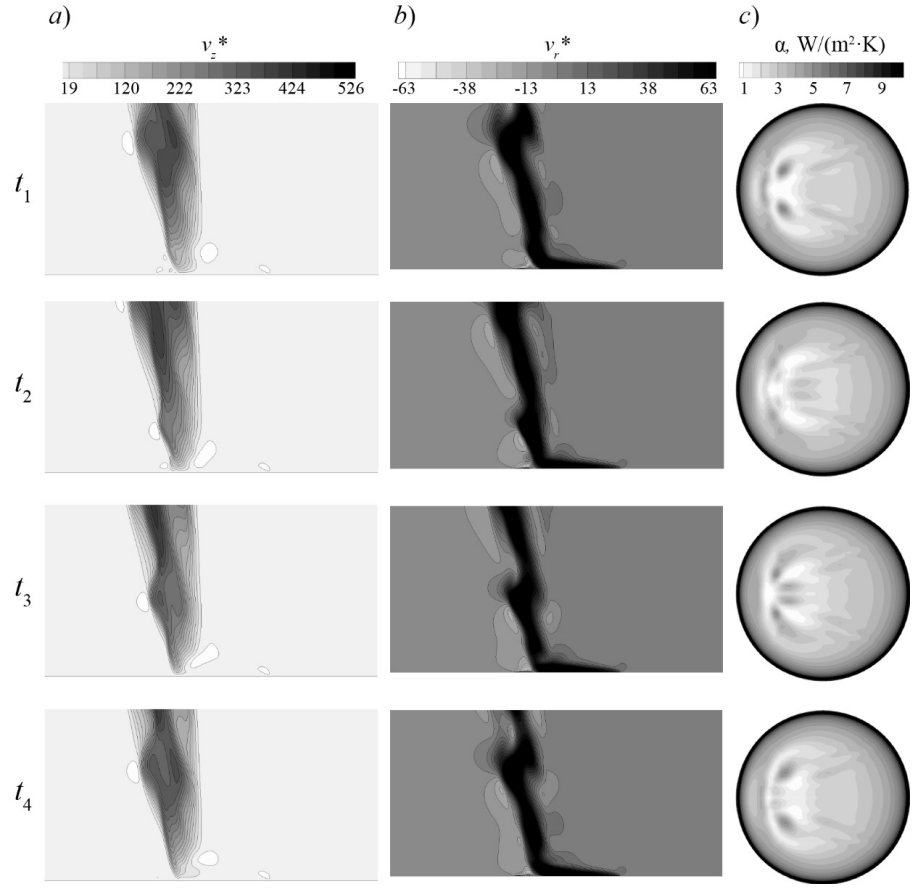


Fig. 5. Variations of axial (*a*) and radial (*b*) velocity components, as well as the heat transfer coefficient (*c*) at $\gamma = 5^{o}$ within one oscillation period

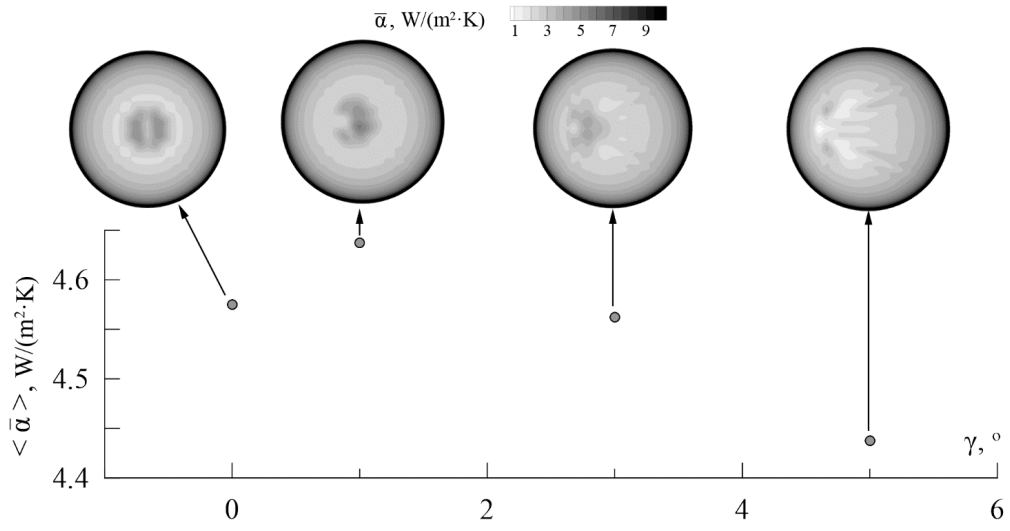


Fig. 6. Effect of surface slope angle γ on the values of the averaged heat transfer coefficients $\langle\bar{\alpha}\rangle$;
the distribution of $\underline{\alpha}$ over the disk surface is shown for each value (symbols $\langle\bar{\alpha}\rangle$)

Violation of the original (observed at $\gamma = 0^{\circ}$) toroidal vortex structure, and, consequently, the patterns by which cold air is forced towards the disk surface leads to the fact that it is impossible to identify a localized region on the disk surface where there is an increase in the coefficient α. Indeed, the regions with increasing values of the heat transfer coefficient turn out to be distributed over the surface (see Fig. 5,*c*). At the same time, there is no sharp increase in the values of α, which is characteristic of smaller angles.

Typically, it is not so much the instantaneous distributions of the local heat transfer coefficient but rather the distribution of the time-averaged coefficient $\underline{\alpha}$ over the disk surface, as well as the value of the time-averaged coefficient $\langle\bar{\alpha}\rangle$ integral over the disk surface that are of interest for practical applications. In this case, brackets denote averaging over the disk surface. The effect of the slope angle of the horizontal surface on the value $\langle\bar{\alpha}\rangle$ is shown in Fig. 6. In addition, the distribution of $\bar{\alpha}$ over the disk surface is shown for each value of $\langle\bar{\alpha}\rangle$.

It can be noted that the dependence of the coefficient $\langle\bar{\alpha}\rangle$ on the slope angle γ is not monotonic, there is a pronounced maximum in the vicinity of the slope angle $\gamma \approx 1^{\circ}$, i.e., an increase in the intensity of heat transfer is observed at this value.

Conclusion

In this paper, the influence of the slope angle on the characteristics of puffing in a pure thermal plume was investigated for the first time. It is established how exactly the change in the properties of the formed vortex structures affects the characteristics of heat transfer.

In our opinion, analysis of unsteady regime of a pure thermal plume over an inclined surface is of interest because such studies provide additional data on the flow model and establish the degree of influence of the plate orientation relative to the free-flow acceleration vector **g** on the characteristics of the unsteady process, whose main peculiarity is puffing.

The study revealed that a small (up to five degrees) inclination of the heated surface has a noticeable effect on the puffing process, which, ultimately, affects the heat transfer between air and the heated surface. There is reason to assume that further study of the influence of the orientation of the heated surface relative to the vector **g** on the given process will allow to develop effective ways to 'control' puffing, influencing the values of the local heat transfer coefficient over the required region of the heated surface.

REFERENCES

1. **Cheesewright R.,** Turbulent natural convection from a vertical plane surface, J. Heat Transfer. 90 (1) (1968) 1–6.

2. **Tsuji T., Nagano Y.,** Characteristics of a turbulent natural convection boundary layer along a vertical flat plate, Int. J. Heat and Mass Transfer. 31 (8) (1988) 1723–1734.

3. **Batchelor G.K.,** Heat convection and buoyancy effects in fluids, Quarterly Journal of the Royal Meteorological Society. 80 (345) (1954) 339–358.

4. **Lewandowski W.M., Kubsky P., Khubeiz J.M.,** Natural convection heat transfer from round horizontal plate, Heat and Mass Transfer (Warme- und Stoffubertragung). 27 (5) (1992) 281–287.

5. **Morton B.R., Taylor G., Turner J.S.,** Turbulent gravitational convection from maintained and instantaneous sources, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences (London). 234 (1196) (1956) 1–23.

6. **Jaluria Y.**, Natural convection: Heat and mass transfer, Pergamon Press, Oxford, 1980.

7. **Plourde F., Pham M.V., Doan K.S., Balachandar S.,** Direct numerical simulations of a rapidly expanding thermal plume: structure and entrainment interaction, J. of Fluid Mechanics. 604 (10 June) (2008) 99–123.

8. **Lopez J. M., Marques F.,** Instability of plumes driven by localized heating, J. of Fluid Mech. 736 (10 December) (2013) 616–640.

9. **Hattori T., Bartos N., Norris S.E., et al.,** Experimental and numerical investigation of unsteady behavior in the near-field of pure thermal planar plumes, Exp. Thermal Fluid Science. 46 (April) (2013) 139–150.

10. **Khrapunov E., Chumakov Y.,** Structure of the natural convective flow above to the horizontal surface with localized heating, Int. J. Heat and Mass Transfer. 152 (May) (2020) 119492 (12 p.)

11. **Rich B.R.,** An investigation of heat transfer from an inclined flat plate in free convection // Transactions – ASME: Journal of Heat Transfer. 75 (May) (1953) 489–499.

12. **Kierkus W.T.,** An analysis of laminar free convection flow and heat transfer about an inclined isothermal plate, Int. J. Heat Mass

Transfer. 11 (2) (1968) 241–253.

13. **Vliet G.C.,** Natural convection local heat transfer on constant-heat-flux inclined surfaces, J. Heat Transfer. 91 (4) (1969) 511–516.

14. **Fujii T., Imura H**., Natural-convection heat transfer from a plate with arbitrary inclination, Int. J. Heat Mass Transfer. 15 (4) (1972) 755–767.

15. **Kimura F.,** Fluid flow and heat transfer of natural convection at a slightly inclined, upward-facing, heated plate, Heat Transfer – Asian Research. 31 (5) (2002) 362–375.

16. **Shaukatullah H., Gebhart B.,** An experimental investigation of natural convection flow on an inclined surface, Int. J. Heat Mass Transfer. 21 (12) (1978) 1481–1490.

17. **Khrapunov E.F., Potechin I.V., Chunakov Y.S.,** Structure of a free convective flow over a horizontal heated surface under conditions of conjugate heat transfer, J. Physics: Conf. Series. 891 (2017) 012081.

Received 08.08.2020, accepted 05.10.2020.

THE AUTHORS

KHRAPUNOV Evgenii F.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
hrapunov.evgenii@yandex.ru

CHUMAKOV Yuriy S.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
chymakov@yahoo.com

DOI: 10.18721/JPM.13407

THE ABDOMINAL AORTA BIFURCATION WITH ILIAC ARTERIES: THE WALL ELASTICITY EFFECT ON THE FLOW STRUCTURE

A.A. Kotmakova[1], Ya.A. Gataulin[2], A.D. Yukhnev[2], D.K. Zaitsev[2]

[1] Scientific Research Institute for Optoelectronic Instrument Engineering, Sosnovy Bor, Leningrad region, Russian Federation;

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

For the first time, a numerical study of the effect of vascular elasticity on blood flow has been carried out using a model of the abdominal aorta bifurcation with iliac arteries of the average statistical geometry. The greatest effect of elasticity was shown to be observed in the abdominal aorta before bifurcation, where there are large deformations of the vessel walls, changing the shape of the cross section from oval to circular. Taking into account the elasticity of the walls, the flow structure remained unchanged, but the maximum flow rate increased by 40 % in the abdominal aorta. In the abdominal-aortic bifurcation region the time-averaged wall shear stress decreased by 20%, and their oscillatory shear index increased by 60%.

Keywords: abdominal aorta's bifurcation, numerical simulation, wall elasticity, blood flow structure

Citation: Kotmakova A.A., Gataulin Ya.A., Yukhnev A.D., Zaitsev D.K., The abdominal aorta bifurcation with iliac arteries: the wall elasticity effect on the flow structure, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 68–76. DOI: 10.18721/JPM.13407

Introduction

Circulatory system diseases are one of the major causes of death worldwide. Numerical simulation of blood flow in 3D artery models provides new data on the flow structure, which helps predict the development of vascular diseases due to hemodynamic causes.

The abdominal aorta is an artery that supplies blood to the abdominal cavity and lower extremities. Currently, studies of flow in the abdominal aortic bifurcation are performed both experimentally [1–3] and by numerical simulation methods [4–9]. Some studies consider flows in the abdominal aortic bifurcation with subsequent bifurcations of iliac arteries [6, 8, 9].

Numerical simulation of blood flow in three different models of abdominal aortic bifurcation with iliac arteries was performed in [6]; the models were built based on anatomical data of magnetic resonance angiography. Blood flow velocities were obtained in [7] from magnetic resonance imaging data; shear stress distributions at rest were analyzed for all three models. The data on pressure and blood flow velocities using both rigid and deformable abdominal aortic wall models were given in [8].

Notably, numerical studies of blood flow in the abdominal aorta are performed mainly for models with rigid walls [10, 11]. The rigid wall approximation is largely used because of the complexity of solving the computational problem of the relationship between blood flow and vessel deformation. Such an approximation is justified by the fact that wall deformations do not significantly change the velocity field under normal conditions. It is used for arteries with small wall pulsations, but it may not be acceptable for arteries with large deformations (e.g., the thoracic or abdominal aorta). Clinical studies [12, 13] obtained the displacements of the abdominal aortic wall reaching 4–10%.

Analysis of the literature showed that most computational studies with 3D models of the abdominal aorta with iliac arteries use patient-specific models [8, 9], constructed based on clinical data. An abdominal aortic bifurcation model was used in [9], providing the flow pattern and the field of averaged shear stresses. However, the authors did not study the dependence of flow characteristics on wall elasticity.

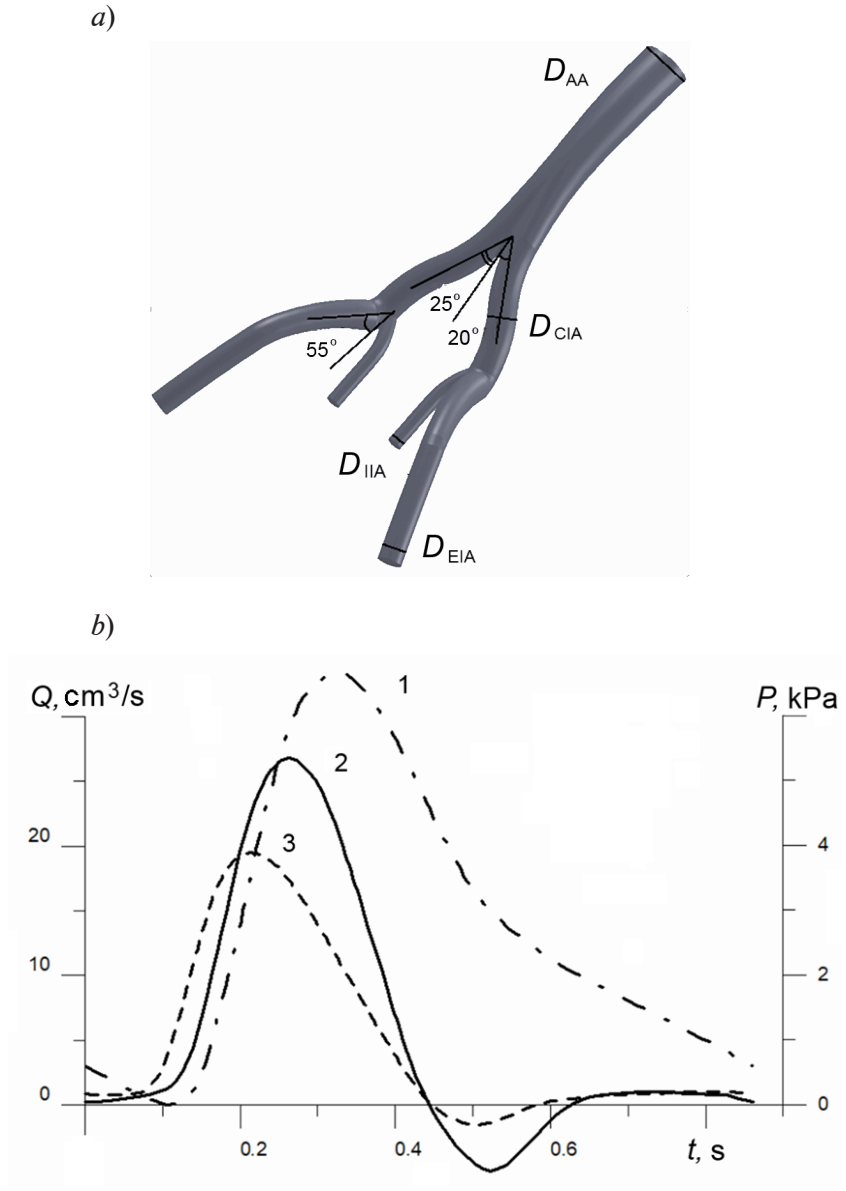


Fig. 1. Three-dimensional model of vessel bifurcation [14] (*a*) and graphical representation of boundary conditions for the mechanical (*1*) and hydrodynamic (*2*, *3*) problems posed (*b*).
The dynamics of the pulsatile component of pressure P on the inner wall of the abdominal aorta (AA) (*1*) and flow rates Q at the outlet of the external (EIA) (*2*) and internal (IIA) (*3*) iliac arteries; D_{CIA} is the diameter of the common iliac artery

Table

Geometric characteristics of the abdominal aortic bifurcation model

Bifurcation segment	Inlet diameter D_{in}, mm	Outlet diameter D_{out}, mm	Vessel length L, mm
Abdominal aorta (AA)	18.0	15.5	85
Left and right iliac arteries:			
common (CIAs)	10.8	10.8	46
external (EIAs)	10.8	9.0	62
internal (IIAs)	5.5	5.5	38

Parametric flow computations in models of blood vessels of average geometry help answer the question whether the wall deformation should be taken into account in patient-specific models or the changes in the flow structure during wall deformation are insignificant and can be neglected without significant loss of accuracy.

This work investigates the influence of wall elasticity on the structure of flow in a model of abdominal aortic bifurcation with iliac arteries of average statistical three-dimensional geometry. We analyzed the influence of the above elasticity on the structure of cross flow as well as on the shear stress magnitude and oscillatory shear index.

Geometric model of abdominal aortic bifurcation

In the present study, we used the model of abdominal aortic bifurcation with iliac arteries (Fig. 1,*a*) from [14], where its construction is described in detail based on averaged clinical data. The model includes the abdominal aorta, right and left common iliac arteries, right and left internal and external iliac arteries.

The geometric characteristics of the model are shown in Table. The deviation from the outlet axis of the abdominal aorta is 25° and 20° for the left and right iliac arteries, respectively. The angle between the internal iliac artery and the plane of the external iliac arteries is 55°.

Mathematical model

The simulation of fluid and wall motion was performed using the Fluid-Structure Interaction (FSI) technology, implemented as a one-way FSI data transfer. The flow in the abdominal aortic bifurcation with the iliac arteries was simulated by solving unsteady 3D Navier-Stokes equations for an incompressible viscous fluid on a deformable mesh. The computations were performed in the ANSYS CFX program with second-order spatial and temporal accuracy. The equation of motion for the solid wall was solved in the ANSYS Transient Structural program based on the finite element method. Data transfer between solvers was performed in the ANSYS System Coupling module, which transfers data on pressure and motion between ANSYS CFX and ANSYS Transient Structural. A mechanical problem is solved at each time step, with the overpressure P, which is the difference between the instantaneous pressure and the minimum (diastolic) pressure, applied to the inner wall (Fig. 1,*b*). Next, wall displacement are automatically transferred to the hydrodynamic problem, where fluid flow in the deformed model is computed, and then the next time step is processed.

The boundary conditions were obtained by averaging clinical data recorded in fifteen patients using magnetic resonance imaging [15–20]. A constant pressure level is set at the inlet to the abdominal aorta, and four pulsatile flow rates, the same for the left and right branches, are set at the outlet (see Fig. 1,*b*). The pulsation period is 0.86 s, which corresponds to a normal resting heart rate (70 bpm). The fluid is considered to be incompressible Newtonian, with the properties similar to those of human blood: density of 1000 kg/m^3, dynamic viscosity of 0.004 Pa·s.

The Reynolds number Re at the inlet to the abdominal aorta is equal to 1600 at the moment of maximum flow, the Womersley number Wo is equal to 24.3. These numbers are expressed as

$$\mathrm{Re} = \frac{Ud}{\nu}, \mathrm{Wo} = d\sqrt{\frac{2\pi}{\nu T}},$$

where U, m/c, is fluid velocity; d, m, is the aortic diameter; ν, m^2/s, is the kinematic viscosity of blood; T, s, is the pulsation period.

Mechanical characteristics of the abdominal aorta

The displacements of the cylindrical tube wall were computed using an isotropic elastic wall model. The wall thickness was taken to be 1 mm [21], its density 1000 kg/m^3, Poisson's ratio $\mu = 0.4$ [22]. It was assumed that the walls at the inlet and outlet were rigidly fixed.

The reference value of Young's modulus of the wall was taken as $E = 1$ MPa, it was chosen so that the wall deformation in the middle of abdominal aorta at maximum applied pressure P (see Fig. 1, *b*) was about 5%, which corresponds to physiological level of wall pulsation [20]. To estimate the value of Young's modulus, we used the following analytical formula [20]:

$$E = P\frac{R_0}{h_0}\left(1-\frac{\mu}{2}\right)\left(\frac{R_0}{R}-\frac{R_0^2}{R^2}\right)^{-1}, \qquad (1)$$

where R, R_0 are radii of the tube's medial surface at maximum (systolic) and minimum (diastolic) pressure, respectively; h_0 is the thickness of the tube wall.

In addition to the basic value of Young's modulus $E = 1$ MPa, computations were carried out for values $E = 2$, 4, 8 MPa and the problem with a rigid wall ($E \to \infty$).

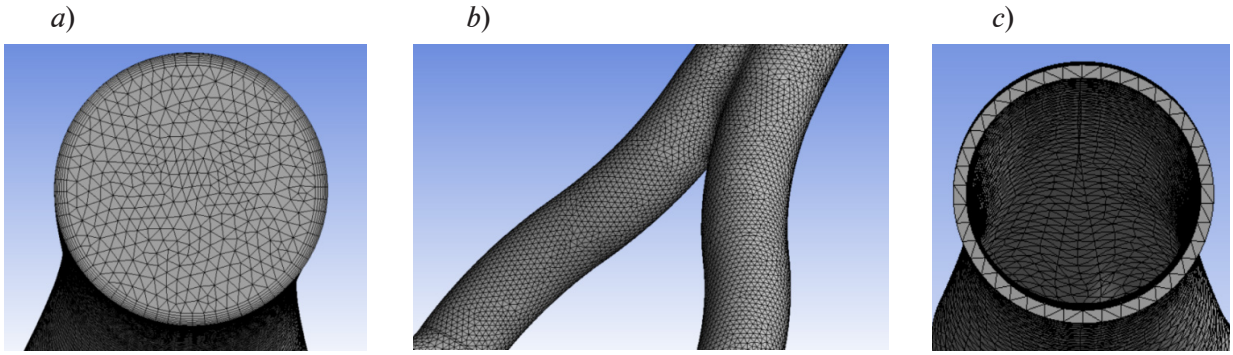


Fig. 2. Computational meshes used for the fluid (*a*) and for the wall (*b*, *c*)

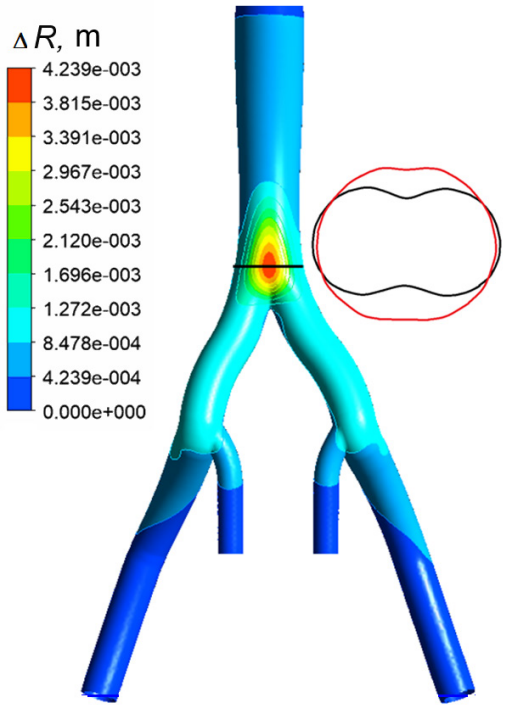


Fig. 3. Field of computed displacements of the elastic model wall at the moment of maximum pressure (rear wall view); boundaries of the vessel cross section at the moments of maximum (red line) and minimum (black line) pressure. Reference value of Young's modulus $E = 1$ MPa

Computational aspects

The CFD mesh (Fig. 2) consists of tetrahedra and has a layer of prismatic cells clustered to the wall. The computational mesh for the vessel wall also consists of tetrahedra. A study of mesh and time convergence was carried out for both meshes; as a result, a mesh of 0.8 million elements was used for CFD computations, and a mesh of 50,000 elements for the vessel wall. The time step was 0.01 s. Three periods were computed for all variants in order to exclude the influence of initial conditions.

Computational results and discussion

Wall displacements. Fig. 3 shows computed wall displacements at $E = 1$ MPa at the moment of maximum pressure P (see Fig. 1,*b*). Evidently, the maximum displacements are observed just before the abdominal aortic bifurcation, where the initial oval cross section of the vessel tends to take a circular shape.

The maximum displacements in the middle of the abdominal aorta are 5%, which corresponds to the preliminary estimate by Eq. (1) and justifies the choice of the basic value of Young's modulus ($E = 1$ MPa).

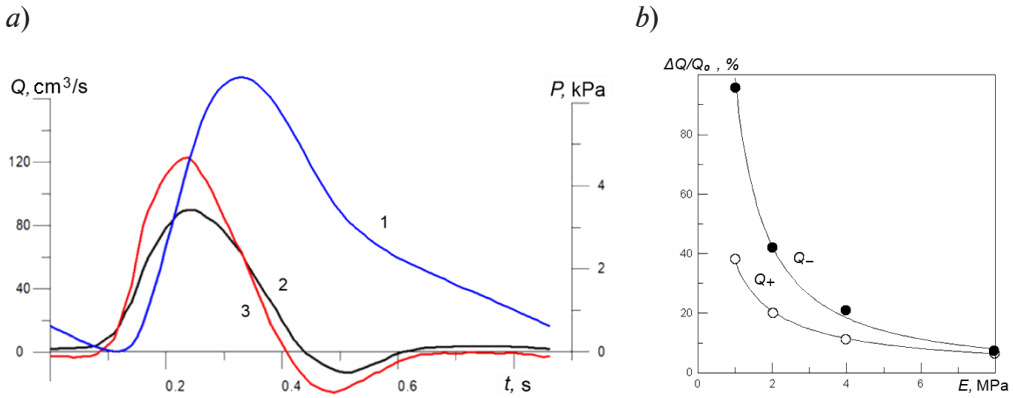


Fig. 4. Evolution of pulsatile component of pressure (*1*) on the inner wall of the vessel model (mechanical problem) and flow at the inlet to the abdominal aorta computed for rigid (*2*) and elastic (*3*) walls at $E \to \infty$ and 1 MPa, respectively (CFD problem) *(a)*; dependences of relative increase in amplitudes of positive (Q_+) and negative (Q_-) flow waves on Young's modulus (*b*)

Flow rate. Fig. 4,*a* shows the time evolution of the computed flow rate at the inlet to the abdominal aorta for the model with rigid ($E \to \infty$) and elastic (E = 1 MPa) walls. The elasticity of the walls increases the amplitude of both positive and negative flow waves. As the pressure increases, the walls of the vessel stretch, the cross-sectional area increases, and therefore the flow rate in the elastic model exceeds that of the rigid-wall model. As the pressure decreases, the walls contract, the cross-sectional area decreases, and the fluid is pushed out of the vessel: the negative flow rate increases, compared to the flow rate for the vessel with rigid walls. In addition, the duration of the positive flow phase decreased by 0.1 s for the deformed vessel (respectively, the duration of the negative flow phase increased by 0.1 s). The effect of Young's modulus on the change in the amplitude of positive (Q_+) and negative (Q_-) flow waves, compared with the results of the model with rigid walls, is shown in Fig. 4,*b*. The increase in amplitudes for the elastic model with E = 1 MPa was 40 and 95%, respectively.

Longitudinal velocity. Fig. 5 shows the fields of longitudinal velocity V_n (in the projection on the direction of the main flow) in three sections for the phases of increasing, maximum and decreasing pressure for models with rigid and elastic (E = 1 MPa) walls. It can be seen that the longitudinal velocity profile has an irregular shape in all cases. The velocity in the abdominal aorta (yellow section) is higher on the anterior wall, since this wall is external to the curved abdominal aorta. In the second part of the decreasing pressure phase, backflow is observed. The pattern of the longitudinal velocity field in the section before the bifurcation is similar to that in the middle of the abdominal aorta, but the velocities are increased. Maximum for the longitudinal velocity in the common iliac artery is shifted toward the inner wall of the artery in the positive flow phase, while in the negative phase, on the contrary, it is shifted toward the outer wall. There is flow separation at the external wall of the common iliac artery at the moment of maximum pressure.

During the phase of increasing pressure (at the moment of maximum flow), the maximum cross-sectional longitudinal velocity *Vn* in the elastic-wall model is higher than in the rigid-wall model by about 15% in the abdominal aorta, and by 2% in the common iliac artery. During the phase of decreasing pressure, the longitudinal backflow velocity in the elastic-wall model is higher for the elastic model: by 25% in the abdominal aorta and by 15% in the common iliac artery.

Transverse velocity. Fig. 6 shows the streamlines in three cross sections for models with rigid ($E \to \infty$) and elastic (E = 1 MPa) walls. Dean's vortices appear in the given sections of bifurcations of the abdominal aorta and iliac arteries due to the curvature of the vessels. In this case, the same vortex structure of the flow is observed in both rigid and elastic models, differing only by the values of the transverse flow velocity due to stretching of the vessel walls. The maximum values of transverse velocity for the elastic model exceed those for the rigid model by about 10%.

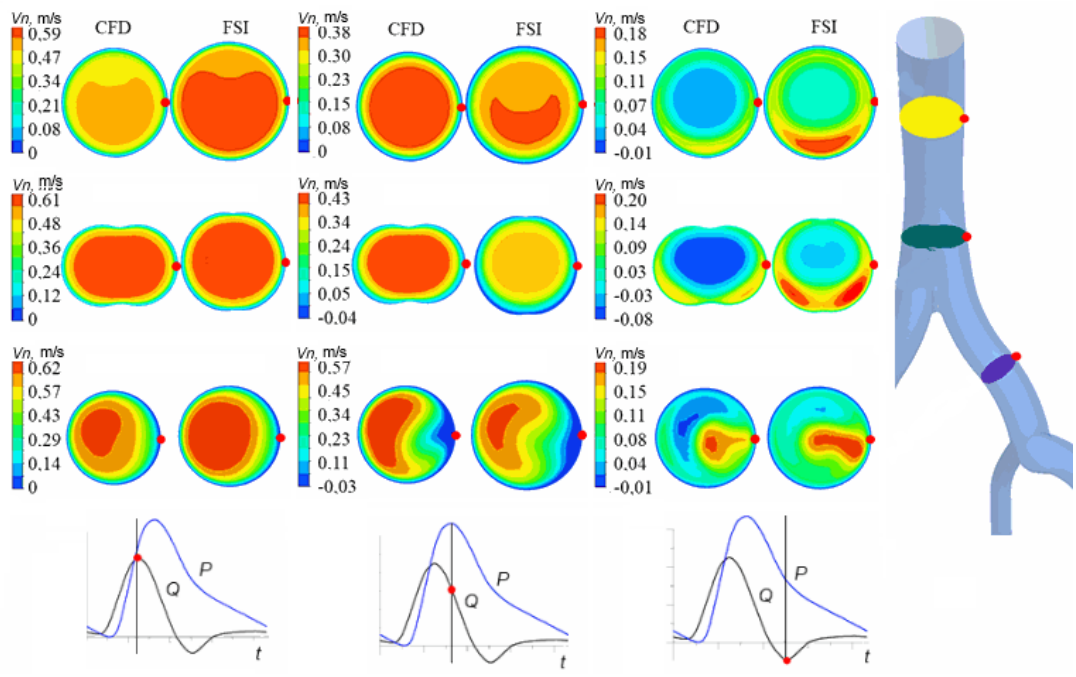


Fig. 5. Patterns of longitudinal velocity fields in three cross sections in different phases of the cycle, computed for rigid (CFD, $E \to \infty$) and elastic (FSI, $E = 1$ MPa) models of the abdominal aorta; the dimensions of the cross sections are conventionally shown as equal, the pressures and flow rates for the three phases of the cycle are marked at the bottom of the figure

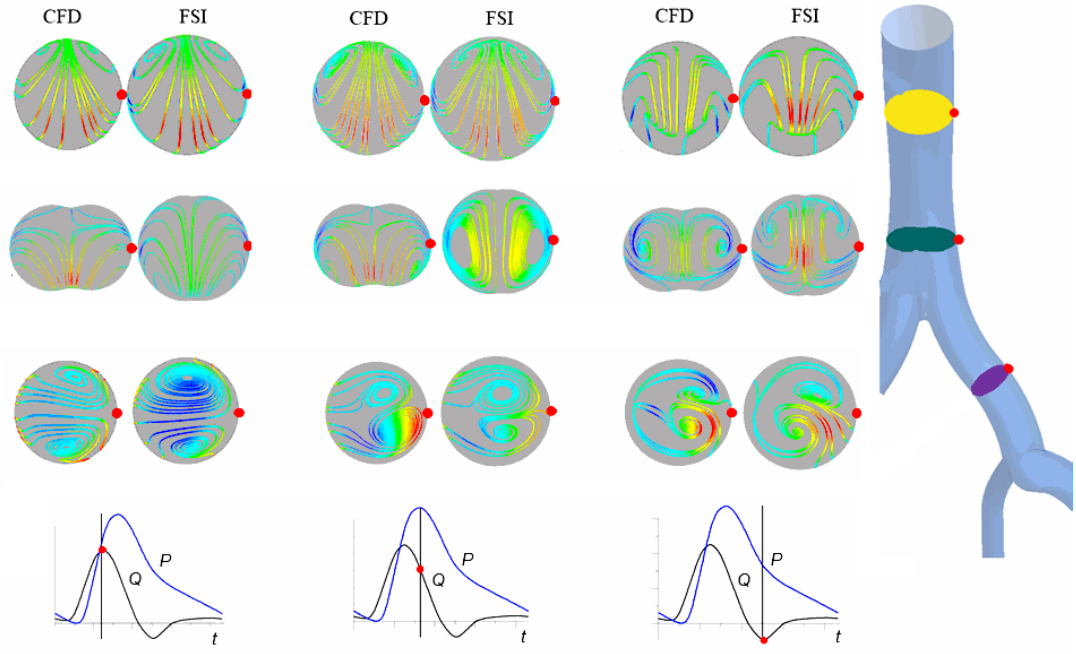


Fig. 6. Streamlines in three cross sections in different phases of the cycle, computed for rigid (CFD, $E \to \infty$) and elastic (FSI, $E=1$ MPa) models of abdominal aortic bifurcation; the dimensions of arterial cross-sections are conventionally shown as equal, the pressures and flow rates for the three phases of the cycle are marked at the bottom of the figure

Wall shear stress. It is known that regions with low shear stresses and high oscillatory shear index on the vessel wall are associated with formation and development of atherosclerosis. Lower (dangerous from the standpoint of atherosclerosis development) shear stresses on the wall are observed in the regions where flow separation occurs. High values of oscillatory shear index are also observed there. To assess the influence of wall elasticity on shear stresses, TAWSS (Time-Averaged Wall Shear Stress) and OSI (Oscillatory Shear Index) values were calculated for rigid and elastic models:

$$\mathrm{OSI} = \frac{1}{2}\left(1 - \frac{\left|\int_0^T \boldsymbol{\tau}_w dt\right|}{\int_0^T |\boldsymbol{\tau}_w| dt}\right),$$

$$\mathrm{TAWSS} = \frac{1}{T}\int_0^T |\boldsymbol{\tau}_w| dt,$$

where $\boldsymbol{\tau}_w$ is the vector of wall shear stress; t, s, is the time; T, s, is the pulsation period.

Fig. 7,*a* shows the distributions of time-averaged wall shear stress and the oscillatory shear index on the abdominal aortic bifurcation wall in the rigid ($E \to \infty$) and elastic (E = 1MPa) models. The effect of Young's modulus on these characteristics is shown in Fig. 7,*b*. The greatest difference in shear stresses (20% less in the elastic model, compared to the rigid model) is observed in the abdominal aorta, in the location of the greatest wall deformation. A minor difference (the difference does not exceed 10% in the elastic model relative to the rigid model) is observed in the external iliac arteries; in the remaining parts of the abdominal aorta, the influence of its wall elasticity on the averaged shear stresses does not exceed 2%.

The greatest difference between the values of the oscillatory shear index (OSI) is observed before the abdominal aortic bifurcation: the OSI value for the elastic model (on average over the circumference) is higher than the data for the rigid model by approximately 60%. Smaller differences from the rigid model are observed in the iliac arteries: OSI values for the elastic model are 10% higher.

Conclusion

The influence of wall elasticity on blood flow in the bifurcation of the abdominal aorta with iliac arteries of the average statistical geometry was investigated using numerical simulation. The maximum wall deformation in the middle of the abdominal aorta was 5%, which corresponds to the physiological level of wall pulsation. The greatest influence of elasticity is observed in the region immediately before the bifurcation,

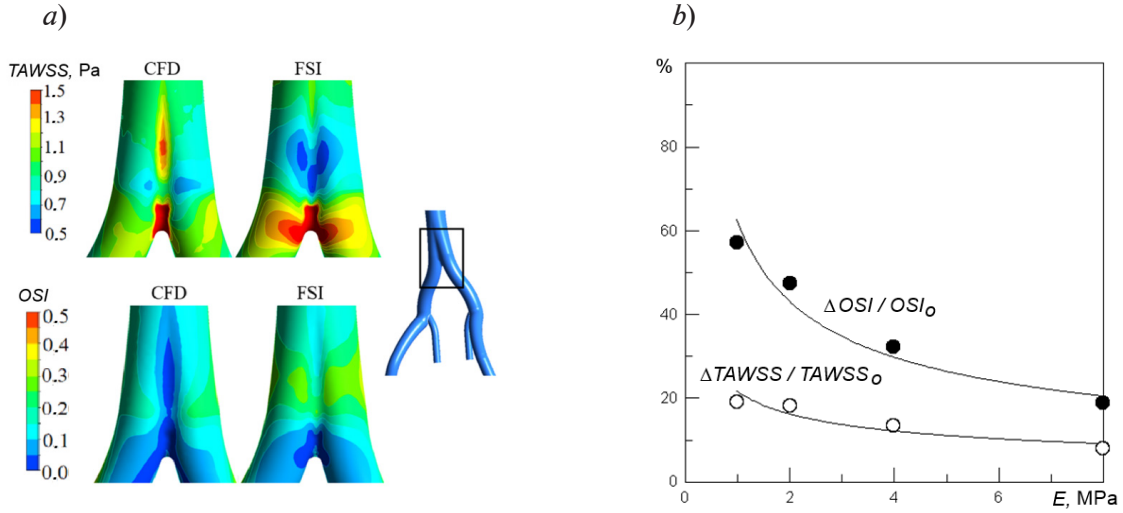


Fig. 7. Computations for two models (CFD and FSI) for the time-averaged wall shear stress (TAWSS (I)) and oscillatory shear index (OSI (II)) in the abdominal aortic bifurcation (*a*); dependences of relative differences of the above characteristics, I and II, on Young's modulus values (*b*)

Rigid (CFD, $E \to \infty$) and elastic (FSI, E = 1 MPa) models were used, the index "o" marks the characteristics of the rigid model

where there are large deformations of the vessel walls, changing the shape of the cross section from oval to circular.

According to the results obtained, the structure of transverse flow along the vessel has no qualitative differences for elastic and rigid bifurcation models. We have found that the maximum values of transverse velocity for the elastic model are about 10% higher than for the rigid model.

Due to the change in the cross-sectional area of the vessel resulting from the deformation of the walls, the flow rate increases at the phase of increasing pressure and decreases with decreasing pressure. Comparative analysis of the obtained results showed that the amplitude of the forward flow rate increased by 40% for the elastic model, while that of the backflow rate almost doubled.

The time-averaged wall shear stress (TAWSS) decreased by 20% in the bifurcation region for the model with elastic walls, and oscillatory shear index (OSI) increased by 60%.

This work was supported financially by the Russian Foundation for Basic Research (grant no. 18-01-00629).

REFERENCES

1. **Chong C.K., How T.V.,** Flow patterns in an endovascular stent-graft for abdominal aortic aneurysm repair, Journal of Biomechanics. 37(1) (2004) 89–97.

2. **Ku D.N., Glagov S., Moore Jr. J.E., Zarins C.K.,** Flow patterns in the abdominal aorta simulated postprandial and exercise conditions: an experimental study, Journal of Vascular Surgery. 9 (2) (1989) 309–316.

3. **Gataulin Ya.A., Zaitsev D.K., Smirnov E.M., Fedorova E.A., Yukhnev A.D.,** Weakly swirling flow in a model of blood vessel with stenosis: Numerical and experimental study, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. (4(230)) (2015) 36–47.

4. **Gataulin Ya.A., Zaitsev D.K., Smirnov E.M., Yukhnev A.D.** Numerical study of spatial-temporal evolution of the secondary flow in the models of a common carotid artery, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. (4 (253)) (2016) 48–55.

5. **Gataulin Y.A., Yukhnev A.D., Zaitsev D.K., et al.,** Structure of the secondary flow in the bifurcation of a blood vessel: patient specific modeling and clinical Doppler measurements, Journal of Physics: Conference Series. 1135 (2018) 012089.

6. **Dalman R.L., Tedesco M.M., Myers J.,** AAA disease: mechanism, stratification, and treatment, Annals of the New York Academy of Sciences. 1085(1) (2006) 92–109.

7. **Tang B.T., Cheng C.P., Tsao P.S., Taylor C.A.,** Subject-specific finite element modeling of three-dimensional pulsatile flow in the human abdominal aorta: comparison of resting and exercise conditions, Proceedings of the 2003 Summer Bioengineering Conference, June 25 – 29, 2003, Sonesta Beach Resort, Key Biscayne, **Florida**, USA (2003) 2.

8. **Figueroa C.A., Vignon-Clementel I.E., Jansen K.E., et al.,** A coupled momentum method for modeling blood flow in three-dimensional deformable arteries, Computer Methods in Applied Mechanics and Engineering. 195(41) (2006) 5685–5706.

9. **Ke L., Wentao J., Yu** C**., et al.,** Fluid-solid interaction analysis on iliac bifurcation artery: a numerical study, International Journal of Computational Methods. Chengdu, China.17. 16 (7) (2019) 1850112.

10. **Javadzadegan A., Simmons A., Barber T.,** Spiral blood flow in aorta-renal bifurcation models, Computer Methods in Biomechanics and Biomedical Engineering. 19(9) (2015) 964–976.

11. **Mesri Y., Niazmand H., Deyranlou A.,** Numerical study on fluid-structure interaction in a patient-specific abdominal aortic aneurysm for evaluating wall heterogeneity and material model effects on its rupture, Journal of Applied Fluid Mechanics. 10(6) (2017) 1699–1709.

12. **Ahlgren A.R., Hansen F., Sonesson B., Lanne T.,** Stiffness and diameter of the common carotid artery and abdominal aorta in wmen, Ultrasound in Medicine and Biology. 23 (7) (1997) 983–988.

13. **Koullias G., Modak R., Tranquilli M., et al.,** Mechanical deterioration underlies malignant behavior of aneurysmal human ascending aorta, The Journal of Thoracic and Cardiovascular Surgery. 130 (3) (2005) 677–683.

14. **Sinitsyna D.E., Yukhnev A.D., Zaitsev D.K., Turkina M.V.,** The flow structure in a three-dimensional model of abdominal aortic bifurcation: ultrasonic and numerical study, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 12 (4) (2019) 50–60.

15. **Yeung J.J., Kim H.J., Abbruzzese T.A., et al.,** Aortoiliac hemodynamic and morphologic adaptation to chronic spinal cord injury, Journal of Vascular Surgery. 44 (6) (2006) 1254–1265.

16. **Malossi A.C.I., Bonnemain J.,** Numerical comparison and calibration of geometrical multiscale models for the simulation of arterial flows, Cardiovascular Engineering and Technology. 4 (4) (2013) 440 –463.

17. **Alimohammadi M., Pichardo-Almarza C., Tomaso G., et al.,** Predicting atherosclerotic plaque location in an iliac bifurcation using a hybrid CFD/biomechanical approach, Lecture Notes in Computer Science: Bioinformatics and Biomedical Engineering, Part II, Eds. Ortuco F., Rojas I. 9044 (2015) 594–606.

18. **Xiao N.,** Simulation of 3-D blood flow in the full systemic arterial tree and computational frameworks for efficient parameter estimation, A dissertation submitted to Stanford University for the degree of PhD, Stanford, 2014. www.isd.ksl.ac.uk/FigueroaLabFiles/papers/NX-thesis.

19. **Long Q., Xu X.Y., Bourne M., Griffith T.M.,** Numerical study of blood flow in an anatomically realistic aortic-iliac bifurcation generated from MRI data, Magnetic Resonance in Medicine. 43 (4) (2000) 565–576.

20. **Tang B.T., Cheng C.P., Draney M.T., et al.,** Abdominal aortic hemodynamics in young healthy adults at rest and during lower limb exercise: quantification using image-based computer modeling, American Journal of Physiology Heart and Circulatory Physiology. 291(2) (2006)668–676.

21. **Schriefl A.J., Zeindlinger G., Pierce D.M., et al.,** Determination of the layer-specific distributed collagen fibre orientations in human thoracic and abdominal aortas and common iliac arteries, Journal of the Royal Society Interface. 9 (71) (2012) 1275–1286.

22. **Mbodj C., Altnji H.E., Bou-Said B., Walter-Le Berre H.,** Analysis of the phenomenon of endoleak of type I A. Influence of the mechanical characterization of the aorta, Journal of Hypertension and Management. 2 (1) (2016) 1510014.

23. **Begun P.I., Afonin P.N.,** Modelirovaniye v biomekhanike [Modeling in biomechanics], Vysshaya shkola, Moscow, 2004 (in Russian).

Received 01.10.2020, accepted 30.10.2020.

THE AUTHORS

KOTMAKOVA Anna A.
Scientific Research Institute for Optoelectronic Instrument Engineering
29 Leningradskaya St., Sosnovy Bor, Leningrad region,188540, Russian Federation
www.anechka_kotmakova@mail.ru

GATAULIN Yakov A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
yakov_gataulin@mail.ru

YUKHNEV Andrey D.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
a.yukhnev@mail.ru

ZAITSEV Dmitri K.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
zaitsev-aero@yandex.ru

DOI: 10.18721/JPM.13408

IDEAL FOCUSING SYSTEMS WITH HOMOGENOUS MAGNETIC FIELDS

K.V. Solovyev

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

Institute for Analytical Instrumentation of the Russian Academy of Sciences, St. Petersburg, Russian Federation

Cases of charged particle pack ideal focusing in the presence of constant magnetic field have been considered. One-directional ideal space-time focusing was shown to remain only on conditions that a magnetic field being homogenous and its direction being the same as the one of quadratic potential growth. Axially symmetric electrostatic fields with superimposed magnetic field were taken as an example because of their practical importance in the mass spectrometry. It was concluded that at least one equation with separated motion should be linear to maintain the ideal space-time focusing.

Keywords: spectrometry, ideal focusing, ion trap, homogenous magnetic field

Citation: Solovyev K.V., Ideal focusing systems with homogenous magnetic fields, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 14 (1) (2021) 77–81. DOI: 10.18721/JPM.13408

Introduction

Field structures providing ideal space-time focusing (IPTF) of a charged particle beam in one of the directions serve as a theoretical basis for synthesizing a wide class of high-resolution devices for mass-spectrometry [1, 8–11]. The entire range of electrostatic fields with a quadratic dependence of the potential on one of the coordinates was described for the first time in [2]. These fields are the ones providing IPVF for the beam. Electrostatic integrable systems with ideal space-time focusing in one of the directions have been sufficiently studied by now [3–7]; one of the non-integrable types considered closely are the so-called Cassinian traps [8–11], combining in some cases the functions of the trap and the ion bunch manipulation system [11].

There are limitless possibilities for synthesis of ideally focusing electrostatic systems. However, to complete the picture, the problem of preserving ISTF in the presence of a magnetic field should be reconsidered (see, for example, an earlier discussion in [12]).

Analysis of Ideal Focusing Systems

Let us first construct a dimensionless model of the problem posed. The motion of a charged particle in an electric field with the potential Φ and a magnetic field with a magnetic flux density vector $\mathbf{B}$ is described by the equation

$$m\frac{d^2\mathbf{R}}{dt^2} = -q\cdot\operatorname{grad}\Phi + q\frac{d\mathbf{R}}{dt}\times\mathbf{B}, \qquad (1)$$

where m, q are the mass and charge of the ion; $\mathbf{R}$ is the radius vector of the particle; t is the time.

We introduce the dimensionless variables $\mathbf{r}$, τ, φ, $\mathbf{b}$, μ in terms of the following relations:

$$\mathbf{R} = l\mathbf{r}, t = T\tau, \Phi = \Phi_0\phi,$$

$$\mathbf{B} = B_0\mathbf{b}, m = m_0\mu,$$

where l is the characteristic size of the system; Φ_0, B_0 are its characteristic potential and magnetic induction; T is the time unit; m_0 is the mass unit;

$$\mathbf{r} = (x, y, z), \mathbf{b} = (b_x, b_y, b_z)$$

are, respectively, the dimensionless radius vector and the dimensionless vector of the magnetic field, which is unit everywhere for a homogeneous field and unit at a given point in space for an inhomogeneous field.

Substituting new variables into Eq. (1), we obtain:

$$m_0\mu \cdot \frac{l}{T^2} \cdot \frac{d^2\mathbf{r}}{d\tau^2} = \\ = -q \cdot \frac{\Phi_0}{l} \cdot \frac{\partial\phi}{\partial\mathbf{r}} + q \cdot \frac{l}{T} \cdot \frac{d\mathbf{r}}{d\tau} \times (B_0\mathbf{b})$$

or

$$\mu \frac{d^2\mathbf{r}}{d\tau^2} = -q \cdot \frac{T^2\Phi_0}{m_0 l^2} \cdot \frac{\partial\phi}{\partial\mathbf{r}} + \\ + q \cdot \frac{TB_0}{m_0} \cdot \frac{d\mathbf{r}}{d\tau} \times \mathbf{b}.$$

One of the three cases has to be chosen taking T as a constant for measuring time: either the coefficient for the first or second term on the right-hand side is set to unity, or a certain relationship is imposed between the values of these coefficients.

Let us consider the first case, taking

$$T = l\sqrt{\frac{m_0}{|q\Phi_0|}}. \qquad (2)$$

Then we obtain the equation of motion in the form

$$\mu\ddot{\mathbf{r}} = -\nabla\phi + \beta \cdot \dot{\mathbf{r}} \times \mathbf{b}, \qquad (3)$$

where the overdot denotes differentiation with respect to τ, and the gradient is taken over the components of the vector $\mathbf{r}$. The parameter β is determined by the expression

$$\beta = lB_0\sqrt{\frac{1}{m_0}\left|\frac{q}{\Phi_0}\right|}. \qquad (4)$$

Evidently, $\mathbf{b} = \mathbf{e}_z$ for a homogeneous magnetic field directed along the axis z. This scheme for introducing dimensionless variables differs from that adopted in [3–7], since the presence of a magnetic field affects the trajectory geometry of particles with different masses. The dimensionless potential of an ideal focusing electrostatic system has the form [2]:

$$\phi(x, y, z) = z^2 + f(x, y), \\ f_{xx} + f_{yy} + 2 = 0. \qquad (5)$$

Due to the additive term z^2 in potential (5), the motion in the direction z in the field with potential (5) with the parameter $\beta = 0$ is separated from the motion in the directions x and y and obeys the equation of harmonic oscillations:

$$\mu\ddot{z} = -2z, \qquad (6)$$

whose solution has the form

$$z = z_0 \cos\left(\sqrt{\frac{2}{\mu}}\tau\right) + \frac{\dot{z}_0}{\sqrt{2}} \sin\left(\sqrt{\frac{2}{\mu}}\tau\right). \qquad (7)$$

Solution (7) guarantees ideal space time focusing of the beam in the planes $z = \pm z_0$ if charged particles start from the plane z_0 with any $\dot{z}_0$ or focusing of the beam in the plane $z = 0$ at their start from different points z_0 with a zero z-component of the initial velocity.

Applying an external magnetic field transforms the equation of motion with respect to z:

$$\mu\ddot{z} = -2z + \beta(\dot{x}b_y - \dot{y}b_x). \qquad (8)$$

Apparently, it is only at $\mathbf{b} = \mathbf{e}_z$ that Eq. (8) retains form (6), typical for electrostatics, guaranteeing that oscillations along the z-axis are independent of motion in the plane orthogonal to z. If $b_x \neq 0$ or $b_y \neq 0$, there is a 'mixing' of coordinates in system (3), and as a result the motion in the direction z is not separated, linearity is violated (unless the potential is a quadratic form and the system of equations of motion is not completely linear; see, for example, [13]). The field must be homogeneous, since the dependence of any component of the vector $\mathbf{b}$ on coordinates implies that (at least) one more component of the vector is different from zero and depends on the coordinates, by virtue of the equation div $\mathbf{b} = 0$.

Thus, it is only possible to expand the class of fields with ideal space time focusing if systems of the following type are included in it:

electrostatic field of type (5) + + homogeneous magnetic field directed along the z-axis.

Recall here that a homogeneous magnetic field provides ideal space time focusing of the beam in the following well-known cases, which we consider in dimensional coordinates assuming $\mathbf{B} = (0, 0, B)$.

1. $\mathbf{E} = 0$, $\mathbf{B} \neq 0$. The period of rotation of particles in a homogeneous magnetic field

$$T = \frac{2\pi m}{qB} \qquad (9)$$

does not depend on the initial data and is determined by the field strength and the charge/mass ratio.

Thus, the beam is ideally focused at the point coinciding with the starting point in the projection onto a plane orthogonal to the magnetic flux density vector, after a time equal to the period T. Particles of all masses are focused at this point but at different times. The motion along the vector **B** is uniform. Given a nonzero z-component of the initial particle velocity, the source becomes spatially separated from the detector (along the coordinate z). A time-of-flight spectrometer based on this principle is described in [14].

2. $\mathbf{E} \neq 0$, $\mathbf{B} \neq 0$, the vector **E** is parallel to the vector **B**. This case is similar to the first one, but the motion along z is uniformly accelerated, so the image size on the detector can be controlled.

3. $\mathbf{E} \neq 0$, $\mathbf{B} \neq 0$, the vector **E** is not parallel to the vector **B**. For definiteness, let the electric field vector be oriented as follows:

$$\mathbf{E} = (-E_x, 0, E_z).$$

This is a generalization of the case of crossed fields (known from elementary physics), which also provides ideal one-dimensional space-time focusing in a time multiple to the period T (see Eq. (9)), in a plane passing through the starting point orthogonally to the axis x. Unlike case 2, the focusing point is shifted relative to the starting point by a distance of $2\pi m E_x/(qB^2)$ in the direction y. Lines of the mass spectrum unfold as segments parallel to the vector **B**. The time-of-flight properties of a cycloidal mass spectrometer were used in patent [15].

A notetworthy combination of an electric field quadratic in one of the coordinates of the potential, and a homogeneous magnetic field directed along the same coordinate is the case of fields with axial symmetry. It was discussed in [16, 17] but in a different context.

Further consideration of the problem will be carried out only in dimensionless coordinates (2)–(4).

The most general form of potential (5) in the axisymmetric case is as follows:

$$\phi(r,z) = z^2 - \frac{r^2}{2} + \alpha \cdot \ln r, \quad \alpha \in \{-1, 0, 1\}. \tag{10}$$

This expression includes the field of the orbital trap [1] ($\alpha = 1$), confining particles at certain initial velocities, as well as the field of the hyperboloid ($\alpha = 0$) and the field of the quadro-logarithmic type ($\alpha = -1$), which fundamentally do not confine the particles along the coordinate r in the absence of a magnetic field. Potential (10) provides harmonic oscillations of the ion in the direction z (guaranteeing ISTF) and its radial-azimuthal motion in the field of the effective potential, which has the form

$$U_0(r) = \mu \frac{r_0^4 \dot{\gamma}_0^2}{2r^2} - \frac{r^2}{2} + \alpha \ln r. \tag{11}$$

The dependence of U_0 on the mass μ generated in this case of non-dimensionalization disappears if we take into account that the total energy of the particle is related to the initial data by the relation

$$\mu\left(\frac{\dot{r}_0^2}{2} + \frac{r_0^2 \dot{\gamma}_0^2}{2} + \frac{\dot{z}_0^2}{2}\right) = E = E_r + E_\gamma + E_z. \tag{12}$$

Then $\mu \frac{r_0^2 \dot{\gamma}_0^2}{2} = E_\gamma$ and

$$U_0(r) = E_\gamma \frac{r_0^2}{r^2} - \frac{r^2}{2} + \alpha \ln r. \tag{13}$$

As a result of immersing system (10) in a homogeneous magnetic field, effective potential (11) is replaced by the potential

$$U_\beta(r) = \left(\dot{\gamma}_0 + \frac{\beta}{2\mu}\right)^2 \frac{\mu r_0^4}{2r^2} + \left(\frac{\beta^2}{4\mu} - 1\right)\frac{r^2}{2} + \alpha \ln r \tag{14}$$

and, accordingly, deformation of the initial radial-azimuthal motion occurs if ideal spatio-temporal focusing of the beam along the coordinate z is maintained. The boundaries of the region of radial motion substantially depend on the mass here.

The well-known Penning trap corresponds to the zero value of the parameter α (see, for example, monograph [18]). This scenario is sufficiently well-understood, so let us focus on the cases when $\alpha = 1\pm$.

Denoting

$$a = \left(\dot{\gamma}_0 + \frac{\beta}{2\mu}\right)^2 \frac{\mu r_0^4}{2}, c = \left(\frac{\beta^2}{4\mu} - 1\right), \tag{15}$$

we write expression (14) in the form

$$U_\beta(r) = \frac{a}{r^2} + \frac{cr^2}{2} + \alpha \ln r. \qquad (16)$$

Since $a > 0$, $U_\beta(r) \to +\infty$ for $r \to 0$. The behavior of $U_\beta(r)$ at $r \to +\infty$ is determined by the sign of the parameter c. For $c > 0$, $U_\beta(r) \to +\infty$ at $r \to +\infty$, and the dependence $U_\beta(r)$ has a minimum, forming a potential well trapping ions. For $c < 0$, $U_\beta(r) \to -\infty$ at $r \to +\infty$ and the effective potential well exists only in the presence of both a minimum and a maximum of $U_\beta(r)$; an ion can only be trapped in the well if certain conditions relating the initial data of the ion to the field parameters similar to the case of electrostatics are met [19].

Thus, the equation $\partial U_\beta(r)/\partial r = 0$ has two real roots for the case when $c < 0$:

$$r^2 = \frac{-\alpha \pm \sqrt{\alpha^2 + 8ac}}{2c}$$

provided that the discriminant is positive, namely, at $8ac > -\alpha^2$.

Then,

$$-\alpha^2/(8a) < c < 0. \qquad (17)$$

The condition $r^2 > 0$ leads to the inequality

$$-\alpha \pm \sqrt{\alpha^2 + 8ac} < 0,$$

which, if relation (17) is satisfied, always holds for $\alpha = 1$ and never for $\alpha = -1$. Using expressions (15), (17) for $\alpha = 1$, we obtain the inequalities

$$-1 < 4\mu r_0^4 \left(\frac{\beta^2}{4\mu} - 1 \right) \cdot \left(\dot{\gamma}_0 + \frac{\beta}{2\mu} \right)^2 < 0. \qquad (18)$$

The requirement that the middle part of expression (18) (that is, the right-hand side of double inequality (18)) be negative is ensured by fulfilling the condition $\beta^2 < 4\mu$, which is always true at zero magnetic field. The left-hand side of double inequality (18) is fulfilled by choosing the corresponding initial data.

The equality $c = 0$ for $\alpha = 1$ makes the existence of the effective potential well (16) possible for any initial data. If $c > 0$, the well is guaranteed for ions with arbitrary starting conditions at any value of α. Thus, the variant $\alpha = 1$ corresponds to a system capable of trapping ions with any masses μ, but only for certain initial data if $\mu \geq \beta^2/4$.

If the condition $c > 0$ is fulfilled, a system with $\alpha = -1$ can confine ions, which is impossible in a purely electric field. In this case, the requirement $\mu < \beta^2/4$, equivalent to the inequality $c > 0$, sets the threshold value of the mass of ions that are radially stable in a magnetic field with the strength β. A device based on such fields can also find applications in mass spectrometry.

In conclusion, let us point out the case of a weakly varying homogeneous magnetic field mentioned in [13], which also expands the given class of fields with ISTF.

Conclusion

It can be concluded from the study that a necessary condition for ideal space-time focusing in the system in at least one directions is either linearity of the equation with separated motion in this direction, or, given mixed coordinates, linearity of the entire system of motion equations (in particular, if the frequency multiplicity of the oscillations providing focusing in several directions is satisfied). The magnetic field must only be homogeneous; there are potential applications for its effect on ions as a holding field, including as a compensator for the repulsive forces of the electric field.

This study was partially supported by NIR 0074-2019-0009, included in State Task No. 075-01073-20-00 of the Ministry of Science and Higher Education of the Russian Federation.

REFERENCES

1. **Zubarev R., Makarov A.,** Orbitrap mass spectrometry, Analytical Chemistry. 85 (11) (20) (2013) 5288–5296.

2. **Gall L.N., Pechalina E.E., Golikov Yu.K.** Ob odnom klasse elektrostaticheskikh poley s prostranstvenno-vremennoy fokusirovkoy [About an electrostatic field class with space-time focusing], Nauchnoye Priborostroyeniye, Elektronno-Ionnaya Optika, Nauka, Leningrad. (1989) 3–7 (in Russian).

3. **Golikov Yu.K., Krasnova N.K., Solovyev K.V., Nikitina D.V.,** Integrating electrostatic traps, Prikladnaya Fizika. (5) (2006) 51–57.

4. **Golikov Yu.K., Solov'ev K.V.,** Electroctrostatic ion traps with separation of variables in parabolic coordinates, Technical Physics Letters. 36 (4)

(2010) 333–336.

5. **Golikov Yu.K., Solov'ev K.V.,** Criterion of transverse stability for ion traps, Technical Physics Letters. 37 (11) (2011) 1062–1064.

6. **Solov'ev K.V., Vinogradova M.V.,** Conditions of ion motion confinement in an electrostatic trap with separation of variables in parabolic coordinates, Technical Physics Letters. 44 (7) (2018) 618–621.

7. **Solovyev K.V., Vinogradova M.V.,** Two-electrode design for electrostatic ion trap integrable in polar coordinates, St. Petersburg State Polytechnical University Journal. Physics and Mathematics 12 (1) (2019) 96–104.

8. **Nikitina D.V.**, Ionnyye lovushki v dinamicheskoy mass-spektrometrii [Ion traps in dynamical mass spectrometry], PhD Thesis autoabstract, St. Petersburg, 2006.

9. **Köster C.,** The concept of electrostatic non-orbital harmonic ion trapping, Int. J. Mass Spectrom. 287 (1–3) (2009) 114–118.

10. **Köster C.,** Twin trap or hyphenation of a 3D Paul- and Cassinian ion trap, J. Am. Soc. Mass Spectrom. 26 (3) (2015) 390–396.

11. **Nikolaev E., Sudakov M., Vladimirov G., et al.,** Multi-electrode harmonized kingdon traps, J. Am. Soc. Mass Spectrom. 29 (11) (2018) 2173–2181.

12. **Makarov A.A.,** Ideal and quasi-ideal time focusing of charged particles, J. Phys. D.: Appl. Phys. 24 (4) (1991) 533–540.

13. **Glikman L.G.,** Fokusiruyushchiye svoystva skreshchennykh odnorodnogo magnitnogo i giperboloidalnogo elektricheskogo poley [Focussing properties of crossed homogeneous magnetic and hyperboloidal electric fields], Soviet Physics – Technical Physics. 41 (10) (1971) 2009–2015 (in Russian).

14. **Hays E.E., Richards P.I., Goudsmit S.A.**, Mass measurements with a magnetic time-of-flight mass spectrometer, Phys. Rev. 84 (4) (1951) 824–829.

15. **Voss G.F., Celo A.B., Duryea A.N.,** Cycloidal mass spectrometer with time of flight characteristics and associated method, Pat. No. 6,617,576 B1, United States, Int. Cl. B01D 59/44, Assignee: Monitor Instruments Company, LLC, Cheswick, PA (US), Appl. No.: 10/085542, Filed: 28.02.2001 (Provisional appl. No. 60/273062, filed: 2.03.2001), Date of Patent: 9.09.2003.

16. **Borodkin A.S.,** Dvizheniye zaryada v spetsialnom sluchaye staticheskogo elektromagnitnogo polya [A charge propagation in the special case of a static electromagnetic field], Soviet Physics – Technical Physics. 31 (5) (1961) 582–587 (in Russian).

17. **Borodkin A.S.**, Dvizheniye zaryazhennoy chastitsy v smeshannom elektricheskom i odnorodnom magnitnom polyakh [A charged particle's propagation in the mixed electric and homogeneous magnetic fields], Soviet Physics – Technical Physics. 41 (9) (1971) 1845–1850 (in Russian).

18. **Major F.G., Gheorghe V.N., Werth G.,** Charged particle traps. Physics and techniques of charged particle field confinement, Springer-Verlag, Berlin, Heidelberg, 2005.

19. **Golikov Yu.K., Solovyev K.V.,** Axisymmetric ion trap, Nauchnoye priborostroyeniye. 24 (1) (2014) 36–49 (in Russian).

Received 05.10.2020, accepted 26.10.2020.

THE AUTHOR

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

DOI: 10.18721/JPM.13409

THERMAL RELAXATION OF OPTICAL NONLINEARITY IN THE POLED GLASSES

I.V. Reshetov[1,2], *V.P. Kaasik*[1,2], *A.A. Lipovskii*[1,2], *D.K. Tagantsev*[1,2], *V.V. Zhurikhina*[1,2]

[1] Alferov University, St. Petersburg, Russian Federation;

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

The thermal relaxation of second-order optical nonlinearity in the subsurface layer of a poled soda-lime silicate glass has been studied. The glass annealing below glass transition temperature was shown to lead to full relaxation of the nonlinearity. At the same time, the measurements of thermostimulated depolarization current demonstrated that spatial electric charge formed in the course of the glass polarization relaxed above the glass transition temperature. This allowed concluding that the second-order optical nonlinearity in the poled glasses was not induced by the spatial electric charge.

Keywords: glass, optical nonlinearity, polarization, poling, relaxation, depolarization, thermostimulated depolarization current

Citation: Reshetov I.V., Kaasik V.P., Lipovskii A.A., Tagantsev D.K., Zhurikhina, V.V., Thermal relaxation of optical nonlinearity in the poled glasses, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 82–87. DOI: 10.18721/JPM.13409

Introduction

Second-order nonlinear optical phenomena (such as second harmonic generation (SHG)), which are observed in multicomponent glasses after their thermal polarization [1, 2], are of clear scientific interest to researchers, and in addition, can find commercial application in photonics, integrated optics, etc. For this reason, studies of thermal relaxation of optical nonlinearity in poled glasses are very relevant today.

Most publications on this problem (see, for example, [3, 4]) associate the appearance of SHG in poled glasses with the fact that a spatial electric charge is formed in glasses during polarization, which in turn creates a 'frozen' electric field in the glass volume. This field leads to breaking of the initially isotropic (centrosymmetric) glass structure, i.e., the latter becomes anisotropic, and the glass itself acquires properties of uniaxial crystals exhibiting second-order optical nonlinearity. However, it was established recently [5] that relaxation of the main part of the spatial electric charge formed in glass during its polarization occurs at temperatures above the glass transition temperature.

In this paper, we present the results of studies comparing the kinetics of SHG relaxation with the data on spatial charge relaxation obtained from measurements of thermally stimulated depolarization currents (TSDC) spectra [6].

These comparative studies were aimed at expanding our understanding of the correlation

Table

Composition of industrial alkali silicate glass

Chemical composition, wt%						
SiO_2	Al_2O	Na_2O	K_2O	MgO	CaO	Other oxides
72.2	1.2	14.3	1.2	4.3	6.4	0.33

between the processes occurring during glass polarization and the appearance of optical nonlinearity in poled glasses. However, spatial charge relaxation and SHG relaxation were found to be independent processes.

Experimental

Industrial alkaline-silicate glass purchased at Agar Scientific (Menzel Gläser slides) was used in this study. The glass composition is given in Table, the glass transition temperature of this object was 530°C [7].

The samples had a thickness of 1 mm. Polarization was carried out at 300°C for 50 min at a constant voltage of 1 kV. After polarization, the samples were cooled to room temperature under applied voltage, after which the voltage was turned off. A schematic of the glass polarization experiment is shown in Fig. 1. We used 10×20 mm stainless steel electrodes pressed against the glass surface.

At this temperature, the mobility of alkaline and alkaline earth cations contained in glass increases so much that the applied electric field forces these cations to shift noticeably towards the cathode [8, 9], with a spatial charge formed as a result. As noted above, the presence of a spatial charge generates an electric field that compensates the external applied field, and the current through the glass decreases. A typical dependence of polarization current on time is shown in Fig. 2. After cooling to room temperature under the applied voltage, the spatial charge 'freezes' in the glass and can no longer disappear (relax) due to kinetic limitations, i.e., extremely low mobility of cations at room temperature.

The SHG efficiency in samples of poled glasses was determined by Maker's fringe method [10]. The scheme of measurements used to determine the SHG characteristics is shown in Fig. 3. We measured the signal of the second harmonic emission of the Litron Nano L laser operating at 1064 nm and generating pulses with a duration of 6 ns. The incidence angle of the laser beam on the sample was equal to 63° and fixed, corresponding to the maximum value of the SHG signal. Fig. 3 also shows the full Maker picture obtained in poled glass, from which it can be seen that the the SHG signal indeed reaches maximum at an angle of 63°.

The figure also shows the measured dependence of SHG on the angle of rotation of the sample relative to laser radiation

Poled glass samples were subjected to isothermal annealing at two temperatures (275 and 400°C). The SHG signal level was measured

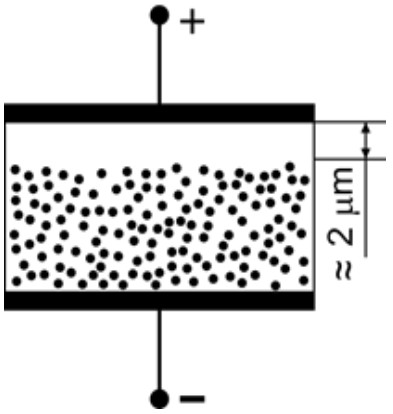


Fig. 1. Scheme of glass polarization: polarization voltage $U \approx 300$ V, polarization temperature $T \approx 250$°C, electric field in the depleted layer formed $E \approx 0.2$ V/nm.
The dots correspond to mobile cations in the sample; its region about 2 μm thick becomes depleted of mobile cations after polarization

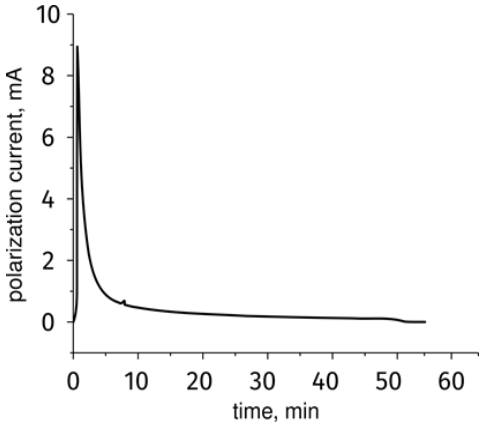


Fig. 2. Typical time dependence of polarization current in the sample glass

during annealing, i.e., the kinetics of isothermal SHG relaxation was determined.

After annealing, TSDC spectra were measured over a wide temperature range (from room temperature to glass transition temperature and higher). These spectra should either point to complete relaxation of the spatial charge after annealing at temperatures below 400°C (absence of any peaks) or determine the region of temperatures where relaxation of the 'frozen' spatial charge actually occurs, i.e., the charge carriers generating the electric field in the bulk of the glass move back to the anode region.

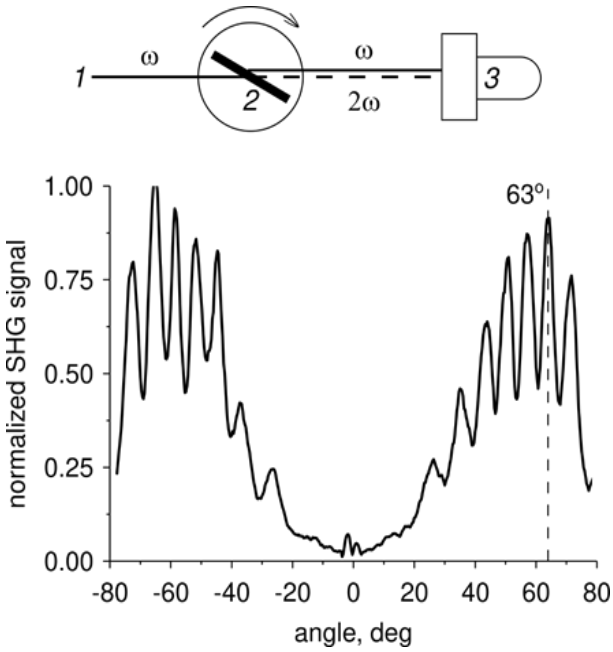


Fig. 3. Schematic for measurement of SHG efficiency in poled glasses: applied pulsed laser radiation (ω) *1*; sample *2*, photodetector *3*; the dashed line shows the second harmonic (2ω)

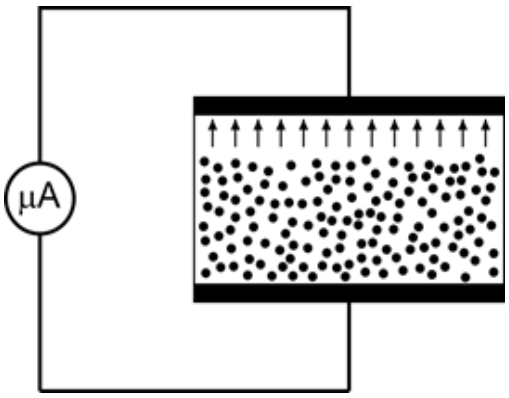


Fig. 4. Schematic for measurement of TSDC spectra of poled glasses; samples were subjected to linear heating, depolarization currents were measured with a microammeter(μA). Arrows show the direction of motion of mobile cations (see Fig. 1)

The scheme of TSDC spectra measurements is shown in Fig. 4. The spectra were measured under linear heating at a rate of 10°C/min; clamped graphite electrodes were used.

Experimental results

Fig. 5 (curve *1*) shows the dependence (on a logarithmic scale) of SHG signal intensity normalized to its maximum value, which has the form ln[$I(t)$/Imax], versus annealing time at 400 °C in poled glass. Apparently, the SHG signal relaxed almost completely in about 35 min and decreased to the level of the SHG signal generated by the non-poled glass surface.

The linear character of this dependence indicates that the SHG relaxation obeys the kinetics of the first order, and only one type of relaxant participates in the relaxation process.

The measurements of SHG signal relaxation kinetics carried out with a poled glass sample at a lower annealing temperature, namely, at 275°C, showed that in this case, the SHG relaxation kinetics is not described by a first-order reaction and the number of relaxors involved in the process is greater than one (see Fig. 5 (curve *2*)). Comparing SHG relaxation kinetics at temperatures of 275 and 400°C, we can conclude that during annealing at 275°C, in contrast to annealing at 400 °C, a process takes place in the glass that cannot be detected at 400°C because it proceeds too quickly at that temperature.

Fig. 6 shows the results of measuring the TSDC spectrum of poled glass annealed at 400°C for 50 min. It can be seen that a noticeable depolarization current appears only at temperatures above the glass transition temperature. The position of the first peak corresponds to 640°C. Since depolarization current is associated with movement of cations creating 'frozen' spatial charge, the peaks of TSDC spectrum should be related precisely to spatial charge relaxation.

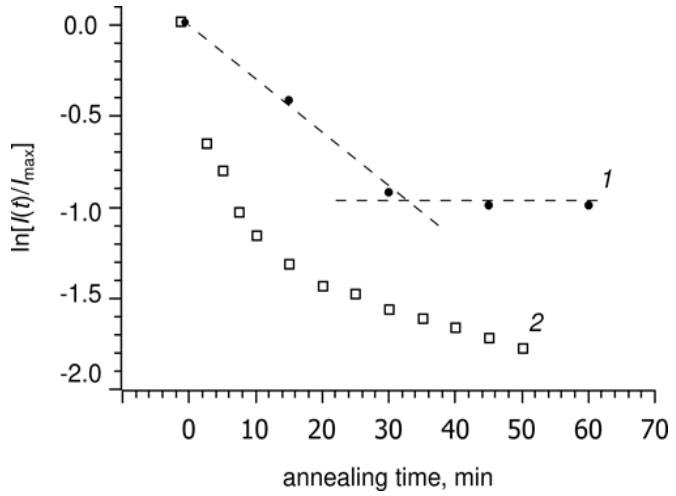


Fig. 5. Relaxation kinetics of normalized SHG at glass annealing temperatures of 400°C (*1*) and 275°C (*2*)

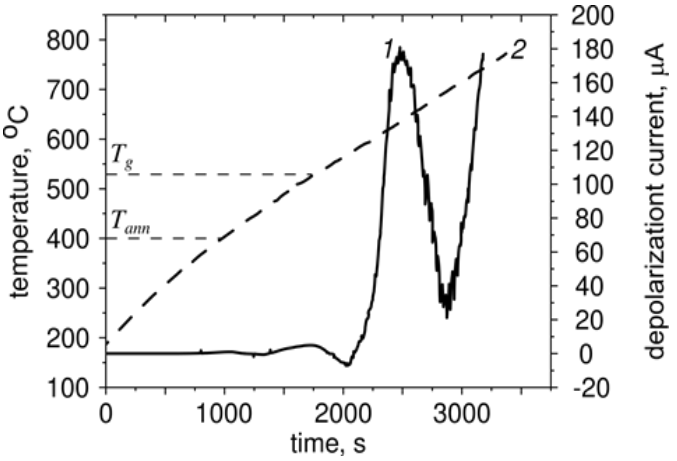


Fig. 6. TSDC spectrum (*1*) of poled glass after annealing at 400°C; the sloping dashed curve (*2*) corresponds to the temperature evolution (annealing and glass transition temperatures T_{ann} and *Tg* are shown by horizontal dashes)

Results and discussion

As mentioned in Introduction, the general consensus is that the appearance of SHG in poled glasses is due to the fact that charged particles (cations) shift towards the cathode during polarization, forming a spatial electric charge and a 'frozen' electric field. It is this field that disturbs the isotropic structure of glass, producing second-order optical nonlinearity. However, we found a high peak in the TSDC spectrum of poled glass that was subsequently annealed at 400°C, i.e., at a temperature much higher than the annealing temperature. This peak is observed at 640°C (see Fig. 6). Recall that the appearance of peaks in the TSDC spectra of poled glasses is related precisely to relaxation of the spatial charge. At the same time, 50-minute annealing of poled glass at 400°C leads to complete degradation of the SHG signal (see Fig. 5). This indicates that SHG in poled glass is not directly related to the spatial charge, since, according to TSDC data, the spatial charge relaxed at temperatures above 530°C (glass transition temperature), while the annealing temperature, which provides complete SHG relaxation, was only 400°C. It can be assumed that another mechanism is responsible for SHG, related to orientation of dipole structures formed by cations of alkaline (or alkaline-earth) elements associated with non-bridging oxygen atoms.

Conclusion

Thus, the experiments showed that 50-minute annealing of the given poled glass at 400°C leads to disappearance of second harmonic generation in the sample, with the characteristic relaxation time of about 30 min. Measurement of thermally stimulated depolarization current through this sample confirms the independence of the processes of relaxation of the spatial charge caused by glass polarization and relaxation of second harmonic generation. These relaxation processes occur at significantly different temperatures and, accordingly, are characterized by different relaxation times.

This study was performed within the framework of the State Task 'Study of structures of micro- and nanophotonics formed in amorphous dielectrics under the action of strong fields' (Theme code 0791-2020-0001).

REFERENCES

1. **Poirier G., Dussauze M., Rodriguez V., et al.,** Second harmonic generation in sodium tantalum germanate glasses by thermal poling, The Journal of Physical Chemistry, C. 123 (43) (2019) 26528–26535.

2. **Lepicard A., Adamietz F., Rodriguez V., et al., et al.,** Demonstration of dimensional control and stabilization of second harmonic electro-optical response in chalcogenide glasses, Optical Materials Express. 8 (6) (2018) 1613–1624.

3. **Nasu H., Suzuki Y., Ohta H., et al.,** Second-harmonic generation from thermally poled mixed alkali silicate glasses containing various alkaline-earth oxides, Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers. 39 (12 A) (2000) 6530–6534.

4. **Le Calvez A., Freysz E., Ducasse A.,** A model for second harmonic generation in poled glasses, The European Physical Journal D: Atomic, Molecular and Optical Physics. 1 (2) (1998) 223–226.

5. **Lipovskii A., Morozova A., Tagantsev D.,** Giant discharge current in thermally poled silicate glasses, The Journal of Physical Chemistry C. 120 (40) (2016) 23129–23135.

6. **Menczel J.D., Prime R.B.** (Eds.), Thermal analysis of polymers: Fundamentals and applications, John Wiley & Sons, Inc., Hoboken (USA, New Jersey), 2009.

7. **Raskhodchikov D., Reshetov I., Brunkov P., et al.,** Mechanism of thermal charge relaxation in poled silicate glasses in a wide temperature range (From liquid nitrogen to glass melting temperature), The Journal of Physical Chemistry B. 124 (36) (2020) 7948–7956.

8. **Smith N., Regier T., Dutta I.,** Structure and composition of surface depletion layers in poled aluminosilicate glasses, Journal of the American Ceramic Society. 102 (6) (2019) 1–26.

9. **Karam L., Adamietz F., Rodriguez V., et al.,** The effect of the sodium content on the structure and the optical properties of thermally poled sodium and niobium borophosphate glasses, Journal of Applied Physics. 128 (4) (2020) 043106.

10. **Maker P.D., Terhune R.W., Nesenoff M., Savage C.M.,** Effects of dispersion and focusing on the production of optical harmonics, Physical Review Letters. 8 (1) (1962) 21–22.

Received 14.09.2020, accepted 20.09.2020.

THE AUTHORS

RESHETOV Ilya V.
Peter the Great St. Petersburg Polytechnic University
Alferov University
8/3 Khlopina, St. Petersburg, 194021, Russian Federation
reshetov_iv@spbstu.ru

KAASIK Vladimir P.
Peter the Great St. Petersburg Polytechnic University
Alferov University
8/3 Khlopina, St. Petersburg, 194021, Russian Federation
vkaasik@yandex.ru

LIPOVSKII Andrey A.
Peter the Great St. Petersburg Polytechnic University
Alferov University
8/3 Khlopina, St. Petersburg, 194021, Russian Federation
lipovskii@mail.ru

TAGANTSEV Dmitry K.
Peter the Great St. Petersburg Polytechnic University
Alferov University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
tagan@dt1386.spb.edu

ZHURIKHINA Valentina V.
Peter the Great St. Petersburg Polytechnic University
Alferov University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
jourikhina@mail.ru

DOI: 10.18721/JPM.13410

EFFECTIVE MODE VOLUME AND ESTIMATION OF HELIUM-NEON LASER OUTPUT POWER

V.A. Kozhevnikov, V.E. Privalov, A.E. Fotiadi

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

The relationship between the generated power and the mode volume of the fundamental mode of a plane-sphere resonator is considered in the article exemplified a He-Ne laser. A novel technique based on the volume of the generating mode has been proposed to estimate the output radiation power of a gas-discharge laser with an arbitrary cross-sectional shape of an active element. It was shown that the results of calculations of the generated power, carried out on the basis of the effective mode volume, which took into account the transverse distribution of the population inversion of the active medium, were in better agreement with the experimental data and the results of estimates using other methods.

Keywords: mode volume, effective mode volume, population inversion, He-Ne laser power

Citation: Kozhevnikov V.A., Privalov V.E., Fotiadi A.E., Effective mode volume and estimation of helium-neon laser output power, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 88–97. DOI: 10.18721/JPM.13410

Introduction

Due to their highest radiation coherence, helium-neon lasers have currently gained widespread application in metrology (time, frequency, and length standards), spectroscopy, industrial and research problems (laser interferometers); searching for methods for increasing the radiation power of such objects is therefore an extremely important problem. Studies [1, 2] indicate that changing the geometry of the discharge tube cross section can provide an increase in both the gain and the output power of the laser.

In this paper, we proposed and implemented a method for estimating the power of gas-discharge lasers based on the concept of effective modal volume of the generating mode; the proposed method takes into account the transverse distribution of inversion in an active element with an arbitrary cross section.

In this paper, we confine ourselves to considering the modal volume of the fundamental mode of a plane-sphere type resonator and the cylindrical geometry of the active element with a different transverse distribution of the population inversion. The concept of a confocal resonator is used for this purpose. However, preliminary analysis of applicability of our method for estimating the level of generated power showed good results for other cross sections as well.

Key theoretical foundations

Let us recall the main results obtained in the classical works [3–7]. For an optical resonator with the curvature radii R_1 and R_2 of mirrors and the distance d between the mirrors, the curvature radius of the corresponding equivalent confocal resonator R_e is found by the formula

$$R_e = \{4S\,(R_1 - S)\}^{1/2},$$

where $S = d\,(R_2 - d)/(R_1 + R_2 - 2d)$.

From here we obtain the following expression from for the plane-sphere resonator ($R_1 = \infty$, $R_2 = R$):

$$R_e = 2\{d\,(R - d)\}^{1/2},$$

where R is the curvature radius of the spherical mirror.

The electric field magnitude $|E_{00}|$ of the fundamental Gaussian mode TEM_{00} in the confocal resonator has the following form in cylindrical coordinates (r, z, φ):

$$|E_{00}| = E_0\sqrt{\frac{2}{1+\xi^2}}\times \exp\left[-\frac{kr^2}{R_e(1+\xi^2)}\right], \quad (1)$$

where $\xi = 2z/R_e$, $k = 2\pi/\lambda$, and the z coordinate is counted from the waist of the Gaussian beam (it is located near the flat mirror for a

plane-sphere resonator); E_0 is the $|E_{00}|$ value for $\xi = 1$ and $r = 0$.

The transverse size of the generated mode (the beam radius at the point w_z with the coordinate z) is determined by the value of the coordinate r to the side surface of the modal volume, where the electric field strength is less than that on the axis, in a given transverse section, by e times:

$$w_z^2 = w_0^2\left[1+\left(\frac{2z}{R_e}\right)^2\right],$$

$$\text{where } w_0^2 = \frac{R_e\lambda}{2\pi} = \frac{R_e}{k}.$$

Accordingly, the radius of the beam on the flat mirror w_1 is expressed as

$$w_0 = \sqrt{R_e\lambda/2\pi} = \{\lambda Re/(2\pi)\}^{1/2},$$

and it is greater by $\{R/(R-d)\}^{1/2}$ times on the spherical one w_2.

The modal volume (MV) of the resonator is assumed to be the energy characteristic of the fundamental mode [6]:

$$MV = \int_0^{2\pi}\int_0^{d}\int_0^{w_z} d\varphi dz r dr, \qquad (2)$$

where w_z is the radius of the beam;

$$w_z = \sqrt{\frac{1}{k}\left(R_e + \frac{4}{R_e}z^2\right)}.$$

It follows from Eqs. (1) and for w_z that the field intensities on the y axis are not identical for the plane and spherical mirrors, differing by $\{R/(R-d)\}^{1/2}$ times, and, accordingly the intensities on lateral boundaries of the modal volume differ as well for the mirrors.

The output power of the laser radiation depends on the level of interaction of the active medium in the modal volume with the field, i.e., both on the field distribution in the modal volume of the generating mode and on the distribution of the population inversion ΔN in the active medium in it. In the first approximation of the perturbation theory, the power of induced radiation can be considered proportional to the product $E^2\Delta N$.

One of the first studies focusing on the relationship between the radiation power and the field distribution in the generating mode was [8], introducing the effective modal volume, denoted as EMV and limited by lines of equal intensity, corresponding to the field intensity at the MV boundary located on a spherical mirror.

Using Eq. (1) and the condition $|E|^2 = E_0^2/e^2$, we can obtain the equation for the EMV boundary. This is a rotation figure with a section in the form of a circle of radius σ, depending on z:

$$\sigma^2 = \frac{w^2}{2}\left(2 + \ln 2 - \ln\frac{w^2 k}{R_e}\right).$$

Then the expression for EMV takes the following form:

$$EMV = \int_0^{2\pi} d\varphi \int_0^{d} dz \int_0^{\sigma} r dr =$$

$$= \left(1+\frac{\ln 2}{2}\right)MV -$$

$$-\frac{\pi R_e^2}{k}\left\{\frac{d}{2R_e}\left(1+\frac{4d^2}{3R_e^2}\right)\ln\left(1+\frac{4d^2}{R_e^2}\right)+\right. \qquad (3)$$

$$\left.+\frac{1}{3}\operatorname{arctg}\frac{2d}{R_e}-\frac{2d}{3R_e}-\frac{4d^3}{9R_e^3}\right\}.$$

Although it was assumed in [8] that the distribution of the population inversion in the active medium influences the size of the modal volume, this distribution was not taken into account in the actual definition of the EMV.

In view of the above, in this paper we propose a new quantity for estimating the output power of laser radiation by introducing a new effective modal volume accounting for both of the above-mentioned facts.

The concept of a new effective modal volume

We define the new effective modal volume (NMV) as a body bounded by a surface where the value of $|E|^2\Delta N$ decreases by e^2 times, compared to the value of $E_0^2\Delta N_0$ (ΔN_0 is the population inversion on the axis).

In other words, $E_0^2\Delta N_0$ is the value of $|E|^2\Delta N$ on the axis at the distance $R_e/2$ from the Gaussian beam waist, i.e., for the semi-confocal resonator, it is the value of $|E|^2\Delta N$ on the y axis of the spherical mirror.

Thus, for a cylindrical discharge tube with axial symmetry of the population inversion, the NMV is a rotation figure, which has a section in the form of a circle of radius ρ, depending on z:

$$NMV = \int_0^{2\pi} d\varphi \int_0^d dz \int_0^\rho r dr; \qquad (4)$$

and the radius ρ must be such that

$$\Delta N(\rho, z)\left|E(\rho, z)\right|^2 = \Delta N_0 E_0^2 / e^2.$$

Example estimation of the laser output power

Let us consider as an example an algorithm for estimating the radiation output power of a helium-neon laser. With an optimal ratio of mixture components and optimal discharge conditions [9], such a laser can be assumed to produce a diffusion discharge. The electron concentration n_e in such a discharge satisfies the following diffusion equation (homogeneous Helmholtz equation):

$$\Delta n_e + \frac{1}{D_a \tau} n_e = \Delta n_e + \frac{1}{\Lambda^2} n_e = \\ = \Delta n_e + \lambda^2 n_e = 0,$$

where D_a is the ambipolar diffusion coefficient, τ is the diffusion escape time.

The distribution of the population inversion ΔN in the helium-neon laser can be considered (in the first approximation) similar to the electron distribution in the active medium. In these media of cylindrical geometry, the radial distribution of the electron concentration and, hence, the inverse population ΔN of the helium-neon mixture, obtained as a solution of the diffusion equation, has the following form [10–14]:

$$\Delta N = \Delta N_0 J_0(\mu_1^{(0)} r / a),$$

where a is the tube radius; r is the distance to the axis; J_0 is a zero-order Bessel function; $\mu_1^{(0)}$ is the first root of the J_0 function, $\mu_1^{(0)} = 2.4048$.

In the first approximation and for the non-cylindrical geometry of the active element, the distribution of the population inversion ΔN in the given laser can be assumed to be similar to the electron distribution in a diffusion discharge. Consequently, it can be approximated that the distribution of the population inversion ΔN in the active medium for the laser also satisfies the homogeneous Helmholtz equation in this case:

$$\Delta(\Delta N) + \lambda^2 \Delta N = 0 \qquad (5)$$

with a homogeneous boundary condition

$$\Delta N\big|_\Gamma = 0, \qquad (6)$$

where Γ is the cross-sectional boundary of the active element.

In [15], we proposed a method for finding an approximate solution of equations (5), (6) for an arbitrary shape of the boundary Γ that has little computational complexity.

The essence of the proposed method for estimating the output power of laser radiation is that this power for a laser with arbitrary geometry of the active element cross section should also be estimated using the modal volume NMV by the following formula:

$$P = \iiint_{NMV} \varepsilon |E|^2 \, \Delta N dV, \qquad (7)$$

where ε is the corresponding proportionality coefficient.

In this case, the power is determined by following steps:

calculating the population inversion ΔN in the active medium by solving equations (5), (6) (for example, by the method proposed in [15]);

finding the NMV boundary;

direct integration by Eq. (7).

Let us test the capabilities of the proposed method for estimating the radiation power using the example of a laser with a cylindrical active element.

Given a cylindrical tube, it is easy to obtain, providing the explicit form of ΔN, that NMV of the fundamental mode of the confocal resonator is a rotation figure which has a section depending on z in the form of a circle of radius ρ, such that

$$\{2 + \ln 2 - \ln(w^2 k / R_e) - \\ -2r^2 / w^2 + \ln(J_0(\mu_1^0 r / a))\}\,|_{r=\rho} = 0. \qquad (8)$$

Equation (8) can be solved numerically and the corresponding function $\rho(z)$ can be found.

For further calculations we take the following parameters of the laser tube, typical for a helium-neon laser:

$a = 2$ mm, $d = 55$ cm, $\lambda = 0.6328$ μm,

R varies from 0.6 to 4.0 m.

Let us represent the Bessel function as

$$J_0(x) \approx 1 - x^2/4 + x^4/64.$$

The dependences of the functions $\rho(z)$, $\sigma(z)$, and $w(z)$ for $R = 1.1$ and 2.0 m are shown in Fig. 1. The dependences of NMV, EMV, and MV for different values of R are shown in Fig. 2, where the relative modal volume η, equal to the ratio of the modal volumes NMV, EMV, and MV to the total tube volume $V = \pi a^2 d$, is plotted along the ordinate axis. It can be seen that the volumes NMV and EMV generally (at $R > 0.8$ m) slightly exceed the value of MV, amounting to several percent of the total volume V.

As noted in [8], the formula can be used to estimate the laser radiation power using the modal volume MV, partially accounting for the population inversion

$$P_{MV} = \int_0^{2\pi} \int_0^{d} \int_0^{w} \varepsilon E^2 \Delta N d\varphi dz r dr.$$

Assuming that

$$\Delta N = \Delta N_0 J_0(\mu_1^{(0)} r / a)$$

and using the decomposition of the Bessel function in the form

$$J_0(x) \approx 1 - x^2/4 + x^4/64,$$

we obtain

$$P_{MV} = \frac{4\pi E_0^2 \Delta N_0 \varepsilon}{k} \left\{ \frac{1-e^{-2}}{2} d\sqrt{d(R-d)} - \right.$$
$$- \frac{\mu_1^{(0)2}}{24}(1-3e^{-2}) \times$$
$$\times \frac{d^2}{a^2 k}(3R-2d) + \qquad (9)$$
$$+ \frac{\mu_1^{(0)4}(1-5e^{-2})}{960} \times$$
$$\left. \times \frac{d^{5/2}(8d^2 - 20Rd + 15R^2)}{a^4 k^2 \sqrt{R-d}} \right\}.$$

Using the calculation method proposed in this paper and the concept of modal volume NMV, the expression for the laser radiation power is written as follows:

$$P_{NMV} = \int_0^{2\pi} \int_0^{d} \int_0^{\rho} \varepsilon E^2 \Delta N d\varphi dz r dr =$$
$$= \frac{4\pi E_0^2 \Delta N_0 \varepsilon}{k} 2\sqrt{d(R-d)} \times \qquad (10)$$
$$\times \int_0^{d} dz \left\{ \frac{1 - \exp(-2\rho^2 / w^2)}{4} - \right.$$

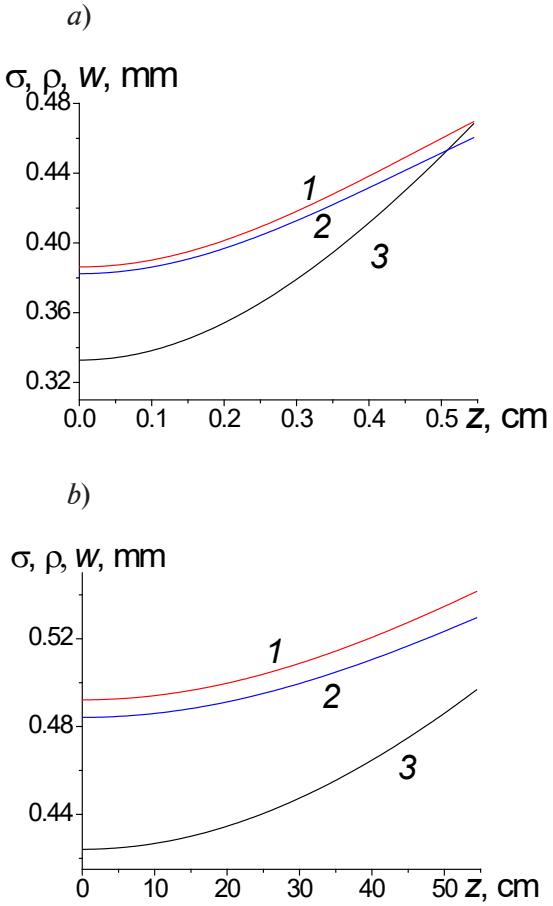


Fig. 1. Dependences for radii σ of the circular cross section of discharge tube (curve *1*), EMV boundary ρ (*2*) and the radiation beam w (*3*) of the helium-neon laser on the coordinate z for values of mirror curvature R, m: 1.1 (*a*) and 2.0 (*b*); laser tube parameters: $a = 2$ mm, $d = 55$ cm, $\lambda = 0.6328$ μm

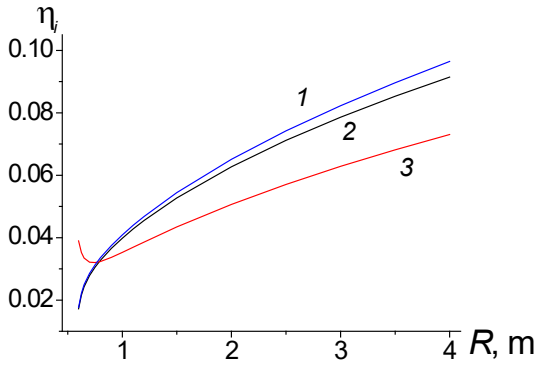


Fig. 2. Dependences of relative modal volumes η_i = EMV/V (*1*), NMV/V (*2*), MV/V (*3*) on the mirror curvature R. Laser tube parameters: $a = 2$ mm, $d = 55$ cm, $\lambda = 0.6328$ μm

$$-\frac{\mu_1^{(0)2}}{32a^2}(w^2-(w^2+2\rho^2)\times$$
$$\times\exp(-2\rho^2/w^2))+$$
$$+\frac{\mu_1^{(0)4}}{512a^4}(w^4-(w^4+2w^2\rho^2+2\rho^4)\times$$
$$\exp(-2\rho^2/w^2))\Big\}.$$

Comparison of the proposed method with traditional methods

It is reasonable to compare this value with an estimate of the laser output power obtained using the modal volume EMV and the total tube volume V:

$$P_{EMV}=\int_0^{2\pi}\int_0^{d}\int_0^{\sigma}\varepsilon E^2\Delta N d\varphi dz r dr=$$
$$=\frac{4\pi E_0^2\Delta N_0\varepsilon}{k}2\sqrt{d(R-d)}\times$$
$$\times\int_0^{d}dz\left(\frac{1-\exp(-2\sigma^2/w^2)}{4}-\right.$$
$$-\frac{\mu_1^{(0)2}}{32a^2}(w^2-(w^2+2\sigma^2)\times$$
$$\times\exp(-2\sigma^2/w^2))+$$
$$+\frac{\mu_1^{(0)4}}{512a^4}(w^4-(w^4+2w^2\sigma^2+2\sigma^4)\times$$
$$\left.\times\exp(-2\sigma^2/w^2))\right);\qquad(11)$$

$$P_V=\int_0^{2\pi}\int_0^{d}\int_0^{a}\varepsilon E^2\Delta N d\varphi dz r dr=$$
$$=\frac{4\pi E_0^2\Delta N_0\varepsilon}{k}2\sqrt{d(R-d)}\times$$
$$\times\int_0^{d}dz\left(\frac{1-\exp(-2a^2/w^2)}{4}-\right.$$
$$-\frac{\mu_1^{(0)2}}{32a^2}(w^2-(w^2+2a^2)\times$$
$$\times\exp(-2a^2/w^2))+$$
$$+\frac{\mu_1^{(0)4}}{512a^4}(w^4-(w^4+2w^2a^2+2a^4)\times$$
$$\left.\exp(-2a^2/w^2))\right).\qquad(12)$$

The results of the calculations of the given values are shown in Fig. 3.

Evidently, the estimates of the laser power obtained using NMV and EMV for the distribution of the population inversion given by the Bessel function are approximately the same. Fig. 3,*b* shows a segment of the P_{EMV}/α and P_{NMV}/α dependences on R on a larger scale in the vicinity of the R values corresponding to the semi-confocal resonator.

One of the reasons for introducing the effective modal volume in [8] was the discrepancy between the experimental data and the calculated power estimate obtained using MV by Eq. (9); it turned out that the experimental values exceeded the calculated ones by about 10%. The results in Fig. 3 indicate that the NMV and EMV estimates of the laser power give good agreement with experiment.

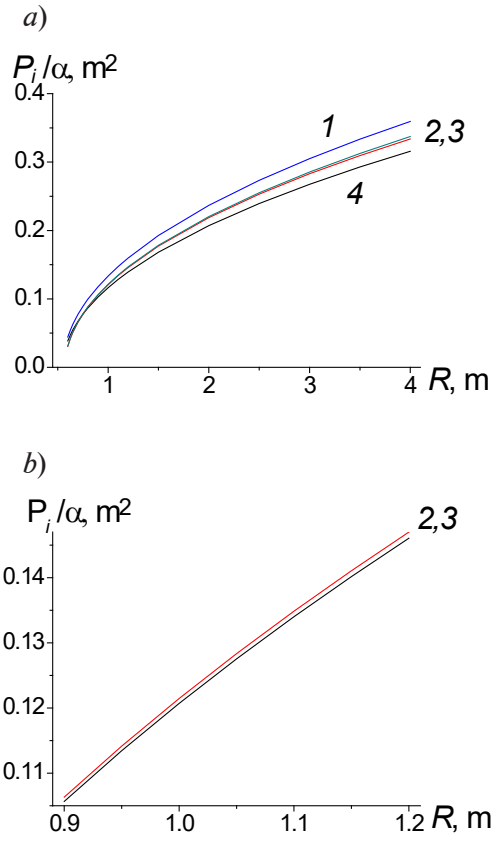


Fig. 3. Comparison of normalized output powers P_V/α (*1*), P_{EMV}/α (*2*), P_{NMV}/α (*3*) and P_{MV}/α (*4*) for a HeNe laser depending on the mirror curvature R calculated by Eqs. (9)–(12) (*a*); data calculated by Eqs. (10), (11) are shown separately (*b*): the graph is enlarged for curves (*2*) and (*3*) to be distinguishable;

$$\alpha=4\pi E_0^2\Delta N_0\varepsilon/k$$

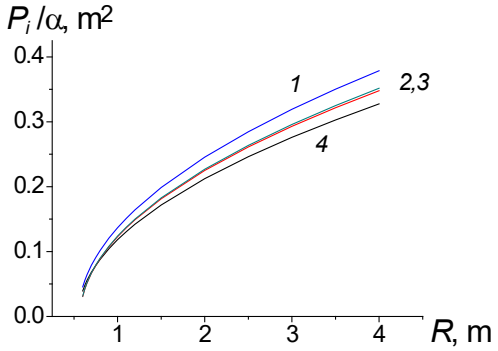


Fig. 4. Comparison of dependences similar to those shown in Fig. 3 but calculated under the condition $\mu_1^{(0)} = 0$

Influence of various factors on the laser power

A correlation between the radial distribution of the population inversion ΔN and the laser output power was observed in the experiments described in [16–19]: a change in ΔN can change the power by the order of 10%.

Radial inversion distribution. Let us assess the effect of the radial inversion distribution with the proposed method, calculating the powers without taking into account the given distribution. To obtain such values, we formally take $\mu_1^{(0)} = 0$ in Eqs. (9)–(12). The results of such calculations are shown in Fig. 4. Apparently, both with and without taking into account the radial distribution of the population inversion, the power estimates using the NMV and EMV values give slightly larger differences than those using MV, which agrees better with the experiments. The result for which the maximum difference between the experimental data and the calculated ones was somewhat larger is explained by the method chosen for assessment: specifically, if $\mu_1^{(0)} = 0$, ΔN remains constant along the cross section; on the other hand, in the experiment, a local minimum of ΔN [13] can also be obtained on the axis, due to a certain ratio of partial pressures of helium and neon. Clearly, the presence of a minimum will give a greater difference in the powers, as compared to the optimal ratio when ΔN is described by a Bessel function.

On the other hand, the output radiation power of a helium-neon laser with a cylindrical tube can be estimated using the well-known technique [20, 21], which gives excellent agreement with the experiment. According to this technique, the radiation power is described by the following formula:

$$P = Aw_0 G_m \left[1 - \left(a_s / G_m\right)^{1/2}\right]^2 \times \times \pi w_1^2 \left[1 - \exp(-2r_0^2 / w_1^2)\right], \quad (13)$$

where Aw_0 is the saturation coefficient, $Aw_0 = 30$ W/cm² [20, 21]; G_m is the total unsaturated gain in the center of the Doppler-broadened amplifying loop for neon atoms, $G_m = 3 \cdot 10^{-4} l/(2r_0)$ [20– 22]; r_0, l are the radius of the capillary and length of active part of the capillary; a_s is the total loss coefficient; w_1 is the radius of the beam on the output mirror, $w_1 = \{\lambda R_e/(2\pi)\}^{1/2}$.

Let us describe the contributions to the total coefficient a_s using the LGN-222 laser as an example [23]:

a_1 are the absorption losses in each Brewster window of parallel active elements;

a_2 are the reflection losses from the resonator mirrors,

a_3 are the diffraction losses,

a_4 are other unaccounted losses.

These contributions to the total coefficient a_s for the LGN-222 laser ($r_0 = 1.5$ mm, $l = 2$ m) can be estimated using the values given in [23]:

$$a_1 = 3 \cdot 10^{-3},\ a_2 = 6 \cdot 10^{-3},\ a_4 = 2 \cdot 10^{-3}.$$

To estimate the diffraction losses a_3, it is reasonable to use an asymptotic representation of generalized radial functions [24, 25] or an empirical correction [26]:

$$\alpha_{pm} = \frac{2\pi(4c - 4.52)^{2p+m+1}}{p!(p+m)!} e^{-2c},$$

where p, m are the mode parameters ($p = m = 0$ for the main Gaussian mode), $c = 2\pi N$ (N is the Fresnel parameter).

The Fresnel parameter for an equivalent confocal resonator follows the expression

$$N_i = a_i^2 \{g_i(1 - g_i g_k) / g_k\}^{1/2} / (\lambda d),$$

where a_i is the aperture of the resonator; i, k are the indices of the resonator halves, i, $k = 1, 2$; g_i is the configuration parameter of the resonator, $g_i = 1 - d/R_i$.

A comparison of the powers calculated by Eqs. (10) and (13) is shown in Fig. 5. For clarity, the values of P_{NMV}/α and P are normalized so that they coincide for the semi-confocal resonator ($d = 2$ m, $R = 4$ m) (an appropriate normalization multiplier γ is introduced). It can be seen that estimating the laser power using NMV gives a very good agreement with the calculation by Eq. (13).

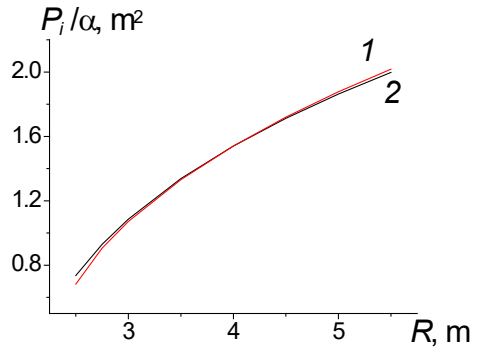


Fig. 5. Comparison of the powers calculated by the traditional method, P/γ (*1*), with that obtained by the proposed method, P_{NMV}/α (*2*) depending on the mirror curvature R (see Eqs. (10) and (13)) for the HeNe laser; laser tube parameters: a = 1.5 mm, d = 2 m, λ = 0.6328 μm

The case of high laser power. A very different kind of dependence should be obtained in the case of high laser power, when nonlinear effects are possible. In this case, it is proposed to estimate the laser power P^{nl} by the following formula [27]:

$$P^{nl} = P_0 \frac{\int\limits_V E^2 \Delta N dV}{\int\limits_V E^4 \Delta N dV}, \qquad (14)$$

where P_0 is a constant.

Let us find the expressions for the cases of high power using MV, NMV, EMV and the total volume V.

Accordingly, we obtain:

$$P^{nl}_{MV} = P_0 \frac{\int\limits_{MV} E^2 \Delta N dV}{\int\limits_{MV} E^4 \Delta N dV} = P_0 \frac{P_{MV}}{F_{1MV}}, \qquad (15)$$

where P_{MV} is described by Eq. (7), and the denominator F_{1MV} is calculated by the formula

$$F_{1MV} = \left\{ \frac{4\pi E_0^2 \Delta N_0}{k} \frac{8E_0^2 d(R-d)}{k} \times \right.$$
$$\times \left(k \frac{1-e^{-4}}{16} \operatorname{arctg} \frac{d}{\sqrt{d(R-d)}} - \right.$$
$$- \frac{\mu_1^{(0)2}}{128a^2}(1-5e^{-4})d +$$
$$+ \frac{\mu_1^{(0)4}(1-13e^{-4})}{6144a^4 k} \times$$
$$\left. \left. \times \frac{d^2(3R-2d)}{\sqrt{d(R-d)}} \right) \right\}.$$

Next,

$$P^{nl}_{NMV} = P_0 \frac{\int\limits_{NMV} E^2 \Delta N dV}{\int\limits_{NMV} E^4 \Delta N dV} = P_0 \frac{P_{NMV}}{F_{1NMV}}, \qquad (16)$$

where P_{NMV} is described by Eq. (8), and the denominator F_{1NMV} is calculated by the formula

$$F_{1NMV} = \frac{4\pi E_0^2 \Delta N_0}{k} \frac{8E_0^2 d(R-d)}{k} \times$$
$$\times \int_0^d dz \left(\frac{1-\exp(-4\rho^2/w^2)}{8w^2} - \right.$$
$$- \frac{\mu_1^{(0)2}}{128a^2 w^2}(w^2 - (w^2+4\rho^2) -$$
$$\times \exp(-4\rho^2/w^2)) +$$
$$+ \frac{\mu_1^{(0)4}}{4096a^4 w^2}(w^4 - (w^4 + 4w^2\rho^2 + 8\rho^4) \times$$
$$\left. \times \exp(-4\rho^2/w^2)) \right).$$

Similarly,

$$P^{nl}_{EMV} = P_0 \frac{\int\limits_{EMV} E^2 \Delta N dV}{\int\limits_{EMV} E^4 \Delta N dV} = P_0 \frac{P_{EMV}}{F_{1EMV}}, \qquad (17)$$

where P_{EMV} is described by Eq. (9), and the denominator F_{1EMV} is calculated by the formula

$$F_{1EMV} = \frac{4\pi E_0^2 \Delta N_0}{k} \frac{8E_0^2 d(R-d)}{k} \times$$
$$\times \int_0^d dz \left(\frac{1-\exp(-4\sigma^2/w^2)}{8w^2} - \right.$$
$$- \frac{\mu_1^{(0)2}}{128a^2 w^2}(w^2 - (w^2+4\sigma^2) \times$$
$$\times \exp(-4\sigma^2/w^2)) +$$
$$+ \frac{\mu_1^{(0)4}}{4096a^4 w^2}(w^4 - (w^4 + 4w^2\sigma^2 + 8\sigma^4) \times$$
$$\left. \times \exp(-4\sigma^2/w^2)) \right).$$

And finally,

$$P_V^{nl} = \frac{P_0 \int\limits_V E^2 \Delta N dV}{\int\limits_V E^4 \Delta N dV} = \frac{P_0 P_V}{F_{1V}}, \qquad (18)$$

where P_V is described by formula (10), and the denominator F_{1V} is calculated by the formula

$$F_{1V} = \frac{4\pi E_0^2 \Delta N_0}{k} \frac{8E_0^2 d(R-d)}{k} \times$$
$$\times \int\limits_0^d dz \left(\frac{1-\exp(-4a^2/w^2)}{8w^2} - \right.$$
$$- \frac{\mu_1^{(0)2}}{128a^2w^2}(w^2 - (w^2+4a^2) \times$$
$$\times \exp(-4a^2/w^2)) +$$
$$+ \frac{\mu_1^{(0)4}}{4096a^4w^2}(w^4 - (w^4 + 4w^2a^2 + 8a^4) \times$$
$$\left. \times \exp(-4a^2/w^2)) \right).$$

The results of calculations by Eqs. (15)–(18) are shown in Fig. 6 ($\beta = P_0/E_0^2$).

The ratios between P^{nl}_{MV}/β, P^{nl}_{NMV}/β, P^{nl}_{V}/β qualitatively correspond approximately to those between P_{MV}, P_{NMV} and P_V, although the dependences on R behave differently. Similar to the estimates made above (see Fig. 4), it is interesting to assess the effect of the radial distribution of the inversion by analyzing the powers without taking into account the radial distribution of the population inversion (taking $\mu_1^{(0)} = 0$). The results of the corresponding calculations are shown in Fig. 7. It can be seen that the obtained relations are qualitatively approximately the same.

Analysis of the obtained results leads us to the conclusion that using the modal volume NMV in calculations of the radiation power of the helium-neon laser yields to a better agreement with the experimental data compared to the calculated results obtained using the standard modal volume MV. The values obtained from the NMV and EMV volumes, taking into account the population inversion distribution based on a given Bessel function (i.e., for a helium-neon laser with a cylindrical discharge tube) are approximately the same but can differ markedly for other discharge tube geometries. The difference is due to the fact that EMV is always a rotation figure (which follows from the EMV definition and Eq. (1)), while the population inversion has no axial symmetry at any discharge tube geometry, and, accordingly, the modal volume NMV also lacks this symmetry.

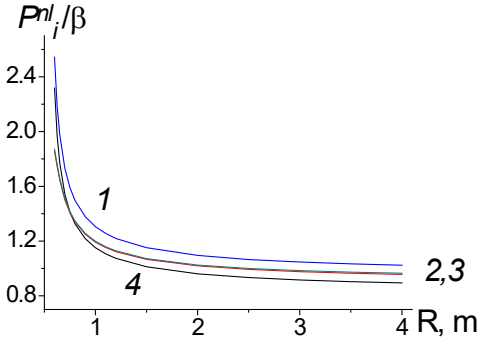


Fig. 6. Comparison of dependences for normalized high power of the helium-neon laser, calculated by different methods, on the mirror curvature R (see Eqs. (15)–(18)): P^{nl}_V/β (*1*), P^{nl}_{EMV}/β (*2*), P^{nl}_{NMV}/β (*3*), P^{nl}_{MV}/β (*4*); $\beta = P_0/E_0^2$
Laser tube parameters: a = 2 mm, d = 55 cm, λ = 0.6328 μm

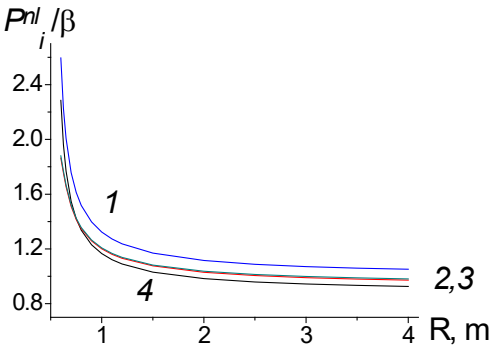


Fig. 7. Comparison of dependences similar to those shown in Fig. 6 but calculated under the condition $\mu_1^{(0)} = 0$

Conclusion

Thus, we have proposed a method for estimating the radiation power of a HeNe laser with an arbitrary cross-sectional shape of the active element, which has been tested for the case of cylindrical geometry and gives good agreement with the experimental data.

Various cross sections of the HeNe laser active element are currently investigated as the search for the optimal one in terms of power output continues. Preliminary calculations of the HeNe laser power using NMV for discharge

tubes with rectangular and elliptical cross sections have shown that the results obtained are in good agreement both with experimental data and with calculated results on the gain of lasers possessing this geometry [15]. The modal volume NMV for these cross sections (in the form of a rectangle and an ellipse) should have an elliptical cross section. At the same time, obtaining formulas similar to (13) for these cross sections causes great difficulties; if the modal volume NMV proposed by us is used, the values of the output power of the helium-neon laser can be estimated by a computationally uncomplicated method.

REFERENCES

1. **Privalov V.E.,** Gas-discharge geometry and studies in laser emission, Russian Physics Journal. 53 (5) (2010) 80–90.

2. **Privalov V.E.,** Some prospects for the development of gas-discharge lasers, Russian Physics Journal. 56 (2-2) (2013) 246–253.

3. **Boyd G.D., Gordon J.P.,** Confocal multimode resonator for millimeter through optical wavelength masers, The Bell Systems Technical Journal. 40 (2) (1960) 489–508.

4. **Boyd G.D., Kogelnik H.,** Generalized confocal resonator theory, The Bell Systems Technical Journal. 41 (4) (1962) 1347–1369.

5. **Gordon J.P., Kogelnik H.,** Equivalence relations among spherical mirror optical resonators, The Bell Systems Technical Journal. 43 (6) (1964) 2873–2886.

6. **Sinclair D.C.,** Choice of mirror curvatures for gas laser cavities, Applied Optics. 3 (9) (1964) 1067–1072.

7. **Kogelnik H., Li T.,** Laser beams and resonators, Applied Optics. 5 (10) (1966) 1550–1567.

8. **Privalov V.E.,** Mode volume and radiation power of a laser, Optics and Spectroscopy. 28 (3) (1970) 524–527.

9. **Field R.L., Jr.,** Operating parameters of dc-excited He-Ne gas lasers, Review of Scientific Instruments. 3 (12) (1967) 1720–1722.

10. **Tako T.,** Self-absorption of spectral line, Journal of the Physical Society of Japan. 15 (10) (1961) 2016–2032.

11. **Bennett W.R., Jr.,** Excitation and inversion mechanisms in gas lasers, Annals New York Academy of Science. 122 (2) (1965) 579–595.

12. **Privalov V.E., Fridrikhov S.A.,** Zavisimost moshchnosti izlucheniya He-Ne lazera ot geometrii secheniya razryadnogo promezhutka [The radiation power of He-Ne laser as a function of a cross-section geometry of a discharge gap], Technical Physics. 38 (12) (1968) 2080–2084 (in Russian).

13. **Privalov V.E., Fridrikhov S.A.,** The ring gas laser, Soviet Physics, Uspekhi. 12 (3) (1969) 153–167.

14. **Herziger G., Holzapfel W., Seelig W.,** Verstärkung einer He-Ne-Gasentladung für die Laserwellenlänge λ = 6328 AE, Zeitschrift für Physik. 189 (4) (1966) 385–400.

15. **Kozhevnikov V.A., Privalov V.E.**, The geometrical effect of an active element cross-section on the laser gain, St. Petersburg State Polytechnical University Journal. Physics and Mathematics. 11 (2) (2018) 84–95.

16. **Golubev Yu.M., Privalov V.E.**, Nekotoryye kharakteristiki He-Ne lazera, generiruyushchego odnovremenno 3.39 i 0.6328 mkm [Some characteristics of a He-Ne laser generating 3.39 and 0.6328 μm simultaneously], Optics and Spectroscopy. 22 (3) (1967) 499–501 (in Russian).

17. **Privalov V.E., Khodovoy V.A.,** Issledovaniye svoystv He-Ne-lazera s malym razryadnym promezhutkom [An investigation of properties of a He-Ne laser with a small discharge gap], Optics and Spectroscopy. 25 (2) (1968) 318–319 (in Russian).

18. **Golubev Yu.M., Privalov V.E., Fridrikhov S.A., Khodovoy V.A.,** O svyazi optimalnogo sootnosheniya komponentov smesi v He-Ne lazere s raspredeleniyem polya v rezonatore [On the relation between an optimal ratio of mixture and a resonator's field distribution in the He-Ne laser], Technical Physics. 38 (6) (1968) 1997–1100 (in Russian).

19. **Golubev Yu.M., Privalov V.E., Fridrikhov S.A., Khodovoy V.A.,** Ob optimalnom sootnoshenii komponent smesi v koltsevom He-Ne lazere [On the optimal ratio of mixture in the ring He-Ne laser], Technical Physics. 38 (11) (1968) 1990–1993 (in Russian).

20. **Smith P.W.,** The output power of a 6328-E He-Ne gas laser, IEEE Journal of Quantum Electronics. 2 (3) (1966) 62–68.

21. **Petrь F., Veselõ Z.,** The output power of a 633 nm He-Ne lasers, Optical and Quantum Electronics. 4 (1) (1972) 1–20.

22. **Li Y., Chen M., Li Z., et al.,** Study of performance of a He-Ne laser having an annular gain zone, Applied Optics. 46 (4) (2007) 591–601.

23. **Lipskii V.V., Privalov V.E.,** Calculating parameters of high-power He-Ne lasers operating at λ = 0.63 μm, Technical Physics Letters. 31 (7) (2005) 608–610.

24. **Slepian D., Pollak H.O.,** Prolate spheroidal wave functions. Fourier analysis and uncertainty. I., The Bell Systems Technical Journal, 40 (1) (1961) 65–84.

25. **Slepian D.,** Prolate spheroidal wave functions. Fourier analysis and uncertainty. IV: Extensions to many dimensions; generalized prolate spheroidal functions, The Bell Systems Technical Journal. 43 (6) (1964) 3009–3057.

26. **Ishchenko E.F.,** Otkrytyye opticheskiye rezonatory: Nekotoryye voprosy teorii i rascheta [Open optical resonators: some issues of theory and calculations], Sovetskoye Radio, Moscow, 1980 (in Russian).

27. **Zeiger S.G., Fradkin E.E.,** Vzaimodeistviye poperechnykh mod [Transverse mode coupling], In "Fizuka gazovykh lazerov" ["The physics of gas lazers", the collected book], Leningrad State University Publishing, Leningrad, (1969) 55–59 (in Russian).

Received 05.10.2020, accepted 20.10.2020.

THE AUTHORS

KOZHEVNIKOV Vadim A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
vadim.kozhevnikov@gmail.com

PRIVALOV Vadim E.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
vaevpriv@yandex.ru

FOTIADI Alexandr E.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
fotiadi@rphf.spbstu.ru

DOI: 10.18721/JPM.13411

CURVED-RAY TENSOR TOMOGRAPHY FOR RESIDUAL STRESS MEASUREMENTS IN THE AXISYMMETRIC GRADED RODS

D.D. Karov[1], *A.E. Puro*[2]

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

[2] Euroacademy, Tallinn, Estonia

For the first time, an algorithm for reconstructing an arbitrary distribution of residual stresses by the polarization tomography method for cylindrical rod structures with a radial distribution of the refractive index has been presented. The reconstruction took into account the ray refraction. The algorithm is based on the expansion of the tensor stress field in angular harmonics (singular value expansion). The case of an axisymmetric tensor field with an arbitrary stress gradient along the cylinder axis was considered. Numerical calculations were carried out for an axially symmetric stress distribution in a gradan for the case of a plane deformation state. The reconstruction was based on the expansion of the stress tensor in eigenfunctions of the boundary value problem. The regularized solution of the resolving equation (of Abelian type) used the expansion in Zernike polynomials. The results of the reconstruction are given with taking into account the additional term due to the deflection of the transmission rays as well as without this doing.

Keywords: residual stress, integrated photoelasticity, transverse translucence, GRIN rod lense

Citation: Karov D.D., Puro A.E., Curved-ray tensor tomography for residual stress measurements in the axisymmetric graded rods, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 98–109. DOI: 10.18721/JPM.13411

Introduction

Residual stress (RS) is one of the key characteristics describing the strength and optical quality of glass and crystal articles. Gradient refractive index (GRIN) structures, like rod lenses, laser rods, ball lenses, fiber preforms, and fibers have numerous applications. Most of them have an axially symmetric distribution of the refractive index that is the highest along its optical axis and decreases toward the periphery. A remarkably diverge range of optical profiling techniques has been developed during the last decades [1]; we assume in the discussion below that the refractive index (RI) n is known. Integrated photoelasticity [2] is a nondestructive method for stress analysis in 3D transparent specimens. This method consists in placing a three-dimensional sample in an immersion bath ($n_{imm} = n_{surf}$; n_{imm}, n_{surf} are the RI of the immersion and product surface, respectively) and passing a beam of polarized light through the sample cross-section. Reconstruction of the spatial distribution of the stress tensor by interpreting the integrated optical effects of the rays that have passed through the medium may be considered a type of tensor field tomography [3]. It is based on solving the problem of optical tomography of the stress tensor field in combination with the resulting problem of elasticity theory. The problem of light propagation is separated into two parts [4] in the case of weak stresses (fiber preforms and fibers, GRIN lenses, laser rods): determining the ray paths in the GRIN structure and determining the change in the polarization of the light passed through a birefringent medium. As a rule, induced birefringence is rather weak in GRIN structures, and it is possible to measure two linear integrals on each ray. One of them is connected with the

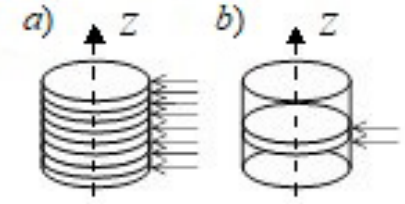


Fig. 1. Schematic drawing of the two raying methods corresponding to the global (*a*) and local (*b*) inverse problems

transversal interaction of the 2D vector field and the other with the transversal interaction of the 2D tensor field. From a mathematical standpoint, we have the special case of tensor tomography with beam deflection [5, 6]. The reconstruction algorithm is based on circular harmonic decomposition (Cormac-type inversion) [7] of ray integrals. The inverse problem of reconstructing the RSs is actually divided into the following two successive stages [3] for rectilinear propagation of rays:

(*i*) reducing the tensor ray integrals to scalar ones,

(*ii*) determining all the stress-tensor components based on inverting the ray integrals and solving the corresponding thermoelasticity problem [8].

In contrast to polarimetric tomography of straight light rays, the tomographical problem and the inverse thermoelastic problem must be solved together in GRIN media [9].

Below we will investigate two types of inverse problems.

The global inverse problem: tomographic measurements are performed in a system of parallel planes over the entire height of the sample (optical axis) (Fig. 1, *a*). The local inverse problem: tomographic measurements are carried out in two closely spaced sections orthogonal to the optical axis (Fig. 1, *b*).

The paper is structured as follows. The basics of tensor field tomography are introduced in the next section. In what follows, we give the algorithm of RS reconstruction. Appendix after the main material of the paper gives the singular decomposition of quasi-plane stresses in a circular cylinder.

Preliminaries

To make the presentation of our paper logical, we will give here the main concepts of tensor tomography of RS: the scalar and 2D-vector field tomography in axially symmetric GRIN media and the main notions of the inverse thermoelasticity problem of integrated photoelasticity.

Ray equation in axially symmetric optical media. We will rely on the results obtained in ray theory of axially symmetric optical media [10]. We use the Cartesian (*x, y, z*) coordinate system, as well as the cylindrical one (*r,* φ*, z*) and the moving one (*s,* α*, z*) associated with the measuring process (Fig. 2).

The axis *z* coincides with the axes of the cylinder. The cylinder is taken to have a unit radius. The ray is localized by its normal vector **s**, such that $s = |\mathbf{s}|$ is the smallest distance from the ray to the origin, and θ is the angle between the **s** and the axis *x* (Fig. 2). An arbitrary point on the ray is determined by its polar angle $\varphi = \theta \pm \alpha$ and by its distances *r* from the origin.

The RI $n(r)$ has rotational symmetry around the *z*-axis and $R(r) = n(r)r$ is a monotonic continuous function of *r* so that there is a one-to-one correspondence between *R* and *r* and only one ray passes through any two points inside the circle. We introduce another moving coordinate systems, (*R*, Ω, *z*) and (*S, T, z*), associated with the RI:

$$S(s) = sn(s), T = \pm\sqrt{R^2 - S^2},$$

$$\sin(\Omega) = \frac{T}{R},$$

where the plus sign holds for the left-hand side and minus for right-hand side of the ray.

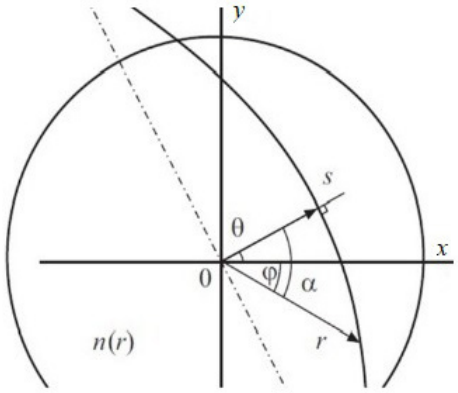


Fig. 2. The Cartesian (*x*, *y*, *z*), the cylindrical (*r*, φ, *z*) and the moving coordinate (*s*, α, *z*) systems used in the paper; the *z*-axis coincides with the axes of the cylinder; $\varphi = \theta \pm \alpha$ is the polar angle, $n(r)$ is a radial distribution of the refractive index

Then, the ray equation can be written in the next form:

$$\begin{aligned}\alpha\left[(R(r),S(s)\right] &= \\ &= \pm\int_{|S|}^{r}\frac{\mathrm{S}(s)d\rho}{\rho\sqrt{\left[R(\rho)\right]^2-\left[S(s)\right]^2}} = \\ &= \pm\int_{|S|}^{R}\frac{SD(K)dK}{K\sqrt{K^2-S^2}} = A(R,S), \\ D(R) &= \frac{d\ln(r(R))}{d\ln(R)},\end{aligned} \tag{1}$$

where $\alpha(r,s) = \pm(\varphi(r) - \theta)$.

We give here also the formula for the differential of length l of an arc on the ray:

$$\begin{aligned}dl &= \pm\sqrt{1+\left[r\frac{d\alpha}{d\mathrm{r}}\right]^2}dr = \\ &= \pm\frac{R(r)dr}{\sqrt{R^2(r)-S^2(s)}} = \\ &= \pm\frac{L(R)RdR}{\sqrt{R^2-S^2}} = L(R)dT, L(R) = \frac{dr(R)}{dR}.\end{aligned} \tag{2}$$

We assume that the RI is known and we can determine the ray equation.

Attenuated Radon transform in axially symmetric optical media. The Radon transform of $\hat{f}(\mathrm{s},\theta,\mu)$ with attenuation is defined here as follows [7]:

$$\begin{aligned}\hat{f}(s,\theta,\mu) &= \int_{-l(s)}^{l(s)} f(s\boldsymbol{\theta}+\mathbf{x})\times \\ &\times\exp\left(\int_0^l \mu(s\boldsymbol{\theta}+\mathbf{x})dk\right)dl,\end{aligned}$$

where $l(s)$ is half the length of the ray in the circle, $\theta = (\cos\theta, \sin\theta)$ is the unit vector along the s-axis, μ is the attenuation.

Let us transform the attenuation accumulated along the ray

$$\begin{aligned}M(R,S) &= \int_0^l ¼(s\vec{\theta}+\vec{x})dk = \\ &= \int_{|S|}^{r}\frac{\mu(k)R(k)dk}{\sqrt{\left[R(k)\right]^2-\left[S(s)\right]^2}} = \\ &= \int_{|S|}^{R}\frac{\mu(K)L(K)Kdk}{\sqrt{K^2-S^2}} = M^0(T,S).\end{aligned}$$

We can transform the ray integral by using polar coordinates (R, φ)

$$\begin{aligned}\hat{f}(S,\theta) &= \int_S^{R_0} f\left[R,\theta+A(R,S)\right]\times \\ &\times\frac{\exp\left[M(R,S)\right]L(R)RdR}{\sqrt{R^2-S^2}} + \\ &+\int_S^{R_0} f\left[R,\theta-A(R,S)\right]\times \\ &\times\frac{\exp\left[-M(R,S)\right]L(R)RdR}{\sqrt{R^2-S^2}}.\end{aligned} \tag{3}$$

Applying the Radon transform for tomography of stress tensor fields in optical glasses is connected with the circular harmonic decomposition (CHD) of stress functions [8, 9]. Expanding $f(r,\varphi)$ and $\hat{f}(S,\theta)$ in a Fourier series with respect to the corresponding angular variables, we obtain

$$\begin{aligned}f(r,\varphi) &= \sum_{m=-\infty}^{\infty} f(r)e^{im\varphi}, \ldots\hat{f}(S,\theta) = \\ &= \sum_{m=-\infty}^{\infty} g_m(S)e^{im\theta},\end{aligned}$$

where

$$\begin{aligned}f_m(r) &= \frac{1}{2\pi}\int_0^{\text{ð}} f(r,\varphi)e^{-im\varphi}d\varphi, \\ g_m(S) &= \frac{1}{2\pi}\int_0^{2\pi}\hat{f}(S,\theta)e^{-im\theta}d\theta.\end{aligned}$$

The absorption is imaginary in the problems of magneto-photoelasticity [11], in the studies of the magnetic field of the Tokamak plasma [12, 13], and in the magnetic resonance imaging [14]. Thus, we give the reconstruction algorithm for the case of imaginary attenuation, $\mu = i\eta$,

$$M(R,S) = \int_{|S|}^{R}\frac{i\eta(K)L(K)KdK}{\sqrt{K^2-S^2}} = iQ(R,S).$$

Inserting relationship (4) into a circular harmonic, we obtain

$$\begin{aligned}g_m(S,\eta) &= \\ &= 2\int_{|S|}^{R_0}\frac{\cos\left[E(m,K,S)\right]F_m(K)KdK}{\sqrt{K^2-S^2}}, \\ E(m,K,S) &= mA(K,S)+Q(K,S),\end{aligned} \tag{4}$$

an integral equation (generalized Cormack equation) for

$$F_m(R) = f_m(R)L(R).$$

The solution of Eq. (4) can be written as follows [7]:

$$f_m(r)n(r) =$$
$$= \frac{-1}{2\pi r}\left[\frac{\partial}{\partial r}\mathrm{Cr1}(qS,R,\eta) + \frac{m}{r}\mathrm{Sr1}(q,R,\eta)\right],$$
$$\mathrm{Cr1}(gS,R,\eta) =$$
$$= \mathrm{Cr}(gS,R,\eta) + \mathrm{Cr}(gS,R,-\eta),$$
$$\mathrm{Cr}(qS,R,\eta) =$$
$$= \int_R^{R0} \frac{\mathrm{ch}\left[G(m,S,R,\eta)\right]}{\sqrt{S^2-R^2}} q_m(S,\eta)SdS,$$
$$G(m,S,R,\eta) =$$
$$= mB(S,R) + N(s,R), B(S,R) = \quad (5)$$
$$= \int_R^{|S|} \frac{SD(K)d\mathrm{K}}{K\sqrt{S^2-K^2}},$$
$$N(S,R) = \int_R^{|S|} \frac{\eta(K)L(K)KdK}{K\sqrt{S^2-K^2}},$$
$$\mathrm{Sr1}(g,R,\eta) =$$
$$= \mathrm{Sr}(g,R,\eta) + \mathrm{Sr}(g,R,-\eta),$$
$$\mathrm{Sr}(q,R,\eta) =$$
$$= \int_R^{R0} \mathrm{sh}\left[G(m,S,R,\eta)\right] q_m(S,\eta)dS.$$

We write it in a more familiar form,

$$f_m(r) = \frac{-1}{2\pi}\frac{dR(r)}{dr}\times$$
$$\times\left\{\left[\mathrm{Cr1}(g'm,R,\eta)\right] +\right.$$
$$\left.+\left[\mathrm{Sr2}(r,\eta) + \mathrm{Sr2}(r,-\eta)\right]\right\},$$
$$\mathrm{Sr2}(R,\eta) =$$
$$= \int_R^{R0} \frac{\mathrm{sh}\left[G(m,S,R,\eta)\right]}{\sqrt{S^2-R^2}} qm(S,\mu)\times$$
$$\left[m\int_R^S \frac{\mathrm{D}'(K)dK}{\sqrt{S^2-K^2}} +\right.$$
$$\left.+\int_R^S \frac{S\left[\eta(K)L(K)\right]' dK}{\sqrt{S^2-K^2}}\right] dS.$$

The operator Sr2 consists of the part connected with the ray curvature (it is proportional to the harmonic number m) and the part connected with attenuation (the second integral in square brackets). It means that this part can be neglected if the angle variation of the field is small (m is small).

Cormack-type inversion of attenuated ray integral of a vector function.

Reconstruction of stresses is connected with the reconstruction of vector fields. According to the Helmholtz theorem, any vector field $\mathbf{W}(r, \varphi)$ can be split up into a curl-free component and a source-free one:

$$\mathbf{W} = W_n\mathbf{e}_n + W_l\mathbf{e}_l =$$
$$= \left(\frac{\partial\tau}{\partial n} + \frac{\partial N}{\partial l}\right)\mathbf{e}_n + \left(\frac{\partial\tau}{\partial l} - \frac{\partial N}{\partial n}\right)\mathbf{e}_l.$$

where τ, N are the potentials of these components, $\mathbf{e}_n$, $\mathbf{e}_l$ denote the unit normal and unit tangent vectors to our curves.

For simplicity, we suppose that

$$\tau(1,\varphi) = N(1,\varphi) = 0$$

have homogeneous boundary conditions, obtained in the case of reconstructions of quasi-plane deformation.

The measurements in linear vector tomography can be represented by a ray path integral in the form of a scalar (inner) product of a "probe" vector $\mathbf{p}$ and vector $\mathbf{W}$ as

$$\int_{-l(s)}^{l(s)} \mathbf{pW}(s\boldsymbol{\theta} + \mathbf{x})\times$$
$$\times\exp\left(i\int_0^l \eta(s\boldsymbol{\theta} + \mathbf{x})dk\right)dl.$$

The special case when $\mathbf{p} = \mathbf{e}_l$ is what H. Braun and A. Hauck [15] call the longitudinal measurements, and this leads to the formula

$$\hat{W}_l = -\int_{-l(s)}^{l(s)}\left[\frac{\partial N}{\partial n} + i\eta\tau\right]\times$$
$$\times\exp\left(i\int_0^l \eta(s\boldsymbol{\theta} + \mathbf{x})dk\right)dl \quad (6)$$

with applications to the Doppler tomography of a velocity field. Here we transform the ray path integral by partial integration. Another special case when $\mathbf{p} = \mathbf{e}_n$ gives the transverse measurements

$$\hat{W}_n = \int_{-l(\theta)}^{l(\theta)}\left[\frac{\partial\tau}{\partial n} - i\eta N\right]\times$$
$$\times\exp\left(i\int_0^l \nu(s\boldsymbol{\theta} + \mathbf{x})dk\right)dl, \quad (7)$$

which is essential for reconstructing the shear stresses.

A simple analogy between Eqs. (6), (7) is evident. Thus, only the application of Cormack-type inversion to the transverse vector integral (7) can be described. It can be written as the sum of two integrals

$$\hat{W}_n = \int_{-l(s)}^{l(s)} \left[\frac{\partial \tau}{\partial n}\right] \times$$
$$\times \exp\left(i\int_0^l \eta(s\boldsymbol{\theta}+\mathbf{x})dk\right)dl - \quad (8)$$
$$-i\int_{-l(s)}^{l(s)} \eta N \exp\left(i\int_0^l \eta(s\boldsymbol{\theta}+\mathbf{x})dk\right)dl.$$

The first one is transformed using cylindrical coordinates:

$$\frac{\partial \tau}{\partial n} = \sin\nu \frac{\partial \tau}{\partial r} + \cos\nu \frac{\partial \tau}{r\partial \varphi},$$
$$\sin\nu = \frac{S(s)}{R(r)} = \cos\Omega,$$
$$\cos\nu = \pm\frac{1}{R(r)}\sqrt{R^2(r) - S^2(s)},$$

where the plus sign holds for the left-hand side and minus for the right-hand side of the ray.

Expanding $\tau(R, \varphi)$, $\eta(R, \varphi)$, $N(R, \varphi)$ and $\hat{W}_n(s,\theta)$ in Fourier series

$$\tau(R,\varphi) = \sum_{m=-\infty}^{\infty} \tau_m(R) e^{im\varphi},$$
$$\eta(R,\varphi)N(R,\varphi) = \sum_{m=-\infty}^{\infty} N_m(R) e^{im\varphi},$$
$$\hat{W}_n(S,\theta) = \sum_{m=-\infty}^{\infty} w_m(S) e^{im\theta},$$

we can transform integral equation (8) to the following form:

$$w_m(S) = w_m^1(S,\tau) + w_m^2(S,N),$$
$$w_m^1(S,\tau) =$$
$$= 2\int_{|S|}^{R_0} \frac{\cos\left[E(m,K,S)\right]S}{\sqrt{K^2-S^2}} \frac{d\tau_m(K)}{dK} +$$
$$+ \frac{m\sin\left[E(m,K,S)\right]}{k(K)} \times \quad (9)$$
$$\times \frac{dk(K)}{dK}\tau_m(K)dK,$$

$$w_m^2(S,N) =$$
$$= 2\int_{|S|}^{R_0} \frac{\cos\left[E(m,K,S)\right]N_m(K)L(K)K}{\sqrt{K^2-S^2}} dK. \quad (10)$$

The reconstruction of potential τ is carried out by the operator [7]:

$$\tau_m(R) = -\frac{1}{\pi}\mathrm{Crl}(w_m^1, R, \eta). \quad (11)$$

As integral (10) coincides with Eq. (4), we can reconstruct $N_m(R)$ by using Eq. (5) for reconstruction of the scalar function [7].

Invariant representation of the stress tensor field in the inverse thermoelasticity problem. Most of the GRIN structures are formed by the ion-exchange method based on diffusion in alkali-containing glasses. There are two possible sources of RSs produced by varying the glass composition: the size difference between the exchange and the diffusing ions, and the radial variation in the coefficient of thermal expansion across the gradient region. As the lens cools from the ion-exchange temperature to room temperature, RS is introduced. In general, the tensor of residual deformation is the reason for RSs; the components of this tensor do not satisfy the compatibility equations. In the case of ordinary glass, the tensor of residual deformation can be considered to be spherical. Such an isotropic dilation field can be described by a certain "fictive" temperature field [16], and determination of RS can be connected with the solution of the thermoelasticity problem. The problem of reconstructing the thermal residual stresses using integrated photoelasticity is called the inverse thermoelasticity problem of optical tomography [8]. It is assumed in this model that the stress tensor σ obeys the equilibrium equations div $\sigma = 0$. The lateral surface of a cylinder is free from loads.

Let us use stress functions Φ, τ, N for the representation of RSs:

$$\sigma_{rr} = \left[\frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}\right]\Phi -$$
$$-\frac{\partial}{\partial z}\tau + 2\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial \varphi}N\right],$$
$$\sigma_{r\varphi} = -\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial \varphi}\Phi\right] - \quad (12)$$
$$-r\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial r}\right]\Phi - \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}N,$$
$$\sigma_{\varphi\varphi} = \left[\frac{\partial^2}{\partial r^2}\right]\Phi - \frac{\partial}{\partial z}\tau -$$

$$-2\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial \varphi}N\right],$$

$$\sigma_{rz}=\frac{\partial}{\partial r}\tau+\frac{\partial}{r\partial \varphi}N,$$

$$\sigma_{\varphi z}=\frac{\partial}{r\partial \varphi}\tau-\frac{\partial}{\partial r}N.$$

It was established [8] that the following stressed states are possible in a sample: (*i*) a quasi-planar stressed state caused by residual deformations ($N = 0$), (*ii*) a torsional stressed state caused by external loads ($\Phi = \tau = 0$), and (*iii*) a superposition of the aforementioned stresses (a quasi-torsional deformation). The state of pure torsion cannot be generated by a thermal source alone. Therefore, the first and third types of stresses are to be reconstructed. We have also two additional equations for the reconstruction of the quasi-planar stress state:

$$\begin{gathered}\Delta_+\Phi=\left[\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}\right]\Phi= \\ =\sigma_{zz}+\Psi(x,y,z), \\ \Delta_+\tau=-\frac{\partial}{\partial z}\sigma_{zz}, \\ \left(\Delta_+ +\frac{\partial^2}{\partial z^2}\right)\Psi=\Delta\Psi=0,\end{gathered} \tag{13}$$

and the equation for determining quasi-torsional stresses $\Delta N = 0$.

Here Ψ is a 3D harmonic function, connected with boundary conditions. It was proved that the quasi-planar stress ($N = 0$, $\Psi = 0$) can be determined locally based on the values of axial stress and its two first derivatives along the direction of the axis. Moreover, it can be seen from Eqs. (13) that in this case,

$$\frac{\partial}{\partial z}\Phi(x,y,z)=-\tau(x,y,z).$$

In particular, it was confirmed that the thermal stresses of the first angular harmonic are completely reconstructed by the local method. The quasi-torsional stresses can be reconstructed by the global method. The magnitude of the quasi-torsional stresses is almost negligibly small in comparison with quasi-planar stresses and we will confine our attention mainly to reconstructing quasi-planar stresses ($N = 0$, $\Psi = 0$).

Algorithm for reconstruction of residual stresses

Since the reconstruction of RS is based on solving the problem of optical tensor tomography of the stress tensor field together with the problem of elasticity theory, let us first consider the tomographical problem.

Ray integrals of polarized tomography. Transformation of light polarization is measured in a plane orthogonal to the axis of the cylinder. Because there is added torsion of the ray in the plane of transillumination, variations in the polarization within the quasi-isotropic approximation are governed by the following system of equations [9]:

$$\begin{gathered}\frac{d}{dl}\mathbf{E}=-iCP\mathbf{E}, \quad \mathbf{E}=\begin{bmatrix}E_z\\E_s\end{bmatrix}, \\ P=\begin{bmatrix}0.5(\sigma_{zz}-\sigma_{nn}) & \sigma_{nz}\\ \sigma_{nz} & 0.5(\sigma_{nn}-\sigma_{zz})\end{bmatrix},\end{gathered} \tag{14}$$

where $\mathbf{E}$ is the vector representing the amplitude of the electric field strength (the Jones vector), σ_{kj} are the stress tensor components in the moving system coordinates $\mathbf{e}_n$, $\mathbf{e}_l$, $\mathbf{e}_z$ (the Frenet – Serret frame), C is the photoelastic constant.

The solution of Eqs. (14) in the linear approximation can be represented as two ray integrals along the ray [2, 3]:

$$\begin{gathered}2\gamma\cos(\psi)=C\int_{-l(s)}^{l(s)}\left(\sigma_{zz}-\sigma_{nn}\right)dl= \\ =CH_0(s,\theta),\end{gathered} \tag{15}$$

$$\begin{gathered}\gamma\sin(\psi)=C\int_{-l(s)}^{l(s)}\sigma_{nz}dl= \\ =CH_1(s,\theta),\end{gathered} \tag{16}$$

where γ is the integrated optical retardation, ψ is the isocline parameter [2].

These parameters can be measured experimentally and form the physical foundation of the tomography. The ray integral (15) is connected with transversal interaction of the 2D tensor field in the plane of transillumination and the other one (16) is connected with transversal interaction of the 2D vector field (σ_{xz}, σ_{yz}). The cylindrical coordinate system is the most convenient for solving the inverse thermoelastic problem of the cylinder. Thus, we rewrite stress tensor components in cylindrical coordinates:

$$\sigma_{nn} = \sigma_{rr}\cos^2\beta + \sigma_{\varphi\varphi}\sin^2\beta + 2\sigma_{r\varphi}\sin\beta\cos\beta,$$
$$\sigma_{nz} = \sigma_{zr}\cos\beta + \sigma_{\varphi z}\sin\beta.$$

where β is the angle between the normal to a ray and the θ-axis.

Further, using the relations,

$$\beta = \nu - \frac{\pi}{2},$$
$$\sin\beta = -\cos\nu,\ \cos\beta = \sin\nu.$$

we transform them to the following form:

$$\sigma_{nn} = \sigma_{rr}\sin^2\nu + \sigma_{\varphi\varphi}\cos^2\nu - 2\sigma_{r\varphi}\sin\nu\cos\nu, \quad (17)$$
$$\sigma_{nz} = \sigma_{zr}\sin\nu - \sigma_{\varphi z}\cos\nu.$$

Let us rewrite stresses (17) in terms of stress functions (12):

$$\sigma_{nn} = -\frac{\partial}{\partial z}\tau + \sigma_{nn}(\Phi) + \sigma_{nn}(N),$$
$$\sigma_{nn}(\Phi) = \sin^2(\nu)\left[\frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}\right]\Phi +$$
$$+\cos^2(\nu)\left[\frac{\partial^2}{\partial r^2}\right]\Phi + \sin(2\nu)\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial\varphi}\Phi\right],$$
$$\sigma_{nn}(N) = 2\sin^2(\nu)\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial\varphi}\right] - \quad (18)$$
$$-2\cos^2(\nu)\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial\varphi}N\right] +$$
$$+\sin(2\nu)\left[r\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial r}\right] - \frac{1}{r^2}\frac{\partial^2}{\partial\varphi^2}\right]N,$$
$$\sigma_{nz} = \sin\nu\left[\frac{\partial}{\partial r}\tau + \frac{1}{r}\frac{\partial}{\partial\varphi}N\right] -$$
$$-\cos\nu\left[\frac{1}{r}\frac{\partial}{\partial\varphi}\tau - \frac{\partial}{\partial r}N\right].$$

The magnitude of quasi-torsion stresses is almost negligibly small in comparison with quasi-plane stresses, and below we consider the case of quasi-plan deformation: $N = 0$, $\Phi(1) = \partial\Phi/\partial r = 0$.

The first ray integral in the space of angular harmonics can be written as

$$\hat{H}_0(m,S) = \hat{\sigma}_{zz}(m,S) - \frac{\hat{\partial}\tau}{\partial z} - \hat{\sigma}_{nn}(m,S),$$
$$\hat{\sigma}_{nn}(S,m) = -2\int_S^{R(1)}\frac{\cos\left[mA(R,S)\right]}{\sqrt{R^2-S^2}}\times \quad (19)$$
$$\times\left\{\sin^2(\nu)\left[\frac{1}{r}\frac{\partial}{\partial r} - \frac{m^2}{r^2}\right]\Phi_m + \right.$$
$$\left. +\cos^2(\nu)\left[\frac{\partial^2}{\partial r^2}\right]\Phi_m\right\}R\mathrm{d}r -$$
$$-2m\int_S^{R(1)}\frac{\sin\left[mA(R,S)\right]}{\sqrt{R^2-S^2}}\times \quad (20)$$
$$\times\sin(2\nu)\frac{\partial}{\partial r}\left[\frac{1}{r}\Phi_m\right]R\mathrm{d}r.$$

Integrating by parts, we transform Eq. (20) to the formula

$$\hat{\sigma}_{nn}(S,m) = -2\int_S^{R(1)}\frac{\cos[mA(R,S)}{\sqrt{R^2-S^2}}\times$$
$$\times\frac{S^2}{R^2}\left(\frac{\partial}{\partial r}\Phi_m\right)\left(\frac{d}{dr}n(r)\right)rdr \quad (21)$$
$$-2m\int_S^{R(1)}\sin\left[mA(R,S)\right]\times$$
$$\times\frac{S}{R^2}\Phi_m\left(\frac{d}{dr}n(r)\right)dr,$$

containing in an explicit form the term $dn(r)/dr$ depending on the curvature of the ray.

The second ray integral (16) is connected with transversal interaction of the two-dimensional vector field and coincides with Eq. (11) in a space of angular harmonics:

$$\hat{H}_1(S,m) = \hat{\sigma}_{nz}(S,m) = w_m^1(S,\tau), \quad (22)$$

Thus, for stress reconstruction, we have four equations: two differential (Eqs. (13)) and two integral (Eqs. (19), (22)) ones.

The problem of reconstructing the RSs in the global and the local forms. Tomographic measurements in the global inverse problem are performed in a system of parallel planes over the entire height of the cylinder. The layer-by-layer reconstruction of stresses is possible starting from the bottom of the cylinder, where the axial stress is zero. Thus, we can determine τ from Eq. (22), the

normal stress σ_{zz} and Φ using Eq. (13) and then all stress components given by Eqs. (12). The first ray integral (15) is not used in this algorithm.

Although the presented algorithm includes, as a special case, rectilinear propagation of rays, numerical implementation of the algorithm may differ for these cases. More detailed analysis shows that the global approach allows for reconstructing stresses completely [8].

Tomographic measurements in the local inverse problem are carried out in two closely spaced sections orthogonal to the optical axis for determining the height derivative of τ. Quasi-plane stresses are determined based on two differential Eqs. (13) and two ray integrals (19), (22).

Reconstruction of the quasi-plane stresses in GRIN structures. For regularization of inverse problem, we use singular decomposition of the solution of Eqs. (13). The angular decomposition using trigonometric functions and radial decomposition using Zernike polynomials were applied to represent the stresses (see Appendix) [8, 17]. MATLAB codes were developed to numerically solve integral Eqs. (19), (22) by the least square method. This method has been previously used for inversion of the Abel transform [18, 19]. We have three types of unknown coefficients:

$$\sigma_{zzm}^{kc,s}, \frac{\partial}{\partial z}\sigma_{zzm}^{kc,s}, \frac{\partial^2}{\partial z^2}\sigma_{zzm}^{kc,s}.$$

The second type of them can be determined from Eq. (22). The integrals

$$w_m^k(S) = 2\int_{|S|}^{1} \frac{\cos\left[mA(R,S)\right]S}{\sqrt{R^2-S^2}} \frac{d\tau_m^{k1}(r)}{dr} + \\ + \frac{m\sin\left[mA(R,S)\right]}{r}\tau_m^{k1}(r)dr,$$

$$\tau_m^{k1}(r) = \frac{1}{2m}\left[C_k^m(r) + D_k^m(r)\right],$$

$$\frac{\partial}{\partial r}\tau_m^{kc,s}(r) = \frac{-1}{2\rho}\left[C_k^m(r) - D_k^m(r)\right]$$

have to be calculated for the inversion Eq. (9).

According to the least-squares method [19], in order to determine $\frac{\partial}{\partial z}\sigma_{zzm}^{kc,s}$ we solved the following equation:

$$\sum_{n=1}^{N}\left[\hat{H}_1(S_n,m) - \right. \\ \left. - w_m^k(S_n)\frac{\partial}{\partial z}\sigma_{zzm}^{kc,s}\right]^2 = \min,$$

where S_n are the n measured points, $\hat{H}_1(S_n,m)$ is the measured ray integral.

The number of polynomials in the representation is connected with noise filtering. The second derivative $\frac{\partial^2}{\partial z^2}\sigma_{zzm}^{kc,s}$ can be approximately determined by measuring integrals in two parallel closely spaced sections. Then we can determine $\sigma_{zzm}^{kc,s}$ using solution (21).

Two types of refractive index profiles have been used in MATLAB codes [20, 21]. The profile of GRIN structures

$$n(r/R) = n_0\sqrt{1-g\left(r/R\right)^2}$$

is known as the parabolic refractive index profile [21]. Here n_0 is the refractive index on the axis and g is a positive constant, R is the radius of the cylinder.

In this case, the ray equation (1) has an analytical solution

$$\alpha(r,s) = \\ = \arccos\left[\frac{s}{r}\sqrt{\frac{R^2-g(r^2+s^2)}{R^2-2gs^2}}\right].$$

The Cartesian coordinates of these rays can be expressed analytically in terms of the parameters r, s:

$$y(r,s) = s\sqrt{\frac{R^2-g(r^2+s^2)}{R^2-2gs^2}},$$

$$x(r,s) = \sqrt{\frac{(r^2-s^2)(R^2-gs^2)}{R^2-2gs^2}}.$$

It is clear that $g < 0.5$ here. These relations are used for ray tracing and solutions (19), (22) if $m > 0$.

Another profile (parabolic) of the form

$$n(r) = n_0\left[1-0.5g^2\left(r/R\right)^2\right]$$

is typically used to approximate the index refraction of GRIN lenses [21, 22] and of multimode fibers. The parameter g varies from 0.01 for fibers up to 0.30 for GRIN rod lenses.

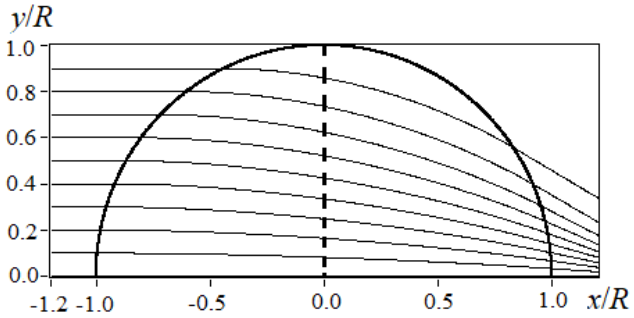


Fig. 3. Ray tracing through the sample (its half-round cross-section of diameter $2R$ is shown) when $n\,(r/R) = n_0[(1 - 0.2(r/R)^2]^{1/2}$ ($g = 0.2$, x is the translucence parameter)

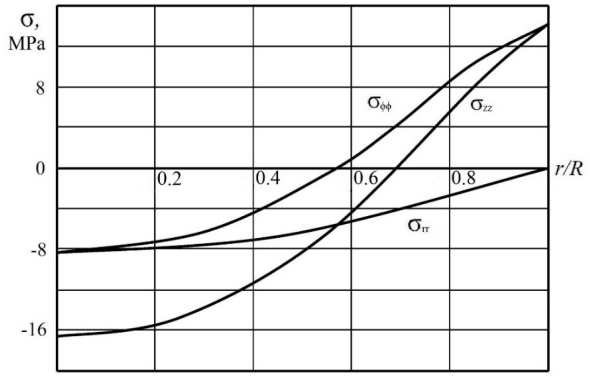


Fig. 4. The reconstructed principal components of the stress tensor over the GRIN lense's radius (the lense was made of zirconium-silicate glass)

Ray tracing is given for the case $g = 0.2$ (Fig. 3). The angle (1) can be expanded using the small parameter g. The expansion to within the first two terms

$$\alpha(r,s) = \pm\int_{|s|}^{r} \frac{S(s)\mathrm{d}\rho}{\rho\sqrt{\left[R(\rho)\right]^2 - \left[S(s)\right]^2}} \approx$$

$$\approx \pm\int_{|s|}^{r} \frac{s\mathrm{d}\rho}{\rho\sqrt{\left[\rho\right]^2 - \left[s\right]^2}} =$$

$$= \pm 0.5\left(\frac{g}{R}\right)^2 s\sqrt{r^2 - s^2}$$

coincides with the expansion of

$$\alpha_1 = \arccos\left[\frac{S_1(s)}{R_1(r)}\right],$$

$$R_1(r) = n_0\left[R^2 + 0.5g^2r^2\right]r,$$

$$S_1(s) = n_0\left[R^2 + 0.5g^2s^2\right]s.$$

This approximation can be used in practice to solve Eqs. (19), (22) if $g < 0.1$ and $m < 15$. The algorithm of reconstruction is simplified [9] for axial symmetric plane-strain state ($m = 0$). In this case, all functions depend only on the radial coordinates, and we take only Eq. (19):

$$\hat{H}_0(s) =$$

$$2\int_{|s|}^{1}\left[\sigma_{zz} + \int_0^r t\sigma_{zz}(t)dt\,\frac{S^2(s)}{R^2(r)}\left(\frac{dn}{dr}\right)\right]\times \quad (23)$$

$$\times\frac{R(r)dr}{\sqrt{R^2(r) - S^2(s)}}.$$

Optical retardation (23) consists of two terms: main retardation due to axial stress and additional one due curving of the ray. It has an additional term connected with ray deflection. The above algorithm is used for stress reconstruction

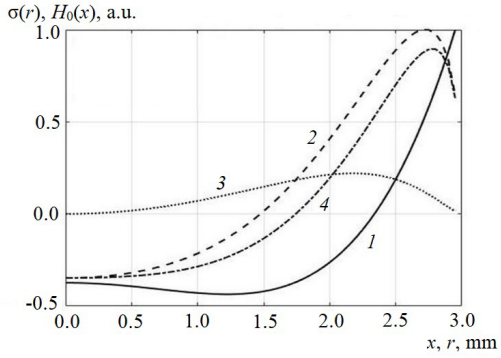


Fig. 5. The result of the residual stresses simulation (an example for using the polynomial datum on the sample of $R = 3$ mm);
in addition to the axial stress curve (*1*), the curves of optical (*2*), additional (*3*) and main (*4*) retardations are presented

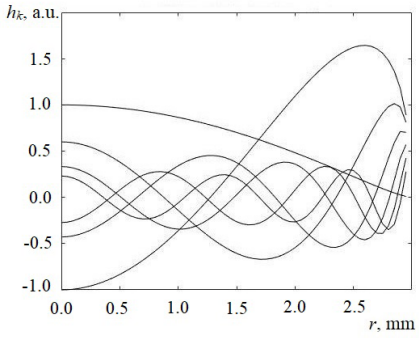


Fig. 6. Integrals $h_k(r)$ obtained on a basis of the first seven Zernike polynomials

$$\sum_{n=1}^{N}\left[\hat{H}_0(s_n) - h_k(s_n)\sigma_{zz0}^{k}\right]^2 = \min.$$

The graphs for the reconstruction of stresses in GRIN lenses made of zirconium-silicate ($n_0 = 1.54$) glasses are shown in Fig. 4 [23].

Fig. 5 shows the result of simulation of residual stresses by the polynomial datum for the case of $R = 3$ mm. Fig. 6 gives the retardations for the first seven Zernike polynomials.

Summary

In recent years, there has been a growing interest towards tomographic reconstruction of vector and tensor fields in the refracting medium [5, 6, 22]. We present the algorithm for reconstruction of residual stresses in GRIN lenses. From an experimental standpoint, implementation of the given method of reconstruction does not pose any particular difficulties and is conducted similarly to the case of the constant refractive index. The simplest examples of applications of this algorithm for the axial symmetric distribution of stresses are presented. The inversion algorithm provides a comparatively smooth stress distribution, produced by the ion-change technique. Including the radial dependences of the photoelasticity coefficients and elastic constants in GRIN structures in the algorithm for reconstructing stresses is the subject for further research.

Appendix

SCD of quasi-plan stresses

Let us consider a representation of quasi-plane stresses in terms of Zernike polynomials $R_k^m(r)$ [12], $m > 0$:

$$\sigma_{zz}(r,\varphi,z) =$$
$$= \sum_{m=1}\sum_{k=0}^{\infty}\left[\sigma_{zzm}^{kc}(z)\cos(m\varphi) + \right.$$
$$\left. + \sigma_{zzm}^{kc}(z)\sin(m\varphi)\, R_{m+2k}^{m}(r)\right],$$
$$\Phi(r,\varphi,z) =$$
$$= \sum_{m=1}\sum_{k=1}^{\infty}\left[f_m^{kc}(r,z)\cos(m\varphi) + \right.$$
$$\left. + f_m^{kc}(r,z)\sin(m\varphi)\right],$$
$$\tau(r,\varphi,z) =$$
$$= \sum_{m=1}\sum_{k=1}^{\infty}\left[\tau_m^{kc}(r,z)\cos(m\varphi) + \right]$$
$$+ \tau_m^{kc}(r,z)\sin(m\varphi)\Big],$$
$$\sigma_{ij}(r,\varphi,z) =$$
$$= \sum_{m=1}\sum_{k=1}^{\infty}\left[\sigma_{ijm}^{kc}(r,z)\cos(m\varphi) + \right.$$
$$\left. + \sigma_{ijm}^{kc}(r,z)\sin(m\varphi)\right].$$

In the following, as evident from the context of the notation, the trigonometric functions are not written out:

$$f_m^{kc,s}(r,z) =$$
$$= \frac{-1}{2m}\left[C_k^m(r) + D_k^m(r)\right]\sigma_{zzm}^{kc,s}(z),$$
$$\tau_m^{kc,s}(\rho,z) =$$
$$= \frac{1}{2m}\left[C_k^m(r) + D_k^m(r)\right]\frac{\partial}{\partial z}\sigma_{zzm}^{kc,s}(z),$$
$$\sigma_{\rho\rho m}^{kc[s]}(r,z) =$$

$$= \frac{1}{2\rho^2}\Big[(m+1)C_k^m(r) +$$

$$+(m-1)D_k^m(r)\Big]\sigma_{zzm}^{kc[s]}(z) -$$

$$-\frac{1}{2m}\Big[C_k^m(r)+D_k^m(r)\Big]\frac{\partial^2}{\partial z^2}\sigma_{zzm}^{kc[s]}(z),$$

$$\sigma_{\rho\varphi m}^{ks[c]}(r,z) =$$

$$= \pm\frac{1}{2r^2}\Big[(m+1)C_k^m(r) -$$

$$-(m-1)D_k^m(r)\Big]\sigma_{zzm}^{kc[s]}(z)\begin{cases}\sin(m\varphi)\\ [\cos(m\varphi)]\end{cases},$$

$$\sigma_{\varphi\varphi m}^{kc[s]}(r,z) =$$

$$= R_{m+2k}^m(r)\sigma_{zzm}^{kc[s]}(z) - \sigma_{\text{ÁÁ}m}^{kc[s]}(r,z) -$$

$$-\frac{1}{m}\Big[C_k^m(r)+D_k^m(r)\Big]\frac{\partial^2}{\partial z^2}\sigma_{zzm}^{kc[s]}(z),$$

$$\sigma_{rzm}^{kc[s]}(r,z) =$$

$$= \frac{\partial}{\partial r}\tau_m^{kc,s}(r,z) = \frac{-1}{2\rho}\Big[C_k^m(r) -$$

$$-D_k^m(r)\Big]\frac{\partial}{\partial z}\sigma_{zzm}^{kc[s]}(z),$$

$$\sigma_{\ddot{o}zm}^{ks[c]}(r,z) = \mp\frac{1}{2r}\Big[C_k^m(r) +$$

$$+D_k^m(r)\Big]\frac{\partial}{\partial z}\sigma_{zzm}^{kc[s]}(z)\begin{cases}\sin(m\varphi)\\ [\cos(m\varphi)]\end{cases},$$

where

$$C_k^m(r) = \frac{r}{2(2k+m+1)}\times$$

$$\times\Big[R_{m+2k+1}^{m+1}(r) - R_{m+2k-1}^{m+1}(r)\Big],$$

$$D_k^m(\rho) = \frac{r}{2(2k+m+1)}\times$$

$$\times\Big[R_{m+2k-1}^{m-1}(r) - R_{m+2k+1}^{m-1}(r)\Big].$$

In the case of an axial symmetric plane strain state (m = 0), the solution of the inverse thermoelastic problem can be represented in terms of Zernike polynomials $R_{2k}^0(r)$:

$$\sigma_{zz}(r) = \sum_{k=0}^{\infty}\sigma_{zz}^k R_{2k}^0(r),$$

$$\sigma_{\rho\rho}(r) = \sum_{k=1}^{\infty}\sigma_{zz}^k\sigma_{\rho\rho}^k(r),$$

$$\sigma_{\varphi\varphi}(r) = \sum_{k=1}^{\infty}\sigma_{zz}^k\sigma_{\varphi\varphi}^k(r),$$

$$\sigma_{\rho\rho}^k(r) =$$

$$= \frac{1}{\rho^2}\int_0^{\rho} tR_{2k}^0(t)\mathrm{d}t =$$

$$= \frac{1}{2(2k+1)\rho}\Big[R_{2k+1}^1(\rho) - R_{2k-1}^1(\rho)\Big],$$

$$\sigma_{\varphi\varphi}^k(r) = R_{2k}^0 - \sigma_{\rho\rho}^k(r).$$

REFERENCES

1. **Yablon A.D.,** Recent progress in optical fiber refractive index profiling, Proceedings of the Optical Fiber Communication Conference/ *National Fiber Optic Engineers Conference*, Los Angeles, USA. 6 – 10 March 2011, Paper OMF1. OSA Technical Digest (CD), Optical Society of America, 2011.

2. **Aben H., Ainola L., Anton J.,** Integrated photoelasticity for nondestructive residual stress measurements in glass, Optics and Lasers in Engineering. 33 (1) (2000) 49–64.

3. **Puro A., Aben H.,** Tensor field tomography for residual stress measurement in glass articles, Proceedings of the 7th European Conference on Non-Destructive Testing, Copenhagen, 26 – 29 May, 1998. NTD net. 3 (8) (1998) 2390–2397.

4. **Fuki A.A., Kravtsov Y.A., Naida O.N.,** Geometrical optics of weakly anisotropic media, Gordon and Breach Science Publishers, Amsterdam, 1998.

5. **Pfitzenreiter T., Schuster T.,** Tomographic reconstruction of the curl and divergence of 2D vector fields taking refractions into account, SIAM Journal on Imaging Sciences. 4 (1) (2011) 40–56.

6. **Svetov I.E., Derevtsov E.Yu., Volkov Yu.S., Schuster T.,** A numerical solver based on B-splines for 2D vector field tomography in a refracting medium, Mathematics and Computers in Simulation. 97 (March) (2014) 207–223.

7. **Puro A.E.,** Tomography in optically axisymmetric media, Optics and Spectroscopy. 124 (2) (2018) 278–284.

8. **Puro A., Karov D.,** Inverse problem of thermoelasticity of fiber gratings, Journal of Thermal Stresses. 39 (5) (2016) 500 – 512.

9. **Puro A., Karov D.,** Polarization tomography of residual stresses in cylindrical gradient-index lenses, Optics and Spectroscopy. 124 (5) (2018) 735–740.

10. **Born M., Wolf E.,** Principles of optics, 7th edition, Cambridge University Press, Cambridge, 2019.

11. **Puro A.,** Cormack-type inversion of exponential Radon transform, Inverse problems. 17 (1) (2001) 179–188.

12. **Puro A.,** Polarization tomography of the tokamak plasma magnetic field, Optics and Spectroscopy. 116 (1) (2014) 122–129.

13. **Bieg B., Chrzanowski J., Kravtsov Y.A., Orsitto F.,** Main physical factors limiting the accuracy of polarimetric measurements in Tokamak plasma, Physics Procedia. 62 (2015) 107–112.

14. **Zeng G.L., Li Y.,** A discrete convolution kernel for No-DC MRI, Inverse Problems. 31 (8) (2015) 085006.

15. **Braun H., Hauck A.,** Tomographic reconstruction of vector fields, IEEE Transactions on Signal Processing. 39 (2) (1991) 464–471.

16. **Gardon R.,** Thermal tempering of glass, In the book "Glass Science and Technology, Vol. 5: Elasticity and Strength in Glasses", Edited by D.R. Uhlmann, N.J. Kreidl. Ch. 5, Academic Press, New York (1980) 145–213.

17. **Pretzier G., Дger H., Neger T., et al.,** Comparison of different methods of Abel inversion using computer simulated and experimental side-on data, Zeitschrift für Naturforschung. A. 47a (1992) 955–970.

18. **Fricker P.,** Zernike polynomials and functions (orthogonal basis on the unit circle), Version 1.3.0.1. (6.84 KB), 2016, https:// se.mathworks.com/ matlabcentral/ fileexchange /7687-Zernike-polynomials.

19. **Killer C.,** Abel inversion algorithm, Fourier-based reconstruction of an unknown radial distribution assuming cylindrical symmetry version 1.5.0.0 (4.17 KB), 2016. https://au.mathworks.com/matlabcentral/ fileexchange/43639-abel-inversion-algorithm.

20. **Wang Y.,** Radial gradient index rod containing Li^+, Acta Photonica Sinica. 31 (7) (2002) 897–900.

21. **Adam J.A., Pohrivchak M.,** Evaluation of ray-path integrals in geometrical optics, International Journal of Applied and Experimental Mathematics. 1 (2) (2016) 108.

22. **Ramadan W.A., Wahba H.H., Shams El-Din M.A.,** Two-dimensional refractive index and birefringence profiles of a graded index bent optical fibre, Optical Fiber Technology. 36 (July) (2017) 115–124.

23. **Karov D.D.,** Polyarizatsionnaya tomografiya napryazhyonnogo sostoyaniya v gradientno-opticheskikh strukturakh [Polarizing tomography of a stressed state in the gradient-optical structures], PhD Thesis, Specialty 01.04.04, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, 2012 (in Russian), http://fizmathim.com/ read/369384/a?#?page=16.

Received 01.10.2020, accepted 14.10.2020.

THE AUTHORS

KAROV Dmitry D.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
dmkarov@yandex.ru

PURO Alfred E.
Euroacademy
4 Mustamäe tee, Tallinn, 10621, Estonia
alfredpuro@gmail.com

DOI: 10.18721/JPM.13412

HYDROGEN FLUORIDE OBTAINING FROM URANIUM HEXAFLUORIDE IN THE COMBUSTION MODE

D.S. Pashkevich[1], A.R. Zimin[1], Yu.I. Alexeev[2], D.A. Mukhortov[2], P.S. Kambur[1], V.B. Petrov[3], D.A. Bazhenov[4], P.A. Smolkin[4], V.V. Kapustin[1]

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

[2] LTD «New chemical products», St. Petersburg, Russian Federation;

[3] Russian Scientific Center «Applied Chemistry», St. Petersburg, Russian Federation;

[4] JSC «Siberian Chemical Plant", Seversk, Tomsk region, Russian Federation

Currently, there are no effective industrial technologies to return fluorine into a technological cycle. The authors have proposed a method for processing uranium hexafluoride depleted in the U-235 (DUHF) isotope in a hydrogen-oxygen flame that can be used as a basis for obtaining hydrogen fluoride and uranium oxides. Using own devising program code and complex ASTRA.4, performed thermodynamic calculations in the U-F-H-O elements system and it was shown that if the number of hydrogen atoms exceeds the number of fluorine atoms, then at temperatures above 1150 K, the only fluorine-containing substance in the thermodynamically equilibrium mixture was HF, while the main uranium-containing was UO_2. The proposed method was experimentally investigated at a pilot plant and it showed that the composition of its products was close to thermodynamic equilibrium. The results can become the basis for industrial technology of the hydrogen fluoride production from DUHF.

Keywords: depleted uranium hexafluoride, uranium dioxide, hydrogen fluoride, combustion

Citation: Pashkevich D.S., Zimin A.R., Alexeev Yu.I., Mukhortov D.A., Kambur P.S., Petrov V.B., Bazhenov D.A., Smolkin P.A., Kapustin, A.V., Hydrogen fluoride obtaining from uranium hexafluoride in the combustion mode, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 110–118. DOI: 10.18721/JPM.13412

Introduction

Uranium hexafluoride is a ^{235}U-depleted by-product of isotopic separation of natural uranium. Uranium hexafluoride is a Class 1 hazardous substance, so its accumulation and storage is an environmental threat. Uranium hexafluoride (DUHF) is also a source of pure ^{235}U-depleted uranium, which can be effectively used to produce mixed-oxide (MOX) fuel for fast-neutron reactors and fluorine. DUHF is currently stored at enrichment facilities of the Rosatom State Atomic Energy Corporation.

The methods for processing DUHF developed in Russia have a number of serious drawbacks, making it impossible to use them commercially [1–4].

The only known industrial technology for processing DUHF is two-stage hydrolysis yielding an aqueous solution of hydrogen fluoride and triuranium octaoxide, developed by Orano (France) [5, 6]:

$$UF_{6\,gas} + 2H_2O_{gas} \rightarrow UO_2F_{2\,sol} + 4HF_{gas}; \quad (1)$$

$$3UO_2F_{2\,sol} + 3H_2O_{gas} + H_{2\,gas} \rightarrow \rightarrow U_3O_{8\,sol} + 6HF_{gas} + H_2O_{gas}. \quad (2)$$

An excess of water is used for complete regeneration of fluorine by this method, with hydrogen fluoride and its azeotropic aqueous 40% solution acting as the main fluorine-containing products, which does not allow to recover fluorine in the nuclear fuel cycle.

In Russia, this technology is used by Electrochemical Plant JSC at the W-ECP facility. Notably, producing hydrogen fluoride at this facility is costlier than producing hydrogen fluoride by the traditional method (sulfuric acid decomposition of high-quality fluorite): the facility however has greater environmental safety [7].

We have proposed a method for regeneration of fluorine in the form of hydrogen fluoride from DUHF by processing it in the flame of a hydrogen-containing fuel and an oxygen-containing oxidizer, for example,

$$UF_{6\,gas} + 3H_{2\,gas} + O_{2\,gas} \rightarrow \rightarrow UO_{2\,sol} + 6HF_{gas} - 583\text{kJ}. \quad (3)$$

Calculations indicate that process (3) is thermodynamically allowed, its thermal effect is that the adiabatic temperature of the products can reach about 2500 K, i.e., the process can proceed under combustion. Calculations also established that if the number of hydrogen atoms in a system of U-F-H-O elements exceeds the number of fluorine atoms, the main fluorine-containing substance at temperatures above 1150 K in a mixture in thermodynamic equilibrium is hydrogen fluoride, and the main uranium-containing substances are uranium oxides. These conclusions are confirmed by experiments on a test setup that we have constructed.

Thermodynamic study of hydrolysis of uranium fluorides

According to the data given in monograph [8], hydrogen fluoride can be obtained from uranium fluorides by hydrolysis:

$$3UO_2F_{2\,sol} + 3H_2O_{gas} \rightarrow \rightarrow U_3O_{8\,sol} + 6HF_{gas} + 0.5O_{2\,gas}, \quad (4)$$

$$UF_{4\,sol} + 2H_2O_{gas} \rightarrow UO_{2\,sol} + 4HF_{gas}, \quad (5)$$

$$UF_{6\,gas} + 3H_2O_{gas} \rightarrow \rightarrow UO_{2\,sol} + 6HF_{gas} + 0.5O_{2\,gas}. \quad (6)$$

To estimate the temperature range where it is optimal to carry out processes (1), (2) and (4)–(6), the values of the Gibbs energy change ΔG were calculated depending on the temperature T (Fig. 1).

Analyzing the data in Fig. 1, we can conclude that a high degree of fluorine recovery from uranium hexafluoride without excess water can be achieved via hydrolysis at a temperature above 1150 K.

This temperature can be obtained by treating UF_6 in the flame of a hydrogen-containing fuel and an oxygen-containing oxidizer:

$$UF_{6\,sol} + x_1C_mH_nO_kN_{lgas} + x_2O_{2gas} \rightarrow \rightarrow UO_{2sol} + 6HF_{gas} + y_1CO_{zgas} + y_1N_2 - Q, \quad (7)$$

where $C_mH_nO_kN_l$ is a volatile hydrogen-containing substance, $n > 0$, $m \geq 0$, $k \geq 0$, $l \geq 0$; $nx_1 \geq 6$, $2x_2 > 2$, $Q \approx 10^2$–10^3 kJ, $T > 10^3$ K.

The amount of fuel and oxidizer is chosen so that fluorine is completely recovered as hydrogen fluoride, without water produced, and the main uranium-containing substance is uranium dioxide.

Notably, the method we propose implies that heat is supplied to the reaction zone as a result of chemical heat release rather than from the reactor wall, so the temperature of the reactor walls can reach about 100°C.

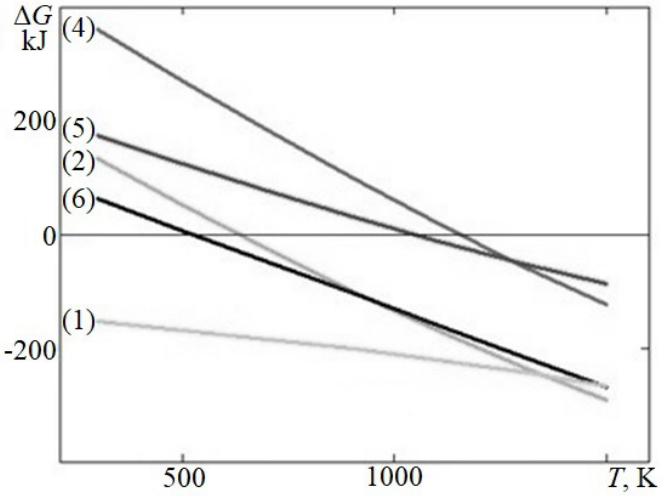


Fig. 1. Gibbs energy changes for hydrolytic reactions of uranium fluorides depending on temperature (curve numbers correspond to reaction numbers)

If hydrogen is used as a hydrogen-containing fuel (reaction (3)), the adiabatic temperature of the products, as well as the temperature accounting for the corrections for radiant heat transfer [9], calculated on the basis of the Kirchhoff law, are 2540 K and 1670 K, respectively. Thus, it is possible to recover fluorine from DUHF as hydrogen fluoride without excess water via scheme (3) in combustion mode.

Thermodynamic equilibrium composition of substances in the system of U-F-H-O elements

The thermodynamic equilibrium composition of substances in the system of U-H-F-O elements was calculated using the method for minimizing the Gibbs energy for a mixture of substances, assuming that the most stable compounds are uranium dioxide, uranium tetrafluoride, uranyl fluoride, hydrogen fluoride and water [10, 11]:

$$x_1 UF_{6\,gas} + x_2 H_{2\,gas} + x_3 O_{2\,gas} \rightarrow \rightarrow y_1 UO_{2\,sol} + y_2 UF_{4\,sol/liq} + + y_3 UO_2F_{2\,sol/liq} + y_4 HF_{gas} + y_5 H_2O_{gas}, \quad (8)$$

where x_i, y_i are the stoichiometric coefficients.

The model was verified by comparing the results obtained by experiment and by calculations in the ASTRA.4 software package. The latter is based on the method for calculating the thermodynamic equilibrium composition by searching for the maximum entropy [11, 12]. The calculated results are given in Table 1. The discrepancy between the calculated results obtained by minimizing the Gibbs energy and maximizing the entropy at temperatures above 1150 K is less than 20%, at temperatures above 1150 K. A possible reason for this discrepancy is that the reference books used give differing values for the thermodynamic functions of the substances considered. For example, the data for calculations by the method for minimizing the Gibbs energy were taken from [10], giving 1854 kJ/mol for the standard enthalpy of formation for UF_4, while the value of the same parameter in [11], used in calculations by the ASTRA.4 software package, was 1920 kJ/mol. Discrepancies of several percent are also observed for other parameters (heat capacity, standard entropy of formation), which is what results in the 20% difference for uranium dioxide and 6% for hydrogen fluoride. Thus, the developed model is suitable for calculating the thermodynamic equilibrium composition.

It follows from the results in Table 1 that HF is the main fluorine-containing substance in the thermodynamic equilibrium mixture at temperatures above 1100 K, and the main uranium-containing substance is UO_2. The content of uranium fluorides is less than 5%, and less than 0.5% at temperatures above 1700 K; the water content at temperatures above 1300 K is several percent.

Thus, apparently, complete recovery of fluorine from DUHF via reaction (3) is not thermodynamically prohibited.

The results of thermodynamic analysis lead us to conclude that the process of obtaining hydrogen fluoride during the interaction of uranium hexafluoride with hydrogen and oxygen in combustion mode deserves further consideration.

Table 1

Calculated thermodynamic equilibrium composition of substances in system of 1U-6F-6H-2O elements

T, K	Compound concentration, C, mol%									
	UO_2		UF_4		UO_2F_2 / UOF_3		H_2O		HF	
	G	Astra	G	Astra	G	Astra	G	Astra	G	Astra
700	< 0.1	< 0.1	20.0	19.5	< 0.1	< 0.1	40.0	39.1	40.0	40.9
900	1.0	< 0.1	19.0	17.8	< 0.1	< 0.1	40.0	35.1	40.0	47.1
1100	17.0	10.3	0.9	4.1	< 0.1	< 0.1	1.5	8.1	80.6	77.6
1300	16.9	13.8	0.8	< 0.1	< 0.1	< 0.1	2.7	1.9	79.6	84.3
1500	16.7	13.4	0.3	0.6	< 0.1	< 0.1	3.5	2.2	79.5	83.8
1700	16.3	12.9	0.3	0.7	< 0.1	0.8	3.9	2.7	79.5	82.9

Note. The calculations were carried out using the in-house program code (G) and the ASTRA.4 software package.

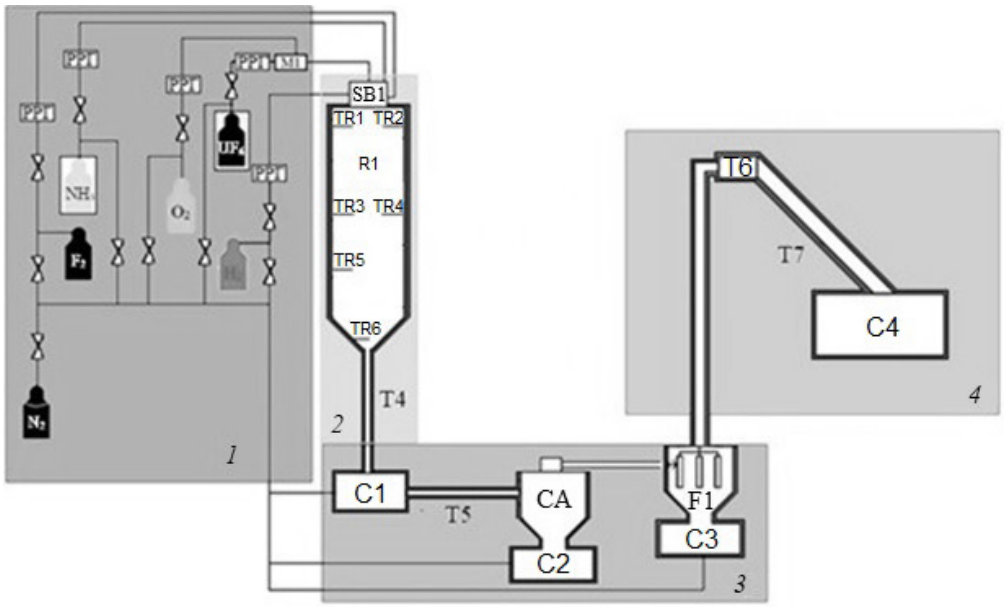


Fig. 2. Schematic of pilot plant:
unit *1* for preparing and supplying initial gaseous substances, reactor unit *2*,
unit *3* for separating solid and gas phases, unit *4* for condensing hydrogen fluoride and water;
nozzle SB1, thermocouples TR1–TR6, solid phase collectors C1–C4,
heat exchangers T4–T7, cyclone separator CA, sintered metal filter F1

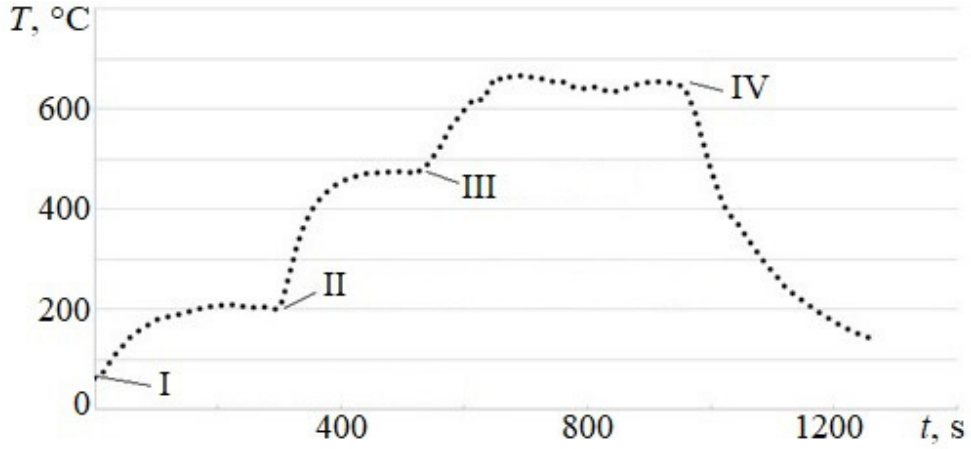


Fig. 3. Time dependence of junction temperature of thermocouple TR2 (see Fig. 2) in the reactor.
The following stages of the process are shown: NH_3 and F_2 are supplied (I), H_2 and O_2 are supplied (II),
DUHF is supplied (III), supply of components is stopped (IV)
Gas flow rates, g/s: 0.010 for F_2, 0.014 for NH_3, 1.50 for DUHF, 0.05 for H_2, 0.27 for O_2

Experimental study of interactions of uranium hexafluoride with hydrogen and oxygen in combustion mode

A pilot plant was constructed to experimentally study the composition of the products resulting from processing DUHF in the flame of a hydrogen-containing fuel and an oxygen-containing oxidizer,

The setup (Fig. 2) consists of four main units.

Unit *1* for preparing and supplying the initial gaseous substances is intended for feeding DUHF, hydrogen, ammonia, fluorine, and nitrogen into the reactor. Ammonia and fluorine were used as a self-igniting mixture to initiate combustion of the main flame. Flow controllers by Bronkhorst were used to monitor the flow rates of all gases.

The main element of reactor unit *2* was a 'tunnel burner' type reactor with a cooling

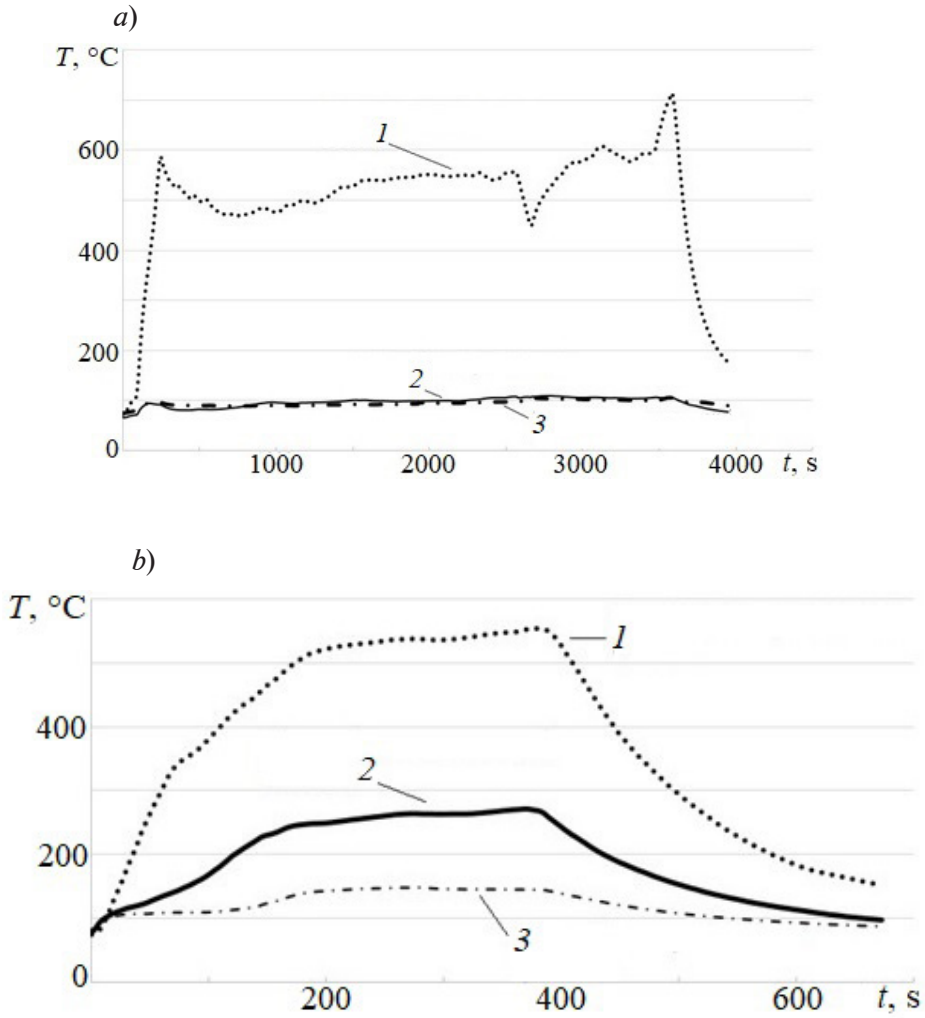


Fig. 4. Time dependences for temperature of thermocouple junctions in the reactor for two cases of component flow rates, g/s: 1.5 DUHF, 0.04 H_2, 0.016 O_2 (a) and 7.0 DUHF, 0.14 H_2, 0.76 O_2 (b); the flow rate of other components remains the same, g/s: F_2 0.010, NH_3 0.014.
Data from thermocouples mounted on top (1), middle (2), and bottom (3) flanges are given

jacket, where water with a temperature of about 90°C was supplied to prevent condensation of hydrogen fluoride. The length and diameter of the reactor were 2650 and 142 mm, respectively. The reactor consisted of two cylindrical sections 1.00 m long and one conical section 0.65 m long. Measuring flanges were mounted between the sections in the reactor, with chromel-alumel thermocouples in protective covers and pressure sensors installed. The components were fed into the reactor using two nozzles installed in nozzle assembly SB1.

Unit *3* for separating the solid and gas phases consisted of solid phase collectors C1, C2, C3, heat exchanger T5, where the flow was cooled to a temperature of 120°C, cyclone separator CA and a sintered metal filter F1 separating particles with the size of more than 0.15 μm.

Unit *4* for condensing hydrogen fluoride and water was a pipe-in-pipe heat exchanger, where the temperature was maintained at 40°C.

The pilot plant was used for studying the interaction of DUHF with hydrogen and oxygen in combustion mode at a ratio of components close to that given in reaction (3). The flow rate of DUHF was varied from 1 to 7 g/s in the experiments, which corresponded to the rate of 0.02–0.20 m/s of the mixture of initial materials in the reactor at 90°C.

Thermocouples were used to observe the formation of the flame front. A typical thermogram of the experiment is shown in Fig. 3, showing the time dependence of the junction temperature for thermocouple TR2 (first measuring flange, 40 mm from the axis) in the reactor. The temperature clearly increases when ammonia

and fluorine (self-ignition), hydrogen and oxygen (combustion of an oxygen-hydrogen flame) and DUHF are supplied; the variation of temperature amounts to hundreds of degrees.

Thus, it was found that a three-component flame consisting of DUHF-hydrogen-oxygen is reliably produced and burns steadily upon initiation by a fluorine-ammonia flame with the flow rate of DUHF varying from 1 to 7 g/s and a component ratio close to that in reaction (3).

Fig. 4 shows thermograms for the thermocouples installed in three measuring flanges for DUHF flow rates close to minimum and maximum.

We can conclude from the data shown in Figs. 3 and 4 that the length of the flame in the reactor did not exceed 1 m in the given range of flow rates.

The main uranium-containing product of process (3) was a powder of uranium oxides, primarily uranium dioxide and triuranium octaoxide. More than 97% of the powder gravitated downwards to collector C1.

The contents of fluorine, covalently bound to uranium and determined using instrumental methods, i.e., X-ray phase analysis and FTIR spectroscopy, varied from 0.1 to 3%.

The liquid phase of reaction products (3) was hydrofluoric acid, with the hydrogen fluoride concentration of more than 90 wt%.

The degree of fluorine recovery from DUHF was calculated by the formula

$$R_F = \left(1 - \frac{1}{6}\sum_i C_{UF_i} N_{UF_i}\right) \cdot 100\%, \qquad (9)$$

where C_{UF} is the molar concentration of uranium fluorides and oxyfluorides, N_{UF} is the number of fluorine atoms in uranium fluorides and oxyfluorides.

The value of the parameter R_F reached 99%.

Table 2 shows the results of experiments determining the concentration of covalently bound fluorine, the degree of fluorine recovery and the concentration of hydrogen fluoride in hydrofluoric acid, depending on the flow rate of DUHF and the molar ratio of the initial components.

Table 3 gives the composition of solid products resulting from processing DUHF, the amount of hydrogen fluoride, water and hydrogen per 1 mole of DUHF; all values were calculated from the solid composition.

It can be concluded from the experimental results in Tables 1 and 2 that using the proposed method for processing DUHF makes it possible to recover more than 95% of fluorine from DUHF in the form of hydrogen fluoride with a slight excess of oxygen and hydrogen relative to the stoichiometry of process (3). The concentration of the obtained hydrofluoric acid (an aqueous solution of hydrogen fluoride) exceeds 90%. The main uranium-containing substances are uranium oxides, and the uranium dioxide content typically exceeds 90%. Impurity

Table 2

Dependence of concentration of fluorine-containing compounds and degree of fluorine recovery on flow rate of DUHF and molar ratio of initial components

#	W_{UF6}, g/s	Ratio of initial components	C_F, wt%	C_{HF}, wt%	R_F,%
1	3.0	$UF_6 + 3.4H_2 + 1.2O_2$	3.28	94.9	87.9
2	4.4	$UF_6 + 3.1H_2 + 1.0O_2$	1.86	96.3	92.0
3	3.0	$UF_6 + 3.3H_2 + 1.2O_2$	1.35	94.6	93.8
4	6.0	$UF_6 + 3.4H_2 + 1.0O_2$	1.18	100.0	95.3
5	3.0	$UF_6 + 3.4H_2 + 1.3O_2$	1.17	96.2	95.3
6	7.0	$UF_6 + 3.4H_2 + 1.2O_2$	0.98	94.2	95.8
7	6.0	$UF_6 + 3.4H_2 + 1.0O_2$	0.91	98.2	96.0
8	1.5	$UF_6 + 4.7H_2 + 1.2O_2$	0.12	98.5	99.5

Notations: W_{UF6} is the flow rate of DUHF, C_F is the amount of fluorine covalently bound with uranium, R_F is the degree of fluorine recovery, C_{HF} is the concentration of hydrogen fluoride.

Note. The experimental values of C_F and C_{HF} were obtained on a pilot plant, the values of RF were calculated by Eq (9).

Table 3

Composition of solid and gaseous products of DUHF processing

#	W_{UF6}, g/s	Chemical composition of solid products, wt%								Gas, mol		
		UO_2	U_3O_8	U_3O_7	UO_3	UO_2F_2	UOF_4	UF_4	UF_5	HF	H_2O	H_2
1	3.0	56.3	5.0	–	18.2	14.1	–	6.4	–	5.27	0.32	0.44
2	4.4	92.3	–	–	–	–	–	7.7	–	5.52	0.24	0.10
3	3.0	79.5	14.6	–	–	–	5.9	–	–	5.63	0.37	0.12
4	6.0	59.3	12.2	23.6	–	–	–	4.9	–	5.72	–	0.54
5	3.0	-	43.5	51.6	–	–	–	4.9	–	5.75	0.26	0.27
6	7.0	86.3	9.4	–	0.8	–	–	–	3.4	5.75	0.4	0.13
7	6.0	96.2	–	–	–	–	–	3.8	–	5.71	0.12	0.55
8	1.5	60.2	39.2	–	–	0.3	0.3	–	–	5.97	0.10	1.62

Notes. 1. All experimental data were obtained on a pilot plant. 2. The amounts of HF, H_2O and H_2 per 1 mole of DUHF were calculated from the composition of solid products.

uranium fluorides are mainly uranyl fluoride and uranium tetrafluoride whose concentrations vary from several percent to tenths of a percent.

Comparing the data in Table 1 with the data in Tables 2 and 3, we can conclude that the results of experiments on the composition of the products resulting from process (3) are in good agreement with the calculation results for the thermodynamically stable composition of substances in the system of U-F-H-O elements.

Conclusion

Thermodynamic calculations have confirmed that the it is expedient to carry out hydrolysis of UF_6 aimed at obtaining uranium oxides and hydrogen fluoride in the absence of excess water at temperatures above 1150 K.

Furthermore, the calculations indicate that the main fluorine-containing substance in the system of U-F-H-O elements in a mixture in thermodynamic equilibrium at temperatures above 1150 K is hydrogen fluoride, and the main uranium-containing one is uranium dioxide.

Based on the experimental data obtained on the setup, we have found that the flame of UF_6-$3H_2$-O_2 is reliably generated and burns steadily in a 'tunnel burner' type reactor (its inner diameter is 142 mm) with combustion initiated by a fluorine-ammonia mixture in the entire range of DUHF flow rates considered (from 1 to 7 g/s).

The main uranium-containing products of the process given an excess of fuel and oxidizer are uranium oxides (about 10%), primarily in the form of uranium dioxide and triuranium octaoxide, while the UF_4 and UO_2F_2 contents in uranium oxides vary from several percent to several tenths of a percent.

The main fluorine-containing substance is hydrogen fluoride, the concentration of the obtained hydrofluoric acid is about 95%.

The degree of fluorine recovery from DUHF reaches 99%.

The results of the experiments on finding the composition of the products resulting from processing DUHF in a hydrogen-oxygen flame are in good agreement with the calculated results of a thermodynamically stable composition of substances in the system of U-F-H-O elements.

Thus, production of hydrogen fluoride from DUHF can be recommended as the basis for an industrial method.

This study was supported by the Ministry of Science and Higher Education of the Russian Federation (grant No. 05.608.21.0277, project ID RFMEFI60819X0277).

REFERENCES

1. **Orekhov V.T., Rybakov A.G., Shatalov V.V.,** Ispolzovaniye obednennogo geksaftorida urana v organicheskom sinteze [Using the depleted uranium hexafluoride in the organic synthesis], Energoatomizdat, Moscow, 2007 (in Russian).

2. **Shatalov V., Seredenko V., Kalmakov D., et al.,** Depleted uranium hexafluoride − the fluorine source for production of the inorganic and organic compounds, Journal of Fluorine Chemistry. 130 (1) (2009) 122−126.

3. **Evdokimov A.N., Gromov O.B., Vdovichenko V.D., et al.,** Uranium hexafluoride reduction with hydrogen in a "Minimodule" unit, Khimicheskaya Technologiya. 10 (2) (2009) 118−125 (in Russian).

4. **Tumanov Yu.N.,** Plazmennyye protsessy v YaTTS. [Plasma processing in the NFC], Fizmatlit, Moscow, 2003 (in Russian).

5. **Borgard J.M., Herbelet F., Gwinner B.,** Recycling hydrofluoric acid in the nuclear industry: The OverAzeotropic Flash process (OVAF), Journal of Fluorine Chemistry. 185 (May) (2016) 17−23.

6. **Morel B., Duperret B.,** Uranium and fluorine cycles in the nuclear industry, Journal of Fluorine Chemistry. 130 (1) (2009) 7−10.

7. **Nikipelova N.V.,** Programma Goskorporatsii "Rosatom" po bezopasnomu obrashcheniyu s OGFU [A program of "The Rosatom" state corporation on safe DUHF waste management (Report)], Fuel Company of "The Rosatom", Moscow, 2019 (in Russian).

8. **Turayev N.S., Zherin I.I.,** Khimiya i tekhnologiya urana [Chemistry and technology of uranium], Atominform, Moscow, 2005 (in Russian).

9. **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.

10. **Mishchenko K.P., Ravdelya A.A.,** Kratkiy spravochnik fiziko-khimicheskikh velichin [Reference book on physical and chemistry quantities], Chemistry Publishing, Leningrad, 1974 (in Russian).

11. **Glushko V.P., Gurvitch L.V., Veits I.V., et al.,** Termodinamicheskiye svoystva individualnykh veshchestv [Thermodynamic properties of the individual substances], Reference Edition (the 3rd) in 6 Vols. Nauka, Moscow, 1978−2004. (in Russian). Vols. 1−4. 1978−1982 (in Russian).

12. **Trusov B.G.,** Programmnaya sistema modelirovaniya fazovykh i khimicheskikh ravnovesiy pri vysokikh temperaturakh [Code system for simulation of phase and chemical equilibriums at higher temperatures], Herald of the Bauman Moscow State Technical University, Series Instrument Engineering, Special Iss. 2: Software Engineering (2012) 240−249 (in Russian).

Received 23.09.2020, accepted 06.10.2020.

THE AUTHORS

PASHKEVICH Dmitrii S.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
pashkevich-ds@yandex.ru

ZIMIN Arsenii R.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
fz1min@yandex.ru

ALEXEEV Yuri I.
LLC «New chemical products»
11, A, Moyka River Emb., St. Petersburg, 191186, Russian Federation
alexeev-588@yandex.ru

MUKHORTOV Dmitrii A.
LLC «New chemical products»
11, A, Moyka River Emb., St. Petersburg, 191186, Russian Federation
dmukhortov@mail.ru

KAMBUR Pavel S.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
pavel.kambur@mail.ru

PETROV Valentin B.
Russian Scientific Center «Applied Chemistry»
26, Bolsheviki Ave., St. Petersburg, 193232, Russian Federation
valentinpetrov46@mail.ru

BAZHENOV Dmitrii A.
JSC «Siberian Chemical Plant»
1 Kurchatov St., Seversk, Tomsk Region, 636039, Russian Federation
dmitriy_bazhenov@mail.ru

SMOLKIN Pavel A.
JSC «Siberian Chemical Plant»
1 Kurchatov St., Seversk, Tomsk Region, 636039, Russian Federation
spa-net@mail.ru

KAPUSTIN Valentin V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
valentin.kapustin.2014@yandex.ru

DOI: 10.18721/JPM.13413

THE STABILITY OF FLAME FRONT AND THE COMPOSITION OF COMBUSTION PRODUCTS OF AMMONIA IN OXYGEN IN THE PRESENCE OF HYDROGEN FLUORIDE

D.S. Pashkevich[1], D.A. Mukhortov[2], V.V. Kapustin[1], P.S. Kambur[1]
E.S. Kurapova[3], A.R Zimin[1], A.S. Maslova[1]

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

[2] LLC «New Chemical Products», St. Petersburg, Russian Federation;

[3] Russian Scientific Center «Applied Chemistry», St. Petersburg, Russian Federation

The authors have proposed a way for producing hydrogen fluoride (the major fluorinating agent in the modern industry) from ammonium bifluoride (by-product in the production of extraction phosphoric acid, tantalum, niobium, etc.) by burning up $NH_4F{\cdot}HF$ in oxygen. Calculations performed via program complex ASTRA.4 showed that the only fluorine-containing substance was hydrogen fluoride, whereas nitric fluorides and oxyfluorides being absent under the ratio of chemical elements N-1.0; H-5.0; F-2.0; O-1.5 (according to the proposed process), in the temperature range 600 − 2500 K, in thermodynamically equilibrium mixture. The combustion of ammonia in oxygen in the presence of hydrogen fluoride was experimentally investigated and it was found the feasibility of forming a stable flame, where a molar ratio of ammonia/hydrogen fluoride consumptions was 1: 2, only by using methane as a fuel. The conditions for the ratio of component consumptions were determined.

Keywords: ammonium bifluoride, ammonia, hydrogen fluoride, combustion, tunnel burner reactor

Citation: Pashkevich D.S., Mukhortov D.A., Kapustin V.V., Kambur P.S., Kurapova E.S., Zimin A.R., Maslova A.S., The stability of flame front and the composition of combustion products of ammonia in oxygen in the presence of hydrogen fluoride, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 119–128. DOI: 10.18721/JPM.13413

Introduction

Hydrogen fluoride (HF) serves as the main fluorinating agent in the modern industry producing fluorine compounds. It is used to obtain uranium fluorides in the nuclear fuel cycle, fluoropolymers, ozone-safe freons, electronic gases, etc. [1]. The primary industrial method for producing hydrogen fluoride is sulfuric acid decomposition of natural calcium fluoride concentrate (CaF_2), or fluorspar [2]. The reserves of fluorspar ores, which are a source for obtaining a concentrate with the required composition, are limited. Therefore, developing methods for producing hydrogen fluoride from alternative raw materials is an important task.

Ammonium bifluoride $NH_4F{\cdot}HF$ (ABF) is a by-product of the production of tantalum, niobium and nitrogen trifluoride [3]. In addition, ABF can be obtained from aqueous solutions of hexafluorosilicic acid (H_2SiF_6), which is a by-product of extraction of phosphoric acid from fluorapatite ($Ca_{10}(PO_4)_6F_2$) [4, 5]. The market for H_2SiF_6 is rather small, so it is neutralized and fluorinated waste is disposed of at industrial landfills. The world reserves of fluorapatite (in terms of calcium fluoride content) exceed the reserves of fluorspar by more than 10 times [2, 6]. Therefore, developing science-based approaches to obtaining hydrogen fluoride from ABF will solve two problems: the environmental one, associated with processing fluorine-containing wastes, and the commercial one, associated with scarce reserves and rising prices of natural CaF_2 [7].

While various methods were used to produce hydrogen fluoride from ABF, in particular, sulfuric acid decomposition [8] and catalytic conversion [9], they are not applied industrially.

Hydrogen fluoride can be obtained from ABF via combustion in oxygen:

$$NH_4F \cdot HF_{liq} + 0.75O_{2\,gas} \rightarrow \rightarrow 0.5N_{2\,gas} + 1.5H_2O_{gas} + + 2.0HF_{gas} - 294\text{kJ}. \quad (1)$$

The subscripts here correspond to the aggregation states of substances: solid (*sol*), liquid (*liq*) and gaseous (*gas*).

The temperature of the products of process (1) can reach about 2400 K. It is known that hydrogen fluoride is one of the most thermally stable fluorides [10].

Combustion of ammonia in oxygen and air is a well-understood phenomenon: the concentration limits of flame propagation are 13–82 and 16–26 vol.%, and the standard velocities are 30–140 and 30–74 $cm \cdot s^{-1}$, respectively [11].

However, we were unable to uncover any reports on the effect of hydrogen fluoride on the interaction of ammonia with oxygen. It is known that hydrogen fluoride is a polar substance (its dipole moment is $5.78 \cdot 10^{-30}$ C·m [12]) that can have both catalytic and inhibitory effects on chemical reactions [13].

The goal of our study was, firstly, to calculate the thermodynamically stable compositions in the systems of N-H-F-O and C-N-H-F-O elements, and, secondly, to analyze the composition of the products resulting from the interaction of ammonia with oxygen in the combustion mode, in the presence of hydrogen fluoride with methane serving as a fuel.

Experimental methods

Calculations of the thermodynamically stable composition in the system of N-H-F-O and C-N-H-F-O elements were carried out using the ASTRA.4 software package [14]. The experimental results previously obtained on the laboratory setup constructed for a carbon-containing system indicate that a stable flame can be produced for combustion of ammonia in oxygen in the presence of hydrogen fluoride only if methane is supplied as a fuel to the reaction zone.

The laboratory setup for experimental studies of combustion of ammonia in oxygen in the presence of hydrogen fluoride is shown in Fig. 1. Vertical reactor *1* of the 'tunnel burner' type was a steel pipe with an inner diameter $d_1 = 41$ mm and a height of 500 mm. A cable connected to a thermostat was attached to the outer surface of the reactor to heat its wall.

Gases were fed into the reactor through a two-component nozzle assembly *2*installed in the lower part of the reactor.

A separate scheme of reactor *1* and assembly *2* is shown in Fig. 2. The nozzle assembly consists of two coaxial channels: axial (outer diameter 3 mm, inner diameter 2 mm) and annular (outer diameter $d_2 = 8$ mm), with a mixing zone (mixer) with a diameter of 8 mm and variable length *h*.

The gas flows and the nozzle were heated to a temperature of about 140 °C so that solid ABF would not form upon contact of ammonia with hydrogen fluoride.

The generation and propagation of the flame front was detected using chromel-alumel thermocouples t_1 and t_2 with identical electrode diameters of 200 μm, installed without protective covers so that the thermocouple junctions were

Fig. 1. Schematic of laboratory setup:
'tunnel burner' type reactor *1*; nozzle assembly *2*; polypropylene absorber *3*;
steel container *4* with pressure sensor for HF and O_2; steel container *5* with pressure sensor for NH_3;
gas flow regulator *6* for HF and O_2 supply lines; gas flow regulator *7* for NH_3 supply line

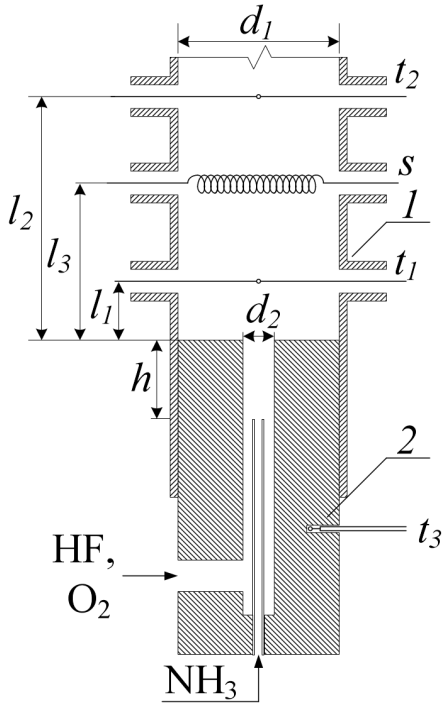


Fig. 2. Schematic of a 'tunnel burner' type reactor and nozzle assembly of the reactor: reactor *1*; nozzle assembly *2*; coiled filament *s*; t_1, t_2 are the chromel-alumel thermocouples for measuring the temperature in the reactor; t_3 is the chromel-copel thermocouple for measuring the temperature in the nozzle assembly; h, l_i, d_i are the geometric parameters

on the reactor axis. Thermocouple t_1 was located at a distance l_1 = 15 mm from the nozzle exit, and t_2 at a distance l_2 = 62 mm from the exit (see Fig. 2).

The chromel-copel thermocouple $t_3$was inserted into nozzle assembly *2* to measure its temperature. Combustion process was initiated by coiled filament *s* made of nichrome wire 0.3 mm in diameter, placed at a distance l_3 = 40 mm from the nozzle exit.

A mixture of oxygen and hydrogen fluoride from the container *4* (see Fig. 1) was fed into the reactor through the annular channel of nozzle assembly *2*, and ammonia was fed through the axial channel. The temperature of the mixture components supplied to the nozzle assembly was 25 °C. To prevent ammonium fluoride NH_4F from forming on the inner wall of the reactor upon contact of ammonia with hydrogen fluoride, the reactor wall was thermostated at 300 ° C.

After the reactor, the products were fed through a heated line into polypropylene vessel *3* filled with water. When the experiment ended, the contents of hydrogen fluoride and ammonia in the liquid from vessel *3* were analyzed by titration.

Several gas samples were collected the outlet of absorber *3* and analyzed by gas chromatography and infrared Fourier spectroscopy (FTIR). The composition of the gas phase was determined using a Crystal-2000M chromatograph with a thermal conductivity detector. Two packed columns 3 m long each with Porapak Q 80/120 adsorbents were used for detecting hydrogen, nitrogen, oxygen, carbon dioxide, carbon tetrafluoride, methane and nitrogen trifluoride; CaA of 0.25-0.40 fraction was used to detect hydrogen, oxygen, nitrogen and carbon monoxide.

FTIR was used to determine the concentration of carbonyl fluoride COF_2 in gaseous products of the process. A Shimadzu FTIR-8400S spectrometer with a resolution of 0.85 cm^{-1} and a cell size of 200 × 230 × 170 mm was used for this purpose.

The efficiency of the method for producing hydrogen fluoride from ammonium bifluoride in combustion mode is characterized by the conversion of ammonia. This value was determined by the material balance; the basis for the calculations were the flow rates of the initial components, as well as the composition of the liquid in absorber *3* after the experiment, K_{liq} and the composition of the gaseous reaction products K_{gas}:

$$K_{liq} = 100\% \cdot \frac{m(\mathrm{NH}_{3gas})_{out}}{m(\mathrm{NH}_{3gas})_{in}}, \qquad (2)$$

$$K_{gas} = 100\% \cdot \frac{W(\mathrm{N}_{2gas})_{out}}{W(\mathrm{NH}_{3gas})_{in}}, \qquad (3)$$

where $m(NH_3)_{out}$, $m(NH_3)_{in}$, mg, are the masses of ammonia collected in absorber *3* and fed into the reactor, respectively; $W(N_2)_{out}$, $cm^3 \cdot s^{-1}$, is the volumetric flow rate of nitrogen at the reactor outlet under normal conditions; $W(NH_3)_{in}$, $cm^3 \cdot s^{-1}$, is the volumetric flow rate of ammonia at the reactor inlet under normal conditions.

Thermodynamic calculations in the system of N-H-F-O elements

Aside from the target product (HF), a large number of fluorine-containing compounds are known in the system of N-H-F-O elements: NF_3, N_2F_2, N_2F_4, NH_4F, NHF_2, $(NO)F$, $(NO_2)F$, NOF_3, etc. [15]. Calculations of

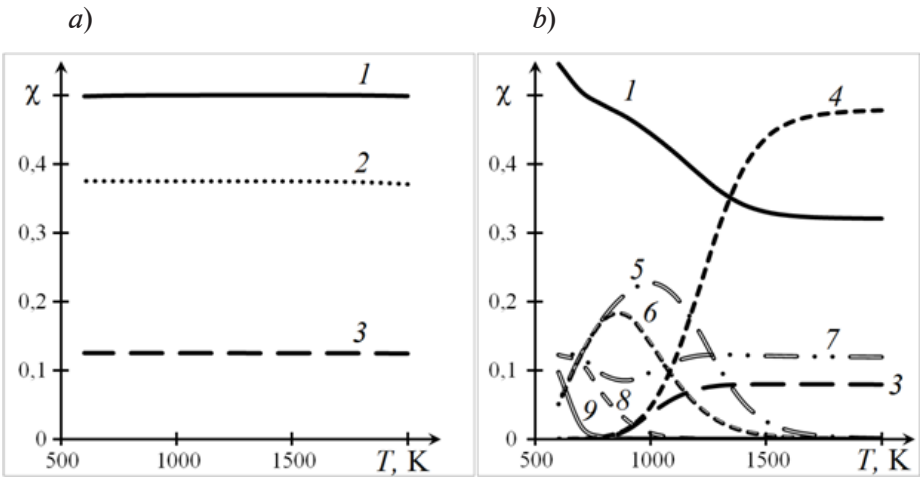


Fig. 3. Temperature dependences of thermodynamically equilibrated composition of substances (in molar fractions) in the system of N-H-F-O elements at various N-H-F-O ratios: 1.0:5,0:2.0:1.5 (*a*) and 1.0:2.0:5.0:1.5 (*b*); $n_F > n_H$.
The data are given for HF (*1*), H_2O (*2*), N_2 (*3*), F (*4*), F_2 (*5*), (NO)F (*6*), O_2 (*7*), NF_3 (*8*), NOF_3 (*9*)

the thermodynamically stable composition for the system of N-H-F-O elements at a pressure P = 0.1 MPa and in the temperature range 600–2000 K established that the thermodynamically stable composition of the reaction products depends on the atomic ratio between hydrogen and fluorine. If the number of fluorine atoms n_F does not exceed the number of hydrogen atoms n_H, then the only fluorine-containing substance in the equilibrated mixture is hydrogen fluoride, with no fluorides or nitrogen oxyfluorides (Fig. 3,*a*). If n_F exceeds n_H, then other fluorine-containing substances besides hydrogen fluoride are present in the thermodynamically stable mixture: NF_3, (NO)F, NOF_3, F and F_2 (Fig. 3,*b*).

Given a ratio of chemical elements corresponding to Eq. (1), hydrogen fluoride is the only fluorine-containing substance in the thermodynamically equilibrated mixture in the temperature range of 600–2000 K and at a pressure P = 0.1 MPa.

We can conclude from the results of thermodynamic calculations that further experimental studies should be conducted on the stability of interaction of ammonia with oxygen in the presence of hydrogen fluoride in combustion mode.

Results of experiments in the system of NH_3-HF-O_2 elements

The purpose of the laboratory study was to confirm the results of thermodynamic analysis on the composition of substances in the system of N-H-F-O elements, as well as to determine the effect of hydrogen fluoride on the combustion of ammonia and oxygen.

Hydrogen fluoride was replaced with argon at the first stage of the experiments; this was aimed at determining the conversion of ammonia in combustion mode under standard conditions of the laboratory setup, and, furthermore, at considering the effect that the length h of the mixing zone had on the conversion of ammonia (Table 1).

A stable flame front could be generated at the given concentrations of the initial substances and the flow velocities. The flame front remained in the reactor and did not penetrate into the mixer during the experiments. The thermal diffusivity coefficients are close for gaseous nitrogen and argon (the difference is 3.6% under normal conditions [16]), so it can be assumed that the concentration limits of combustion only change insignificantly if N_2 in the NH_3-O_2-N_2 system is replaced with an equimolar amount of Ar [17]. Therefore, it can be argued that the ammonia concentration was within the concentration limits of combustion for the NH_3+O_2+Ar mixture (Fig. 4). The linear velocity of the gas in the reactor was significantly lower than the normal velocity of flame propagation in the NH_3+O_2+N_2 (Ar) mixture. The degree of ammonia conversion depended on the length of the mixing zone, and the ratio of the component flow rates and, reaching over 95%.

When argon was replaced with an equimolar amount of hydrogen fluoride (corresponding to its content in ABF), a stable flame front could not be produced for component flow rates in the range of 37–64 $cm^3 \cdot s^{-1}$ withthe length of

Table 1

Composition of products of interaction of NH_3 with O_2 in the presence of Ar in combustion mode (excluding unreacted NH_3) and conversion of NH_3 depending on the flow rate of initial components and mixer length

h, mm	Initial mixture, mol			W, cm^3/s	V_R, cm/s	V_M, cm/s	Composition of reaction products, mol				K_{gas}, %	K_{liq}, %
	NH_3	O_2	Ar				N_2	Ar	H_2	O_2		
0	1,00	1.06	2.03	32.0	2.4	–	0.23	2.03	0.0	0.69	47	56
35		1.20		20.0	1.5	39.5	0.49			0.44	98	93
50		1.20		20.0	1.5	39.5	0.49			0.44	99	89
		1.03		38.0	2.9	76.3	0.42			0.24	84	88
		0.96		63.0	4.8	124.8	0.41			0.17	83	87

Notations: h is the length of the mixing zone; W is the total volumetric flow rate of components at flow temperature t = 140 °C; V_R, V_M are the flow rates in the reactor and in the mixing zone (respectively) at a temperature t = 140°C; K_{gas}, K_{liq} are the ammonia conversion rates, determined from the gaseous reaction products and the composition of the liquid in the absorber, respectively.

the mixing zone varying from 0 to 40 mm. The linear gas velocity in the reactor varied in the range of 2.9–4.9 $cm \cdot s^{-1}$ and was lower than the normal flame propagation velocity in the $NH_3+O_2+N_2$ mixture. The concentration of ammonia in the experiments was within the concentration limits of flame propagation in the $NH_3+O_2+N_2$ mixture (see Fig. 4).

The absence of a stable flame front may be due to the inhibitory effect of hydrogen fluoride on the rate of ammonia oxidation or to the formation of liquid ABF upon contact of ammonia with hydrogen fluoride in a mixer.

To generate a stable flame front of ammonia and oxygen in the presence of hydrogen fluoride, fuels such as methane can be added to the initial system and oxygen consumption can be increased for complete oxidation of methane.

Our further goal was to determine the minimum amount of methane at which the flame front would be stable and ammonia conversion would be close to 100%.

Thermodynamic calculations in the system of N-H-F-O elements

Calculations of the thermodynamic equilibrated composition for the system of N-H-F-O-C elements at a pressure P = 0.1 MPa in the system and in the temperature range of 600–2000 K confirmed that the only fluorine-containing substance at $n_H \geq n_F$ was hydrogen fluoride, the same as for the system of N-H-F-O

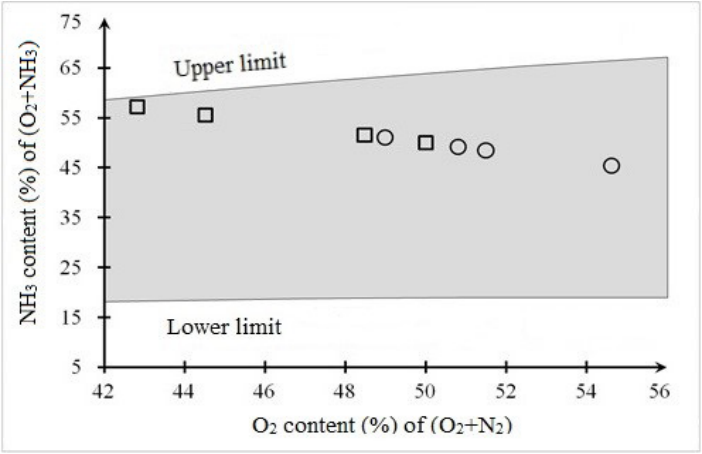


Fig. 4. Propagation region of NH_3-O_2 flame (shaded trapezoid), its concentration limits (upper and lower border lines) and O_2/NH_3 ratios in the experiments; the data are given for experiments with Ar (circles) and HF (squares)

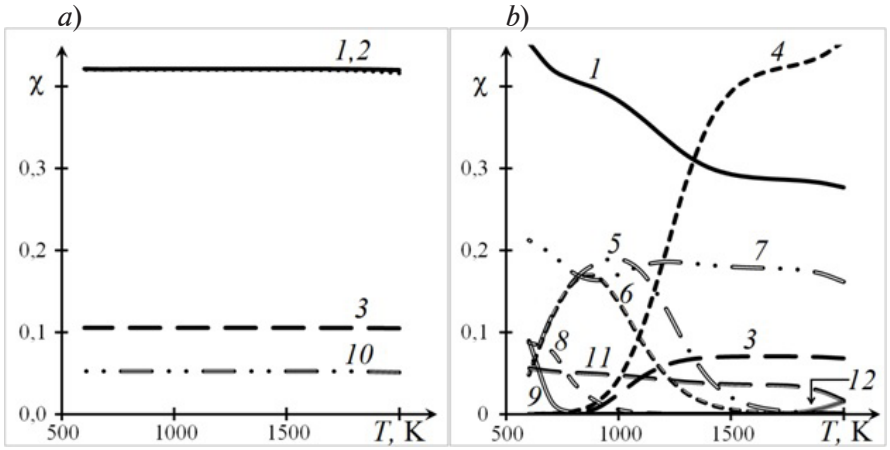


Fig. 5. Temperature dependences of thermodynamically equilibrated composition of substances (in molar fractions) in the system of N-H-F-O-C elements at various N-H-F-O-C ratios: 1.00:6.00:2.00:2.50:0.25 (*a*) and 1.00:2.00:6.00:2.50:0.25 (*b*); $n_F > n_H$.
The data are given for HF (*1*), H_2O (*2*), N_2 (*3*), F (*4*), F_2 (*5*), (NO)F (*6*), O_2 (*7*), NF_3 (*8*), NOF_3 (*9*), CO_2 (*10*), CF_4 (*11*), COF_2 (*12*)

elements, (Fig. 5,*a*). If $n_H <; n_F$, then other fluorine-containing substances besides hydrogen fluoride are present in the thermodynamically equilibrated mixture: NF_3, (NO)F, NOF_3, F, F_2, CF_4 и COF_2 (Fig. 5,*b*).

Results of experiments in the NH_3-HF-O_2-CH_4 system

Methane was used to stabilize the flame front in the NH_3-HF-O_2 system, with a mixture of ammonia and methane supplied to the reaction zone. The oxygen content in the system corresponded to the minimum amount necessary for complete combustion of ammonia producing nitrogen and water vapors, and for complete combustion of methane producing water vapors and carbon dioxide. To prevent liquid ABF from forming upon contact of ammonia with the hydrogen fluoride in the mixer, the experiments were carried out at h = 0 mm. The results of experiments are given in Table 2.

It was established that a stable flame front could be formed successfully at the concentrations of methane and oxygen in the initial mixture above 5.6 and 27.8 vol.%, respectively. Neither methane nor carbon

Table 2

Composition of reaction products of NH_3, O_2 and CH_4 in the presence of HF in combustion mode and NH_3 conversion depending on the flow rates of the initial components

Initial mixture, mol				W, cm³/s	V_R, cm/s	V_{RC}, cm/s	V_{AC}, cm/s	Composition of reaction products, vol.%				K_{gas}, %	K_{liq}, %
NH_3	HF	O_2	CH_4					H_2	N_2	O_2	CO_2		
1,00	2.00	0.97	0.11	51.0	3.8	85.5	439.4	Impossible to form a stable flame front					
		1.25	0.25	50.0	3.7	82.7	437.5	0	0.50	0.01	0.25	95	97
		1.65	0.42	52.0	4.0	87.2	466.2		0.50	0.07	0.42	98	98
		1.59	0.42	53.0	4.0	88.4	480.7	0.03	0.50	0.02	0.42	96	97
	1.99	1.70	0.48	52.0	4.0	82.6	475.2	0	0.49	0.01	0.48	98	98

Notations: W is the total volumetric flow rate of components at the flow temperature t = 140 °C; V_R, V_{RC}, V_{AC} are the flow rates in the reactor, annular channel and axial channel (respectively) at a temperature t = 140 °C; K_{gas}, K_{liq} are the ammonia conversion rates, determined from the composition of gaseous reaction products and the composition of the liquid in the absorber after the experiment, respectively.

monoxide were found in the reaction products (within to 10^{-2}%). Conversion of ammonia in combustion mode reached 98% and practically did not depend on the concentration of methane in the mixture. The minimum contents of methane and oxygen in the mixture, at which it was possible to carry out the reaction in combustion mode correspond to the equation:

$$NH_{3gas} + 2.0HF_{gas} + 1,25O_{2gas} + \\ + 0,25CH_{4gas} \rightarrow 0.5N_{2gas} + 2.0H_2O_{gas} + \\ + 2.0HF_{gas} + 0.25CO_{2gas}. \quad (4)$$

To reduce the flow rate of methane and the water content in the products of process (4), gaseous fluorine F_2 can be used to initiate reaction chains, stabilizing combustion. Elemental fluorine is known to have a low dissociation energy equal to $\Delta H_{dis.} = 157.5$ kJ·mol^{-1}, and the presence of fluorine radicals in the reacting mixture can stabilize the flame front [1]. Therefore, gaseous fluorine mixed with oxygen and hydrogen fluoride was fed into the reaction zone in an amount corresponding to a concentration of 0.5–1.3 vol.% in a NH_3+HF+O_2+CH_4+F_2 mixture. However, a stable flame front could not be formed in the presence of gaseous fluorine without methane (Table 3). The flame front was stable in the presence of methane in an amount corresponding to a concentration of 2.7 vol.% and gaseous fluorine in an amount corresponding to a concentration of 0.5%; ammonia conversion reached 98% in this mode. Thus, using gaseous fluorine allows to reduce the concentration of methane at which the flame front is stable.

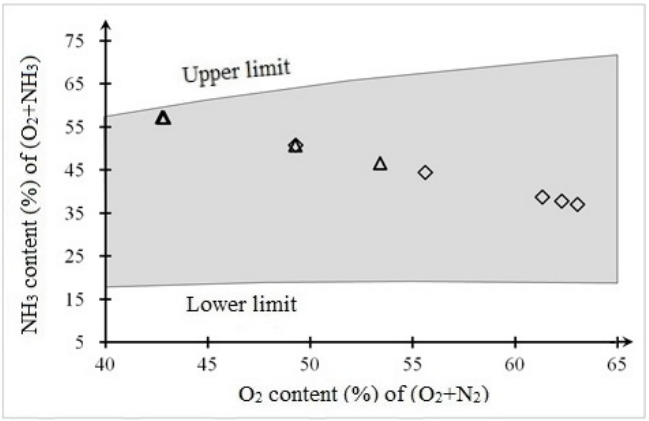


Fig. 6. Propagation region of NH_3-O_2 flame (shaded trapezoid), its concentration limits (upper and lower border lines) and O_2/NH_3 ratios in the experiments; the data are given for experiments with CH_4 (triangles) and F_2 (diamonds)

Table 3

Composition of reaction products of NH_3, O_2, CH_4 and F_2 in the presence of HF in combustion mode and NH_3 conversion depending on the flow rates of initial components

Initial mixture, vol.%					W, cm^3/s	V_R, cm/s	V_{RC}, cm/s	V_{AC}, cm/s	Composition of reaction products, vol.%			K_{gas}, %	K_{liq}, %
NH_3	HF	O_2	CH_4	F_2					N_2	O_2	CO_2		
1.00	2.00	0.97	0.11	0.02	50.0	3.8	84.2	430.0	0.49	0.03	0.11	97	98
		0.75	–	0.03	51.0	3.8	86.4	427.3	Impossible to form a stable flame front				
		0.75	–	0.04	50.0	3.8	85.0	419.1					
		0.75	–	0.05	50.0	3.8	84.9	416.8					
		1.15	–	0.04	47.0	3.6	82.7	356.6					

The notations are identical to those given in Table 2.

Flame propagation was possible in the experiments carried out, both in a mixture of ammonia with oxygen and in a mixture of methane with oxygen. The concentration of ammonia was within the concentration limits of flame propagation of ammonia and oxygen (Fig. 6) in all experiments, and the concentration of methane was within the concentration limits of flame propagation of methane and oxygen [18]. The normal velocity of flame propagation in the CH_4+O_2 mixture was 330 cm·s^{-1}, which is 2.4 times higher than that for the NH_3+O_2 mixture. Therefore, it can be assumed that methane oxidation is the main stage of wave process (4) [19].

Main results and conclusions

Calculations via the ASTRA.4 software package have confirmed that hydrogen fluoride is the only fluorinated substance in a thermodynamically stable mixture in the system of N-H-F-O elements in the temperature range from 600 to 2000 K , given a ratio of chemical elements corresponding to Eq. (1), , while other nitrogen fluorides and oxyfluorides are absent.

It was established in experiments with a laboratory reactor of the 'tunnel burner' type and a nozzle assembly that it was impossible to generate a stable combustion front of ammonia in oxygen with hydrogen fluoride present in the reaction zone, with ammonia concentrations varying from 25.0 to 27.5%, hydrogen fluoride from 50.0 to 55.1% and oxygen from 17.4 to 25.0%, i.e., when the molar ratio is NH_3-2HF-O_2 and the concentration of ammonia is within the concentration limits of flame propagation, if hydrogen fluoride is replaced by nitrogen.

For methane used in a mixture with ammonia, when the concentration of methane in a four-component mixture CH_4-NH_3-HF-O_2 was not less than 5.6%, the molar ratio NH_3-H corresponded to$NH_4F{\cdot}HF$, and the amount of oxygen was sufficient for complete oxidation of both ammonia and methane:

$$NH_3 + 2.0\ HF + 1.25\ O_2 + 0.25\ CH_4.$$

A stable combustion front could be maintained under these conditions. The conversion of ammonia in combustion mode reached 98%.

Using fluorine in the amount corresponding to about 1% concentration in the mixture did not allow to generate a stable combustion flame of ammonia in oxygen in presence of hydrogen fluoride. Using fluorine in the amount corresponding to a concentration of about 0.5% made it possible to reduce the concentration of methane providing a stable flame front to 2.7%.

The method proposed can be used as a basis for obtaining hydrogen fluoride from ammonium fluorides, which are by-products in some industries.

This study was supported by the Ministry of Science and Higher Education of the Russian Federation (grant No. 05.608.21.0277, project ID RFMEFI60819X0277).

REFERENCES

1. **Maksimov B.N., Barabanov V.G., Serushkin I.L. et al.,** Promyshlennye ftororganicheskie produkty. Spravochnik [Industrial fluoroorganic products. Handbook], Khimiya, Leningrad, 1990.

2. **Tumanov V.V., Ostrovskij S.V., Starostin A.G.,** The mineral resource fluorine base of Russia and other countries, Vestnik of Perm National Research Polytechnic University, Chemical Technology & Biotechnology. (4) (2015) 124–136.

3. **Pashkevich D.S., Mukhortov D.A., Alexeev Yu.I., et al.,** New aspects of waste-free technology for producing high-purity nitrogen trifluoride by fluorination of a melt of acidic ammonium bifluoride with fluorine, Proceedings of 2nd International Siberian Seminar "Modern Inorganic Fluoride" (Intersibfluorine-2006), June 11–16, 2006, Tomsk, TPU (2006) 222–225.

4. **Dahlke T., Ruffiner O., Cant R.,** Production of HF from H_2SiF_6, Procedia Engineering. 138 (2016) 231–239.

5. **Pashkevich D.S., Mukhortov D.A., Kambur P.S., et al.,** Method of obtaining hydrogen fluoride, Pat. No. 2669838, Russian Federation, MPK C 01 B 7/19; SPb., "New Chemical Product" LLC is a declarant and a patentee, No. 2018101617; declar. 16.01.18; publ. 16.10.18, Bull. No. 29, 9 p.

6. Mineral commodity summaries 2019: U.S. Geological Survey, United States Government Printing Office (2019).

7. **Pashkevich D.S., MamaevA.V.,** Production of hydrogen fluoride by processing fluorine-containing wastes and by-products of modern industries, WIT Transactions on Ecology and the Environment, 231 (2019) 111–123.

8. **Rodin V.I., Kharitonov V.P., Seredenko V.A.,** Hydrogen fluoride technology by sulfuric acid decomposition of ammonium bifluoride, Russian Chemical Industry. 76 (12) (1999) 753–757 (in Russian).

9. **Kurapova E., Mukhortov D., Pashkevich D., et al.,** Production of hydrogen fluoride using fluorine-containing waste of manufacturing phosphate fertilizers, Chimica Oggi – Chemistry Today. 35 (3) (2017) 24–26.

10. **Pashkevich D.S.,** Regeneration of fluorine in the form of hydrogen fluoride from volatile fluorine-containing wastes in a hydrogen-containing fuel-oxygen-containing oxidant flame, Russian Journal of Physical Chemistry B: Focus on Physics. 13 (6) (2019) 993–1003.

11. **Shebeko Yu.N., Trunev A.V., Tsarichenko S.G., Zaitsev A.A.,** Investigation of concentration limits of flame propagation in ammonia-based gas mixtures, Combustion, Explosion and Shock Waves. 32 (5) (1996) 477–480.

12. **Morrison R., Boyd R.,** Organic chemistry, 2nd Ed., Allyn and Bacon Inc., Boston, 1970.

13. **Pashkevich D.S., Kostyaev R.A.,** Rol' ftorovodoroda i diftorida kobal'ta v protsessakh ftorirovaniya uglevodorodov triftoridom kobal'ta [The role of hydrogen fluoride and cobalt difluoride in the processes of fluorination of hydrocarbons with cobalt trifluoride], Russian Journal of Applied Chemistry. 67 (12) (1994) 2012–2017 (in Russian).

14. **Trusov B.G.,** Programmnaya sistema modelirovaniya fazovykh i khimicheskikh ravnovesiy pri vysokikh temperaturakh [Code system for simulation of phase and chemical equilibriums at higher temperatures], Herald of the Bauman Moscow State Technical University, Series Instrument Engineering, Special Iss. 2: Software Engineering (2012) 240–249 (in Russian).

15. **Pankratov A.V.,** Khimiya ftoridov azota [Chemistry of the nitrogen fluorides], Khimiya, Moscow, 1973 (in Russian).

16. **Grigoriev I.S., Meilikhov E.Z.** (Eds.), Fizicheskie velichiny [Handbook of Physical Quantities], Energoatomizdat, Moscow, 1991 (in Russian).

17. **Snegiryov A.Yu.,** Osnovy teorii goreniya [Principles of combustion theory], Polytechnical University Publishing, St. Petersburg, 2014 (in Russian).

18. **Antonov V.N., Lapidus A.S.,** Proizvodstvo atsetilena [Acetylene production], Khimiya, Moscow, 1970 (in Russian).

19. **Han J.W., Lee C.E., Kum S.M., Hwang Y.S.,** Study on the improvement of chemical reaction mechanism of methane based on the laminar burning velocities in OEC, Energy & Fuels. 21 (6) (2007) 3202–3207.

Received 11.09.2020, accepted 21.09.2020.

THE AUTHORS

PASHKEVICH Dmitrii S.
Peter the Great St. Petersburg Polytechnic University
29, Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
pashkevich-ds@yandex.ru

MUKHORTOV Dmitrii A.
LLC «New chemical products»
11, A, Moyka River Emb., St. Petersburg, 191186, Russian Federation
dmukhortov@mail.ru

KAPUSTIN Valentin V.
Peter the Great St. Petersburg Polytechnic University
29, Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
valentin.kapustin.2014@yandex.ru

KAMBUR Pavel S.
Peter the Great St. Petersburg Polytechnic University
29, Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
pavel.kambur@mail.ru

KURAPOVA Ekaterina S.
Russian Scientific Center «Applied Chemistry»
26, Bolsheviki Ave., St. Petersburg, 193232, Russian Federation
e_kurapova@mail.ru

ZIMIN Arsenii R.
Peter the Great St. Petersburg Polytechnic University
29, Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
fz1min@yandex.ru

MASLOVA Anastasiia S.
Peter the Great St. Petersburg Polytechnic University
29, Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
anastasiiamaslova5@gmail.com

DOI: 10.18721/JPM.13414

MODIFICATION OF GLASS SURFACE BY A HIGH ELECTRIC FIELD

E.S. Babich[1], *I.V. Reduto*[2], *A.V. Redkov*[3], *I.V. Reshetov*[4,1], *V.V. Zhurikhina*[1], *A.A. Lipovskii*[4,1]

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

[2] University of Eastern Finland, Joensuu, Finland;

[3] Institute for Problems of Mechanical Engineering RAS, St. Petersburg, Russian Federation;

[4] Alferov University, St. Petersburg, Russian Federation

The work has studied an effect of a high DC field on the composition and properties of the subsurface region of a multicomponent silicate glass. The concentration of alkali ions in the subsurface glass region was shown to drastically decrease under the high electric field. This led to a change in the ion-exchange characteristics of glasses and their resistance to etching. The effect allows employing the poled regions of the glass surface as dielectric masks in the formation of gradient optical structures, as well as relief microstructures, e.g., grooves for microfluidics, in glass substrates. The advantage of this approach is that there is no need in multiple lithography using liquid chemicals. This reduces the cost and makes this technological process environment friendly.

Keywords: high electric field, chemical etching, ion etching, ion-exchange

Citation: Babich E.S., Reduto I.V., Redkov A.V., Reshetov I.V., Zhurikhina V.V., Lipovskii A.A., Modification of glass surface by a high electric field, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 129–134. DOI: 10.18721/JPM.13414

Introduction

In recent years, glasses have been actively used to fabricate integral elements of waveguide optics and photonics: lenses, multiplexers, modulators, optical amplifiers, as well as microfluidics and lab-on-a-chip devices [1]. In most cases, the change in the optical characteristics of glass is achieved by ion-exchange (thermal or electrostimulated), in which ions from the glass matrix are replaced by ions from the salt melt [2]. Chemical or reactive ion etching is used to change the morphology of the glass surface, for example to form microgrooves. Both the ion-exchange process and the etching process require masks, which are usually formed from dielectric or metallic films deposited on the surface of the glass substrate by lithographic methods. The advantage of dielectric masks is the absence of a contact potential difference [3], so that they do not affect the ion distribution during ion-exchange. However, metal masks are widely used for technological reasons.

This paper presents a new approach to forming a dielectric mask on the glass surface, namely, local modification of its near-surface region by a high electric field. The electric field is applied by means of a relief electrode, which is produced using lithography. The proposed approach has an advantage over the traditional one due to its low cost, since the same electrode can be used multiple times to modify an almost unlimited number of substrates, i.e., to form masks on their surface for ion-exchange or etching.

Experimental

We have considered thermal poling of glass, that is, modification of multicomponent silicate glass by high electric field at elevated temperature with the help of a pressed anode electrode [4]. The process is shown schematically in Fig. 1. A flat (*1*) or comb-shaped (*2*) structure made of glass graphite was used as the anode electrode [5]. After anode *1* (or *2*) and cathode *3* were pressed to glass *4*, it was heated and a DC voltage was applied to it. Patterned electrode *2* was used to form alternating poled (*5*) and unpoled (*6*) regions in the sample.

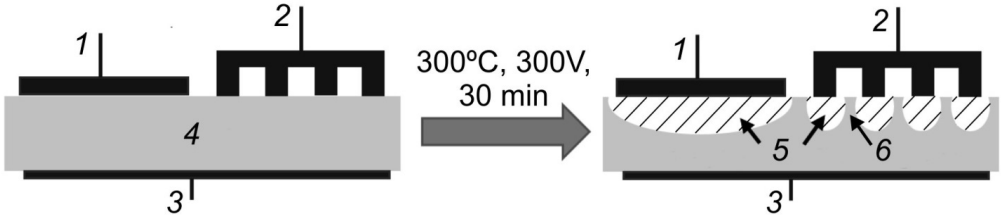


Fig. 1. Schematic of thermal poling:
flat (*1*) and comb-shaped (*2*) anode, respectively; cathode *3*; glass sample *4*;
the same sample with alternating poled (*5*) and non-poled (*6*) regions, respectively

During the thermal poling, the electric field in the sub-anode region of the glass can reach values of tenths of a volt per nanometer [6]. The alkaline ions (sodium ions in our case) located near the glass surface are displaced deep into the bulk under the influence of such a high field. In addition, water vapor penetrating into the sample from the atmosphere partially decomposes to form hydronium ions [7], and replaces the alkaline ions in the near-surface region. As a result, the near-surface region is depleted of alkaline ions and there is subsequently a local change in the physical properties of glass and its chemical resistance. For example, a periodic change of glass properties is observed for a comb-shaped electrode: the non-poled region under the electrode hollow retains the original properties, and the poled region changes its characteristics.

In the experiment, 1 mm thick soda-lime glass plates were used. The content of the main network-modifying oxide in the glass was as follows (wt%): 14.3 Na_2O, 6.4 CaO and 4.3 MgO. Poling was carried out at atmospheric pressure, temperature of 300°C and an applied DC voltage of 300 V; the process duration was 30 min (see Fig. 1). The concentration of sodium, calcium and magnesium in the poled glass was measured by energy dispersive X-ray spectroscopy (EDS) using an automated EDS Oxford Instruments AZtecLive system, installed on a TESCAN LYRA3 scanning electron microscope (SEM) [8]. The measurements showed that sodium ions were absent in the area about 2 μm deep from the glass surface, while the concentration of calcium and magnesium ions remains unchanged (Fig. 2, *a*). EDS is not sensitive to hydronium/hydrogen, but a bright line can be seen in the SEM image (Fig. 2,*b*) at the depth corresponding to the sodium concentration step. Presumably, this corresponds to electrons scattered by the spatial charge accumulated at the interface between hydronium and sodium ions. This charge arises due to the difference in the mobility of hydronium and sodium ions: the mobility of the former is more than three orders of magnitude lower than that of the latter [7]. Together with the known data on detecting the recoil of helium ions (ERDA) by poled glass [9], this allows to conclude that the deficit of sodium ions is mainly compensated by hydronium ions.

The deficit of sodium ions in the near-surface region of glass has a significant impact on the process where sodium ions are displaced by silver ions from the molten salt (ion-exchange). This process is most often used for formation of integral optical structures: due to higher polarizability of silver ions as compared to sodium ions, the $Ag^+ \rightarrow Na^+$ exchange leads to an increase in the refractive index by Δn in the ion-exchange region and formation of optical waveguides [2].

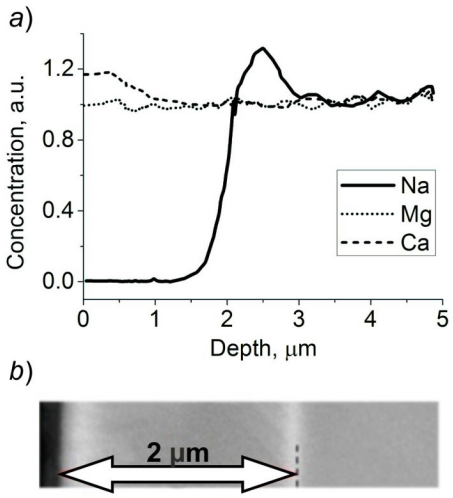


Fig. 2. Distributions of normalized metal concentration along the depth near the surface of poled glass sample (*a*) and SEM-image of its cross section (*b*)

To study the effect of thermal poling on the ion-exchange properties of glass, we placed poled glass in a melt of $Ag_{0.05}Na_{0.95}NO_3$ at 325°C. The treatment lasted for 60 min. The refractive index behavior in the poled and non-poled regions of the glass was investigated after ion-exchange. For this purpose, the modal spectra of gradient optical waveguides formed as a result of ion-exchange in the corresponding glass regions were measured on the MetriconM-2010 setup using prism coupling of light. Both poled and non-poled glass regions exhibited light wave grooveing after ion-exchange, but the modal spectra of these waveguides differed significantly. We calculated the refractive index profiles from the measured modal spectra using the Heidrich-White algorithm [9]. The results of the calculations are presented in Fig. 3.

Fig. 3 shows significant differences both in the magnitude of the change in the refractive index and in the depth of the area with increased refractive index for poled and non-poled glasses. It should be noted that the magnitude of the change in the refractive index linearly depends on the concentration of silver in the glass [11]. Thus, it can be concluded that the maximum concentration of silver in the poled region of glass is about 3.5–4 times less compared to the non-poled region, and the depth of silver penetration is about 4 times less (with the ion-exchange duration of 60 minutes). Since the ion penetration depth is proportional to the square root of the diffusion coefficient, the difference in depths allows to conclude that the diffusion coefficient of silver ions in poled glass is at least 10–15 times smaller than in virgin glass. The decrease of both concentration and depth of penetration of silver ions is presumably due to replacement of sodium ions by less mobile hydronium ions and structural changes caused by thermal poling. Thus, poled regions of glass are more stable with respect to ion-exchange treatment. This makes it possible to use them as dielectric masks for producing gradient integral optical structures on glass substrates or diffraction gratings.

The change in the glass structure after treatment with a high electric field is also confirmed by Raman scattering (RS) data. Fig. 4 shows Raman spectra of the virgin glass and glass after thermal poling (the spectra were obtained on a WITec Alpha 300R Raman microscope with a laser excitation wavelength of 532 nm). Fig. 4 shows that the poled glass lacks a peak in the 1100 cm^{-1} region, whose intensity is determined by the presence of modifying ions in

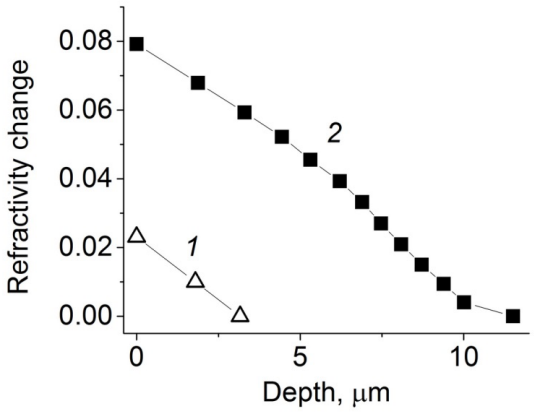


Fig. 3. Calculated profiles of refractive index increase in the ion-exchanged region of poled (*1*) and non-poled (*2*) glasses.
The duration of ion-exchange was 60 min

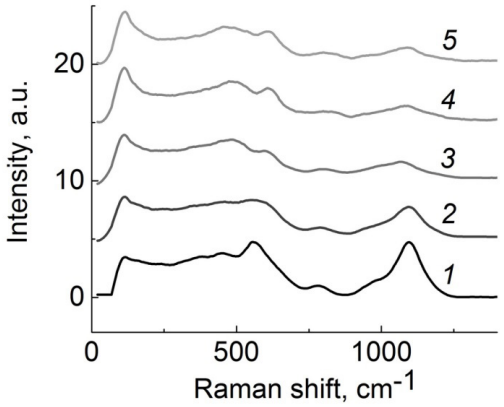


Fig. 4. Raman scattering spectra of virgin (*1*) and poled (*2–5*) glasses.
As the passed charge increases (this is reflected in the numbering of curves, *2* → *5*), the glass structure becomes more silica-like

the glass. Comparing the Raman spectra for the samples of the virgin glass and the glass poled at different applied voltages shows that the higher the electric field or the duration of thermal poling, the closer the structure of the poled region to that of fused glass without modifying alkali oxides [12]. Consequently, the number of non-bridging oxygen atoms decreases in poled regions [12]. The latter are associated in particular with alkaline ions which are capable of exchange, so the number of positions in the glass structure that can be occupied by ions involved in the process also decreases. This confirms the conclusion made above that thermal poling increases the resistance of glass to ion-exchange.

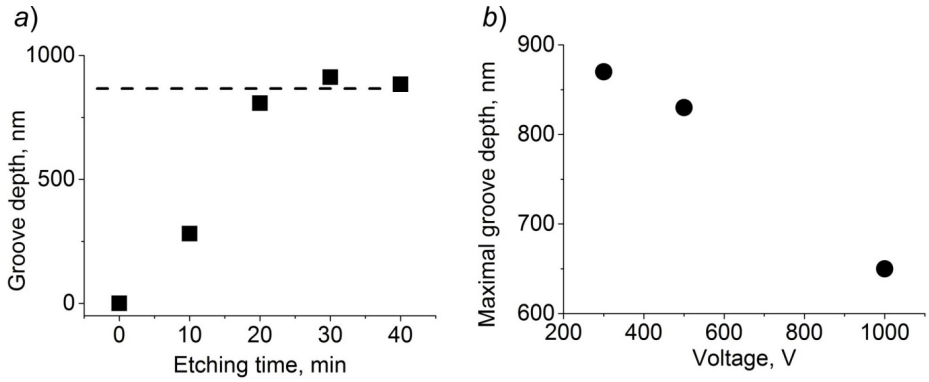


Fig. 5. Dependences of depth on the glass surface on the etching time in $NH_4F : 8H_2O$ solution (*a*) and the maximum depth of this groove on the applied voltage (*b*). The width of the groove is 60 μm; its maximum depth corresponds to complete etching of the poled layer

It has recently been shown that poled regions of soda-lime glass have better resistance to acid etchants and worse resistance to ion etching compared to non-poled regions [13].

We considered chemical etching of glass poled with a comb-shaped electrode (see Fig. 1) in an aqueous solution of ammonium fluoride ($NH_4F : 8H_2O$). It was found that etching of non-poled areas of glass under the electrode hollow occurs faster than that of poled ones, and grooves form on the glass surface as a result. The maximum depth of grooves is determined by the difference of etching rates of poled and non-poled regions and the thickness of the poled layer. The etching dynamics of poled and non-poled regions was studied by plotting the dependence of the groove depth on the etching time (Fig. 5, *a*); the groove width corresponding to the hollow width in the electrode was 60 μm. Fig. 5,*a* shows that the groove depth first increases linearly with time, and the dependence reaches saturation after 20 minutes of etching, at a depth of about 900 nm. Thus, the poled region in the given sample is completely etched after 20 minutes.

Notably, the glass in the region below the electrode hollow is not poled for grooves wider than 10 μm. Reducing the groove width leads to substantial edge effects and, consequently, glass poling even under the electrode hollow. A similar effect is observed when the applied voltage is increased (Fig. 5,*b*).

The picture will be reversed for reactive ion etching: poled regions are etched faster than non-poled ones, and the relief formed on the glass surface repeats the geometry of the electrode [13]. The maximum depth of the grooves is also determined by the depth of the poled region.

Thus, thermal poling of glass with a patterned electrode makes it possible to form a dielectric mask on the glass surface for chemical or ion etching. Depending on the task, forward or reverse (mirror) relief of the electrode can be formed in the glass. This method can be used to form, for example, lab-on-a-chip devices, as well as grooves for microfluidics.

Conclusion

Based on the results obtained, we can conclude that the changes in the composition and structure of multicomponent silicate glasses induced by thermal poling can be used to form dielectric masks on their surface for silver-sodium ion-exchange, chemical and reactive ion etching. The reason for this is that the escape of mobile alkaline ions from the sub-anode glass region during its thermal poling significantly slows down the process of ion-exchange, while increasing the cohesion of the glass network makes its structure more ‘silica-like’. This property increases the resistance of poled glass regions to chemical etchants used for multicomponent glasses and decreases the resistance of these regions to the process of reactive ion etching traditionally used for quartz glass.

Acknowledgements

The authors would like to thank A.I. Dergachev, a technical specialist of Tescan Ltd, for conducting measurements by energy dispersive X-ray spectroscopy (EDS).

This study was financially supported by the Ministry of Education and Science of the Russian Federation, Topic FSRM-2020-001.

REFERENCES

1. **DeCorby R., Irannejad M.,** Springer handbook of electronic and photonic materials: glasses for photonic integration, Springer, New York City, 2017. p.1536

2. **Righini G.C., Chiappini A.,** Glass optical waveguides: a review of fabrication techniques, Optical Engineering. 53 (7) (2014) 071819.

3. **Honkanen S., West B.R., Yliniemi S., et al.,** Recent advances in ion-exchanged glass waveguides and devices, Physics and Chemistry of Glasses: European Journal of Glass Science and Technology, Part B, Part B, 47 (2) (2006) 110–120.

4. **Brunkov P.N., Melekhin V.G., Goncharov V.V., et al.,** Submicron-resolved relief formation in poled glasses and glass-metal nanocomposites, Technical Physics Letters. 34 (12) (2008) 1030–1033.

5. **Babich E.S., Gangrskaia E.S., Reduto I.V., et al.,** Self-assembled silver nanoparticles in glass microstructured by poling for SERS application, Current Applied Physics. 19(10) (2019) 1088–1095.

6. **Lipovskii A.A., Melehin V.G., Petrov M.I., et al.,** Bleaching versus poling: Comparison of electric field induced phenomena in glasses and glass-metal nanocomposites, Journal of Applied Physics, 109 (1) (2011) 011101.

7. **Doremus R.H.,** Mechanism of electrical poling of silica glass, Applied Physics Letters. 87 (23) (2005) 1–2.

8. **Dergachev A., Kaasik V., Lipovskii A., et al.,** Control of soda-lime glass surface crystallization with thermal poling, Journal of Non-Crystalline Solids. 533 (1 April) (2020) 119899.

9. **Lepienski C.M., Giacometti J.A., Leal Ferreira G.F., et al.,** Electric field distribution and near-surface modifications in soda-lime glass submitted to a dc potential, Jдournal of Non-Crystalline Solids. 159 (3) (1993) 204–212.

10. **White J.M., Heidrich P.F.**, Optical waveguide refractive index profiles determined from measurement of mode indices: a simple analysis, Applied Optics. 15 (1) (1976) 151–155.

11. **Licares J., Sotelo D., Lipovskii A.A., et al.,** New glasses for graded-index optics: Influence of non-linear diffusion in the formation of optical microstructures, Optical Materials. 14 (2) (2000) 145–153.

12. **Redkov A.V., Melehin V.G., Lipovskii A.A.,** How does thermal poling produce interstitial molecular oxygen in silicate glasses? The Journal of Physical Chemistry, C. 119 (30) (2015) 17298–17307.

13. **Reduto I., Kamenskii A., Brunkov P., et al.,** Relief micro- and nanostructures by the reactive ion and chemical etching of poled glasses, Optical Materials Express. 9 (7) (2019) 3059–3068.

Received 14.09.2020, accepted 20.09.2020.

THE AUTHORS

BABICH Ekaterina S.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
babich.katherina@gmail.com

REDUTO Igor V.
University of Eastern Finland
Joensuu, 80101, Finland
igor.reduto@uef.fi

REDKOV Alexey V.
Institute for Problems of Mechanical Engineering RAS
61 Bolshoi Ave. V.O., St. Petersburg, 199178, Russian Federation
red-alex@mail.ru

RESHETOV Ilya V.
Alferov University,
Peter the Great St. Petersburg Polytechnic University
8/3 Khlopina St., St. Petersburg, 194021, Russian Federation
reshetov_iv@spbstu.ru

ZHURIKHINA Valentina V.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
jourikhina@mail.ru

LIPOVSKII Andrey A.
Alferov University,
Peter the Great St. Petersburg Polytechnic University
8/3 Khlopina St., St. Petersburg, 194021, Russian Federation
lipovskii@mail.ru

DOI: 10.18721/JPM.13415

AN OUTPUT POWER LEVEL OF THE FIBER-OPTIC INTERFEROMETRIC SCHEMES WITH FIBER BRAGG GRATINGS FOR MULTIPLEXING THE SENSITIVE ELEMENTS

L.B. Liokumovich[1], *A.O. Kostromitin*[2,1]

Ph.V. Skliarov[2,1], *O.I. Kotov*[1]

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

[2] Concern CSRI "Elektropribor", St. Petersburg, Russian Federation

The paper continues a series of articles devoted to the procedure proposed by the authors how to calculate the required elements' parameters and the output optical power level of the fiber-optic interferometric schemes with time-division multiplexing (TDM) of sensing elements. In particular, the schemes based on fiber Bragg gratings have been analyzed. The proposed method enables to ensure uniformity of powers of optical signals from individual sensing elements, as well as to evaluate how the deviations of elements' parameters from the nominal ones influence the optical powers. According to the calculation methodology, a sequence of applying the required expressions was substantiated and some obtained results and an analysis of characteristic behaviors for the considered optical fiber circuits were exemplified. The proposed approach was recommended for design of interferometric optical fiber sensors with multiplexed sensing elements.

Keywords: fiber-optic interferometric sensor, Bragg grating, optical power, time-division multiplexing

Citation: Liokumovich L.B., Kostromitin A.O., Skliarov Ph.V., Kotov O.I., An output power level of the fiber-optic interferometric schemes with fiber Bragg gratings for multiplexing the sensitive elements, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 135–147. DOI: 10.18721/JPM.13415

Introduction

Fiber optic measurement systems and devices have a number of advantages, associated primarily with the possibility of using fully optical electroneutral and passive fiber-optic sensing circuits, as well as circuits for coupling the sensing elements to the optoelectronic part [1]. Interferometric systems are an attractive option of fiber optic measurement systems [2], due to the fact that they offer the best resolution. In recent decades, such measurement systems have been actively developed and implemented in various fields, for example, in navigation systems, in hydroacoustics, in the oil and gas industry, in measuring equipment for medical purposes [3–6].

Practical implementations of such measurement systems mainly involve multiplexing of a set of fiber optic sensing elements (SE) in a single path connected to the optical source and the photodetector. The principle of Time Division Multiplexing (TDM) is most commonly used for this purpose [6]: instead of continuous light radiation, optical pulses are fed into a fiber optical circuit, and a sequence of output pulses is generated at the output of the circuit in response to each input pulse carrying information about the corresponding SE and the impact on it. The structure of such an optical circuit and the parameters of all elements (sensitive fiber sections, beam splitting elements, reflectors, etc.) must correspond to a whole range of relations in order to ensure the operability of such a measurement system and its efficiency.

An important component in the set of parameters for fiber-optic circuits with multiplexing SE in the interferometric measurement system is the so-called energy efficiency, namely, the level of power of the

output pulses relative to the pulse power at the input of the circuit. The energy efficiency directly affects the achieved signal-to-noise ratio in the generated series of interference signals and, consequently, the level of output noise in the measuring system. Nevertheless, the available literature does not present comprehensive methods for calculating the energy efficiency of such fiber-optic circuits. Only simplified estimates are typically given for particular cases, when it is impossible to take into account important factors (losses of optical power in the elements, possible deviations of actual parameters from the required optimal values, including random deviations) [7]. The approximations used often perform poorly in practice, for example, too large or even an infinite number of multiplexed SEs is assumed [7], while up to 16 or 32 SEs are generally multiplexed in real systems, and circuits with eight or even four SEs are frequently constructed.

An example of a consistent method of energy calculation for fiber optic circuits with multiplexed SEs is given in our earlier study [8], which also discusses the possibilities of analyzing the characteristics of the given circuits based on the developed calculation method.

In this paper, we propose a similar consideration of the energy calculation procedure and a brief analysis of the results obtained for a circuit of another type, where fiber Bragg gratings (FBG) are used as light-splitting elements [10]. This option is attractive for practical applications, in particular because it yields improved energy characteristics.

However, such gratings have a serious drawback: they lead to a crosstalk between signals from different SEs [10].

In view of the above, it seems extremely important to produce a well-founded comprehensive calculation of the optimal system of FBG parameters and the power level of the output optical radiation in such circuits taking into account all key element parameters, as well as parameters that characterize the level of parasitic crosstalk.

Circuit under consideration and principles of calculations

The subject of the analysis is an optical fiber path for a multiplexed interferometric measuring system. Its sensing elements are fiber sections mounted in certain structural units; the latter are optimized to convert the target perturbation passing through this section into a change in the phase delay of light. A key part of such a system is an optical circuit where N fiber SEs are combined into a certain path using splitters and reflectors [11].

In reflection-type circuits, the same fiber output and input can actually be used as the input and output of the radiation (an additional Y-splitter or circulator can be used in the instrumentation unit of the system for separation). In feed-through circuits, the input and output are at different ends of the path with the SE. If a short optical pulse of power P_{in} is fed to the circuit input, then a sequence of output pulses of power p_n should be generated at the circuit output, where the nth pulse goes through the SEs from the first to the nth and does not go through other SEs. Typically, $N + 1$ pulses with numbers from 0 to N are generated, where the pulse with the number zero ($n = 0$) does not pass through any SEs. These pulses will be delayed relative to each other due to the delay in passing the corresponding system of links in the SE circuit.

For the circuits considered in our previous study [8], as well as those considered in this paper (including the circuits given in Appendix 1), the light delay in passing the circuit links is primarily determined by the delay in propagation through fiber SEs. Next, because the so-called compensation interferometer is used in the interrogation unit, the pulses are separated, shifted, and each initial pulse is combined with the previous one. As a result, due to separate processing of pulses, the result of interference of pulses whose phase difference is given by the phase delay of light in the nth SE can be recorded at the output of the compensated interferometer, and, therefore, the data about the target impact on this element can be collected. The input pulses are repeated during operation of the measuring system, with a sequence of interference signal samples from each SE (from each channel) produced as a result.

The analysis of such circuits implies different types of description and calculations associated with different types of factors, which should be considered and coordinated to maintain acceptable performance of the system and optimize its characteristics. Here we consider the energy efficiency factor of the circuit, i.e., analyze the optical power levels of the pulses generated by the circuit.

By analogy with the approach used and described in [8], this work implies the choice of parameters of dividing and reflecting elements that ensure the same power of all pulses

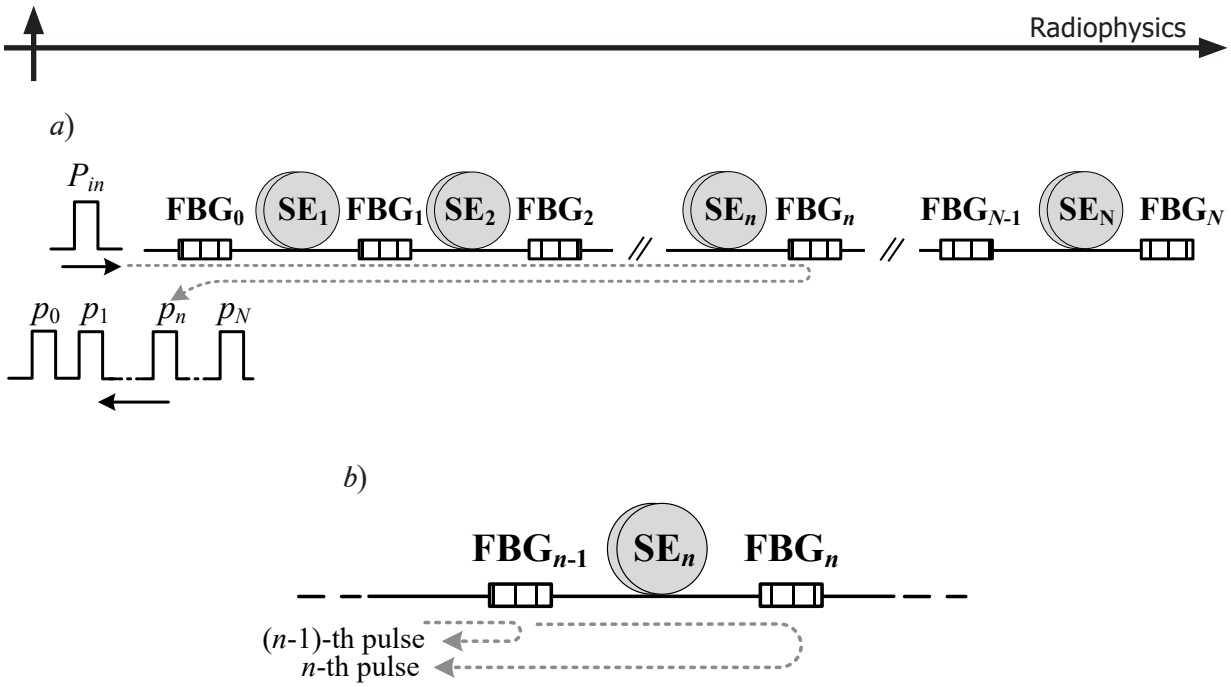


Fig. 1. Example of a circuit based on FBG with N sensing elements (*a*), *nth* link of this circuit (*b*); Sensitive Element SE*i*; Fiber Bragg Grating FBG*i*; P_{in}, p_n are the input and output pulse powers, respectively

generated by the circuit: $p_n = P_0$. In this case, the main parameter that characterizes the energy efficiency of the optical circuit is the relative level of the optical power of the generated series of pulses $P_{norm} = P_0/P_{in}$.

The types of fiber optic circuits with multiplexing SEs analyzed in [8] are arranged so that the nth reverse pulse reflected from the nth SE is split twice by the splitter and, consequently, the P_{norm} value decreases rapidly with increasing N (at best, the relation is expressed approximately as $1/N^2$).

Let us also consider another configuration of circuits with multiplexing of fiber SEs, more advantageous from the standpoint of the P_{norm} level. An example is a circuit attractive for practical use (Fig. 1), with partially reflecting fiber elements (semi-transparent mirrors). Fiber Bragg gratings (FBG) are primarily used as the latter. Compared with the circuits described in [8], the configuration that includes FBG has a different form (see Appendix 1), which has two important features (positive and negative):

on the one hand, the circuit generates a significantly higher relative power level of pulses P_{norm}, which decreases rather slowly with increasing value of N;

on the other hand, in addition to the main sequence of output pulses, additional parasitic pulses are generated in the circuit, which cause crosstalk between the target signals from different SEs.

As shown in Appendix 1, not only circuits with feed-through reflectors but also with splitters can have such properties, and they can serve as a basis for implementing circuits of both reflective and feed-through types. The presence of significant crosstalk makes practical applications of the described variants considerably less attractive, effectively eliminating the benefits of such schemes in terms of energy efficiency. To reduce the level of crosstalk, it is necessary to choose the parameters so that the relative level of power P_{norm} decreases. All of this makes it the more important to develop adequate methods for calculating such circuits.

The general principle and basic stages of calculating the parameters of the elements (in this case, these are primarily the reflectivities of the FBG) and the power of the output optical pulses are completely similar to the approach described in [8]. Part of the calculation procedure is related to obtaining and using multiplicative expressions to calculate p_n, assuming that all circuit element parameters are given. The other part of the procedure implies the derivation of recurrence relations for the main elements of the nth circuit, making it possible to fulfill the condition that the powers of all output pulses be equal ($p_n = P_0$). These parts of the calculation expressions allow not only to determine the required values of circuit element parameters but also to take into account

and analyze the influence of various factors, such as losses in elements (including random ones), deviations of actual element parameters from nominal or prescribed values, and other factors inevitably present in real-life design of circuits.

Procedure for calculating the parameters of FBG and general analysis of power level of output pulses

To describe the operating principles of an FBG as a reflecting element, we will use the reflectivity R, equal to the ratio of power of the optical radiation entering the FBG to that of the radiation directed back by the grating. The power of the radiation that has passed through the FBG is described by the transmittance K at the generation wavelength, which is given by the expression

$$K = (1 - R)(1 - \alpha_{el}),$$

where the first factor accounts for the fact that part of the power is directed backwards by the grating, and the second factor accounts for the presence of internal (excessive) optical power losses; these losses are characterized by a small parameter α_{el}.

We also use the power transfer coefficient K_f for a section of fiber with an SE, taking into account the optical power loss in this section. Since the fiber in the sensitive elements can be twisted into a coil or otherwise assembled into the SE structure, the magnitude of these losses can be significant, even despite the relatively small length of this fiber optic section. If there are spliced or detachable connections in the circuit, losses in them should also be taken into account in the values of the coefficients K_f.

In practice, it is convenient to characterize losses in decibels, using standard ratios

$$\alpha_{el[\mathrm{dB}]} = 10 \cdot \lg(1 - \alpha_{el}),$$

$$\alpha_f = 10 \cdot \lg K_f.$$

The values of K_f and α can be found from a_f and $\alpha_{el[\mathrm{dB}]}$ based on trivial inverse relations.

Given the parameters and notations introduced, if we consider the passage of the input pulse to the nth FBG and back (see Fig. 1, *a*), it is easy to formulate the multiplicative expression for p_n:

$$p_n = P_{in} \cdot R_n \cdot \prod_{q=1}^{n} \left(K_{q-1}^2 \cdot K_{\mathrm{f}q}^2 \right). \qquad (1)$$

Eq. (1) implies that if the upper limit of the product is less than the lower limit (which is the case when $n = 0$), then the product is equal to unity. In practice, we can often assume that all SEs are equivalent and K_f does not depend on the value of n. Then this parameter can be excluded from the product in expression (1) and an additional factor K_f^{2n} can be used.

Analyzing one link in the circuit and comparing the path difference of the $(n - 1)$th and the nth pulse (see Fig. 1, *b*), we can obtain the equation corresponding to the balance $p_{n-1} = p_n$; it has the following form:

$$R_{(n-1)} = K_{n-1}^2 \cdot K_{\mathrm{f}n}^2 \cdot R_n = \left(1 - \alpha_{\mathrm{el}(n-1)}\right)^2 \left(1 - R_{n-1}\right)^2 K_{\mathrm{f}n}^2 \cdot R_n, \qquad (2)$$

where n can vary from 1 to N.

Ratio (2) is a quadratic equation with respect to $R_{(n-1)}$, which has a solution in the form

$$R_{(n-1)} = \frac{a + 1 - \sqrt{2a + 1}}{a}, \qquad (3)$$

where the notation $a = 2[(1 - \alpha_{el\,(n-1)}) K_{fn}]^2 R_n$ is used.

The second solution of Eq. (2) has a similar structure, but with a plus sign before the square root, and does not correspond to the physical sense in the given situation, because it leads to obtaining a value $R > 1$.

The recurrence relation (3) can be applied if two conditions are fulfilled.

First, it is necessary to give some value for the reflectivity of the closing mirror R_N, relative to which we can then calculate the reflectivities of other FBGs. The choice of this value and its influence will be discussed below.

Second, relation (3) includes the parameters K_{fn} (for $n = 1, 2, \ldots, N$) and α_{eln} (for $n = 1, 2, \ldots, N - 1$) related to the optical power loss in the circuit elements. The transfer coefficients K_{fn} depend on the level of losses in the SE, and they are not related to the choice of FBG, but must be known. However, the internal losses α_{el} in the FBG can be generally related to the value of R of the grating, and this relationship may be different depending on the technologies by which the grating was fabricated and installed in the optical circuit.

For a rigorous formal solution of the given problem, the dependence of α_{el} on R must be given in some form, Eq. (2) must be solved relative to R_{n-1} with this dependence taken into account, and the result may differ significantly from the one given by Eq. (3).

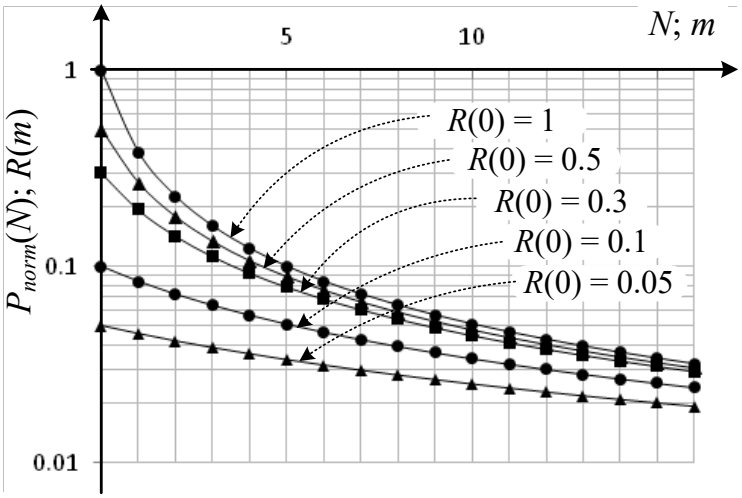


Fig. 2. Dependences of relative pulse power P_{norm} on the number of elements N (N varies from 1 to 16) and reflectivity R on the lattice number m (numbering from the end of the path, n is the circuit link number, $m = N - n$ and also varies from 0 to 16) for different reflectivities $R(0)$ of the final element

In view of the above, it is problematic to obtain a specific solution of Eq. (2) for the general case of the $\alpha_{el}(R)$ dependence. Therefore, it is proposed to specifically use relation (3) in the calculations, with some fixed value independent of n serving as α_{eln} (a typical or average value for the FBG intended to be used). In this case, if the condition of small losses $\alpha_{eln} << 1$ is fulfilled rigorously enough, then the influence of the difference of α_{el} due to the difference of R_n will have little effect on the obtained values of p_n. Moreover, after the reflectivities R_n of the gratings are calculated, the data on the actual losses in the FBGs with these reflectivities can be subsequently taken into account when analyzing the influence of this factor on the p_n values.

If we take $\alpha_{eln} = \alpha_{el}$, the recurrence relation (3) can be written in a simpler form taking into account the above explanations:

$$R_{(n-1)} = \frac{K_{fn}^2 \cdot R_n + b - \sqrt{K_{fn}^2 \cdot R_n + b^2}}{K_{fn}^2 \cdot R_n}, \quad (4)$$

introducing a constant $b = 0.5/(1 - \alpha_{el})$.

Choosing some level of reflectivity of the last reflector R_N, we can further obtain values for other reflectivities Rn from $n = N - 1$ to $n = 0$ by recurrence expression (4) in turn.

After the set of required grating reflectivities Rn is determined, the final stage of calculations involves finding the pulse powers p_n based on expression (1). If the parameters K_{fn} and α_{eln} do not depend on n, then, according to the principle by which expression (4) is derived, all p_n will be the same and equal to P_0. Therefore, to estimate P_0, we can use, for example, the simplest relation

$$P_0 = p_0 = P_{in} \cdot R_0.$$

On the other hand, if there is a priori information about the actual values of K_{fn} and α_{eln}, p_n can be found by expression (1) to consider their mean and standard deviations from P_0. If there is data on the relation between the excess losses in the FBG and the grating reflectance, α_{eln} values can be found from the calculated R_n to account for the n-dependent losses when calculating p_n based on expression (1).

In addition, for various reasons, the values of R'_n (the real reflectivities of the FBGs that are intended to be used) may differ from the values of R_n (initially calculated). If such data are available, they can be taken into account at this stage by using R'_n rather than R_n in expression (1). An example is the case, when the reflectivities R'_n of real FBGs are set with rounding to some level, then the calculated R_n values should be rounded; another case is when the calculations are performed to study the influence of random deviations of parameters of real elements from the preset ones.

For general analysis of the energy efficiency of circuits, it is usually sufficient to determine the relative level P_{norm}; to calculate this parameter, substitute $P_{in} = 1$ in Eq. (1).

To illustrate the form of general dependences, we will give examples of calculations

for the case when losses are neglected, assuming that $K_{fn} = K_n = 1$. For this purpose, Fig. 2 shows the result of the calculation of R_n when $n = 0, 1, ..., 16$, when $R_N = 100\%$, 50%, 30%, 10% and 5%. To simplify further analysis, the dependences $R(m)$ are given, where m is the number of the FBG from the end of the path, i.e., $m = N - n$. Given this numbering, $R_N = R(0)$. According to the principle of recurrence calculation, the presented dependences reflect the result for any number of sensitive elements N in the range from 0 to 15. For a given N, we should use the first $N + 1$ values of the coefficients $R(m)$, where m takes values from 0 to N. The remaining calculation coefficients with values $m > N$ are not needed in this case.

As already mentioned, if we do not introduce the differences between the actual values of reflectivities and the initial calculated ones into the power calculations by formula (1), the power P_0 can be estimated as $P_0 = P_{in} \cdot R_0$. Then

$$P_{norm} = R_0 = R(m)|_{m=N}.$$

It follows that the presented $R(m)$ dependences simultaneously characterize the relationship between the values of P_{norm} and N, which is also evident from Fig. 2. To analyze the specific values of $R(m)$ or $P_{norm}(N)$ shown in Fig. 2, these values are given as a table in Appendix 3.

If $R_N = 1$ (it corresponds to $R(0) = 1$ in Fig. 2), the dependence $P_{norm}(N)$ is proportional to $1/N$ with good accuracy; approximation gives the dependence $P_{norm} = 0.45 / N^{0.95}$ for the values of N from 1 to 15. The dependences for the reflectances of 50%, 30% and 10% ($R(0) = 0.5$, 0.3 and 0.1) are also in good agreement with the power dependences, but the exponent decreases, and the dependency at $R_N = 10\%$ is already proportional to $1/\sqrt{N}$. The dependence at $R(0) = 0.05$ ($R_N = 5\%$) is better described by a logarithmic rather than a power dependence; the approximation gives the function

$$P_{norm} = 0.051 - 0.011 \cdot \ln N.$$

If we compare these results with similar calculated data for circuits based on splitters, which are free of channel crosstalk [9], a significant gain in the level of output optical pulses is evident for all values of N, even if relatively low values of R_N, such as 5%, are used. Moreover, in the case of an alternative type of circuit, the $P_{norm}(N)$ dependences can be approximated by power dependences, with exponents close to 2 or higher [8].

Obviously, for any value of N, the maximum level of P_{norm} is provided at $R_N = 1$. However, reducing the reflectivity of the final FBG and, accordingly, of the other gratings allows to reduce the level of channel crosstalk. A quantitative characteristic of this coupling is the coefficient K_{cr} (see Appendix 2). If losses in the fiber elements are neglected, then this coefficient is introduced by expression (a4) for the given circuit and has the form

$$K_{cr} = R^2_{N-1} R_{N-2} / [R_N (1 - R_{N-1})^2].$$

Given the numbering $R(m)$ used in Fig. 2, we can write

$$K_{cr} = R(1)^2 \cdot R(2) / [R(0)\,(1 - R(1)^2].$$

Taking into account the data obtained (see, for example, Appendix 3), the crosstalk coefficients K_{cr} for the cases $R_N = 100\%$, 10% and 5% take values of 0.087; $6.06 \cdot 10^{-3}$ and $1.91 \cdot 10^{-3}$, respectively.

Thus, a 20-fold decrease in the reflectivity R_N of the final reflector leads to a 45-fold decrease in the crosstalk coefficient K_{cr}. However, the relative power level P_{norm} of optical pulses formed by the circuit will decrease. This is demonstrated

Table 1

Dependence of optical power level of generated pulse system on the values of main circuit parameters

R_N	K_{cr}	P_{norm}		
		$N = 4$	$N = 8$	$N = 16$
1.00	0.087	0.123	0.063	0.032
0.10	6.06·10–3	0.056	0.039	0.024
0.05	1.91·10–3	0.036	0.028	0.019

Notations: N is the number of sensitive elements, R_N is the reflectivity of the last reflector, K_{cr} is the crosstalk coefficient.

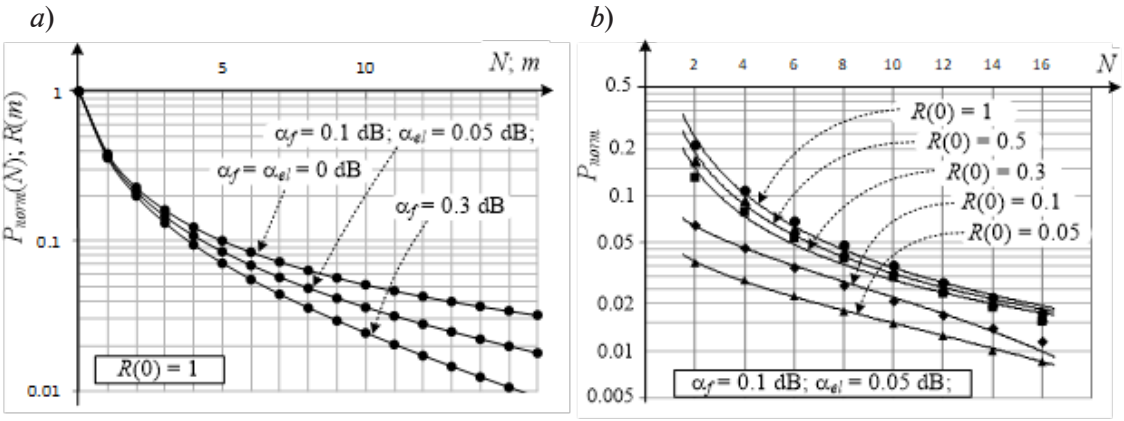


Fig. 3. Examples of calculations accounting for optical power losses in the circuit elements. The $R(m)$ dependences (or $P_{norm}(N)$, the dependences coincide) are given for $R(0) = 1$ and different values of loss factors (*a*) and $P_{norm}(N)$ for different values of R_N (*b*)

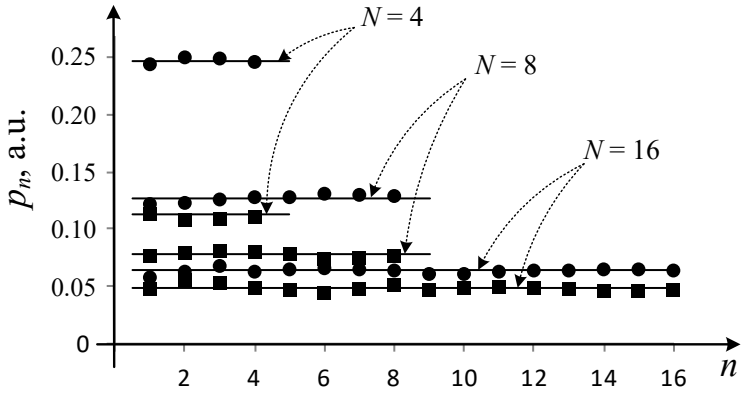


Fig. 4. Dependences of output pulse power on circuit link number for different N and two R_N values, %: 99 and 10 (circles and squares, respectively). The values of p_n are normalized to the input pulse power $_{Pin}$. We consider the case when R'_n values are obtained with R_n rounded to 1%; solid lines correspond to P_{norm} levels with no rounding.

in Table 1, where the values of R_N, K_{cr}, and P_{norm} are compared. At the same time, according to Table 1, a decrease in crosstalk by almost two orders of magnitude is achieved with a relatively small decrease in P_{norm}. The power level is reduced by 3.4 times for a circuit with four SEs. For more attractive configurations with $N = 8$ and $N = 16$, the power decreases less dramatically: the P_{norm} value decreases by 2.3 times for eight SEs, and by only 1.6 times for sixteen.

Effect of power loss in circuit elements and differences between calculated and actual reflectivities

The given calculation procedure allows to account for the losses in the circuit elements. Comprehensive analysis of the influence of losses on the $R(m)$ and $P_{norm}(N)$ dependences is somewhat complicated due to too high variability of possible parameter values with their significant influence on the circuit characteristics. Therefore, we will confine ourselves to examples of calculations accounting for such losses and their influence. For example, Fig. 3,*a* shows the $R(m)$ or $P_{norm}(N)$ dependences (with the numbering introduced, these dependences will coincide, see explanations above). The graph demonstrates the reduction in the reflectivities of the FBG required to equal the power of the output pulses, as well as the reduction in the relative level of this power due to accounting for the losses in the fiber optic SE and the excess losses in the FBG. In this example, the element losses of all circuit links are assumed

Table 2

Effect of rounding the calculated R_n values to 1% for calculations of optical power of the output pulses using Eq. (1)

N	R_N,%	P_{norm} (no rounding)	$^{mean}p_n$	($^{SD}p_n$/ $^{mean}p_n$), %
4	99	0.246	0.247	1.1
	10	0.113	0.11	2.0
8	99	0.1267	0.1266	2.5
	10	0.0783	0.0776	3.0
16	99	0.006357	0.006369	6.0
	10	0.004842	0.004863	3.6

Notations: P_{norm} is the relative level of output power with no rounding; $^{mean}p_n$ is the mean value of p_n, $^{SD}p_n$ is the standard deviation *of* p_n, normalized to the input pulse power $_{Pin}$; other notations are identical to those given in Table 1.

to be the same. Apparently, accounting for insignificant losses has a noticeable effect on the calculated dependences, especially when the number of SEs is large.

Fig. 3, *b* show the results of calculation of $P_{norm}(N)$ at different values of $R_N = R(0)$ for the case of losses expressed in terms of $a_f = 0.10$ dB и $\alpha_{el} = 0.05$ dB. In spite of noticeable influence of losses on the dependences shown (if they are compared with the curves in Fig. 2), the general conclusion about a relatively small decrease in the output power level with a significant decrease in the values of R_N and K_{cr} remains valid.

As already noted, an important advantage of the given procedure is that it allows to account for the differences between the calculated and actual parameter values of the circuit elements, as well as to analyze the impact of such differences.

As an example, consider the case where reflection coefficients for real FBGs can be given with a finite accuracy, for example, 1%. Appendix 3 shows the result of such rounding of values of R_n obtained neglecting losses. Referring to the data in the table in Appendix 3, we can see that rounding can make significant differences between R'_n and R_n, since the initial values of R_n themselves are small, especially with relatively large values of N and small reflectivities R_N. The effect of rounding on the irregularity of p_n is illustrated by the graphs in Fig. 4, where, as an example, the calculated values of p_n values are given with rounded coefficients R'_n substituted into Eq. (1). It can be seen that the variations of p_n introduced are noticeable but not significant. More accurate quantitative data characterizing the dependences shown in Fig. 4 are given in Table 2. We can also see that the deviation of the mean pulse powers differs from the initial estimate of P_{norm} only in the third significant digit. Even for $N = 16$, where the influence is more significant due to the low initial values *of* R_n, the standard deviation of the set of p_n values is 6 %, which can be deemed acceptable.

Similarly, it is possible to consider other cases of deviations of actual parameters from the calculated ones when calculating initial values of R_n (including actual values of losses in elements) and to analyze their influence on the obtained values of p_n. In this case it is possible to account for and analyze both regular and random deviations of parameter values.

Similarly, other factors of deviation of R'_n from R_n can be taken into account. The last stage of calculations by Eq. (1) can also involve accounting for the differences of actual values of other parameters and the values used in the first stage of calculations, calculating the initial values of R_n, and analyze their influence on the obtained values of p_n. Both regular and random deviations of parameter values can be taken into account and analyzed.

Conclusion

We have proposed an ideology for calculating the parameters of the elements of interferometric fiber optic schemes of multiplexed fiber optic sensors using FBG, which makes it possible to optimize the circuit to achieve the maximum level and visibility of the generated interference signals. The procedure for obtaining expressions to calculate the reflectivities of FBGs that provide the same output pulse power is described. The proposed calculation procedure accounts for the optical power losses both in the FBG and in the fiber sections of the sensing

elements. It is established that the calculations can account for the influence of differences between the parameters of the elements used and obtained in the initial calculation and the actual parameters of the elements that will be used for creating a real optical circuit.

We have provided examples of calculating the FBG parameters in the given circuit for different numbers of sensitive elements and different levels of crosstalk in the circuit. The examples confirm that it is possible to analyze the effect of losses in the circuit elements and use rounded values of the FBG reflectivities.

The presented methods and calculation results can be directly used to design fiber optic interferometric measurement systems based on multiplexing of sensing elements in a circuit with FBG.

Appendix 1

Two classes of fiber optic circuits with multiplexed sensitive elements

To explain the peculiarities of the analyzed circuit, let us briefly describe two classes of fiber optic circuits for interferometric systems with multiplexing of sensing elements (SEs). Characteristic examples of circuits of the first class include the ones considered in [8] and shown in Fig. A1. The first one is a reflective circuit, the second one is a feed-through type.

Another class of optical circuits is possible; examples are shown in Fig. A2. This class also includes reflective (*a*, *b*) and feed-through (*c*) circuits. In such circuits, low-loss reflective fiber elements, such as FBGs, can be used in partial reflection/transmission mode.

In all these circuits, the *n*th output optical pulse is formed by passing through the sensing elements from the 1st to the *n*th. In reflective circuits with combined input/output fiber output, the pulse passes through these SEs twice; in circuits with separated input and output fibers, the pulse passes through the SEs once.

Evidently, for circuits in Fig. A1, the *n*th pulse that passes through each SE from the first to the *n*th also passes through the *n*th splitter of this circuit along the forward path (where the larger share of power is transmitted) and passes twice through the splitter of the *n*th link along the crosstalk direction (where a smaller share of power is transmitted). The situation is similar in circuits in Fig. A2 (*b* and *c*), but the *n*th pulse passes through the *n*th splitter in the crosstalk direction only once. The same happens in the case of the FBG scheme (*a*), if we assume that the passage of light through the FBG corresponds to the forward path through the splitter, and the reflection from the grating corresponds to the crosstalk path through it. It is the additional crosstalk passage through the splitter in the circuits of the first class that leads to a significant reduction in the pulse power level, compared with the circuit of the second class, because the required crosstalk transmittance of the splitters is small and decreases with increasing number N of the SEs.

However, it is clear that the *n*th output pulse can also be formed by other propagation paths in circuits of the second class, which is described in more detail in Appendix 2. For example, in the circuit in Fig. A1,*b*, the pulse may not pass through the *n*th SE, but pass through any previous element twice. The pulse in the circuit

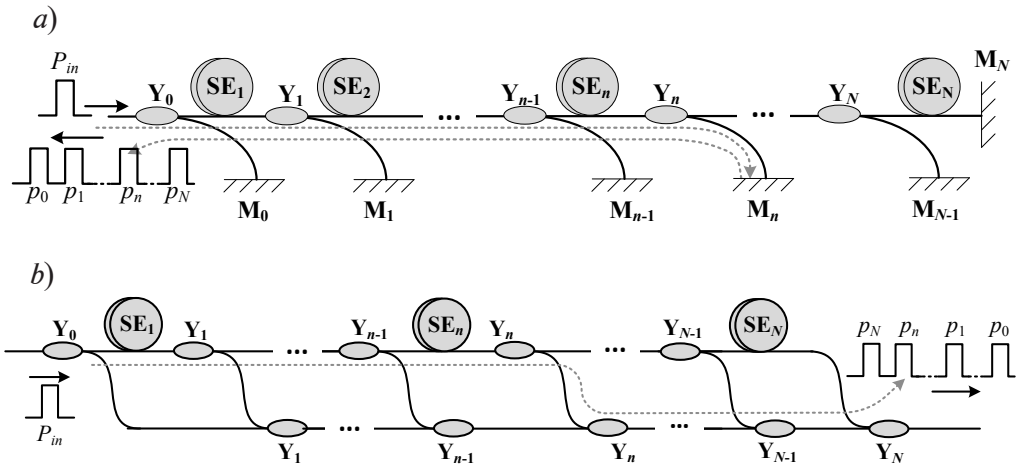


Fig. A1. Fiber optic multiplexing circuits of the first class: reflective type based on splitters and mirrors (*a*) and feed-through type based on splitters (*b*); Sensitive Element SE*i*; Y-type fiber splitter Y_i; mirror M_i; input and output pulse powers P_{in} and p_n, respectively

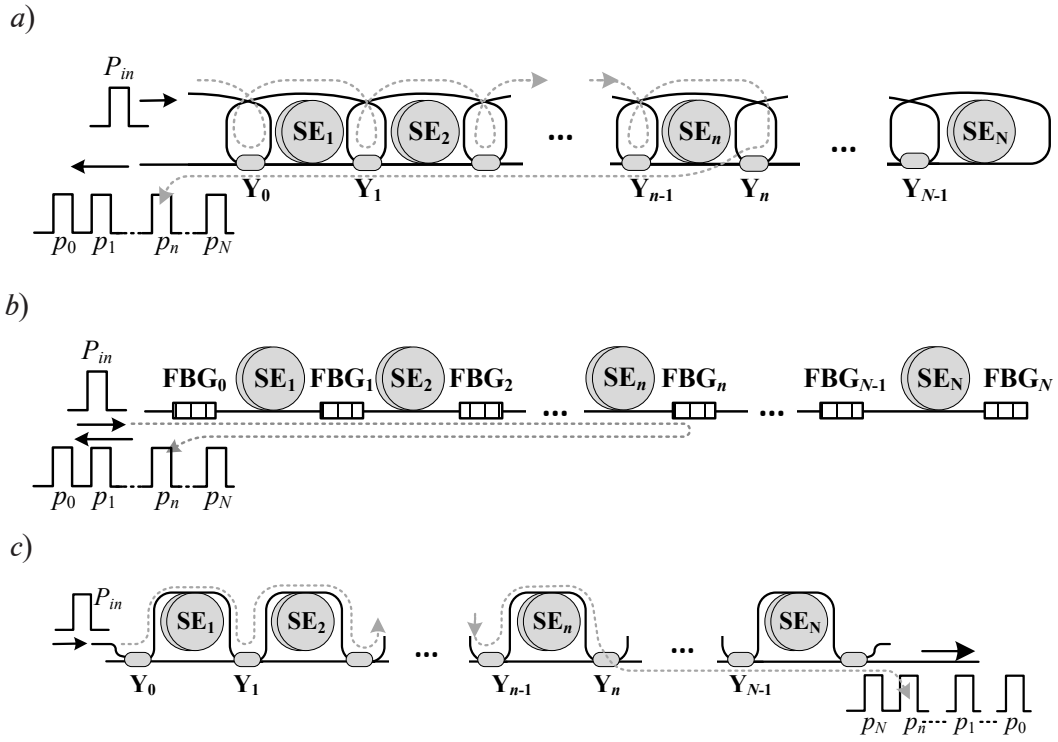


Fig. A2. Fiber optic multiplexing circuits of the second class: reflective based on FBG (*a*) and splitters (*b*); feed-through type with splitters (*c*);
FBG_i are Fiber Bragg Grating, other notations are identical to those in Fig. A1

in Fig. A1,*c* can pass the SEs from 1 through $(n-1)$, then pass by the *n*th SE, and then pass through any of the subsequent ones. If the total number of SEs passed is *n*, then this pulse will be in the *nth* position in the output sequence of pulses. As a result, pulses whose phases differ not only by the delay in the *n*th SE, but also by phase delays in other SEs subsequently participate in generating the interference signal of the *n*th channel. Generation of an additional *n*th pulse is connected with additional crosstalk passes of splitters and significant additional attenuation. Nevertheless, the presence of not only main but also additional pulses in the output sequence leads to cross-influence of the measured target perturbations in different channels. The generation of additional pulses in the FBG circuit is analyzed in more detail in Appendix 2.

Note that there are no alternative ways for forming the *n*th pulse in the output sequence in circuits of the first class (see, for example, Fig. A1). However, in order to exclude such a possibility in the configuration of the circuit, an additional crosstalk pass of the splitter is inevitably introduced, with the energy efficiency of the circuit considerably deteriorating.

Thus, reducing the relative power level of the output pulses can be regarded as payback of sorts for eliminating crosstalk of channels, and vice versa, it is possible to increase the pulse power level if crosstalk of channels is allowed.

Appendix 2

Generation of parasitic output pulses in the multiplexing circuit based on FBG

Let us consider the generation of additional output pulses, which determine the crosstalk of channels in the given fiber optic scheme with FBG. The main sequence of output pulses is formed when the *n*th pulse passes to the *n*th grating and back, and reflection occurs once (this path is denoted by number *1* in Fig. A3). However, other paths of the circuit, which have the same length, are also possible, due to which additional pulses coinciding with the main *n*th pulse are generated. First of all, such additional cases are associated with a single additional repeated reflection between the FBGs, where the reflection occurs three times, with the light reflected once from the $(n-1)$ th FBG. Such cases are denoted by numbers *2* to *4* in Fig. A3. Number *3* corresponds to

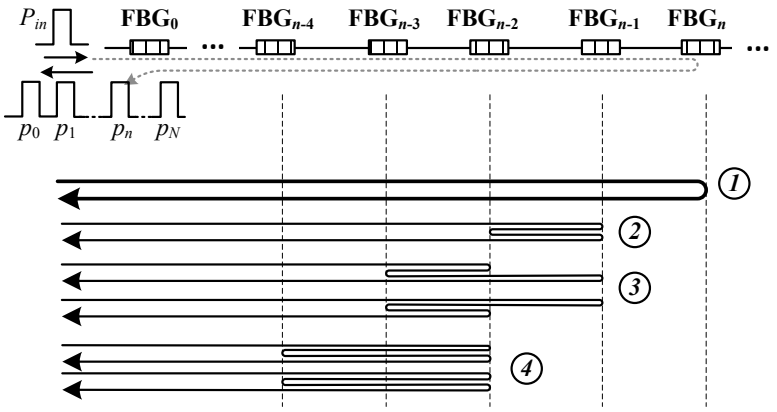


Fig. A3. Cases (*1* – *4*) of the light passing through the FBG circuit with additional *n*th pulse generated at the output

two cases with reflections from the FBG and numbers $(n-2)$ and $(n-3)$, which differ in the direction of the passage. Similar paths with repeated reflections between any two adjacent FBGs can be proposed. Importantly, in this case the additional pulse is attenuated and its power differs from the power of the main pulse by the multiplier

$$k_{crn}(k) = R_k R_{k-1} K_{fk} R_{n-1} \times \\ \times [R_n (1-R_{n-1})^2 (1-\alpha_{el(n-1)})^2 K_{fn}]^{-1}, \quad \text{(a1)}$$

assuming an additional repeated reflection from the FBG with the numbers k and $(k-1)$ (it is assumed that $k \le n$).

The first four multipliers take into account the attenuation factors of parasitic pulse added to the attenuation factors of the main pulse, and the multipliers in square brackets take into account the attenuation factors that the main pulse has but the parasitic pulse does not. Additional reflection can occur not only between neighboring FBGs, but also several positions apart (e.g., see Fig. A3, case *4*). In this more complex case, the attenuation coefficient has additional components:

$$k_{crn}(k, m) = R_k R_m (K_{fk} K_{f(k+1)} \dots K_{fm}) \times \\ \times \{R_{n-D} \cdot [R_n (1-R_{n-\Delta})^2 \times \\ \times (1-\alpha_{el(n-\Delta)})^2 K_{fn} \dots K_{f(n-\Delta)}]^{-1}\}, \quad \text{(a2)}$$

where $\Delta = k - m$; any different k and m $(k \ge m)$ less than n can be chosen.

Case *4* in Fig. A3 is the variant for $k = n-2$, $m = n-4$.

As follows from the explanations, the *n*th can travel by many additional trajectories. Although their number is finite for a particular value of N, it is difficult to comprehensively consider the general case. It seems appropriate to focus on the variant when the additional pulse has the maximum power. According to the procedure developed in this paper, equalizing the output pulse powers requires that the condition $R_n > R_{n-1}$ be satisfied. If we also assume that the losses in the elements are small and approximately equal, it is easy to see that the highest level of additional pulses with a single repeated reflection will be in case *2*, shown in Fig. A3 for $k = n$ and $n = N$. Therefore, taking into account expression (*a*1), we can introduce the following coefficient as some quantitative measure characterizing the relative level of power of additional pulses and crosstalk:

$$K_{cr} = k_{crN}(N-1) = R^2_{N-1} R_{N-2} K_{f(N-1)} \times \\ \times [R_N (1-R_{N-1})^2 (1-\alpha_{el(N-1)})^2 K_{fN}]^{-1}. \quad \text{(a3)}$$

The coefficient K_{cr} for the case where losses in circuit elements can be neglected can be written as

$$K_{cr} = k_{crN}(N-1) = R^2_{N-1} R_{N-2} K_{f(N-1)} \times \\ \times K_{cr} = R^2_{N-1} R_{N-2} / [R_N (1-R_{N-1})^2]. \quad \text{(a4)}$$

In addition to the trajectories shown in Fig. A3, the *n*th pulse can be generated by different means, due to double reflection. In this case, the pulse is reflected, for example, from the $(n-2)$th FBG and is re-reflected twice more between it and the $(n-3)$th FBG. However, it is evident that in this case the power level will have a significant additional attenuation, primarily due to additional multiplication by two FBG reflectances. Therefore, it is more logical to assume the coefficient (*a*3) to be the measure of crosstalk.

Appendix 3

Results of calculating the $R(m)$ dependences

The results of calculations of $R(m)$ dependences are shown graphically in Fig. 2 (m is the number of FBGs from the end of the circuit). However, for more accurate understanding and analysis of these results, consider the numerical values of the calculated reflectivities $R(m)$. For the case without losses, the results of the $R(m)$ calculations for the three R_N values are given in the table. The table also shows the case for $R'(m)$ with rounding to 1% to illustrate the use of an FBG with rounded coefficients R'.

Table

Calculated dependence of FBG reflectivities R on the grating number m, with and without rounding

	R, without rounding			R, rounded to 1 %	
m	$R_N = R(0)=1$	$R_N = R(0)=0.1$	$R_N = R(0)=0.05$	$R_N = R(0)=1$	$R_N = R(0)=0.1$
18	0.0318	0.0242	0.01940	0.03	0.03
15	0.0339	0.0254	0.02018	0.03	0.03
14	0.0363	0.0268	0.02102	0.04	0.03
13	0.0391	0.0283	0.02193	0.04	0.03
12	0.0424	0.0299	0.02293	0.04	0.03
11	0.0462	0.0318	0.02402	0.05	0.03
10	0.0508	0.0339	0.02521	0.05	0.04
9	0.0564	0.0364	0.02653	0.06	0.04
8	0.0633	0.0392	0.02800	0.06	0.04
7	0.0722	0.0424	0.02964	0.07	0.05
6	0.0839	0.0462	0.03147	0.08	0.05
5	0.0999	0.0508	0.03355	0.10	0.06
4	0.1234	0.0564	0.03592	0.12	0.06
3	0.1605	0.0634	0.03865	0.16	0.07
2	0.2278	0.0722	0.04182	0.23	0.08
1	0.3820	0.0839	0.04555	0.38	0.10
0	1.0000	0.1000	0.05000	1.00	0.03

This study was performed within the framework of State Assignment for Basic Research (Code FSEG-2020-0024)

REFERENCES

1. **Jones J.D.C.,** 30 Years of the OFS conference: the origins and directions of our subject, Proceedings of SPIE – The International Society for Optical Engineering. 9157 (2014) 9157C2.

2. **Udd E., Spillman Jr W. B.** (Eds.), Fiber optic sensors: an introduction for engineers and scientists, John Wiley & Sons, Hoboken, New Jersey, USA, 2011.

3. **Measures R.M.,** Structural monitoring with fiber optic technology, Academic Press, Cambridge, Massachusetts, USA, 2001.

4. **Langhammer J., Eriksrud M., Berg C., Nakstad H.,** Fiber optic permanent seismic system for increased hydrocarbon recovery, In: Proceedings of the 11th International Congress of the Brazilian Geophysical Society, European Association of Geoscientists & Engineers, Salvador, Brasil, 24–28 Aug 2009, cp-195-00408.

5. **Plotnikov M.Y., Lavrov V.S., Dmitraschenko P.Y., et al.,** Thin cable fiber-optic hydrophone array for passive acoustic surveillance applications, IEEE Sensors Journal. 19 (9) (2019) 3376–3382.

6. **Akkaya O.C., Digonnet M.J.F., Kino G.S., Solgaard O.,** Time-division-multiplexed interferometric sensor arrays, Journal of Lightwave Technology. 31 (16) (2013) 2701–2708.

7. **Kersey A.D.,** Multiplexed interferometric fiber sensors, Proceedings of the 7th Optical Fiber Sensors Conference, December 2 – 6, 1990, The Institution of Radio and Electronics

Engineers, Australia, Sydney, New South Wales (1990) 313–319.

8. **Kostromitin A.O., Liokumovich L.B., Skliarov P.V., Kotov O.I.,** The fiber-optic interferometric schemes with multiplexed sensitive elements: an analysis of output optical power level, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (2) (2020) 126–141.

9. **Liao Y., Austin E., Nash P.J., et al.,** Highly scalable amplified hybrid TDM/DWDM array architecture for interferometric fiber-optic sensor systems, Journal of Lightwave Technology. 31 (6) (2013) 882–888.

10. **Ren Z., Cui K., Zhu R., et al.,** Efficient and compact inline interferometric fiber sensor array based on fiber Bragg grating and rectangular-pulse binary phase modulation, IEEE Sensors Journal. 18 (23) (2018) 9556–9561.

11. **Yoshida M., Hirayama Y., Takahara A.,** Real-time displacement measurement system using phase-shifted optical pulse interferometry: Application to a seismic observation system, Japanese Journal of Applied Physics. 55 (2) (2016) 022701.

Received 30.09.2020, accepted 14.10.2020.

THE AUTHORS

LIOKUMOVICH Leonid B.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
leonid@spbstu.ru

KOSTROMITIN Aleksey O.
Concern CSRI "Elektropribor";
Peter the Great St. Petersburg Polytechnic University
30 Malaya Posadskaya St., St. Petersburg, 197046, Russian Federation
kostromitin.aleksei@yandex.ru

SKLIAROV Philip V.
Concern CSRI "Elektropribor";
Peter the Great St. Petersburg Polytechnic University
30 Malaya Posadskaya St., St. Petersburg, 197046, Russian Federation
sklyarov.fil@gmail.com

KOTOV Oleg I.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
kotov@rphf.spbstu.ru

DOI: 10.18721/JPM.13416

FUNCTIONAL-TYPE A POSTERIORI ERROR ESTIMATES FOR SOLUTIONS OF PROBLEMS IN DEFORMABLE SOLID MECHANICS

J. Valdman[1], *M.E. Frolov*[2]

[1] University of South Bohemia in České Budějovice, Czech Republic;

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

The paper provides a historical review and recent developments on theoretical justification and numerical implementations of functional a posteriori error estimates and adaptive algorithms for approximate solutions to problems in deformable solid mechanics. The efficient practical implementation of such methods is a relevant objective, including for modern engineering practice.

Keywords: error estimates, finite element method, adaptive algorithms, solid mechanics

Citation: Valdman J., Frolov M.E., Functional-type a posteriori error estimates for solutions of problems in deformable solid mechanics, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 13 (4) (2020) 148–155. DOI: 10.18721/JPM.13416

Introduction

Different issues related to construction of methods for numerically solving boundary-value problems and for a posteriori control of the error of the obtained approximate solutions have been the focus of attention from scientists and engineers for several decades. Today, there are thousands of publications on this subject. Due to high significance of the topic cluster *Finite Element Method*, *Galerkin Methods*, and *Errors*, it is very important to continue investigations in this direction. The issues addressed by the theory of a posteriori error control go beyond computational mathematics or engineering analysis, raising questions of mathematical modeling in general, and even philosophy.

Modern engineering practices now customarily rely on huge computational resources; the physical processes simulated become incredibly sophisticated. Problems are considered in a complex, multidisciplinary manner, at different scales of physical phenomena, with practically no simplifications, generating large amounts of data and solutions. Such problems are being successfully solved, in particular, at Peter the Great St. Petersburg Polytechnic University (see, for example, [1–3]). Since closed-source commercial software products by well-known developers are often used for these purposes, it is impossible to have a proper 'look inside' such systems to assess all the mathematical and algorithmic aspects of their implementation. This motivates experts working at the intersection of computational mathematics, physics, mechanics, engineering sciences to search for universal methods for objective estimation of the results obtained. Both technical methods and rigorous, mathematically sound approaches are being developed for this purpose, concerning in one way or another two key issues:

1. What is the accuracy of the solution obtained for a specific problem and is it always possible to estimate it reliably?

2. In which regions does the solution need to be refined (it is highly preferable that such a procedure is automated)?

Unfortunately, there can be simply no common, universal approach to estimating the error, which would have low computational complexity, making it easy to obtain new theoretical results and implement them algorithmically. Therefore, the choice of priorities is hardly a simple matter, since it is a problem extending to the philosophical domain.

Classical a posteriori estimates

It is generally agreed that modern methods for constructing a posteriori estimates have been intensively developing thanks to the groundwork laid in [4, 5], published in the late 1970s. The interest aroused by these studies was so great that several approaches aimed at estimating the accuracy of the results obtained by mathematical

and engineering simulation were formulated in the early to mid-1980s. In the decade following the mid-1990s, several monographs have been published by various authors (see, in particular, [6–9] and references therein), summarizing the experience accumulated over this time.

Thus, a separate branch evolved within the theory on the finite element method, considering the verification of numerical solutions and the validation of mathematical models. Since a posteriori estimates allow to evaluate the deviation of the approximate solution from the unknown exact solution, the error introduced by the mathematical model can be separated from the error of the numerical solution. It is only in this case that it is possible to correctly assess whether the mathematical model is suitable and adequate for analysis of a particular object.

Several groups of methods became the most popular at that time and were even implemented in commercial software products (ANSYS, MATLAB, etc.) for a number of simple boundary-value problems. Let us consider, in particular, explicit and implicit residual methods, and methods for post-processing of derivatives of an approximate solution.

While the first group of methods is much simpler than the second one, it requires for auxiliary constants to be computed so that the estimates adequately reflect not only the local distribution of the error over the computational domain (i.e., serve as indicators), but also its global value (i.e., provide reliable guaranteed upper bounds of the error). In practice, this simplistic approach faces significant difficulties, so the methods of this group are applied specifically to the adaptation of meshes.

Implicit methods circumvent the problem of computing constants by introducing the solution of a sequence of local problems with boundary conditions of different types into the structure of the methods, which increases their computational complexity but does not always yield an upper bound of the error norm.

Finally, another group of classical methods is based on the superconvergence phenomenon, which is used to construct error indicators for different problems and provides the theoretical justification for this approach. These methods involve averaging derivatives or stresses at the points of superconvergence. They are not computationally expensive and are quite simple to implement, yielding good local indication of the regions with large errors even when they do not have a rigorous mathematical justification; however, these methods do not give reliable upper bounds of the error.

In addition to the above monographs, a more detailed review of the results in Russian can be found in [10].

The key feature of the classical methods for a posteriori error control is that their rigorous mathematical justification is fundamentally based on the assumption that the given solution is not arbitrary and approximate but rather the exact solution of the corresponding finite-dimensional problem, the Galerkin approximation. This is by no means always the case in commercial packages, where the solution may not have this property due to the specifics of numerical algorithms. For example, to solve problems in the theory of plates, it is necessary to overcome the computational locking phenomenon. This is achieved, in particular, by using numerical integration formulas that are coarser in accuracy on some of the elements, producing non-Galerkin solutions.

Classical a posteriori estimates in deformable solid mechanics

Finding effective solutions to engineering problems is no less important on its own than developing the theory of numerical methods. For this reason, along with the general theory of a posteriori accuracy control as an independent branch of computational mathematics, researchers have long focused on constructing methods for adaptively solving mechanical problems. As advances were made almost simultaneously, the methods were proposed and theoretically justified based on the example of classical elliptical boundary-value problems (Poisson's equation or stationary diffusion problem), while successful concepts were extended to plane or spatial problems of linear elasticity theory, plate theory, etc. (see, for example, [11–13] and the collection in [14]).

The method of hypercircles appeared in the late 1940s [15]; it was later generalized in other studies and called the method of errors in constitutive relations (see, for example, monograph [16]). The gradient averaging method [12, 17, 18] was also extended to problems of deformable solid mechanics in 1987–1992 and is currently included in the ANSYS package as a built-in procedure for analyzing the accuracy of solutions in a number of simple problems. Different aspects of numerical implementation of the averaging procedures are discussed in detail in the literature. References to early

publications extending the explicit and implicit residual-based methods to mechanical problems are given, in particular, in monographs [6, 19] and review [20]. Details can also be found in later monographs [21, 22] and the relatively recent work [23].

Functional-type a posteriori estimates

The functional approach is based on rigorous mathematical methods, including mathematical physics, functional analysis, calculus of variations (in particular, the duality theory), theory of partial differential equations, using weak (generalized) formulations of the problems under consideration. By design, it includes an arbitrary solution of the problem from the corresponding functional space as an element whose error must be estimated. This makes the approach sufficiently general and reliable, since it allows analyzing errors in solutions regardless of how the procedures for computing these solutions are specifically implemented in commercial packages.

The Galerkin orthogonality, which is the key property for classical methods, does not play much of a role in the theory here, although it can be used in practical implementations. The estimates remain (guaranteed) upper bounds for all conforming approximations, which is true not only in theory but also in numerical implementation: this important property implies that the a posteriori error control method is reliable. This versatile and general approach is characteristic for Mikhlin's studies [24]. The functional approach was proposed and is currently developed by Prof. Sergey Repin (principal investigator at the St. Petersburg Department of the Steklov Mathematical Institute of the Russian Academy of Sciences and professor of Peter the Great St. Petersburg Polytechnic University) together with colleagues in Russia and several European countries, including Finland, Czech Republic, Austria, Switzerland and Germany. A detailed review of the advances in this direction can be found in monograph [22].

The key aspect of the approach is that it not only yields guaranteed upper bounds (that are not violated in practical implementation) but is also suitable for a wide range of approximate solutions. At the same time, the functional approach differs from the geometric arguments of the method of hypercircles, does not require construction of equilibrated fields, and has been successfully extended to nonlinear problems in theoretical aspects.

The formal starting point for this direction is 1996, when the first study [25] appeared; the approach was subsequently developed by Repin in a series of works [26–29]. Further history of the approach can be traced by the sources described in monographs [9, 22, 30].

Without going into the specifics of each particular problem, the estimate has the following general form:

$$\| u - \tilde{u} \| \leq \mathcal{M}(\tilde{u}, D, \tilde{\tilde{y}}_1, \tilde{\tilde{y}}_2, ..., C_1, C_2, ...), \quad (1)$$

where the left-hand side contains a norm of the difference between the exact solution u and an approximation $\tilde{u}$, while the right-hand side includes the functional $\mathcal{M}$ which is the deviation majorant.

The arguments of the functional are the approximate solution $\tilde{u}$, the parameters of the problem D, the set of real constants $(C_1, C_2, ...)$, which are determined by the data of the problem but not by the properties of the solution or discretization. Finally, $(\tilde{\tilde{y}}_1, \tilde{\tilde{y}}_2, ...)$ is a set of free elements that control the accuracy of the upper bound. The functional $\mathcal{M}$ vanishes if and only if $\tilde{u}$ is the exact solution; it must be explicitly computable in practice, providing an accurate estimate of the error given a reasonable (from a physical viewpoint) choice of the set $(\tilde{\tilde{y}}_1, \tilde{\tilde{y}}_2, ...)$.

Notably, there are local and global characteristics introduced in the literature for comparing various methods but only one of them, the efficiency index of an a posteriori estimate, has gained the widest popularity and is universally acclaimed as irrefutable. The efficiency index for the functional-type error majorants is calculated as the ratio of the functional $\mathcal{M}$ to the norm of the error estimated by it. Thus, for guaranteed upper bounds, its value is never lower than 1.0 that is the optimal value at which the method exactly reproduces the unknown norm of error over the entire region. There are also various local characteristics (see, for example, [22] and the references therein) but they are not as common as the efficiency index.

Practically speaking, numerical implementation of the methods of this group is not as simple as for the classical error indicators discussed above. Greater computational complexity and the additional effort it takes to construct efficient algorithms for computing functional-type a posteriori estimates are the price for the versatility and reliability of the approach. Research in this direction is still ongoing and is far from completeness. Numerical

results for plane problems in the framework of classical linear elasticity were first obtained in [31], and in [32] using an alternative, more efficient implementation. An estimate for plane problems of the Cosserat elasticity theory was obtained in [33], modified in [34]; an overview of numerical studies can be found in [23].

The first theoretical result for problems of elastoplasticity was obtained in 1996 [25]; sufficient experience was subsequently accumulated in both theoretical and practical development of the estimates (see [35–38] and the references therein, as well as related studies [39, 40]).

Finally, studies into the functional approach to solving problems in the theory of plates started with [41], considering the theory of Kirchhoff–Love thin plates. The first version of such an estimate for Reissner–Mindlin plates was obtained in 2004 [42]. A review of further results, including those concerning numerical implementation, can be found in the recent work [43].

New interesting results have been obtained for obstacle problems [44–46] and for problems in exterior domains [47], as well as for parabolic equations [48, 49] and for norms differing from the energy norm [50]. A particularly noteworthy study [46] discussed a class of nonlinear problems with free boundaries, also confirming that majorants can control not only the error of the solution in the energy space but also some measures of the distance to the exact solution for nonlinear problems. Sufficiently efficient and explicitly computable two-sided estimates are obtained in this case, which is described in more detail in the next section.

Methodology for applying the approach

Effective numerical methods for obtaining solutions that do not have additional smoothness and contain singularities should involve adaptive approaches based on a posteriori estimates. In this case, the following steps, well known in the Western literature starting from the widely cited pioneering study [51] (see also [52]), are repeated cyclically:

$$\text{SOLVE} \rightarrow \text{ESTIMATE} \rightarrow \rightarrow \text{MARK} \rightarrow \text{REFINE}. \qquad (2)$$

This approach is aimed at obtaining an approximate solution to the problem of higher accuracy with less computational efforts. However, the efficiency of the algorithm has to be substantiated for each class of problems, which does not automatically follow from the known a priori estimates of the convergence rate, because the latter were obtained under the assumption of higher regularity of the exact solution and smoothness of the boundary of the computational domain.

It is well known that the Finite Element Method (FEM) can converge slowly even for typical elliptical problems. Thus, the true convergence rate of an adaptive numerical method largely depends on the algorithmic details of the implementation. Notably, the foundations of research into adaptive FEM (AFEM) were laid by the well-known work [53].

Algorithm (2) provides wide range of opportunities for implementation. The methods can differ in finite elements for the SOLVE step, engineering or rigorous approaches to estimating the error and indicating its local distribution over the computational domain for the ESTIMATE step; an a posteriori estimate of type (1) can be used in this case. The key difference at the REFINE step is that either the algorithm for mesh refinement (*h*-version) is used, or the order of approximation (*p*-version) is increased, or a more efficient combined method (*hp*-version) is used. There is another type of mesh improvement (*rp*-version) for nonlinear problems with unknown free boundaries (for example, unknown elastoplastic zones), when the element boundary is fitted to the free boundary. In this case, the nodes of the existing finite-element mesh are shifted so that discretization is improved without having to introduce additional degrees of freedom (see, for example, [54]). Finally, different criteria for selecting mesh elements at the MARK step also have an impact.

Implementation of the functional approach in software packages can be fairly sophisticated; successful examples done in the MATLAB package are described in recent works [55–57]. The generated codes implement low-order nodal and boundary elements, vectorized and providing reasonable assembly times in both plane and spatial cases, even for discretization with a large number of unknowns (up to several millions).

It seems interesting to apply the accumulated experience to various problems of mechanics, in particular, to computations of a posteriori estimates for plane problems of the Cosserat elasticity theory, also implemented in the MATLAB package [23].

In conclusion, let us focus more closely on the the effectiveness of the functional approach.

To compute the majorant in linear problems, it is necessary to additionally minimize the quadratic functional $\mathcal{M}^2$, which leads to solving a second system of linear algebraic

equations. In this case, it is not simple standard finite elements that are of interest but rather the elements related to mixed FEM (see, for example, [23, 32, 43] for more detail). The problem of estimating the error is much more difficult mathematically than the original boundary-value problem. For this reason, it is difficult to expect that it can always be solved with low computational costs compared to the original one. In some cases, this can be achieved, but estimating the error typically takes effort (requiring both resources and complex algorithmic solutions). The complexity of computing the a posteriori estimate is rarely less than that of the solution of the initial problem.

In most cases, it is possible to obtain the estimates that majorize the unknown error with an efficiency index no greater than 2.5 (see [22, 23, 32, 43] and the references therein). With certain additional computational costs, the overestimation obtained is by no more than 1.2–1.3 times (see [23, 36], etc.). In simple elliptical boundary-value problems, it is even possible to obtain an efficiency index close to the optimal value, namely 1.0. Then the functional approach provides a result close to the exact representation of the error that has been theoretically proved to be obtainable.

Consider an example of what this implies in practice. Let the criterion for stopping the computational process be chosen when a guaranteed global solution accuracy equal to 90% is generally reached. Then, with a true error of the approximate solution (which is actually unknown) of 7% and an efficiency index of 1.3, the computations on this mesh stop, since the majorant $\mathcal{M}$, computed in this case, shows an error of less than 10%. No additional computational costs are required for analysis of mesh convergence, since there is no need to solve the problem on a finer mesh and make a comparison. With an efficiency index of up to 2.5, the estimate yields an error level of up to 20%, and a decision is made about one redundant iteration of uniform mesh refinement in order to reduce the error by half. This result is also considered acceptable, since, as noted above, the problem of estimating the accuracy is crucially more complicated than the original problem without such control.

As for adaptive algorithms, adapting the meshes allows to obtain solutions of the same level of accuracy as for uniform partitions, for a considerably smaller number of nodes. The difference reaches several tens of times even when compared with uniform partitions up to 100 thousand nodes [23]. This difference only grows subsequently, which fully compensates all computational costs required to construct estimates and error indicators.

Conclusion

The paper outlines the main historical and modern approaches to constructing reliable (guaranteed) estimates of the accuracy of approximate solutions to the problems of deformable solid mechanics. The discussion largely revolves around the methods developed within the framework of the functional approach. Efficient implementation of the methods of this group is a complex and challenging task directly applicable to modern engineering practices. This is a direction demanding much further effort.

This study was supported by the Academic Excellence Project 5-100 proposed by Peter the Great St. Petersburg Polytechnic University.

REFERENCES

1. **Nemov A., Novokshenov A., Lagutkina A., et al.,** Multiphysics engineering analysis for high field side reflectometry, Fusion Engineering and Design. 124 (November) (2017) 501–506.

2. **Antonova O., Borovkov A., Boldyrev Yu., Voynov I.,** Variational problem for hydrogenerator thrust bearing, Materials Physics and Mechanics. 34 (1) (2017) 97–102.

3. **Alekseev S., Tarasov A., Borovkov A., et al.,** Validation of EURONCAP frontal impact of frame off-road vehicle: road traffic accident simulation, Materials Physics and Mechanics. 34 (1) (2017) 59–69.

4. **Babuška I., Rheinboldt W.C.,** A-posteriori error estimates for the finite element method, International Journal for Numerical Methods in Engineering. 12 (10) (1978) 1597–1615.

5. **Babuška I., Rheinboldt W.C.,** Error estimates for adaptive finite element computations, SIAM Journal on Numerical Analysis. 15 (4) (1978) 736–754.

6. **Verfürth R.,** A review of a posteriori error estimation and adaptive mesh-refinement techniques, John Wiley & Sons, B.G. Teubner, Chichester, Stuttgart, 1996.

7. **Ainsworth M., Oden J.T.,** A posteriori error estimation in finite element analysis, John Wiley & Sons, New York, 2000.

8. **Babuška I., Strouboulis T.,** The finite element method and its reliability, The Clarendon Press Oxford University Press, New York, 2001.

9. **Neittaanmäki P., Repin S.,** Reliable methods for computer simulation: Error control and a posteriori estimates, Elsevier, Amsterdam, 2004.

10. **Repin S.I., Frolov M.E.,** Implementation of methods of mathematical physics to error control in mechanical problems, St. Petersburg State Polytechnical University Journal. 1 (49) (2007) 203–209 (in Russian).

11. **Ladevèze P., Leguillon D.,** Error estimate procedure in the finite element method and applications, SIAM Journal on Numerical Analysis. 20 (3) (1983) 485–509.

12. **Zienkiewicz O.C., Zhu J.Z.,** A simple error estimator and adaptive procedure for practical engineering analysis, International Journal for Numerical Methods in Engineering. 24 (2) (1987) 337–357.

13. **Johnson C., Hansbo P.,** Adaptive finite element methods in computational mechanics, Computer Methods in Applied Mechanics and Engineering. 101 (1–3) (1992) 143–181.

14. **Ladevèze P., Oden J.T.** (Eds.), Advances in adaptive computational methods in mechanics, Elsevier, Amsterdam, 1998.

15. **Prager W., Synge J.L.,** Approximations in elasticity based on the concept of function space, Quarterly of Applied Mathematics. 5 (3) (1947) 241–269.

16. **Ladevèze P., Pelle J.-P.,** Mastering calculations in linear and nonlinear mechanics, Springer, New York, 2005.

17. **Zhu J.Z., Zienkiewicz O.C.,** Adaptive techniques in the finite element method, Communications in Applied Numerical Methods. 4 (2) (1988) 197–204.

18. **Zienkiewicz O.C., Zhu J.Z.,** The superconvergent patch recovery (SPR) and adaptive finite element refinement, Computer Methods in Applied Mechanics and Engineering. 101 (1–3) (1992) 207–224.

19. **Bangerth W., Rannacher R.,** Adaptive finite element methods for differential equations, Birkhдuser, Basel, 2003.

20. **Verfürth R.,** A review of a posteriori error estimation techniques for elasticity problems, Computer Methods in Applied Mechanics and Engineering. 176 (1–4) (1999) 419–440.

21. **Verfürth R.,** A-posteriori error estimation techniques for finite element methods, Oxford University Press, Oxford, 2013.

22. **Mali O., Neittaanmдki P., Repin S.,** Accuracy verification methods: Theory and algorithms, Springer Science & Business Media, Dordrecht, 2014.

23. **Churilova M.A., Frolov M.E.,** Comparison of adaptive algorithms for solving plane problems of classical and Cosserat elasticity, Materials Physics and Mechanics. 32 (3) (2017) 370–382.

24. **Mikhlin S.G.,** Variational methods in mathematical physics, Translated by T. Boddington, Pergamon Press, Oxford, 1964.

25. **Repin S.I., Xanthis L.S.,** A posteriori error estimation for elastoplastic problems based on duality theory, Computer Methods in Applied Mechanics and Engineering. 138 (1–4) (1996) 317–339.

26. **Repin S.I.,** A posteriori error estimates for approximate solutions of variational problems with power growth functionals, Zapiski Nauchnykh Seminarov POMI. 249 (1997) 244–255.

27. **Repin S.I.,** A posteriori error estimation for nonlinear variational problems by duality theory, Zapiski Nauchnykh Seminarov POMI. 243 (1997) 201–214.

28. **Repin S.I.,** A unified approach to a posteriori error estimation based on duality error majorants, Mathematics and Computers in Simulation. 50 (1–4) (1999) 305–321.

29. **Repin S.I.,** A posteriori error estimation for variational problems with uniformly convex functionals, Mathematics of Computation. 69 (230) (2000) 481–500.

30. **Repin S.I.,** A posteriori estimates for partial differential equations, de Gruyter, Berlin, 2008.

31. **Muzalevsky A.V., Repin S.I.,** On two-sided error estimates for approximate solutions of problems in the linear theory of elasticity, Russian Journal of Numerical Analysis and Mathematical Modelling. 18 (1) (2003) 65–85.

32. **Frolov M.E.,** Application of functional error estimates with mixed approximations to plane problems of linear elasticity, Computational Mathematics and Mathematical Physics. 53 (7) (2013) 1000–1012.

33. **Repin S., Frolov M.,** Estimates for deviations from exact solutions to plane problems in the Cosserat theory of elasticity, Journal of Mathematical Sciences. 181 (2) (2012) 281–291, Translated from "Problems in Mathematical Analysis". (62) (2011) 153–161.

34. **Frolov M.,** Functional a posteriori estimates of the error in the solutions of plane problems in Cosserat elasticity theory, Journal of Applied Mathematics and Mechanics. 78 (4) (2014) 425–431.

35. **Repin S., Valdman J.,** Functional a posteriori error estimates for incremental models in elasto-plasticity, Central European Journal on Mathematics. 7 (3) (2009) 506–519.

36. **Neittaanmäki P., Repin S., Valdman J.,** Estimates of deviations from exact solutions of elasticity problems with nonlinear boundary conditions, Russian Journal of Numerical Analysis and Mathematical Modelling. 28 (6) (2013) 597–630.

37. **Repin S., Sysala S., Haslinger J.,** Computable majorants of the limit load in Hencky's plasticity problems, Computers and Mathematics with Applications. 75 (1) (2018) 199–217.

38. **Sysala S., Haslinger J., Repin S.,** Reliable computation and local mesh adaptivity in limit analysis, Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Hejnice, June 24–29, 2018. Institute of Mathematics CAS, Prague. (2019) 149–158.

39. **Carstensen C., Orlando A., Valdman J.,** A convergent adaptive finite element method for the primal problem of elastoplasticity, International Journal for Numerical Methods in Engineering. 67 (13) (2006) 1851–1887.

40. **Gruber P., Kienesberger J., Langer U., et al.,** Fast solvers and a posteriori error estimates in elastoplasticity, Numerical and symbolic scientific computing. Progress and prospects, Springer, New York. (2012) 45–63.

41. **Neittaanmäki P., Repin S.I.,** A posteriori error estimates for boundary-value problems related to the biharmonic operator, East-West Journal of Numerical Mathematics. 9 (2) (2001) 157–178.

42. **Repin S., Frolov M.,** Estimation of deviations from the exact solution for the Reissner – Mindlin plate problem, Journal of Mathematical Sciences. 132 (3) (2006) 331–338. Translated from Zapiski Nauchnykh Seminarov POMI, 310 (35) (2004) 145–157.

43. **Frolov M., Chistiakova O.,** Adaptive algorithm based on functional-type a posteriori error estimate for Reissner – Mindlin plates, Advanced Finite Element Methods with Applications. Selected papers from the 30th Chemnitz Finite Element Symposium, Lecture Notes in Computational Science and Engineering (LNCSE).128 (7) (2019) 131–141.

44. **Apushkinskaya D., Repin S.,** Biharmonic obstacle problem: guaranteed and computable error bounds for approximate solutions. arXiv:2003.09261v2 (2020).

45. **Apushkinskaya D., Repin S.,** Thin obstacle problem: Estimates of the distance to the exact solution, Interfaces and Free Boundaries. 20 (4) (2018) 511–531.

46. **Repin S., Valdman J.,** Error identities for variational problems with obstacles, Zeitschrift fur Angewandte Mathematik und Mechanik. 98 (4) (2018) 635–658.

47. **Pauly D., Repin S.,** A posteriori estimates for the stationary Stokes problem in exterior domains, St. Petersburg Mathematical Journal. 31 (3) (2020) 533–555.

48. **Langer U., Matculevich S., Repin S.** Guaranteed error bounds and local indicators for adaptive solvers using stabilised space–time IgA approximations to parabolic problems, Computers & Mathematics with Applications. 78 (8) (2019) 2641–2671.

49. **Matculevich S.V., Repin S.I.,** Estimates for the difference between exact and approximate solutions of parabolic equations on the basis of Poincarй inequalities for traces of functions on the boundary, Differential Equations. 52 (10) (2016) 1355–1365.

50. **Repin S.I.,** Estimates of the deviation from exact solutions of boundary value problems in measures stronger than the energy norm, Computational Mathematics and Mathematical Physics. 60 (5) (2020) 749–765.

51. **Dцrfler W.,** A convergent adaptive algorithm for Poisson's equation, SIAM Journal on Numerical Analysis. 33 (3) (1996) 1106–1124.

52. **Canuto C., Nochetto R.H., Stevenson R., Verani M.,** On p-robust saturation for hp-AFEM, Computers & Mathematics with Applications. 73 (9) (2017) 2004–2022.

53. **Babuška I., Vogelius M.,** Feedback and adaptive finite element solution of one-dimensional boundary value problems, Numerische Mathematik. 44 (1) (1984) 75–102.

54. **Franke D., Düster A., Nübel V., Rank E.,** A comparison of the *h*-, *p*-, *hp*-, and *rp*-version of the FEM for the solution of the 2D Hertzian contact problem, Computational Mechanics. 45 (5) (2010) 513–522.

55. **Rahman T., Valdman J.,** Fast MATLAB assembly of FEM matrices in 2D and 3D: nodal elements, Applied Mathematics and Computation. 219 (13) (2013) 7151–7158.

56. **Anjam I., Valdman J.,** Fast MATLAB assembly of FEM matrices in 2D and 3D: edge elements, Applied Mathematics and Computation. 267 (15 September) (2015) 252–263.

57. **Čermák M., Sysala S., Valdman J.,** Efficient and flexible MATLAB implementation of 2D and 3D elastoplastic problems, Applied Mathematics and Computation. 355 (2019) 595–614.

Received 29.09.2020, accepted 14.10.2020.

THE AUTHORS

VALDMAN Jan
University of South Bohemia in Českŭ Budejovice
1645/31A, Branišovská, České Budějovice 2. 37005, Česká Republika
jan.valdman@gmail.com

FROLOV Maxim E.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
frolov_me@spbstu.ru