MATHEMATICA MONTISNIGRI
Vol XLVII (2020) 5-212

CONTENTS

Mathematics

Computational mathematics

Mathematical modeling

Computational material research

MATHEMATICA MONTISNIGRI
Vol XLVII (2020) 5-212

Computer science applications

Equations of mathematical physics

Person

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

ON (APPROXIMATE) HOMOLOGICAL NOTIONS OF CERTAIN BANACH ALGEBRAS

AMIR SAHAMI[1*], ROSHANAK LOTFIKAR[1], AND AHMAD SOLTANMOHAMADI[2]

[1]Department of Mathematics Faculty of Basic Sciences Ilam University
P.O. Box 69315-516 Ilam, Iran.

[2]Faculty of Mathematics and Computer Science, Amirkabir University of Technology, 424 Hafez Avenue, 15914 Tehran, Iran

*Corresponding author. E-mail: a.sahami@ilam.ac.ir

DOI: 10.20948/mathmontis-2020-47-1

Summary. In this paper, we study the notion of ϕ-biflatness, ϕ-biprojectivity, approximate biprojectivity and Johnson pseudo-contractibility for a new class of Banach algebras. Using this class of Banach algebras, we give some examples which are approximately biprojective. Also some Banach algebras are given among matrix algebras which are never Johnson pseudo-contractible.

1 INTRODUCTION

Given a Banach algebra A, Kamyabi-Gol *et al.* in [4] defined a new product on A which is denoted by $*$. In fact $a * b = aeb$, for each $a,b \in A$ where e is an element of the closed unit ball $\overline{B_1^0}$ of A. A Banach algebra A equipped with $*$ as its product is denoted by A_e. They studied some properties like amenability and Arens regularity of A_e. In [6] some homological properties of A_e like biflatness, biprojectivity and ϕ −amenability discussed.

New notions of ϕ −amenability and approximate notions of homological Banach theory introduced and studied for Banach algebras see[14], [15] and [5]. In fact a Banach algebra a Banach algebra A is called approximate ϕ −contractible if there exists a net (m_α) in A such that $am_\alpha - \phi(a)m_\alpha \to 0$. and $\phi(m_\alpha) = 1$, for every $a \in A$. where ϕ is a multiplicative linear functional on A. For more information see [2]. Also a Banach algebra A is called approximate biprojective if there exists a net of bounded linear maps from A into $A \otimes_p A$, say $(\rho_\alpha)_{\alpha\in I}$, such that

1. $a \cdot \rho_\alpha(b) - \rho_\alpha(ab) \xrightarrow{||\cdot||} 0$,
2. $\rho_\alpha(ba) - \rho_\alpha(b) \cdot a \xrightarrow{||\cdot||} 0$,
3. $\pi_A \circ \rho_\alpha(a) - a \to 0$,

for every $a,b \in A$. In [1] the structure of approximate biprojective Banach algebras and its nilpotent ideals and also the relation with other notions of amenability are discussed.

We present some standard notations and definitions that we shall need in this paper. Let A be a Banach algebra. Throughout this work, the character space of A is denoted by $\Delta(A)$, that

2010 Mathematics Subject Classification: 46M10, 46H20, 46H05.
Key words and Phrases: Approximate ϕ −contratiblity, Approximate biprojectivity, ϕ −biflatness, ϕ −biprojectivity, Johnson pseudo-contractibility, ϕ −biflatness, Banach algebra.

is, all non-zero multiplicative linear functionals on A. For each $\phi \in \Delta(A)$ there exists a unique extension $\tilde{\phi}$ to A^{**} which is defined $\tilde{\phi}(F) = F(\phi)$. It is easy to see that $\tilde{\phi} \in \Delta(A^{**})$. The projective tensor product $A \otimes_p A$ is a Banach A-bimodule via the following actions

$$a \cdot (b \otimes c) = ab \otimes c. (b \otimes c) \cdot a = b \otimes ca \quad (a.b.c \in A).$$

The product morphism $\pi_A: A \otimes_p A \to A$ is given by $\pi_A(a \otimes b) = ab$, for every $a.b \in A$. Let A and B be Banach algebras. We denote by $\phi \otimes \psi$ a map defined by $\phi \otimes \psi(a \otimes b) = \phi(a)\psi(b)$ for all $a \in A$ and $b \in B$. It is easy to see that $\phi \otimes \psi \in \Delta(A \otimes_p B)$.
Let X and Y be Banach A −bimodules. The map $T: X \to Y$ is called A −bimodule morphism, if

$$T(a \cdot x) = a \cdot T(x). \quad T(x \cdot a) = T(x) \cdot a. \qquad (a \in A, x \in X).$$

Also a net of (T_α) of maps from X into Y is called approximate A −bimodule morphism, if

$$T_\alpha(a \cdot x) - a \cdot T_\alpha(x) \to 0. \quad T_\alpha(x \cdot a) - T_\alpha(x) \cdot a \to 0. \qquad (a \in A. x \in X).$$

The content of the paper is as follows. In section 2 we study ϕ −homological properties of A_e like ϕ −biflatness and ϕ −biprojectivity. Approximate biprojectivity and Johnson pseudo-contractibility are two important notions of Banach homology theory, which we discuss for A_e in section 3. We give some examples of matrix algebras to illustrate the paper.

2. ϕ - HOMOLOGICAL PROPERTIES OF CERTAIN BANACH ALGEBRAS

This section is devoted to the concepts of Banach homology related to a character ϕ.

Proposition 2.1 *[4, Proposition 2.3] Let A be a Banach algebra and $e \in \overline{B_1^0}$. Then A_e is unital if and only if A is unital and e is invertible.*

Proposition 2.2 *[4, Proposition 2.4] Let A be a Banach algebra and $e \in \overline{B_1^0}$ .Then the followings hold:*

1. If ϕ is a multiplicative linear functional on A, then $\phi(e)\phi$ is a multiplicative linear functional on A_e.
2. If A_e is unital and ψ is a multiplicative linear functional on A_e, then $\phi(a) = \psi(e^{-1}a)$ is a multiplicative linear functional on A.

Proposition 2.3 *[6, Proposition 2.3] Let A be a Banach algebra and $e \in \overline{B_1^0}$. If A_e is unital then $(A_e)_{e^{-2}}$ = A, (isometrically isomorphism).*

Proposition 2.4 *Suppose that A is a Banach algebra and also suppose that $e \in \overline{B_1^0}$ and $\phi \in \Delta(A)$. Then the followings hold:*

1. If A is approximate ϕ −contractible and $\phi(e) \neq 0$, then A_e is approximately ψ-contractible, where $\psi = \phi(e)\phi$.
2. If A_e is unital and approximate ψ −contractible, then A is approximate ϕ-contractible, where $\phi(a) = \psi(e^{-1}a)$ for each $a \in A$.

Proof. Suppose that A is approximately ϕ −contractible. So there is a net (m_α) in A such that

$$am_\alpha - \phi(a)m_\alpha \to 0. \quad \phi(m_\alpha) = 1. \qquad (a \in A).$$

Define $n_\alpha = \frac{m_\alpha}{\phi(e)}$. Since $\psi(a) = \phi(ae) = \phi(ea)$, we have

$$\begin{aligned} a * n_\alpha - \psi(a)n_\alpha &= aen_\alpha - \psi(a)n_\alpha \\ &= ae\frac{m_\alpha}{\phi(e)} - \psi(a)\frac{m_\alpha}{\phi(e)} \\ &= ae\frac{m_\alpha}{\phi(e)} - \phi(ae)\frac{m_\alpha}{\phi(e)} + \phi(ae)\frac{m_\alpha}{\phi(e)} - \psi(a)\frac{m_\alpha}{\phi(e)} \to 0. \qquad (a \in A_e). \end{aligned}$$

Also

$$\psi(n_\alpha) = \psi(\tfrac{m_\alpha}{\phi(e)}) = \phi(e)\phi(\tfrac{m_\alpha}{\phi(e)}) = \phi(m_\alpha) = 1.$$

It follows that A_e is approximate ϕ −contractible.

Suppose that $\phi(a) = \psi(e^{-1}a)$ and also suppose that A_e is unital and approximately left ψ −contractible. It is easy to see that $\psi(a) = \phi(ea)$. Let (m_α) be a net in A_e such that

$$a * m_\alpha - \psi(a)m_\alpha \to 0. \quad \psi(m_\alpha) = 1. \qquad (a \in A_e).$$

Since

$$\begin{aligned} a * m_\alpha - \psi(a)m_\alpha &= aem_\alpha - \psi(a)m_\alpha \\ &= aem_\alpha - \phi(ea)m_\alpha \\ &= aem_\alpha - \phi(e)\phi(a)m_\alpha \\ &= aem_\alpha - \phi(a)\phi(e)m_\alpha \\ &= aem_\alpha - \phi(ae)m_\alpha, \end{aligned}$$

we have

$$a * m_\alpha - \psi(a)m_\alpha = aem_\alpha - \phi(ae)m_\alpha \to 0$$

for each $a \in A$. Replacing a with ae^{-1} we have $am_\alpha - \phi(a)m_\alpha \to 0$. Regarding

$$1 = \psi(m_\alpha) = \phi(em_\alpha) = \phi(e)\phi(m_\alpha).$$

we may suppose that $\phi(m_\alpha) \neq 0$, for each α. Now define $n_\alpha = \frac{m_\alpha}{\phi(m_\alpha)}$. Clearly $\phi(n_\alpha) = 1$. Also

$$an_\alpha - \phi(a)n_\alpha = a\frac{m_\alpha}{\phi(m_\alpha)} - \phi(a)\frac{m_\alpha}{\phi(m_\alpha)} \to 0.$$

It finishes the proof.

Example 2.5 *In this example we show that there exists a Banach algebra A_e which is not approximate ψ-contractible. Let* $A = \{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix} | a_{ij} \in \mathbb{C}\}$ *and suppose that* $e =$

$\begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{1}{6} \end{pmatrix}$. *Clearly A with matrix operations and ℓ^1-norm is a Banach algebra. We know that e is invertible and by Proposition 2.1, A_e is unital. Define $\phi: A \to \mathbb{C}$ by*

$$\phi(\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}) = a_{33}.$$

Clearly ϕ is a character(multiplicative linear functional) and $\phi(e) \neq 0$. Suppose conversely that A_e is approximate ψ −contractible. By previous Proposition(2), A becomes approximate ϕ −contractible. On the other hand by the same arguments as in the proof of [7, Theorem 5.1] A is not approximate ϕ −contractible, which is a contradiction.

Let A be a Banach algebra and $\phi \in \Delta(A)$. A is called ϕ-biprojective, if there exists a bounded A-bimodule morphism $\rho: A \to A \otimes_p A$ such that $\phi \circ \pi_A \circ \rho = \phi$. Also A is called ϕ-biflat if there exists a bounded A-bimodule morphism $\rho: A \to (A \otimes_p A)^{**}$ such that $\tilde{\phi} \circ \pi_A^{**} \circ \rho = \phi$. For more information about ϕ −biflatness and ϕ −biprojectivity, the reader refers to [8] and [9].

Theorem 2.6 *Let A be a Banach algebra and $\phi \in \Delta(A)$. Suppose that $e \in \overline{B_1^0}$ and $\phi(e) \neq 0$. If A is ϕ-biprojective, then A_e is $\psi = \phi(e)\phi$-biprojective.*

Proof. Since A is ϕ-biprojective, there exists a bounded A −bimodule morphism $\rho: A \to A \otimes_p A$ such that $\phi \circ \pi_A \circ \rho = \phi$. Define $\tilde{\rho} = \frac{1}{\phi(e)}\rho$. We show that $\tilde{\rho}$ is a bounded A_e-bimodule morphism. To see this, consider

$$\begin{aligned} \tilde{\rho}(a * b) = \frac{1}{\phi(e)}\rho(a * b) = \frac{1}{\phi(e)}\rho(aeb) \quad &= ae\frac{1}{\phi(e)}\rho(b) \\ &= a * \frac{1}{\phi(e)}\rho(b) \\ &= a * \tilde{\rho}(b). \qquad (a.b \in A_e). \end{aligned}$$

Also

$$\begin{aligned} \tilde{\rho}(a * b) = \frac{1}{\phi(e)}\rho(a * b) = \frac{1}{\phi(e)}\rho(aeb) \quad &= \frac{1}{\phi(e)}\rho(a)be \\ &= \frac{1}{\phi(e)}\rho(a) * b \\ &= \tilde{\rho}(a) * b. \qquad (a.b \in A_e). \end{aligned}$$

On the other hand, since

$$\psi \circ \pi_{A_e} \circ \tilde{\rho} = \phi(e)\phi \circ \pi_A \circ \rho.$$

we have

$$\psi \circ \pi_{A_e} \circ \tilde{\rho}(a) = \phi(e)\phi \circ \pi_A \circ \rho(a) = \phi(e)\phi(a) = \psi(a). \quad (a \in A_e).$$

So A_e is ψ −biprojective.

Using the similar arguments as in the proof of the previous theorem, we have the following corollary:

Corollary 2.7 *Let A be a Banach algebra and $\phi \in \Delta(A)$. Suppose that $e \in \overline{B_1^0}$ and $\phi(e) \neq 0$. If A is ϕ-biflat, then A_e is $\psi = \phi(e)\phi$-biflat.*

Let A be a Banach algebra and $\phi \in \Delta(A)$. A is called ϕ −amenable if there exists a bounded net (m_α) in A such that $am_\alpha - \phi(a)m_\alpha \to 0$ and $\phi(m_\alpha) = 1$, for every $a \in A$. see [5].

Corollary 2.8 *Let A be a Banach algebra and $\phi \in \Delta(A)$. Suppose that $e \in \overline{B_1^0}$ and $\phi(e) \neq 0$. If A is ϕ-biflat and A has a left approximate identity, then A_e is approximate $\psi = \phi(e)\phi$-contractible.*

Proof. Since A is ϕ-biflat and A has a left approximate identity, by similar arguments as in the proof of [7, Theorem 2.2] A is ϕ-amenable. It is easy to see that ϕ-amenability of A implies that A is approximate ϕ −contractible. Applying Proposition 2.4, A_e becomes approximate ψ −contractible.

Let A b a Banach algebra and $\phi \in \Delta(A)$. Then A is called approximate left ϕ-biprojective if there exists a net of bounded linear maps from A into $A \otimes_p A$, say $(\rho_\alpha)_{\alpha \in I}$, such that

1. $\rho_\alpha(ab) - \phi(a)\rho_\alpha(b) \xrightarrow{||\cdot||} 0$,
2. $\rho_\alpha(ba) - \rho_\alpha(b) \cdot a \xrightarrow{||\cdot||} 0$,
3. $\pi_A \circ \rho_\alpha(a) - a \to 0$,

for every $a, b \in A$, see [12].

Theorem 2.9 *Let A be a Banach algebra and $\phi \in \Delta(A)$. Suppose that $e \in \overline{B_1^0}$ and $\phi(e) \neq 0$. If A is approximate left ϕ-biprojective, then A_e is approximate left $\psi = \phi(e)\phi$-biprojective.*

Proof. Since A is approximate left ϕ-biprojective, there exists a net of bounded linear maps (ρ_α) from A into $A \otimes_p A$ such that

$$\rho_\alpha(ab) - \phi(a)\rho_\alpha(b) \to 0, \quad \rho_\alpha(ab) - \rho_\alpha(a) \cdot b \to 0, \quad \phi \circ \pi_A \circ \rho(a) - \phi(a) \to 0.$$

Define $\tilde{\rho}_\alpha = \frac{1}{\phi(e)}\rho_\alpha$. We show that there exists a net of bounded linear maps $(\tilde{\rho}_\alpha)$ from A_e in to $A_e \otimes_p A_e$ such that

$$\tilde{\rho}_\alpha(a * b) - \psi(a)\tilde{\rho}_\alpha(b) \to 0. \quad \tilde{\rho}_\alpha(a * b) - \tilde{\rho}_\alpha(a) * b \to 0. \quad \psi \circ \pi_A \circ \tilde{\rho}(a) - \psi(a) \to 0.$$

To see this, consider

$$\begin{aligned}
\tilde{\rho}_\alpha(a * b) - \psi(a)\tilde{\rho}_\alpha(b) &= \tilde{\rho}_\alpha(aeb) - \phi(a)\phi(e)\tilde{\rho}_\alpha(b) \\
&= \frac{1}{\phi(e)}(\rho_\alpha(aeb) - \phi(a)\phi(e)\rho_\alpha(b)) \\
&= \frac{1}{\phi(e)}(\rho_\alpha(aeb) - \phi(ae)\rho_\alpha(b) + \phi(ae)\rho_\alpha(b) - \phi(a)\phi(e)\rho_\alpha(b)) \\
&\to 0
\end{aligned}$$

Also

$$\tilde{\rho}_\alpha(a * b) - \tilde{\rho}_\alpha(a) * b = \frac{1}{\phi(e)}\rho_\alpha(aeb) - \frac{1}{\phi(e)}\rho_\alpha(a)eb \to 0$$

On the other hand, since

$$\psi \circ \pi_{A_e} \circ \tilde{\rho}_\alpha = \phi(e)\phi \circ \pi_A \circ \rho_\alpha.$$

we have for $(a \in A_e$

$$\psi \circ \pi_{A_e} \circ \tilde{\rho}_\alpha(a) - \psi(a) = \phi(e)\phi \circ \pi_A \circ \rho_\alpha(a) - \phi(e)\phi(a) \to \phi(e)\phi(a) - \phi(e)\phi(a) = 0.$$

So A_e is approximate left ψ −biprojective.

Remark 2.10 *Let A and B be Banach algebras and* $e_A \in \overline{B_1^0}^A$ *and* $e_B \in \overline{B_1^0}^B$ *. Then there exist two sequences* (x_n) *and* (y_n) *in the unit ball A and the unit ball B such that* $x_n \to e_A$ *and* $y_n \to e_B$*. respectively. Since*

$$||x_n \otimes y_n - e_A \otimes e_B|| \leq ||x_n \otimes y_n - e_A \otimes y_n|| + ||e_A \otimes y_n - e_A \otimes e_B|| \to 0.$$

we have $e_A \otimes e_B \in \overline{B_1^0}^{A\otimes_p B}$. Define $T: A_{e_A} \otimes_p B_{e_B} \to A \otimes_p B_{e_A \otimes e_B}$ by $T(a \otimes b) = a \otimes b$ for every $a \in A$ and $b \in B$. It is easy to see that T is an isometric algebra isomorphism. Also T is a bounded $A \otimes_p B_{e_A \otimes e_B}$ −bimodule morphism.

Proposition 2.11 *Let A and B be Banach algebras and* $e_A \in \overline{B_1^0}^A$ *and* $e_B \in \overline{B_1^0}^B$ *. Suppose that* $\phi_A \in \Delta(A)$ *and* $\phi_B \in \Delta(B)$ *which* $\phi_A(e_A) \neq 0$ *and* $\phi_B(e_B) \neq 0$*. If A and B are* ϕ_A *−biprojective and* ϕ_B *−biprojective, respectively, then* $A \otimes_p B_{e_A \otimes e_B}$ *is* $\phi_A(e_A)\phi_A \otimes \phi_B(e_B)\phi_B$ *−biprojective.*

Proof. Since A and B are ϕ_A −biprojective and ϕ_B −biprojective, respectively, then by Theorem 2.9, A_e and B_e are $\phi_A(e_A)\phi_A$ −biprojective and $\phi_B(e_B)\phi_B$ −biprojective, respectively. So there exist a A_{e_A} −bimodule morphism $\rho_0: A_{e_A} \to A_{e_A} \otimes_p A_{e_A}$ and a B_{e_B}-bimodule morphism $\rho_1: B_{e_B} \to B_{e_B} \otimes_p B_{e_B}$ such that $\phi_A(e_A)\phi_A \circ \pi_A \circ \rho_0 = \phi_A(e_A)\phi_A$ and $\phi_B(e_B)\phi_B \circ \pi_B \circ \rho_1 = \phi_B(e_B)\phi_B$.

Define $\theta: (A_{e_A} \otimes_p A_{e_A}) \otimes_p (B_{e_B} \otimes_p B_{e_B}) \to (A_{e_A} \otimes_p B_{e_B}) \otimes_p (A_{e_A} \otimes_p B_{e_B})$ by

$$(a_1 \otimes a_2) \otimes (b_1 \otimes b_2) \mapsto (a_1 \otimes b_1) \otimes (a_2 \otimes b_2).$$

where $a_1.a_2 \in A$ and $b_1.b_2 \in B$. Clearly θ is an isometric algebra isomorphism. Set $\rho = (T \otimes T) \circ \theta \circ (\rho_0 \otimes \rho_1) \circ T^{-1}$, where T is the map defined as in Remark 2.10. We know that ρ is a bounded linear map from $A \otimes_p B_{e_A \otimes e_B}$ into $(A \otimes_p B_{e_A \otimes e_B}) \otimes_p (A \otimes_p B_{e_A \otimes e_B})$. Consider

$$\pi_{A\otimes_p B_{e_A\otimes e_B}} \circ \theta(a_1 \otimes a_2 \otimes b_1 \otimes b_2) = \pi_{A\otimes_p B_{e_A\otimes e_B}}(a_1 \otimes b_1 \otimes a_2 \otimes b_2)$$
$$= \pi_{A_{e_A}}(a_1 \otimes a_2) \otimes \pi_{B_{e_B}}(b_1 \otimes b_2).$$

then clearly one can show that $\pi_{A\otimes_p B_{e_A\otimes e_B}} \circ \theta = \pi_{A_{e_A}} \otimes \pi_{B_{e_B}}$. Hence,

$$\pi_{A\otimes_p B_{e_A\otimes e_B}} \circ \theta(\rho_0(a) \otimes \rho_1(b)) = \pi_{A_{e_A}} \circ \rho_0(a) \otimes \pi_{B_{e_B}} \circ \rho_1(b)$$

and it is easy to see that

$$\phi_A(e_A)\phi_A \otimes \phi_B(e_B)\phi_B \circ \pi_{A\otimes_p B} \circ \theta(\rho_0 \otimes \rho_1)(a \otimes b) = \phi_A(e_A)\phi_A \otimes \phi_B(e_B)\phi_B(a \otimes b).$$

the proof is complete.

3 APPROXIMATE HOMOLOGICAL PROPERTIES OF CERTAIN BANACH ALGEBRAS

In this section we investigate approximate biprojectivity and Johnson pseudo-contractibility of A_e.

Theorem 3.1 *Suppose that A is a Banach algebra and also suppose that $e \in \overline{B_1^0}$. Then the followings hold:*

1. If A is approximately biprojective and A_e is unital then A_e is approximately biprojective.
2. If A_e is unital and approximately biprojective, then A is approximately biprojective.

Proof. To show (1), suppose that A is approximately biprojective and A_e is unital. It follows that there is an approximately A −bimodule morphism (ρ_α) from A into $A \otimes_p A$ such that $\pi_A \circ \rho_\alpha(a) - a \to 0$ for each $a \in A$. Note that

$$\begin{aligned}\rho_\alpha(a * b) - a * \rho_\alpha(b) &= \rho_\alpha(aeb) - a * \rho_\alpha(b)\\ &= \rho_\alpha(aeb) - ae\rho_\alpha(b) + ae\rho_\alpha(b) - a * \rho_\alpha(b) \to 0.\end{aligned}$$

and

$$\begin{aligned}\rho_\alpha(a * b) - \rho_\alpha(a) * b &= \rho_\alpha(aeb) - \rho_\alpha(a) * b\\ &= \rho_\alpha(aeb) - \rho_\alpha(a)eb + \rho_\alpha(a)eb - \rho_\alpha(a) * b \to 0,\end{aligned}$$

for each $a \in A_e$. It implies that (ρ_α) from A_e into $A_e \otimes_p A_e$ is an approximately A_e-bimodule morphism. Define $T: A_e \otimes_p A_e \to A_e \otimes_p A_e$ by $T(a \otimes b) = ae^{-1} \otimes b$. Note that using **Proposition 2.1**, the definition of T makes sense. It is easy to see that

$$T(a * (c \otimes d)) = a * T(c \otimes d). \quad T((c \otimes d) * a) = T(c \otimes d) * a. \qquad (a.c.d \in A).$$

Set $\tilde{\rho}_\alpha = T \circ \rho_\alpha$. Using direct calculations we can see that

$$\pi_{A_e} \circ \tilde{\rho}_\alpha = \pi_A \circ \rho_\alpha$$

It follows that $\pi_{A_e} \circ \tilde{\rho}_\alpha - a = \pi_A \circ \rho_\alpha - a \to 0. \qquad (a \in A_e)$.

Thus A_e is approximately biprojective.

To show (2), suppose that A_e is unital and approximately biprojective. By Proposition 2.3, we know that $A = (A_e)_{e^{-2}}$. Now applying (1) it is easy to see that A is approximately biprojective.

A Banach algebra A is called biprojective if there exists a bounded A −bimodule morphism $\rho: A \to A \otimes_p A$ such that $\pi_A \circ \rho(a) = a$ for each $a \in A$. see [13].

Example 3.2 *In this example we give a Banach algbra A_e which is approximately biprojective. Let $A = \{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} | a_{ij} \in \mathbb{C}\}$. With the matrix operations and ℓ^1-norm, A becomes a Banach algebra. Suppose that $e = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix}$. Clearly e is invertible and A is unital. So by Proposition 2.1, A_e is unital. It is well-known that A is biprojective, see [13]. So A is approximately biprojective. Applying previous theorem A_e becomes approximately biprojective.*

Definition 3.3 *We say that a Banach algebra A has approximate (F)-property(or A is AFP) if there is an approximate A −bimodule morphsim (ρ_α) from A into $(A \otimes_p A)^{**}$ such that $\pi_A^{**} \circ \rho_\alpha(a) - a \to 0$. for each $a \in A$.*

For the motivation of this definition see [3].

Proposition 3.4 *If A is AFP and A_e is unital, then A_e is approximately biprojective.*

Proof. Since A is AFP, there exists an approximate A −bimodule morphsim (ρ_α) from A into $(A \otimes_p A)^{**}$ such that $\pi_A^{**} \circ \rho_\alpha(a) - a \to 0$, for each $a \in A$. It is easy to see that (ρ_α) is an approximate A_e −bimodule morphsim from A_e into $(A_e \otimes_p A_e)^{**}$ such that $\pi_{A_e}^{**} \circ \rho_\alpha(a) - a \to 0$. for each $a \in A_e$. Let $T: A_e \otimes_p A_e \to A_e \otimes_p A_e$ be the same map as in the proof of Theorem 3.1. Clearly T is A_e-module morphism, so is T^{**}. Similar to the proof of Theorem 3.1, for the net $(T^{**} \circ \rho_\alpha)$ is an approximate A_e −bimodule morphism from A_e into $(A_e \otimes_p A_e)^{**}$ such that

$$\pi_{A_e}^{**} \circ T^{**} \circ \rho_\alpha(a) - a = \pi_A^{**} \circ \rho_\alpha(a) - a \to 0. \quad (a \in A).$$

We denote the identity of A_e with a_0 and define $m_\alpha = \rho_\alpha(a_0)$. Clearly (m_α) is a net in $(A_e \otimes_p A_e)^{**}$ which satisfies

$$a * m_\alpha - m_\alpha * a \to 0. \qquad \pi_{A_e}^{**}(m_\alpha) * a - a \to 0. \quad (a \in A_e).$$

Take $\epsilon > 0$ and arbitrary finite subsets $F \subseteq A_e$, $\Lambda \subseteq (A_e \otimes_p A_e)^*$ and $\Gamma \subseteq A_e^*$. Then we have

$$||a * m_\alpha - m_\alpha * a|| < \epsilon. \quad ||\pi_{A_e}^{**}(m_\alpha) * a - a|| < \epsilon. \quad (a \in F).$$

It is well-known that for each α, there exists a net $(n_\beta^\alpha)_\beta$ in $A_e \otimes_p A_e$ such that $n_\beta^\alpha \xrightarrow{w^*} m_\alpha$. Since $\pi_{A_e}^{**}$ is a w^*-continuous map, we have $\pi_{A_e}(n_\beta^\alpha) = \pi_{A_e}^{**}(n_\beta^\alpha) \xrightarrow{w^*} \pi_A^{**}(m_\alpha)$.

Thus we have $|a * n_\beta^\alpha(f) - a * m_\alpha(f)| < \frac{\epsilon}{K_0}$. $|n_\beta^\alpha * a(f) - m_\alpha * (f)| < \frac{\epsilon}{K_0}$

and
$$|\pi_{A_e}(n_\beta^\alpha)(g) - \pi_{A_e}^{**}(m_\alpha)(g)| < \frac{\epsilon}{K_1}.$$

for each $a \in F$, $f \in \Lambda$ and $g \in A^*$, where $K_0 = \sup\{||f||: f \in \Lambda\}$ and $K_1 = \sup\{||g||: g \in \Gamma\}$. Since $a * m_\alpha - m_\alpha * a \to 0$ and $\pi_{A_e}^{**}(m_\alpha) * a - a \to 0$. we can find $\beta = \beta(F.\Lambda.\Gamma.\epsilon)$ such that

$$|a * n_\beta^\alpha(f) - n_\beta^\alpha * a(f)| < c\frac{\epsilon}{K_0} \quad |\pi_{A_e}(n_\beta^\alpha) * a(g) - a(g)| < \frac{\epsilon}{K_1}. \quad (a \in F.f \in \Lambda.g \in \Gamma)$$

for some $c \in \mathbb{R}^+$. Using Mazur's lemma, we have a net $(n_{(F.\Lambda.\Gamma.\epsilon)})$ in $A_e \otimes_p A_e$ such that

$$||a * n_{(F.\Lambda.\Gamma.\epsilon)} - n_{(F.\Lambda.\Gamma.\epsilon)} * a|| \to 0. \quad ||\pi_A(n_{(F.\Lambda.\Gamma.\epsilon)}) * a - a|| \to 0. \quad (a \in F).$$

Define $\rho_{(F.\Lambda.\Gamma.\epsilon)}: A_e \to A_e \otimes_p A_e$ by $\rho_{(F.\Lambda.\Gamma.\epsilon)}(a) = a * n_{(F.\Lambda.\Gamma.\epsilon)}$ for each $a \in A_e$. It is clear that $\rho_{(F.\Lambda.\Gamma.\epsilon)}(a * b) = a * \rho_{(F.\Lambda.\Gamma.\epsilon)}(b)$ for each $a, b \in A$. Also

$$\begin{aligned} ||\rho_{(F.\Lambda.\epsilon)}(a * b) - \rho_{(F.\Lambda.\Gamma.\epsilon)}(a) * b|| &= ||ab * n_{(F.\Lambda.\Gamma.\epsilon)} - a * (n_{(F.\Lambda.\Gamma.\epsilon)} * b)|| \\ &\leq ||a||||b * n_{(F.\Lambda.\Gamma.\epsilon)} - n_{(F.\Lambda.\Gamma.\epsilon)} * b|| \to 0. \end{aligned} \tag{3.1}$$

for each $a.b \in A_e$. Also

$$\begin{aligned} ||\pi_{A_e} \circ \rho_{(F.\Lambda.\Gamma.\epsilon)}(a) - a|| &= ||\pi_{A_e}(a * n_{(F.\Lambda.\Gamma.\epsilon)}) - a|| \\ &= ||\pi_{A_e}(a * n_{(F.\Lambda.\Gamma.\epsilon)}) - \pi_{A_e}(n_{(F.\Lambda.\Gamma.\epsilon)} * a) + \pi_{A_e}(n_{(F.\Lambda.\Gamma.\epsilon)} * a) - a|| \\ &\leq ||\pi_{A_e}(a * n_{(F.\Lambda.\Gamma.\epsilon)}) - \pi_{A_e}(n_{(F.\Lambda.\Gamma.\epsilon)} * a)|| + ||\pi_{A_e}(n_{(F.\Lambda.\Gamma.\epsilon)}) * a - a|| \\ &\to 0. \end{aligned}$$

for each $a \in F$. Thus with respect to the net $(\rho_{(F.\Lambda.\Gamma.\epsilon)})_{(F.\Lambda.\Gamma.\epsilon)}$. A_e becomes approximately biprojective.

A Banach algebra A is called Johnson pseudo-contractible, if there exists a not necessarily bounded net (m_α) in $(A \otimes_p A)^{**}$ such that $a \cdot m_\alpha = m_\alpha \cdot a$ and $\pi_A^{**}(m_\alpha)a - a \to 0$. for every $a \in A$, see [11] and [10].

A Banach algebra A is called biflat, if there is a bounded A −bimodule morphsim ρ from A into $(A \otimes_p A)^{**}$ such that $\pi_A^{**} \circ \rho_\alpha(a) = a$, for each $a \in A$, see [13].

Proposition 3.5 *Let A be a Banach algebra and $e \in \overline{B_1^0}$. Suppose that A_e is unital. Then A is Johnson pseudo-contractible if and only if A_e is Johnson pseudo-contractible.*

Proof. Since A_e is unital, by Proposition 2.1 A is unital. So using [3, Theorem 2.1], Johnson pseudo-contractibility of A implies that A is amenable. Thus by [13, Exercise 4.3.15], A is biflat. Then by [6, Theorem 2.4] A_e is biflat. Since A_e is unital, biflatness of A_e gives the amenability of A_e.

For converse, suppose that A_e is Johnson pseudo-contractible. Since A_e is unital by [3, Theorem 2.1] A_e is amenable, so is biflat. Applying [6, Theorem 2.4] follows that A is biflat.

Using Proposition 2.1, A is unital, thus by [13, Exercise 4.3.15] A is amenable. So [11, Lemma 2.1] implies that A is Johnson pseudo-contractible.

Example 3.6 *We give a Banach algebra A_e which is not Johnson pseudo-contractible. Let*

$A = \{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix} | a_{ij} \in \mathbb{C}\}$ *and suppose that* $e = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix}$. *Clearly e is invertible and A is unital. So by Proposition 2.1 A_e is unital. Using [11, Theorem 2.5] we know that A is not Johnson pseudo-contractible. So by previous proposition A_e is not Johnson pseudo-contractible.*

REFERENCES

[1] H. P. Aghababa, "Approximately biprojective Banach algebras and nilpotent ideals". *Bull. Aust. Math. Soc.* **87,** 1-12 (2012).

[2] H. P. Aghababa, L. Y. Shi and Y. J. Wu, "Generalized notions of character amenability". *Acta Math. Sin,* **29**, 1329-1350 (2013).

[3] M. Askari-Sayah, A. Pourabbas and A. Sahami, "Johnson pseudo-contractibility of certain Banach algebras and its nilpotent ideals". *Analysis Mathematica,* **3,** 461-473 (2019).

[4] R. A. Kamyabi-Gol and M. Janfada, "Banach algebras related to the elements of the unit ball of a Banach algebra.", *Taiwan. J. Math.,* **12**, 1769-1779 (2008).

[5] E. Kaniuth, A. T. Lau and J. Pym, "On ϕ-amenability of Banach algebras", *Math. Proc. Camb. Philos. Soc.* **44,** 85-96 (2008).

[6] A. R. Khoddami, " Biflatness, biprojectivity, ϕ −amenability and ϕ −contractibility of certain Banach algebras", *U.P.B. Sci. Bull., Series A,* **80** (2), 81-86 (2018).

[7] A. Sahami, "On biflatness and ϕ −biflatness of some Banach algebras", *U.P.B. Sci. Bull., Series A*, Vol. **80**, Iss. **2**, 112-123 (2018).

[8] A. Sahami and A. Pourabbas, "On ϕ-biflat and ϕ-biprojective Banach algebras", *Bull. Belg. Math. Soc. Simon Stevin*, **20,** 789-801 (2013).

[9] A. Sahami and A. Pourabbas, "Approximate biprojectivity and ϕ-biflatness of some Banach algebras", *Colloq. Math,* **145,** 273-284 (2016).

[10] A. Sahami and A. Pourabbas, " Johnson pseudo-contractibility of various classes of Banach algebras", *Bull. Belg. Math. Soc. Simon Stevin*, **25,** 171-182 (2018).

[11] A. Sahami and A. Pourabbas, " *Johnson pseudo-contractibility of certain semigroup algebras*", *Semigroup Forum,* **97,** 203-207 (2018).

[12] A. Sahami and A. Pourabbas, "On approximate left ϕ-biprojective Banach algebras", *Glasnik Matematicki*, Vol. **53**, 187-203 (2018).

[13] V. Runde, *Lectures on Amenability*, Springer, New York, (2002).

[14] E. Samei, N. Spronk and R. Stokke, "Biflatness and pseudo-amenability of Segal algebras", *Canad. J. Math*. **62,** 845-869 (2010).

[15] Y. Zhang, " Nilpotent ideals in a class of Banach algebras", *Proc. Amer. Math. Soc.* **127** , 3237-3242 (1999).

Received October 31, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

CONGRUENCES INVOLVING SUMS OF HARMONIC NUMBERS AND BINOMIAL COEFFICIENTS

LAID ELKHIRI[1*], MILOUD MIHOUBI[2] and ABDELLAH DERBAL[3]

[1]Univercity of Tiaret, EDTNLHM Laboratory of ENS Kouba, Algiers, Algeria
[2]USTHB, RECITS Laboratory, Bab-Ezzouar, Algiers, Algeria
[3]ENS Old Kouba, EDTNLHM Laboratory, Algiers, Algeria

*Corresponding author. E-mail: elkhirialjebre@hotmail.fr or laid@univ-tiaret.dz

DOI: 10.20948/mathmontis-2020-47-2

Summary. Congruences involving sums of Harmonic numbers and binomial coefficients are considered in this paper. Recently, many great mathematicians have been interested to find congruences and relationships between these numbers such Sun & Tauraso, Koparal & Ömür, Mao & Sun and Meštrović & Andjić. In the present paper, some new combinatorial congruences are proved. These congruences are mainly determined modulo p^2 or p^3 (p in any prime) and they are motivated by a recent paper by Meštrović and Andjić. The first main result (Theorem 1) presents the congruence modulo p^2 ($p > 3$ is any prime) involving sum of products of two binomial coefficients and Harmonic numbers. Two interesting congruences modulo a prime $p > 3$ (Corollary 2) involving Harmonic numbers H_k, Catalan numbers C_k and Fermat quotient $q_2 := (2^{p-1} - 1)/p$ are obtained as consequences of Theorem 1. The second main result (Theorem 2) presents the congruence modulo p^3 ($p > 3$ is any prime) involving sum of products of two binomial coefficients and Harmonic numbers.

1. INTRODUCTION AND MAIN RESULTS

The harmonic number and the congruence in the ring of p integer $\mathbb{Z}_p$ ply important role in mathematics. Recall that harmonic numbers are to be

$$H_0 = 0, \qquad H_n = \sum_{k=1}^{n} \frac{1}{k}, \qquad n \geq 1,$$

$\mathbb{Z}_p$ is the set of rational numbers having denominators not divisible by p and the unit group $U(\mathbb{Z}_p)$ is the set of rational numbers having denominators and numerators not divisible by p.

We define, for all prime number p and for all numbers $x, y \in \mathbb{Z}_p$

$$x \equiv y \ (mod\ p) \Leftrightarrow numerator(x - y) \equiv 0 \ (mod\ p).$$

This shows when $x, y \in U(\mathbb{Z}_p)$ that

$$x \equiv y \ (mod\ p) \Leftrightarrow \frac{1}{x} \equiv \frac{1}{y} \ (mod\ p).$$

Congruences involving sums of Harmonic numbers and binomial coefficients in the ring of p integer have been studied recently by many mathematicians and a considerable amount of research results has been produced, such in 2011 Sun and Tauraso [9] proved, that for any prime $p \geq 5$, the following congruences hold

2010 Mathematics Subject Classification: 11B39, 05A10, 05A19.
Key words and Phrases: Binomial coefficients, Harmonic numbers, Congruences, Catalan numbers.

$$\sum_{k=1}^{p-1}\binom{2k}{k} \equiv -\left(\frac{p}{3}\right) \ (mod\ p^2), \qquad (1)$$

$$\sum_{k=1}^{p-1}\binom{2k}{k} H_k \equiv -\left(\frac{p}{3}\right) q_3 \ (mod\ p), \qquad (2)$$

where $\left(\frac{p}{3}\right)$ denotes the Legendre symbol and $q_a := (a^{p-1}-1)/p$ is the Fermat quotient with a prime p and an integer a. Also, in 2016 Mao and Sun [5] established, that for a prime $p > 3$, the following congruences

$$\sum_{k=1}^{p-1}\binom{2k}{k}\frac{H_k}{k} \equiv \frac{1}{3}\left(\frac{p}{3}\right) B_{p-2}\left(\frac{1}{3}\right) \ (mod\ p), \qquad (3)$$

$$\sum_{k=1}^{p-1}\binom{2k}{k}\frac{H_{2k}}{k} \equiv \frac{7}{12}\left(\frac{p}{3}\right) B_{p-2}\left(\frac{1}{3}\right) \ (mod\ p), \qquad (4)$$

where $B_n(.)$ is the n-th Bernoulli polynomial. In 2016 Koparal and Ömür [2] proved that

$$\sum_{k=1}^{(p-1)/2}(-1)^k\binom{2k}{k} H_{k-1} \equiv \frac{2^p}{p}\left(2F_{p+1}-5^p\right) \ (mod\ p), \qquad (5)$$

$$\sum_{k=1}^{(p-1)/2}\frac{C_k H_k}{(-4)^k} \equiv 2\frac{Q_{p+1}}{p} - \frac{2^{p+1}}{p}\left(1+2^{(p+1)/2}\right) \ (mod\ p) \qquad (6)$$

and if $\left(\frac{5}{p}\right) = 1$ they also proved that

$$\sum_{k=1}^{(p-1)/2}\binom{2k}{k}\frac{H_{k-1}}{(-4)^k} \equiv \frac{1}{p}\left(F_{2p+1}-F_{p+2}\right) - \frac{2^p}{p}F_{p-1} \ (mod\ p), \qquad (7)$$

where F_n is the Fibonacci numbers, $\left(\frac{a}{p}\right)$ denotes the Legendre symbol, $\{Q_n\}$ is the Pell-Lucas sequence and $C_n = \frac{1}{n+1}\binom{2n}{n}$ is the n-th Catalan number.

We have the following two theorems and corollaries.

Theorem 1. Let $p > 3$ be a prime number and $m \in \{1,2,\cdots,(p-1)/2\}$. We have

$$\sum_{k=1}^{(p-1)/2}(-1)^{k-1}\binom{k+m}{k}^{-1}\binom{(p-1)/2}{k} H_k \equiv \frac{m}{2m-1}\left(T(m)+S(m)p\right) \ (mod\ p^2), \qquad (8)$$

where

$$T(m) = -4q_2 + H_{2m-2} - 4H_{m-1} \; and$$

$$S(m) = 2q_2^2 + H_{m-1,2} - 4H_{2m-2,2} + 4\frac{q_2 + H_{m-1} - H_{2m-2}}{2m-1}.$$

Reducing the modulus in this congruence to get

$$\sum_{k=1}^{(p-1)/2} \binom{k+m}{k}^{-1} \binom{2k}{k} \frac{H_k}{4^k} \equiv \frac{4m}{2m-1}(q_2 - H_{2m-2} + H_{m-1}) \; (mod\ p). \qquad (9)$$

By the congruence (16), the congruence (9) for $m \in \{1,2\}$ and by the fact that

$$\binom{k+2}{k}^{-1} = \frac{2}{k+1} - \frac{2}{k+2},$$

we may state:

Corollary 2. For each prime number $p > 3$ we have

$$\sum_{k=1}^{(p-1)/2} \frac{C_k H_k}{4^k} \equiv 4q_2 \pmod{p}, \qquad (10)$$

$$\sum_{k=1}^{(p-1)/2} \binom{2k}{k} \frac{H_k}{4^k(k+2)} \equiv \frac{2}{3} + \frac{8}{3}q_2 \pmod{p}. \qquad (11)$$

Theorem 3. Let $p > 3$ be a prime number and $m \in \{1,2,\cdots,(p-3)/2\}$. We have

$$\sum_{k=1}^{(p-1)/2} (-1)^{k-1} \binom{(p-1)/2}{k} \binom{k}{m} H_{2k} \equiv \frac{(-1)^m}{p-1-2m}\left(1 + \frac{(-1)^{(p-1)/2}}{2^{2m+p-1}}\binom{2m}{m}\right) \pmod{p^3}. \quad (12)$$

Reducing this modulus to obtain

$$\sum_{k=1}^{(p-1)/2} \binom{2k}{k}\binom{k}{m} \frac{H_{2k}}{4^k} \equiv \frac{(-1)^m}{2m+1}\left(1 + \frac{(-1)^{(p-1)/2}}{2^{2m}}\binom{2m}{m}\right) \pmod{p}. \qquad (13)$$

Corollary 4. For each prime $p > 3$ we have

$$\sum_{k=1}^{(p-1)/2} (-1)^{k-1} \binom{(p-1)/2}{k} H_{2k} \equiv \frac{1}{p-1}\left(1 + \frac{(-1)^{(p-1)/2}}{2^{p-1}}\right) \pmod{p^3},$$

$$\sum_{k=1}^{(p-1)/2} (-1)^{k-1} \binom{(p-1)/2}{k} kH_{2k} \equiv -\frac{1}{p-3}\left(1 + \frac{(-1)^{(p-1)/2}}{2^{p}}\right) \pmod{p^3}$$

and

$$\sum_{k=1}^{(p-1)/2} \binom{2k}{k} \frac{H_{2k}}{4^k} \equiv 1 + (-1)^{(p-1)/2} \pmod{p},$$

$$\sum_{k=1}^{(p-1)/2} \binom{2k}{k} \frac{kH_{2k}}{4^k} \equiv -\frac{1}{3}\left(1 + \frac{1}{2}(-1)^{(p-1)/2}\right) \pmod{p}.$$

To prove Theorem 1, we give the following two lemmas.

Lemma 5. [8, Eq. 19] Let $n \geq 0$ and $m \geq 1$ be integers. The following identity holds

$$\sum_{k=1}^{n} (-1)^{k-1} \binom{k+m}{k}^{-1} \binom{n}{k} H_k = \frac{m}{m+n}(H_{m+n-1} - H_{m-1}). \tag{14}$$

Lemma 6. Let $p > 3$ be a prime number. Then for $m \in \{1,2,\cdots,(p-1)/2\}$ we have

$$H_{\frac{p-1}{2}+m-1} \equiv -2q_2 + 2H_{2m-2} - H_{m-1} - p\left(2H_{2m-2,2} - \frac{1}{2}H_{m-1,2} - q_2^2\right) (mod\ p^2). \tag{15}$$

Reducing the modulus in this congruence to obtain

$$H_{\frac{p-1}{2}+m-1} \equiv -2q_2 + 2H_{2m-2} - H_{m-1} (mod\ p).$$

Proof. We have

$$\begin{aligned}
H_{\frac{p-1}{2}+m-1} &= \sum_{k=1}^{(p-1)/2} \frac{1}{k} + \sum_{k=1}^{m-1} \frac{1}{\frac{p-1}{2}+k} \\
&= \sum_{k=1}^{(p-1)/2} \frac{1}{k} + 2\sum_{k=1}^{m-1} \frac{1}{p-(1-2k)} \\
&= H_{\frac{p-1}{2}} + 2\sum_{k=1}^{m-1} \frac{p+(1-2k)}{p^2-(1-2k)^2} \\
&\equiv H_{\frac{p-1}{2}} - 2\sum_{k=1}^{m-1} \frac{p+(1-2k)}{(1-2k)^2} \\
&= H_{\frac{p-1}{2}} - 2p\sum_{k=1}^{m-1} \frac{1}{(2k-1)^2} + 2\sum_{k=1}^{m-1} \frac{1}{2k-1} \\
&= H_{\frac{p-1}{2}} - 2p\left(H_{2m-2,2} - \frac{1}{4}H_{m-1,2}\right) + 2\left(H_{2m-2} - \frac{1}{2}H_{m-1}\right),
\end{aligned}$$

which, by the result congruence of Lehmer [4] $H_{\frac{p-1}{2}} \equiv -2q_2 + pq_2^2\ (mod\ p^2)$, the proof is complete.

Proof of Theorem 1. From the relation (14), we can write

$$\sum_{k=1}^{(p-1)/2}(-1)^{k-1}\binom{k+m}{k}^{-1}\binom{(p-1)/2}{k}H_k=\frac{m}{m+\frac{p-1}{2}}\left(H_{\frac{p-1}{2}+m-1}-H_{m-1}\right)$$
$$=\frac{2m}{2m+p-1}\left(H_{\frac{p-1}{2}+m-1}-H_{m-1}\right).$$

So, by the congruences (15) and

$$\frac{1}{2m+p-1}\equiv\frac{1}{2m-1}-\frac{p}{(2m-1)^2}\,(mod\ p^2)$$

we obtain

$$\sum_{k=1}^{(p-1)/2}(-1)^{k-1}\binom{k+m}{k}^{-1}\binom{(p-1)/2}{k}H_k$$
$$=\left(\frac{2m}{2m-1}-\frac{2mp}{(2m-1)^2}\right)\left(-2q_2+2H_{2m-2}-H_{m-1}-p\left(2H_{2m-2,2}-\frac{1}{2}H_{m-1,2}-q_2^2\right)-H_{m-1}\right)$$
$$=\left(\frac{2m}{2m-1}-\frac{2mp}{(2m-1)^2}\right)\left(-2q_2+2H_{2m-2}-2H_{m-1}-p\left(2H_{2m-2,2}-\frac{1}{2}H_{m-1,2}-q_2^2\right)\right)$$
$$=\frac{m}{2m-1}(-4q_2+4H_{2m-2}-4H_{m-1})$$
$$+\frac{m}{2m-1}\left(2q_2^2+H_{m-1,2}-4H_{2m-2,2}+4\frac{q_2+H_{m-1}-H_{2m-2}}{2m-1}\right)p\ (mod\ p^2).$$

To prove the relationship (9) we use the known congruence [2]

$$\binom{(p-1)/2}{k}\equiv\frac{1}{(-4)^k}\binom{2k}{k}\ (mod\ p). \tag{16}$$

Proof of Theorem 2. Let $n=((p-1)/2)$ in the identity of Corollary 2.2 [1]

$$\sum_{k=m}^{n}(-1)^{k-1}\binom{n}{k}\binom{k}{m}H_{2k}=\frac{(-1)^m}{n-m}\left(\frac{1}{2}+\frac{2^{2n-2m-2}\binom{2m}{m}}{\binom{2n-1}{n-1}}\right).$$

Then, we have

$$\sum_{k=1}^{(p-1)/2}(-1)^{k-1}\binom{(p-1)/2}{k}\binom{k}{m}\mathrm{H}_{2k}=\frac{(-1)^m}{\frac{p-1}{2}-m}\left(\frac{1}{2}+2^{p-2m-3}\binom{2m}{m}\binom{p-1-1}{(p-1)/2-1}^{-1}\right)$$
$$=\frac{2(-1)^m}{p-1-2m}\left(\frac{1}{2}+2^{p-2m-2}\binom{2m}{m}\binom{p-1}{(p-1)/2}^{-1}\right)$$
$$=\frac{(-1)^m}{p-1-2m}\left(1+2^{p-2m-1}\binom{2m}{m}\binom{p-1}{(p-1)/2}^{-1}\right).$$

From the known congruence [7]

$$\binom{p-1}{(p-1)/2} \equiv (-1)^{\frac{p-1}{2}} 4^{p-1} \ (mod\ p^3),$$

we have

$$\sum_{k=1}^{(p-1)/2} (-1)^{k-1} \binom{(p-1)/2}{k} \binom{\mathrm{k}}{\mathrm{m}} \mathrm{H}_{2\mathrm{k}} = \frac{(-1)^{\mathrm{m}}}{\mathrm{p}-1-2\mathrm{m}} \left(1 + (-1)^{\frac{p-1}{2}} \frac{2^{p-2m-1}\binom{2m}{m}}{4^{p-1}} \right)$$

$$= \frac{(-1)^{\mathrm{m}}}{\mathrm{p}-1-2\mathrm{m}} \left(1 + (-1)^{\frac{p-1}{2}} 2^{-2m} \binom{2m}{m} \left(\frac{1}{2}\right)^{p-1} \right) (\mathrm{mod}\ \mathrm{p}^3).$$

To prove the relationship (13), use the congruence (16) and Fermat little theorem.

2. CONCLUSION

The principal results of this paper given by Theorems 1 and 2 represent an interesting contribution in congruences. They are obtained upon using technical operations on the binomial coefficients, harmonic numbers and Catalan numbers. To extend our results using the useful technics or methods to study congruences in the ring of p-integers may be, in general, difficult. A first question on the extension of these congruences is: how can us generalize the obtained congruences modulo some successive powers of a prime number p?. A second question on such extensions of Theorems 1 and 2 can be viewed as generalizations on using the q-Binomial coefficients instead of the binomial coefficients or the hyper-harmonic numbers instead of the harmonic numbers. These seem to be interesting and require technical calculus and some mathematical tools based on number theory and on complex integration.

Acknowledgements: The authors thank the anonymous referee for his/her careful reading and valuable suggestions that led to an improved version of this manuscript.

REFERENCES

[1] H.-T. Jin and D.K. Du, “Abel's lemma and identities on harmonic numbers”, *Integers* **15**, #A22 (2015).

[2] S. Koparal and N. Ömür, “Congruences related to central binomial coefficients, harmonic and Lucas numbers”, *Turk. J. Math.*, **40**, 973-985 (2016).

[3] S. Koparal and N. Ömür, “On congruences involving special numbers”, *J. Ramanujan Math.*, **34**, 151-167 (2019).

[4] E. Lehmer, “On congruences involving Bernoulli numbers and the quotients of Fermat and Wilson”, *Ann. Math*,. **39**, 350-360 (1938).

[5] G.S. Mao and Z.W. Sun, “Tow congruences involving harmonic numbers with applications”, *Int. J. Number Theory*, **12** (2), 527-539, (2016).

[6] R. Meštrović and M. Andjić, “Certain congruence for harmonic numbers”, *Mathematica Montisnigri*, **38,** (2017).

[7] F. Morley, “Note on the congruence $2^{4n} \equiv (-1)^n (2n)!/(n!)^2$, where *2n+1* is a prime”, *Ann. Math.*, **9**, 168-170 (1895).

[8] B .M. Tuladhar, J. López-Bonilla and R. López-Vázque, “Identities for harmonic numbers and binomial relations via Legendre polynomials, Khatamudu university”, *J. of science, engineering and technology*, **13**, 92-97 (2017).
[9] Z.W. Sun and R. Tauraso, “On some new congruences for binomial coefficients”, *J. Number Theory*, **7**, 645-662, (2011).

Received November 3, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

STUDY ON LA-RING BY THEIR INTUITIONISTIC FUZZY IDEALS

NASREEN KAUSAR[1*], MESHARI ALESEMI[2], SALAHUDDIN[2], MOHAMMAD MUNIR[3]

[1]Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistán.

[2] Department of Mathematics, Jazan University, Jazan, K. S. A.

[3] Department of Mathematics, Government Postgraduate College, Abbottabad, Pakistán.

*Corresponding author. E-mail: kausar.nasreen57@gmail.com

DOI: 10.20948/mathmontis-2020-47-3

Summary. In this paper, we extend the characterizations of Kuroki [17], by initiating the concept of intuitionistic fuzzy left (resp. right, interior, quasi-, bi-, generalized bi-) ideals in a class of non-associative and non-commutative rings (LA-ring). We characterize regular (intra-regular, both regular and intra-regular) LA-rings in terms such ideals.

1 INTRODUCTION

In ternary operations, the commutative law is given by $abc = cba.$ Kazim and Naseerudin [7], have generalized this notion by introducing the parenthesis on the left side of this equation to get a new pseudo associative law, that is $(ab)c = (cb)a.$ This law $(ab)c = (cb)a$ is called the left invertive law. A groupoid S is called a left almost semigroup (abbreviated as LA-semi-group) if it satisfies the left invertive law. An LA-semi-group is a midway structure between a commutative semigroup and a groupoid.

A groupoid S is said to be medial (resp. paramedial) if $(ab)(cd) = (ac)(bd)$ (resp. $(ab)(cd) = (db)(ca)$). An LA-semi-group is medial, but in general an LA-semi-group needs not to be paramedial. Every LA-semi-group with left identity is paramedial and also satisfies $a(bc) = b(ac),\ (ab)(cd) = (dc)(ba).$

Kamran [16], extended the notion of LA-semi-group to the left almost group (LA-group). An LA-semi-group G is called a left almost group, if there exists a left identity $e \in G$ such that $ea = a$ for all $a \in G$ and for every $a \in G$ there exists $b \in G$ such that $ba = e.$

Shah et al. [22], by a left almost ring, mean a non-empty set R with at least two elements such that $(R,+)$ is an LA-group, $(R,\cdot)$ is an LA-semi-group, both left and right distributive laws hold. For example, from a commutative ring $(R,+,\cdot)$, we can always obtain an LA-ring $(R,\oplus,\cdot)$ by defining for all $a,b \in R,\ a \oplus b = b - a$ and $a \cdot b$ is same as in the ring. Although the structure is non-associative and non-commutative, nevertheless, it possesses many interesting properties which we usually find in associative and commutative algebraic structures.

A non-empty subset A of R is called an LA-subring of R if $a - b \in A$ and $ab \in A$ for all

2010 Mathematics Subject Classification: 97H20, 97H40, 94D05.

Key words and Phrases: Intuitionistic fuzzy left (right, interior, quasi-, bi-, generalized bi-) ideals, regular (intra-regular) LA-rings.

$a, b \in A$. A is called a left (resp. right) ideal of R if $(A,+)$ is an LA-group and $RA \subseteq A$ (resp. $AR \subseteq A$). A is called an ideal of R if it is both a left ideal and a right ideal of R.

A non-empty subset A of R is called an interior ideal of R if $(A,+)$ is an LA-group and $(RA)R \subseteq A$. A non-empty subset A of R is called a quasi-ideal of R if $(A,+)$ is an LA-group and $AR \cap RA \subseteq A$. An LA-subring A of R is called a bi-ideal of R if $(AR)A \subseteq A$. A non-empty subset A of R is called a generalized bi-ideal of R if $(A,+)$ is an LA-group and $(AR)A \subseteq A$.

We will define the concept of intuitionistic fuzzy left (resp. right, interior, quasi-, bi-, generalized bi-) ideals of an LA-ring R. We will establish a study by discussing the different properties of such ideals. We will characterize regular (resp. intra-regular, both regular and intra-regular) LA-rings by the properties of intuitionistic fuzzy (left, right, quasi-, bi-, generalized bi-) ideals such ideals.

2 INTUITIONISTIC FUZZY IDEALS IN LA-RINGS

After, the introduction of fuzzy set by Zadeh [24], several researchers explored on the generalization of the notion of fuzzy set. The concept of intuitionistic fuzzy set was introduced by Atanassov [1], as a generalization of the notion of fuzzy set. Liu [18], introduced the concept of fuzzy subrings and fuzzy ideals of a ring. Many authors have explored the theory of fuzzy rings (for example [3, 9, 11-15, 18, 19-20, 23]). Gupta and Kantroo [4], gave the idea of intrinsic product of fuzzy subsets of a ring. Kuroki [17], characterized regular (intra-regular, both regular and intra-regular) rings in terms of fuzzy left (right, quasi, bi-) ideals.

An intuitionistic fuzzy set (briefly, IFS) A in a non-empty set X is an object having the form $A = \{(x, \mu_A(x), \gamma_A(x)) : x \in X\}$, where the functions $\mu_A : X \to [0,1]$ and $\gamma_A : X \to [0,1]$ denote the degree of membership and the degree of non-membership, respectively and $0 \leq \mu_A(x) + \gamma_A(x) \leq 1$ for all $x \in X$ [1].

An intuitionistic fuzzy set $A = \{(x, \mu_A(x), \gamma_A(x)) : x \in X\}$ in X can be identified to be an ordered pair (μ_A, γ_A) in $I^X \times I^X$, where I^X is the set of all functions from X to $[0,1]$. For the sake of simplicity, we shall use the symbol $A = (\mu_A, \gamma_A)$ for the IFS $A = \{(x, \mu_A(x), \gamma_A(x)) : x \in X\}$.

Banerjee and Basnet [2] and Hur et al. [6], initiated the notion of intuitionistic fuzzy subrings and intuitionistic fuzzy ideals of a ring. Subsequently many authors studied the intuitionistic fuzzy subrings and intuitionistic fuzzy ideals of a ring by describing the different properties (see [5]). Shah et al. [21, 22] initiated the concept of intuitionistic fuzzy normal subrings over a non-associative ring and also characterized the non-associative rings by their intuitionistic fuzzy bi-ideals in [8]. Kausar [10] explored the notion of direct product of finite intuitionistic anti fuzzy normal subrings over non-associative rings.

We initiate the notion of intuitionistic fuzzy left (resp. right, interior, quasi-, bi-, generalized bi-) ideals of an LA-ring R.

An intuitionistic fuzzy set (IFS) $A=(\mu_A,\gamma_A)$ of an LA-ring R is called an intuitionistic fuzzy LA-subring of R if

(1) $\mu_A(x-y) \geq \min\{\mu_A(x), \mu_A(y)\}$,

(2) $\gamma_A(x-y) \leq \max\{\gamma_A(x), \gamma_A(y)\}$,

(3) $\mu_A(xy) \geq \min\{\mu_A(x), \mu_A(y)\}$,

(4) $\gamma_A(xy) \leq \max\{\gamma_A(x), \gamma_A(y)\}$, for all $x, y \in R$.

An IFS $A=(\mu_A,\gamma_A)$ of an LA-ring R is called an intuitionistic fuzzy left ideal of R if

(1) $\mu_A(x-y) \geq \min\{\mu_A(x), \mu_A(y)\}$,

(2) $\gamma_A(x-y) \leq \max\{\gamma_A(x), \gamma_A(y)\}$,

(3) $\mu_A(xy) \geq \mu_A(y)$,

(4) $\gamma_A(xy) \leq \gamma_A(y)$, for all $x, y \in R$.

An IFS $A=(\mu_A,\gamma_A)$ of an LA-ring R is called an intuitionistic fuzzy right ideal of R if

(1) $\mu_A(x-y) \geq \min\{\mu_A(x), \mu_A(y)\}$,

(2) $\gamma_A(x-y) \leq \max\{\gamma_A(x), \gamma_A(y)\}$,

(3) $\mu_A(xy) \geq \mu_A(x)$,

(4) $\gamma_A(xy) \leq \gamma_A(x)$, for all $x, y \in R$.

An IFS $A=(\mu_A,\gamma_A)$ of R is called an intuitionistic fuzzy ideal of an LA-ring R if it is both an intuitionistic fuzzy left ideal and an intuitionistic fuzzy right ideal of R.

Let A be a non-empty subset of an LA-ring R. Then the intuitionistic characteristic of A is denoted by $\chi_A = \langle \mu_{\chi_A}, \gamma_{\chi_A} \rangle$ and defined by

$$\mu_{\chi_A}(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{cases} \text{ and } \gamma_{\chi_A}(x) = \begin{cases} 0 \text{ if } x \in A \\ 1 \text{ if } x \notin A \end{cases}$$

We note that an LA-ring R can be considered an intuitionistic fuzzy set of itself and we write $R = I_R$, i.e., $R(x) = (\mu_R, \gamma_R) = (1,0)$ for all $x \in R$.

Let A and B be two intuitionistic fuzzy sets of an LA-ring R. Then

(1) $A \subseteq B \Leftrightarrow \mu_A \subseteq \mu_B$ and $\gamma_A \supseteq \gamma_B$,

(2) $A = B \Leftrightarrow A \subseteq B$ and $B \subseteq A$,

(3) $A^c = (\gamma_A, \mu_A)$,

(4) $A \cap B = (\mu_A \wedge \mu_B, \gamma_A \vee \gamma_B) = (\mu_{A \wedge B}, \gamma_{A \vee B})$,

(5) $A \cup B = (\mu_A \vee \mu_B, \gamma_A \wedge \gamma_B) = (\mu_{A \vee B}, \gamma_{A \wedge B})$,

(6) $0 \approx= (0,1), 1 \approx= (1,0)$.

The product of $A = (\mu_A, \gamma_A)$ and $B = (\mu_B, \gamma_B)$ is denoted by $A \circ B = (\mu_A \circ \mu_B, \gamma_A \circ \gamma_B)$ and defined as:

$$(\mu_A \circ \mu_B)(x) = \begin{cases} \vee_{x=\sum_{i=1}^{n} a_i b_i} \{\wedge_{i=1}^{n} \{\mu_A(a_i) \wedge \mu_B(b_i)\}\} & \text{if } x = \sum_{i=1}^{n} a_i b_i,\ a_i, b_i \in R \\ 0 & \text{if } x \neq \sum_{i=1}^{n} a_i b_i \end{cases}$$

$$\text{and } (\gamma_A \circ \gamma_B)(x) = \begin{cases} \wedge_{x=\sum_{i=1}^{n} a_i b_i} \{\vee_{i=1}^{n} \{\gamma_A(a_i) \vee \gamma_B(b_i)\}\} & \text{if } x = \sum_{i=1}^{n} a_i b_i,\ a_i, b_i \in R \\ 1 & \text{if } x \neq \sum_{i=1}^{n} a_i b_i \end{cases}$$

An IFS $A = (\mu_A, \gamma_A)$ of an LA-ring R is called an intuitionistic fuzzy interior ideal of R if

(1) $\mu_A(x-y) \geq \mu_A(x) \wedge \mu_A(y)$,

(2) $\gamma_A(x-y) \leq \gamma_A(x) \vee \gamma_A(y)$,

(3) $\mu_A((xy)z) \geq \mu_A(y)$,

(4) $\gamma_A((xy)z) \leq \gamma_A(y)$, for all $x, y, z \in R$.

An IFS $A = (\mu_A, \gamma_A)$ of an LA-ring R is called an intuitionistic fuzzy quasi-ideal of R if

(1) $(\mu_A \circ R) \cap (R \circ \mu_A) \subseteq \mu_A$,

(2) $(\gamma_A \circ R) \cup (R \circ \gamma_A) \supseteq \gamma_A$,

(3) $\mu_A(x-y) \geq \mu_A(x) \wedge \mu_A(y)$,

(4) $\gamma_A(x-y) \leq \gamma_A(x) \vee \gamma_A(y)$, for all $x, y \in R$.

An Intuitionistic fuzzy LA-subring $A = (\mu_A, \gamma_A)$ of an LA-ring R is called an intuitionistic fuzzy bi-ideal of R if

(1) $\mu_A((xy)z) \geq \mu_A(x) \wedge \mu_A(z)$,

(2) $\gamma_A((xy)z) \leq \gamma_A(x) \vee \gamma_A(z)$, for all $x, y, z \in R$.

An IFS $A = (\mu_A, \gamma_A)$ of an LA-ring R is called an intuitionistic fuzzy generalized bi-ideal of R if

(1) $\mu_A(x-y) \geq \mu_A(x) \wedge \mu_A(y)$,

(2) $\gamma_A(x-y) \leq \gamma_A(x) \vee \gamma_A(y)$,

(3) $\mu_A((xy)z) \geq \mu_A(x) \wedge \mu_A(z)$,

(4) $\gamma_A((xy)z) \leq \gamma_A(x) \vee \gamma_A(z)$, for all $x, y, z \in R$.

An intuitionistic fuzzy ideal $A = (\mu_A, \gamma_A)$ of an LA-ring R is called an intuitionistic fuzzy idempotent if $\mu_A \circ \mu_A = \mu_A$ and $\gamma_A \circ \gamma_A = \gamma_A$.

Now we give some imperative properties of such ideals of an LA-ring R, which will be very helpful in later sections.

Lemma 2.1: Let R be an LA-ring. Then the following properties hold:

(1) $(A \circ B) \circ C = (C \circ B) \circ A$,

(2) $(A \circ B) \circ (C \circ D) = (A \circ C) \circ (B \circ D)$,

(3) $A \circ (B \circ C) = B \circ (A \circ C)$,

(4) $(A \circ B) \circ (C \circ D) = (D \circ B) \circ (C \circ A)$,

(5) $(A \circ B) \circ (C \circ D) = (D \circ C) \circ (B \circ A)$, for all intuitionistic fuzzy sets A, B, C and D of R.

Proof: Obvious.

Theorem 2.2: Let A and B be two non-empty subsets of an LA-ring R. then the following properties hold:

(1) If $A \subseteq B$ then $\chi_A \subseteq \chi_B$.

(2) $\chi_A \circ \chi_B = \chi_{AB}$.

(4) $\chi_A \cap \chi_B = \chi_{A \cap B}$.

Proof: (1) Suppose that $A \subseteq B$ and $a \in R$. If $a \in A$, this implies that $a \in B$. Thus $\mu_{\chi_A}(a) = 1 = \mu_{\chi_B}(a)$ and $\gamma_{\chi_A}(a) = 0 = \gamma_{\chi_B}(a)$, i.e., $\chi_A \subseteq \chi_B$.

If $a \notin A$, and $a \notin B$. Thus $\mu_{\chi_A}(a) = 0 = \mu_{\chi_B}(a)$ and $\gamma_{\chi_A}(a) = 1 = \gamma_{\chi_B}(a)$, i.e., $\chi_A \subseteq \chi_B$.

If $\alpha \notin A$ and $\alpha \in B$. Thus $\mu_{\chi A}(\alpha) = 0$ and $\mu_{\chi B}(\alpha) = 1$ and $\gamma_{\chi A}(\alpha) = 1$ and $\gamma_{\chi B}(\alpha) = 0$, i.e.,

(2) Let $x \in R$ and $x \in AB$. This means that $x=ab$ for some $\alpha \in A$ and $b \in B$.

Now

$$(\mu_{\chi_A} \circ \mu_{\chi_B})(x) = \vee_{x = \sum_{i=1}^{n} a_i b_i} \{\wedge_{i=1}^{n} \{\mu_{\chi_A}(a_i) \wedge \mu_{\chi_B}(b_i)\}\}$$

$$\geq \mu_{\chi_A}(a) \wedge \mu_{\chi_B}(b) = 1 \wedge 1 = 1 = \mu_{\chi_{AB}}(x)$$

and $$(\gamma_{\chi_A} \circ \gamma_{\chi_B})(x) = \wedge_{x = \sum_{i=1}^{n} a_i b_i} \{\vee_{i=1}^{n} \{\gamma_{\chi_A}(a_i) \vee \gamma_{\chi_B}(b_i)\}\}$$

$$\leq \gamma_{\chi_A}(a) \vee \gamma_{\chi_B}(b) = 0 \vee 0 = 0 = \gamma_{\chi_{AB}}(x).$$

If $x \notin AB$, i.e., $x \neq ab$ for all $a \in A$ and $b \in B$. Then there are two cases.

(i) If $x = uv$ for some $u, v \in R$, then

$$(\mu_{\chi_A} \circ \mu_{\chi_B})(x) = \vee_{x = \sum_{i=1}^{n} a_i b_i} \{\wedge_{i=1}^{n} \{\mu_{\chi_A}(a_i) \wedge \mu_{\chi_B}(b_i)\}\}$$
$$= 0 \wedge 0 = 0 = \mu_{\chi_{AB}}(x)$$
$$\text{and } (\gamma_{\chi_A} \circ \gamma_{\chi_B})(x) = \wedge_{x = \sum_{i=1}^{n} a_i b_i} \{\vee_{i=1}^{n} \{\gamma_{\chi_A}(a_i) \vee \gamma_{\chi_B}(b_i)\}\}$$
$$= 1 \vee 1 = 1 = \gamma_{\chi_{AB}}(x).$$

(ii) If $x \neq uv$ for all $u, v \in R$, then obviously $(\chi_A \circ \chi_B)(x) = 0 = \chi_{AB}(x)$. Hence $\chi_A \circ \chi_B = \chi_{AB}$.

Similarly, we can prove (3) and (4).

Theorem 2.3: Let A be a non-empty subset of an LA-ring R. then the following properties hold.

(1) A is an LA-subring of R if and only if χ_A is an intuitionistic fuzzy LA-subring of R.

(2) A is a left (resp. right, two-sided) ideal of R if and only if χ_A is an intuitionistic fuzzy left (resp. right, two-sided) ideal of R.

Proof: (1) Let A be an LA-subring of R and $a, b \in R$. If $a, b \in A$, then by definition $\mu_A(a) = 1 = \mu_A(b)$ and $\gamma_A(a) = 0 = \gamma_A(b)$. Since $a - b$ and $ab \in A$, A being an LA-subring of R, this implies that $\mu_A(a-b) = 1 = \mu_A(ab)$ and $\gamma_A(a-b) = 0 = \gamma_A(ab)$. Thus

$$\mu_A(a-b) \geq \mu_A(a) \wedge \mu_A(b), \quad \mu_A(ab) \geq \mu_A(a) \wedge \mu_A(b)$$
$$\text{and } \gamma_A(a-b) \leq \gamma_A(a) \vee \gamma_A(b), \gamma_A(ab) \leq \gamma_A(a) \vee \gamma_A(b).$$

Similarly, we have

$$\mu_A(a-b) \geq \mu_A(a) \wedge \mu_A(b), \quad \mu_A(ab) \geq \mu_A(a) \wedge \mu_A(b).$$
$$\text{and } \gamma_A(a-b) \leq \gamma_A(a) \vee \gamma_A(b), \gamma_A(ab) \leq \gamma_A(a) \vee \gamma_A(b).$$

when $a, b \notin A$. Hence χ_A is an intuitionistic fuzzy LA-subring of R.

Conversely, suppose that χ_A is an intuitionistic fuzzy LA-subring of R and let $a, b \in A$. This means that $\mu_A(a) = 1 = \mu_A(b)$ and $\gamma_A(a) = 0 = \gamma_A(b)$. Since

$$\mu_A(a-b) \geq \mu_A(a) \wedge \mu_A(b) = 1 \wedge 1 = 1,$$
$$\mu_A(ab) \geq \mu_A(a) \wedge \mu_A(b) = 1 \wedge 1 = 1,$$
$$\gamma_A(a-b) \leq \gamma_A(a) \vee \gamma_A(b) = 0 \vee 0 = 0,$$
$$\gamma_A(ab) \leq \gamma_A(a) \vee \gamma_A(b) = 0 \vee 0 = 0,$$

χ_A being an intuitionistic fuzzy LA-subring of R. Thus $\mu_A(a-b) = 1 = \mu_A(ab)$ and $\gamma_A(a-b) = 0 = \gamma_A(ab)$, i.e., $a - b$ and $ab \in A$. Hence A is an LA-subring of R.

(2) Let A be a left ideal of R and $a, b \in R$. If $a, b \in A$, then by definition $\mu_A(a) = 1 = \mu_A(b)$ and $\gamma_A(a) = 0 = \gamma_A(b)$. Since $a - b$ and $ab \in A$, A being a left ideal of R, this implies that $\mu_A(a-b) = 1 = \mu_A(ab)$ and $\gamma_A(a-b) = 0 = \gamma_A(ab)$. Thus

$$\mu_A(a-b) \geq \mu_A(a) \wedge \mu_A(b), \quad \mu_A(ab) \geq \mu_A(b)$$
$$\text{and } \gamma_A(a-b) \leq \gamma_A(a) \vee \gamma_A(b), \gamma_A(ab) \leq \gamma_A(b).$$

Similarly, we have

$$\mu_A(a-b) \geq \mu_A(a) \wedge \mu_A(b), \quad \mu_A(ab) \geq \mu_A(b)$$
$$\text{and } \gamma_A(a-b) \leq \gamma_A(a) \vee \gamma_A(b), \gamma_A(ab) \leq \gamma_A(b).$$

when $a, b \notin A$. Therefore χ_A is an intuitionistic fuzzy left ideal of R.

Conversely, assume that χ_A is an intuitionistic fuzzy left ideal of R and let $a, b \in A$ and $z \in R$. This means that $\mu_A(a) = 1 = \mu_A(b)$ and $\gamma_A(a) = 0 = \gamma_A(b)$. Since

$$\mu_A(a-b) \geq \mu_A(a) \wedge \mu_A(b) = 1 \wedge 1 = 1,$$
$$\mu_A(zb) \geq \mu_A(b) = 1,$$
$$\gamma_A(a-b) \leq \gamma_A(a) \vee \gamma_A(b) = 0 \vee 0 = 0,$$
$$\gamma_A(zb) \leq \gamma_A(b) = 0,$$

χ_A being an intuitionistic fuzzy left ideal of R. Thus $\mu_A(a-b) = 1 = \mu_A(zb)$ and $\gamma_A(a-b) = 0 = \gamma_A(zb)$, i.e., $a - b$ and $zb \in A$. Therefore A is a left ideal of R.

Remark 2.4: (i) A is an additive LA-subgroup of R if and only if χ_A is an intuitionistic fuzzy additive LA-subgroup of R.

(ii) A is an LA-subsemigroup of R if and only if χ_A is an intuitionistic fuzzy LA-subsemigroup of R.

Lemma 2.5: If A and B are two intuitionistic fuzzy LA-subrings (resp. (left, right, two-sided) ideals) of an LA-ring R, then $A \cap B$ is also an intuitionistic fuzzy LA-subring (resp. (left, right, two-sided) ideal) of R.

Proof: Obvious.

Lemma 2.6: If A and B are two intuitionistic fuzzy LA-subrings of an LA-ring R, then $A \circ B$ is also an intuitionistic fuzzy LA-subring of R.

Proof: Let $A = (\mu_A, \gamma_A)$ and $B = (\mu_B, \gamma_B)$ be two intuitionistic fuzzy LA-subrings of R. We have to show that $A \circ B$ is also an intuitionistic fuzzy LA-subring of R. Now

$$(\mu_A \circ \mu_B)^2 = (\mu_A \circ \mu_B) \circ (\mu_A \circ \mu_B) = (\mu_A \circ \mu_A) \circ (\mu_B \circ \mu_B) \subseteq \mu_A \circ \mu_B$$
$$\text{and } (\gamma_A \circ \gamma_B)^2 = (\gamma_A \circ \gamma_B) \circ (\gamma_A \circ \gamma_B) = (\gamma_A \circ \gamma_A) \circ (\gamma_B \circ \gamma_B) \supseteq \gamma_A \circ \gamma_B.$$

Since $\mu_B - \mu_B \subseteq \mu_B$ and $\gamma_B - \gamma_B \supseteq \gamma_B$, $B = (\mu_B, \gamma_B)$ being an intuitionistic fuzzy LA-subring of R. This implies that $\mu_A \circ (\mu_B - \mu_B) \subseteq \mu_A \circ \mu_B$ and

$\gamma_A \circ (\gamma_B - \gamma_B) \supseteq \gamma_A \circ \gamma_B$, i.e., $\mu_A \circ \mu_B - \mu_A \circ \mu_B \subseteq \mu_A \circ \mu_B$ and $\gamma_A \circ \gamma_B - \gamma_A \circ \gamma_B \supseteq \gamma_A \circ \gamma_B$. Therefore $A \circ B$ is an intuitionistic fuzzy LA-subring of R.

Remark 2.7: If A is an intuitionistic fuzzy LA-subring of an LA-ring R, then $A \circ A$ is also an intuitionistic fuzzy LA-subring of R.

Lemma 2.8: Let R be an LA-ring with left identity e. Then $RR = R$ and $eR = R = \mathrm{Re}$.

Proof: Since $RR \subseteq R$ and $x = ex \in RR$, where $x \in R$, i.e., $RR = R$. Since e is the left identity of R, i.e., $eR = R$. Now $\mathrm{Re} = (RR)e = (eR)R = RR = R$.

Lemma 2.9: Let R be an LA-ring with left identity e. Then every intuitionistic fuzzy right ideal of R is an intuitionistic fuzzy ideal of R.

Proof: Let $A = (\mu_A, \gamma_A)$ be an intuitionistic fuzzy right ideal of R and $x, y \in R$. Now

$$\mu_A(xy) = \mu_A((ex)y) = \mu_A((yx)e) \geq \mu_A(yx) \geq \mu_A(y)$$
$$\text{and } \gamma_A(xy) = \gamma_A((ex)y) = \gamma_A((yx)e) \leq \gamma_A(yx) \leq \gamma_A(y).$$

Thus A is an intuitionistic fuzzy ideal of R.

Lemma 2.10: If A and B are two intuitionistic fuzzy left (resp. right) ideals of an LA-ring R with left identity e, then $A \circ B$ is also an intuitionistic fuzzy left (resp. right) ideal of R.

Proof: Let $A = (\mu_A, \gamma_A)$ and $B = (\mu_B, \gamma_B)$ be two intuitionistic fuzzy left ideals of R. We have to show that $A \circ B$ is also an intuitionistic fuzzy left ideal of R. Since $\mu_A \circ \mu_B - \mu_A \circ \mu_B \subseteq \mu_A \circ \mu_B$ and $\gamma_A \circ \gamma_B - \gamma_A \circ \gamma_B \supseteq \gamma_A \circ \gamma_B$. Now

$$R \circ (\mu_A \circ \mu_B) = (R \circ R) \circ (\mu_A \circ \mu_B) = (R \circ \mu_A) \circ (R \circ \mu_B) \subseteq (\mu_A \circ \mu_B)$$
$$\text{and } R \circ (\gamma_A \circ \gamma_B) = (R \circ R) \circ (\gamma_A \circ \gamma_B) = (R \circ \gamma_A) \circ (R \circ \gamma_A) \supseteq (\gamma_A \circ \gamma_B).$$

Hence $A \circ B$ is an intuitionistic fuzzy left ideal of R.

Remark 2.11: If A is an intuitionistic fuzzy left (resp. right) ideal of an LA-ring R with left identity e, then $A \circ A$ is an intuitionistic fuzzy ideal of R.

Lemma 2.12: If A and B are two intuitionistic fuzzy ideals of an LA-ring R, then $A \circ B \subseteq A \cap B$.

Proof: Let $A = (\mu_A, \gamma_A)$ and $B = (\mu_B, \gamma_B)$ be two intuitionistic fuzzy ideals of R and $x \in R$. If x cannot expressible as $x = \sum_{i=1}^{n} a_i b_i$, where $a_i, b_i \in R$ and n is any positive integer, then obviously $A \circ B \subseteq A \cap B$, otherwise we have

$$\begin{aligned}
(\mu_A \circ \mu_B)(x) &= \vee_{x = \sum_{i=1}^{n} a_i b_i} \{\wedge_{i=1}^{n} \{\mu_A(a_i) \wedge \mu_B(b_i)\}\} \\
&\leq \vee_{x = \sum_{i=1}^{n} a_i b_i} \{\wedge_{i=1}^{n} \{\mu_A(a_i b_i) \wedge \mu_B(a_i b_i)\}\} \\
&= \vee_{x = \sum_{i=1}^{n} a_i b_i} \{\wedge_{i=1}^{n} (\mu_A \cap \mu_B)(a_i b_i)\} = (\mu_A \cap \mu_B)(x). \\
&\Rightarrow \mu_A \circ \mu_B \subseteq \mu_A \cap \mu_B.
\end{aligned}$$

Similarly, we can prove $\gamma_A \circ \gamma_B \supseteq \gamma_A \cup \gamma_B$.

Therefore $A \circ B \subseteq A \cap B$ for all intuitionistic fuzzy ideals A and B of R.

Remark 2.13: If A is an intuitionistic fuzzy ideal of an LA-ring R, then $A \circ A \subseteq A$.

Lemma 2.14: Let R be an LA-ring. Then $A \circ B \subseteq A \cap B$ for every intuitionistic fuzzy right ideal A and every intuitionistic fuzzy left ideal B of R.

Proof: Same as Lemma 2.12

Theorem 2.15: Let A be a non-empty subset of an LA-ring R. Then A is an interior (resp. quasi-, bi-, generalized bi-) ideal of R if and only if χ_A is an intuitionistic fuzzy interior (resp. quasi-, bi-, generalized bi-) ideal of R.

Proof: Let A be an interior ideal of R, this implies that A is an additive LA-subgroup of R. Then χ_A is an intuitionistic fuzzy additive LA-subgroup of R by the Remark 2.4. Let $x, y, a \in R$. If $a \in A$, then by definition $\mu_{\chi_A}(a) = 1$ and $\gamma_{\chi_A}(a) = 0$. Since $(xa)y \in A$, A being an interior ideal of R, this means that $\mu_{\chi_A}((xa)y) = 1$ and $\gamma_{\chi_A}((xa)y) = 0$. Thus $\mu_{\chi_A}((xa)y) \geq \mu_{\chi_A}(a)$ and $\gamma_{\chi_A}((xa)y) \leq \gamma_{\chi_A}(a)$. Similarly, we have $\mu_{\chi_A}((xa)y) \geq \mu_{\chi_A}(a)$ and $\gamma_{\chi_A}((xa)y) \leq \gamma_{\chi_A}(a)$, when $a \notin A$. Hence χ_A is an intuitionistic fuzzy interior ideal of R.

Conversely, suppose that χ_A is an intuitionistic fuzzy interior ideal of R, this means that χ_A is an intuitionistic fuzzy additive LA-subgroup of R. Then A is an additive LA-subgroup of R by the Remark 2.4. Let $x, y \in R$ and $a \in A$, so $\mu_{\chi_A}(a) = 1$ and $\gamma_{\chi_A}(a) = 0$. Since $\mu_{\chi_A}((xa)y) \geq \mu_{\chi_A}(a) = 1$ and $\gamma_{\chi_A}((xa)y) \leq \gamma_{\chi_A}(a) = 0$, χ_A being an intuitionistic fuzzy interior ideal of R. Thus $\mu_{\chi_A}((xa)y) = 1$ and $\gamma_{\chi_A}((xa)y) = 0$, i.e., $(xa)y \in A$. Hence A is an interior ideal of R. Similarly, we can prove for (quasi-, bi-, generalized bi-) ideal.

Lemma 2.16: If A and B are two intuitionistic fuzzy bi-(resp. generalized bi-, quasi-, interior) ideals of an LA-ring R, then $A \cap B$ is also an intuitionistic fuzzy bi- (resp. generalized bi-, quasi-, interior) ideal of R.

Proof: Obvious.

Lemma 2.17: If A and B are two intuitionistic fuzzy bi- (resp. generalized bi-, interior) ideals of an LA-ring R with left identity e, then $A \circ B$ is also an intuitionistic fuzzy bi- (resp. generalized bi-, interior) ideal of R.

Proof: Let $A = (\mu_A, \gamma_A)$ and $B = (\mu_B, \gamma_B)$ be two intuitionistic fuzzy bi-ideals of R. We have to show that $A \circ B$ is also an intuitionistic fuzzy bi-ideal of R. Since A and B are two intuitionistic fuzzy LA-subrings of R, then $A \circ B$ is also an intuitionistic fuzzy LA-subring of R by the Lemma 2.6. Now

$$\begin{aligned}((\mu_A \circ \mu_B) \circ R) \circ (\mu_A \circ \mu_B) &= ((\mu_A \circ \mu_B) \circ (R \circ R)) \circ (\mu_A \circ \mu_B) \\ &= ((\mu_A \circ R) \circ (\mu_B \circ R)) \circ (\mu_A \circ \mu_B) \\ &= ((\mu_A \circ R) \circ \mu_A) \circ ((\mu_B \circ R) \circ \mu_B) \\ &\subseteq \mu_A \circ \mu_B .\end{aligned}$$

Similarly, we have $((\gamma_A \circ \gamma_B) \circ R) \circ (\gamma_A \circ \gamma_B) \supseteq \gamma_A \circ \gamma_B$. Therefore $A \circ B$ is an intuitionistic fuzzy bi-ideal of R.

Lemma 2.18: Every intuitionistic fuzzy ideal of an LA-ring R is an intuitionistic fuzzy interior ideal of R. The converse is not true in general.

Proof: Let $A = (\mu_A, \gamma_A)$ be an intuitionistic fuzzy ideal of R and $x, y, z \in R$. Thus $\mu_A((xy)z) = \mu_A(xy) \geq \mu_A(y)$ and $\gamma_A((xy)z) = \gamma_A(xy) \leq \gamma_A(y)$. Hence A is an intuitionistic fuzzy interior ideal of R.

The converse is not true in general, giving an example:

Example 2.19: Let $R = \{0, 1, 2, 3, 4, 5, 6, 7\}$ is an LA-ring.

+	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	2	0	3	1	6	4	7	5
2	1	3	0	2	5	7	4	6
3	3	2	1	0	7	6	5	4
4	4	5	6	7	0	1	2	3
5	6	4	7	5	2	0	3	1
6	5	7	4	6	1	3	0	2
7	7	6	5	4	3	2	1	0

and

·	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	4	4	0	0	4	4	0
2	0	4	4	0	0	4	4	0
3	0	0	0	0	0	0	0	0
4	0	3	3	0	0	3	3	0
5	0	7	7	0	0	7	7	0
6	0	7	7	0	0	7	7	0
7	0	3	3	0	0	3	3	0

Let $A = (\mu_A, \gamma_A)$ be an IFS of an LA-ring R. We define

$$\mu_A(0) = \mu_A(4) = 0.7, \quad \mu_A(1) = \mu_A(2) = \mu_A(3) = \mu_A(5) = \mu_A(6) = \mu_A(7) = 0$$

and $\gamma_A(0) = \gamma_A(4) = 0, \quad \gamma_A(1) = \gamma_A(2) = \gamma_A(3) = \gamma_A(5) = \gamma_A(6) = \gamma_A(7) = 0.7.$

$A = (\mu_A, \gamma_A)$ is an intuitionistic fuzzy interior ideal of R, but not an intuitionistic fuzzy ideal of R, because A is not an intuitionistic fuzzy right ideal of R, as

$$\begin{aligned} \mu_A(41) &= \mu_A(3) = 0. \\ \mu_A(4) &= 0.7. \\ &\Rightarrow \mu_A(41) \ngeq \mu_A(4). \\ \text{and } \gamma_A(41) &= \gamma_A(3) = 0.7. \\ \gamma_A(4) &= 0. \\ &\Rightarrow \gamma_A(41) \nleq \gamma_A(4). \end{aligned}$$

Proposition 2.20: Let $A = (\mu_A, \gamma_A)$ be an IFS of an LA-ring R with left identity e. Then A is an intuitionistic fuzzy ideal of R if and only if A is an intuitionistic fuzzy interior ideal of R.

Proof: Let $A = (\mu_A, \gamma_A)$ be an intuitionistic fuzzy interior ideal of R and $x, y \in R$. Thus $\mu_A(xy) = \mu_A((ex)y) \geq \mu_A(x)$ and $\gamma_A(xy) = \gamma_A((ex)y) \leq \gamma_A(x)$, i.e., A is an intuitionistic fuzzy right ideal of R. Hence A is an intuitionistic fuzzy ideal of R by the Lemma 2.9. Converse is true by the Lemma 2.18.

Lemma 2.21: Every intuitionistic fuzzy left (resp. right, two-sided) ideal of an LA-ring R is an intuitionistic fuzzy bi-ideal of R.

Proof: Suppose that $A = (\mu_A, \gamma_A)$ is an intuitionistic fuzzy right ideal of R and $x, y, z \in R$. Thus

$$\mu_A((xy)z) = \mu_A(xy) \geq \mu_A(x) \text{ and } \mu_A((xy)z) = \mu_A((zy)x) \geq \mu_A(zy) \geq \mu_A(z),$$

this implies that $\mu_A((xy)z) \geq \mu_A(x) \wedge \mu_A(z)$. Similarly, we have $\gamma_A((xy)z) \leq \gamma_A(x) \vee \gamma_A(z)$. Therefore A is an intuitionistic fuzzy bi-ideal of R.

The converse is not true in general, giving an example:

Using Example 2.19, $A = (\mu_A, \gamma_A)$ is an intuitionistic fuzzy bi-ideal of R, but not an intuitionistic fuzzy right ideal of R, as

$$\begin{aligned} \mu_A(41) &= \mu_A(3) = 0. \\ \mu_A(4) &= 0.7. \\ &\Rightarrow \mu_A(41) \ngeq \mu_A(4). \\ \text{and } \gamma_A(41) &= \gamma_A(3) = 0.7. \\ \gamma_A(4) &= 0. \\ &\Rightarrow \gamma_A(41) \nleq \gamma_A(4). \end{aligned}$$

Lemma 2.22: Every intuitionistic fuzzy bi-ideal of an LA-ring R is an intuitionistic fuzzy generalized bi-ideal of R.

Proof: Obvious.

Lemma 2.23: Every intuitionistic fuzzy left (resp. right, two-sided) ideal of an LA-ring R is an intuitionistic fuzzy quasi-ideal of R.

Proof: Assume that $A = (\mu_A, \gamma_A)$ is an intuitionistic fuzzy left ideal of R. Now $\mu_A \circ R \cap R \circ \mu_A \subseteq R \circ \mu_A \subseteq \mu_A$ and $\gamma_A \circ R \cup R \circ \gamma_A \supseteq R \circ \gamma_A \supseteq \gamma_A$. So A is an intuitionistic fuzzy quasi-ideal of R.

Lemma 2.24: Let R be an LA-ring with left identity e, such that $(xe)R = xR$ for all $x \in R$. Then every intuitionistic fuzzy quasi-ideal of R is an intuitionistic fuzzy bi-ideal of R.

Proof: Let $A = (\mu_A, \gamma_A)$ be an intuitionistic fuzzy quasi-ideal of R and $A \circ A \subseteq A$ by the Proposition 2.20. Now

$$\begin{aligned} (\mu_A \circ R) \circ \mu_A &\subseteq (R \circ R) \circ \mu_A \subseteq R \circ \mu_A \\ \text{and } (\mu_A \circ R) \circ \mu_A &\subseteq (\mu_A \circ R) \circ R = (\mu_A \circ R) \circ (e \circ R) \\ &= (\mu_A \circ e) \circ (R \circ R) \subseteq (\mu_A \circ e) \circ R = \mu_A \circ R. \\ &\Rightarrow (\mu_A \circ R) \circ \mu_A \subseteq \mu_A \circ R \cap R \circ \mu_A \subseteq \mu_A. \end{aligned}$$

Similarly, $(\gamma_A \circ R) \circ \gamma_A \supseteq \gamma_A \circ R \cup R \circ \gamma_A \supseteq \gamma_A$. Hence A is an intuitionistic fuzzy bi-ideal of R.

3 REGULAR LA-RINGS

An LA-ring R is called a regular if for every $x \in R$, there exists an element $a \in R$ such that $x = (xa)x$. In this section, we characterize regular LA-rings by the properties of intuitionistic fuzzy left (right, quasi-, bi-, generalized bi-) ideals.

Lemma 3.1: Every intuitionistic fuzzy right ideal of a regular LA-ring R is an intuitionistic fuzzy ideal of R.

Proof: Suppose that $A=(\mu_A,\gamma_A)$ is an intuitionistic fuzzy right ideal of R. Let $x, y \in R$, this implies that there exists an element $a \in R$, such that $x = (xa)x$. Thus

$$\mu_A(xy) = \mu_A(((xa)x)y) = \mu_A((yx)(xa)) \geq \mu_A(yx) \geq \mu_A(y)$$

and

$$\gamma_A(xy) = \gamma_A(((xa)x)y) = \gamma_A((yx)(xa)) \leq \gamma_A(yx) \leq \gamma_A(y).$$

Hence A is an intuitionistic fuzzy ideal of R.

Lemma 3.2: Let $A = (\mu_A, \gamma_A)$ be an IFS of a regular LA-ring R. Then A is an intuitionistic fuzzy ideal of R if and only if A is an intuitionistic fuzzy interior ideal of R.

Proof: Consider that $A = (\mu_A, \gamma_A)$ is an intuitionistic fuzzy interior ideal of R. Let $x, y \in R$, then there exists an element $a \in R$, such that $x = (xa)x$. Thus

$$\mu_A(xy) = \mu_A(((xa)x)y) = \mu_A((yx)(xa)) \geq \mu_A(x)$$

and

$$\gamma_A(xy) = \gamma_A(((xa)x)y) = \gamma_A((yx)(xa)) \leq \gamma_A(x),$$

i.e., A is an intuitionistic fuzzy right ideal of R. So A is an intuitionistic fuzzy ideal of R by the Lemma 3.1. Converse is true by the Lemma 2.18.

Remark 3.3: The concept of intuitionistic fuzzy (interior, two-sided) ideals coincides with the same concept in regular LA-rings.

Proposition 3.4: Let R be a regular LA-ring. Then $(A \circ R) \cap (R \circ A) = A$ for every intuitionistic fuzzy right ideal A of R.

Proof: Suppose that $A = (\mu_A, \gamma_A)$ is an intuitionistic fuzzy right ideal of R. This implies that $(A \circ R) \cap (R \circ A) \subseteq A$, because every intuitionistic fuzzy right ideal of R is an intuitionistic fuzzy quasi-ideal of R by the Lemma 2.23. Let $x \in R$, this implies that there exists an element $a \in R$, such that $x = (xa)x$. Thus

$$\begin{aligned}
(\mu_A \circ R)(x) &= \vee_{x = \sum_{i=1}^{n} a_i b_i} \{\wedge_{i=1}^{n} \{\mu_A(a_i) \wedge R(b_i)\}\} \\
&\geq \mu_A(xa) \wedge R(x) \geq \mu_A(x) \wedge 1 = \mu_A(x) \\
\text{and } (\gamma_A \circ R)(x) &= \wedge_{x = \sum_{i=1}^{n} a_i b_i} \{\vee_{i=1}^{n} \{\gamma_A(a_i) \vee R(b_i)\}\} \\
&\leq \gamma_A(xa) \vee R(x) \leq \gamma_A(x) \vee 0 = \gamma_A(x). \\
&\Rightarrow A \subseteq A \circ R.
\end{aligned}$$

Similarly, we have $A \subseteq R \circ A$, i.e., $A \subseteq (A \circ R) \cap (R \circ A)$. Hence $(A \circ R) \cap (R \circ A) = A$.

Lemma 3.5: Let R be a regular LA-ring. Then $D \circ L = D \cap L$ for every intuitionistic fuzzy right ideal D and every intuitionistic fuzzy left ideal L of R.

Proof: Since $D \circ L \subseteq D \cap L$, for every intuitionistic fuzzy right ideal $D = (\mu_D, \gamma_D)$ and every intuitionistic fuzzy left ideal $L = (\mu_L, \gamma_L)$ of R by the Lemma 2.14. Let $x \in R$, this means that there exists an element $a \in R$ such that $x = (xa)x$. Thus

$$\begin{aligned}
(\mu_D \circ \mu_L)(x) &= \vee_{x = \sum_{i=1}^{n} a_i b_i} \{\wedge_{i=1}^{n} \{\mu_D(a_i) \wedge \mu_L(b_i)\}\} \\
&\geq \mu_D(xa) \wedge \mu_L(x) \geq \mu_D(x) \wedge \mu_L(x) = (\mu_D \cap \mu_L)(x) \\
\text{and } (\gamma_D \circ \gamma_L)(x) &= \wedge_{x = \sum_{i=1}^{n} a_i b_i} \{\vee_{i=1}^{n} \{\gamma_D(a_i) \vee \gamma_L(b_i)\}\} \\
&\leq \gamma_D(xa) \vee \gamma_L(x) \leq \gamma_D(x) \vee \gamma_L(x) = (\gamma_D \cup \gamma_L)(x).
\end{aligned}$$

Therefore $D \circ L = D \cap L$.

Lemma 3.6: Let R be an LA-ring with left identity e. Then Ra is the smallest left ideal of R containing a.

Proof: Let $x, y \in Ra$ and $r \in R$. This implies that $x = r_1 a$ and $y = r_2 a$, where $r_1, r_2 \in R$. Now

$$\begin{aligned}
x - y &= r_1 a - r_2 a = (r_1 - r_2)a \in Ra \\
\text{and } rx &= r(r_1 a) = (er)(r_1 a) = ((r_1 a)r)e = ((r_1 a)(er))e \\
&= ((r_1 e)(ar))e = (e(ar))(r_1 e) = (ar)(r_1 e) \\
&= ((r_1 e)r)a \in Ra.
\end{aligned}$$

Since $a = ea \in Ra$. Thus Ra is a left ideal of R containing a. Let I be another left ideal of R containing a. Since $ra \in I$, where $ra \in Ra$, i.e., $Ra \subseteq I$. Hence Ra is the smallest left ideal of R containing a.

Lemma 3.7: Let R be an LA-ring with left identity e. Then aR is a left ideal of R.

Proof: Straight forward.

Proposition 3.8: Let R be an LA-ring with left identity e. Then $aR \cup Ra$ is the smallest right ideal of R containing a.

Proof: Let $x, y \in aR \cup Ra$, this means that $x, y \in aR$ or Ra. Since aR and Ra both are left ideals of R, so $x - y \in aR$ and Ra, i.e., $x - y \in aR \cup Ra$. We have to show that $(aR \cup Ra)R \subseteq (aR \cup Ra)$. Now

$$\begin{aligned}
(aR \cup Ra)R &= (aR)R \cup (Ra)R = (RR)a \cup (Ra)(eR) \\
&\subseteq Ra \cup (\mathrm{Re})(aR) = Ra \cup R(aR) \\
&= Ra \cup a(RR) \subseteq Ra \cup aR = aR \cup Ra. \\
&\Rightarrow (aR \cup Ra)R \subseteq aR \cup Ra.
\end{aligned}$$

Since $a \in Ra$, i.e., $a \in aR \cup Ra$. Let I be another right ideal of R containing a. Since $aR \in IR \subseteq I$ and $Ra = (RR)a = (aR)R \in (IR)R \subseteq IR \subseteq I$, i.e., $aR \cup Ra \subseteq I$. Therefore $aR \cup Ra$ is the smallest right ideal of R containing a.

Theorem 3.9: Let R be an LA-ring with left identity e, such that $(xe)R = xR$ for all $x \in R$. Then the following conditions are equivalent.

(1) R is a regular.

(2) $D \cap L = D \circ L$ for every intuitionistic fuzzy right ideal D and every intuitionistic fuzzy left ideal L of R.

(3) $C = (C \circ R) \circ C$ for every intuitionistic fuzzy quasi-ideal C of R.

Proof: Suppose that (1) holds and $C = (\mu_C, \gamma_C)$ be an intuitionistic fuzzy quasi-ideal of R. Then $(C \circ R) \circ C \subseteq C$, because every intuitionistic fuzzy quasi-ideal of R is an intuitionistic fuzzy bi-ideal of R by the Lemma 2.24. Let $x \in R$, this implies that there exists an element a of R such that $x = (xa)x$. Thus

$$\begin{aligned}((\mu_C \circ R) \circ \mu_C)(x) &= \vee_{a = \sum_{i=1}^{n} a_i b_i} \{\wedge_{i=1}^{n} \{(\mu_C \circ R)(a_i) \wedge \mu_C(b_i)\}\} \\ &\geq (\mu_C \circ R)(xa) \wedge \mu_C(x) \\ &= \vee_{xa = \sum_{i=1}^{n} p_i q_i} \{\wedge_{i=1}^{n} \{\mu_C(p_i) \wedge R(q_i)\}\} \wedge \mu_C(x) \\ &\geq \mu_C(x) \wedge R(a) \wedge \mu_C(x) = \mu_C(x). \\ &\Rightarrow \mu_C \subseteq (\mu_C \circ R) \circ \mu_C.\end{aligned}$$

Similarly, we have $\gamma_C \supseteq (\gamma_C \circ R) \circ \gamma_C$, i.e., $C = (C \circ R) \circ C$. Hence (1) implies (3). Assume that (3) holds. Let D be an intuitionistic fuzzy right ideal and L be an intuitionistic fuzzy left ideal of R. This means that D and L be intuitionistic fuzzy quasi-ideals of R by the Lemma 2.23, so $D \cap L$ be also an intuitionistic fuzzy quasi-ideal of R. Then by our assumption, $D \cap L = ((D \cap L) \circ R) \circ (D \cap L) \subseteq (D \circ R) \circ L \subseteq D \circ L$, i.e., $D \cap L \subseteq D \circ L$. Since $D \circ L \subseteq D \cap L$. Therefore $D \circ L = D \cap L$, i.e., $(3) \Rightarrow (2)$. Suppose that (2) is true and $a \in R$. Then Ra is a left ideal of R containing a by the Lemma 3.7 and $aR \cup Ra$ is a right ideal of R containing a by the Proposition 3.8. This implies that χ_{Ra} is an intuitionistic fuzzy left ideal and $\chi_{aR \cup Ra}$ is an intuitionistic fuzzy right ideal of R, by the Theorem 2.3. Then by our supposition

$$\chi_{aR \cup Ra} \cap \chi_{Ra} = \chi_{aR \cup Ra} \circ \chi_{Ra}, \text{ i.e., } \chi_{(aR \cup Ra) \cap Ra} = \chi_{(aR \cup Ra)Ra}$$

by the Theorem 2.2. Thus $(aR \cup Ra) \cap Ra = (aR \cup Ra)Ra$. Since $a \in (aR \cup Ra) \cap Ra$, i.e., $a \in (aR \cup Ra)Ra$, so $a \in (aR)(Ra) \cup (Ra)(Ra)$. This implies that

$$a \in (aR)(Ra) \text{ or } a \in (Ra)(Ra).$$

If $a \in (aR)(Ra)$, then

$a = (ax)(ya) = ((ya)x)a = (((ey)a)x)a = (((ay)e)x)a = ((xe)(ay))a = (a((xe)y))a$ for any $x, y \in R$.

If $a \in (Ra)(Ra)$, then $(Ra)(Ra) = ((\mathrm{Re})a)(Ra) = ((ae)R)(Ra) = (aR)(Ra)$, i.e., $a \in (aR)(Ra)$. So a is a regular, i.e., R is a regular. Hence $(2) \Rightarrow (1)$.

Theorem 3.10: Let R be an LA-ring with left identity e, such that $(xe)R = xR$ for all $x \in R$. Then the following conditions are equivalent.

(1) R is a regular.

(2) $A = (A \circ R) \circ A$ for every intuitionistic fuzzy quasi-ideal A of R.

(3) $B = (B \circ R) \circ B$ for every intuitionistic fuzzy bi-ideal B of R.

(4) $C = (C \circ R) \circ C$ for every intuitionistic fuzzy generalized bi-ideal C of R.

Proof: $(1) \Rightarrow (4)$, is obvious. Since $(4) \Rightarrow (3)$, every intuitionistic fuzzy bi-ideal of R is an intuitionistic fuzzy generalized bi-ideal of R by the Lemma 2.22. Since $(3) \Rightarrow (2)$, every intuitionistic fuzzy quasi-ideal of R is an intuitionistic fuzzy bi-ideal of R by the Lemma 15. $(2) \Rightarrow (1)$, by the Theorem 3.9.

Theorem 3.11: Let R be an LA-ring with left identity e, such that $(xe)R = xR$ for all $x \in R$. Then the following conditions are equivalent.

(1) R is a regular.

(2) $A \cap I = (A \circ I) \circ A$ for every intuitionistic fuzzy quasi-ideal A and every intuitionistic fuzzy ideal I of R.

(3) $B \cap I = (B \circ I) \circ B$ for every intuitionistic fuzzy bi-ideal B and every intuitionistic fuzzy ideal I of R.

(4) $C \cap I = (C \circ I) \circ C$ for every intuitionistic fuzzy generalized bi-idea C and every intuitionistic fuzzy ideal I of R.

Proof: Assume that (1) holds. Let $C = (\mu_C, \gamma_C)$ be an intuitionistic fuzzy generalized bi-ideal and $I = (\mu_I, \gamma_I)$ be an intuitionistic fuzzy ideal of R. Now $(C \circ I) \circ C \subseteq (R \circ I) \circ R \subseteq I \circ R \subseteq I$ and $(C \circ I) \circ C \subseteq (C \circ R) \circ C \subseteq C$, i.e., $(C \circ I) \circ C \subseteq C \cap I$. Let $x \in R$, this means that there exists an element $a \in R$ such that $x = (xa)x$. Now $xa = ((xa)x)a = (ax)(xa) = x((ax)a)$. Thus

$$\begin{aligned}
((\mu_C \circ \mu_I) \circ \mu_C)(x) &= \vee_{x = \sum_{i=1}^{n} a_i b_i} \{\wedge_{i=1}^{n} \{(\mu_C \circ \mu_I)(a_i) \wedge \mu_C(b_i)\}\} \\
&\geq (\mu_C \circ \mu_I)(xa) \wedge \mu_C(x) \\
&= \vee_{xa = \sum_{i=1}^{n} p_i q_i} \{\wedge_{i=1}^{n} \{\mu_C(p_i) \wedge \mu_I(q_i)\}\} \wedge \mu_C(x) \\
&\geq \mu_C(x) \wedge \mu_I((ax)a) \wedge \mu_C(x) \\
&\geq \mu_C(x) \wedge \mu_I(x) = (\mu_C \cap \mu_I)(x). \\
\Rightarrow \mu_C \cap \mu_I &\subseteq (\mu_C \circ \mu_I) \circ \mu_C .
\end{aligned}$$

Similarly, we have $\gamma_C \cup \gamma_I \supseteq (\gamma_C \circ \gamma_I) \circ \gamma_C$. Hence $C \cap I = (C \circ I) \circ C$, i.e., $(1) \Rightarrow (4)$. It is clear that $(4) \Rightarrow (3)$ and $(3) \Rightarrow (2)$. Suppose that (2) is true. Then

$A \cap R = (A \circ R) \circ A,$ where R itself is an intuitionistic fuzzy two-sided ideal of $R.$ So $A = (A \circ R) \circ A,$ because every intuitionistic fuzzy two-sided ideal of R is an intuitionisitc fuzzy quasi-ideal of $R.$ Hence R is a regular by the Theorem 3.9, i.e., $(2) \Rightarrow (1).$

Theorem 3.12: Let R be an LA-ring with left identity $e,$ such that $(xe)R = xR$ for all $x \in R.$ Then the following conditions are equivalent.

(1) R is a regular.

(2) $A \cap D \subseteq D \circ A$ for every intuitionistic fuzzy quasi-ideal A and every intuitionistic fuzzy right ideal D of $R.$

(3) $B \cap D \subseteq D \circ B$ for every intuitionistic fuzzy bi-ideal B and every intuitionistic fuzzy right ideal D of $R.$

(4) $C \cap D \subseteq D \circ C$ for every intuitionistic fuzzy generalized bi-ideal C and every intuitionistic fuzzy right ideal D of $R.$

Proof: $(1) \Rightarrow (4),$ is obvious. It is clear that $(4) \Rightarrow (3)$ and $(3) \Rightarrow (2).$ Suppose that (2) holds, this implies that $D \cap A = A \cap D \subseteq D \circ A,$ where A is an intuitionistic fuzzy left ideal of $R.$ Since $D \circ A \subseteq D \cap A,$ i.e., $D \cap A = D \circ A.$ Hence R is a regular by the Theorem 3.9, i.e., $(2) \Rightarrow (1).$

4 IINTRA-REGULAR LA-RINGS

An LA-ring An LA-ring R is called an intra-regular if for every $x \in R,$ there exist elements $a_i, b_i \in R$ such that $x = \sum_{i=1}^{n} (a_i x^2) b_i$. In this section, we characterize intra-regular LA-rings by the properties of intuitionistic fuzzy left (right, quasi-, bi-, generalized bi-) ideals.

Lemma 4.1: Every intuitionistic fuzzy left (right) ideal of an intra-regular LA-ring R is an intuitionistic fuzzy ideal of $R.$

Proof: Suppose that $A = (\mu_A, \gamma_A)$ is an intuitionistic fuzzy right ideal of $R.$ Let $x, y \in R,$ this implies that there exist elements $a_i, b_i \in R,$ such that $x = \sum_{i=1}^{n} (a_i x^2) b_i$. Thus

$$\begin{aligned} \mu_A(xy) &= \mu_A(((a_i x^2) b_i) y) = \mu_A((y b_i)(a_i x^2)) \\ &\geq \mu_A(y b_i) \geq \mu_A(y) \\ \text{and } \gamma_A(xy) &= \gamma_A(((a_i x^2) b_i) y) = \gamma_A((y b_i)(a_i x^2)) \\ &\leq \gamma_A(y b_i) \leq \gamma_A(y). \end{aligned}$$

Hence A is an intuitionistic fuzzy ideal of $R.$

Proposition 4.2: Let A be an IFS of an intra-regular LA-ring R with left identity $e.$ Then A is an intuitionistic fuzzy ideal of R if and only if A is an intuitionistic fuzzy interior ideal of $R.$

Proof: Suppose that $A = (\mu_A, \gamma_A)$ is an intuitionistic fuzzy interior ideal of $R.$ Let $x, y \in R,$ this implies that there exist elements $a_i, b_i \in R,$ such that $x = \sum_{i=1}^{n} (a_i x^2) b_i$. Thus

$$\begin{aligned}\mu_A(xy) &= \mu_A(((a_i x^2)b_i)y) = \mu_A((yb_i)(a_i x^2)) \\ &= \mu_A((yb_i)(a_i(xx))) = \mu_A((yb_i)(x(a_i x))) \\ &= \mu_A((yx)(b_i(a_i x))) \geq \mu_A(x).\end{aligned}$$

Similarly, we have $\gamma_A(xy) \leq \gamma_A(x)$, i.e., A is an intuitionistic fuzzy right ideal of R. Hence A is an intuitionistic fuzzy ideal of R by the Lemma 4.1. Converse is true by the Lemma 2.18.

Remark 4.3: The concept of intuitionistic fuzzy (interior, two-sided) ideals coincides in intra-regular LA-rings with left identity.

Lemma 4.4: Let R be an intra-regular LA-ring with left identity e. Then $D \cap L \subseteq L \circ D$ for every intuitionistic fuzzy left ideal L and every intuitionistic fuzzy right ideal D of R.

Proof: Let $L = (\mu_L, \gamma_L)$ be an intuitionistic fuzzy left ideal and $D = (\mu_D, \gamma_D)$ be an intuitionistic fuzzy right ideal of R. Let $x \in R$, this means that there exist elements such that $a_i, b_i \in R$ such that $x = \sum_{i=1}^{n}(a_i x^2)b_i$. Now

$$\begin{aligned}x &= (a_i x^2)b_i = (a_i(xx))b_i = (x(a_i x))b_i \\ &= (x(a_i x))(eb_i) = (xe)((a_i x)b_i) = (a_i x)((xe)b_i).\end{aligned}$$

Thus

$$\begin{aligned}(\mu_L \circ \mu_D)(x) &= \vee_{x=\sum_{i=1}^{n} p_i q_i} \{\wedge_{i=1}^{n}\{\mu_L(p_i) \wedge \mu_D(q_i)\}\} \\ &\geq \mu_L(a_i x) \wedge \mu_D((xe)b_i) \geq \mu_L(x) \wedge \mu_D(x) \\ &= \mu_D(x) \wedge \mu_L(x) = (\mu_D \cap \mu_L)(x) \\ \text{and } (\gamma_L \circ \gamma_D)(x) &= \wedge_{x=\sum_{i=1}^{n} p_i q_i} \{\vee_{i=1}^{n}\{\gamma_L(p_i) \vee \gamma_D(q_i)\}\} \\ &\leq \gamma_L(a_i x) \vee \gamma_D((xe)b_i) \leq \gamma_L(x) \vee \gamma_D(x) \\ &= \gamma_D(x) \vee \gamma_L(x) = (\gamma_D \cup \gamma_L)(x). \\ &\Rightarrow D \cap L \subseteq L \circ D.\end{aligned}$$

Theorem 4.5: Let R be an LA-ring with left identity e, such that $(xe)R = xR$ for all $x \in R$. Then the following conditions are equivalent.

(1) R is an intra-regular.

(2) $D \cap L \subseteq L \circ D$ for every intuitionistic fuzzy left ideal L and every intuitionistic fuzzy right ideal D of R.

Proof: $(1) \Rightarrow (2)$ is true by the Lemma 4.4. Suppose that (2) holds. Let $a \in R$, then Ra is a left ideal of R containing a by the Lemma 3.6 and $aR \cup Ra$ is a right ideal of R containing a by the Proposition 3.8. So χ_{Ra} is an intuitionistic fuzzy left ideal and $\chi_{aR \cup Ra}$ is an intuitionistic fuzzy right ideal of R, by the Theorem 1.3. By our supposition

$$\chi_{aR \cup Ra} \cap \chi_{Ra} \subseteq \chi_{Ra} \circ \chi_{aR \cup Ra}, \text{ i.e., } \chi_{(aR \cup Ra) \cap Ra} \subseteq \chi_{(Ra)(aR \cup Ra)}$$

by the Theorem 1.2. Thus $(aR \cup Ra) \cap Ra \subseteq Ra(aR \cup Ra)$. Since $a \in (aR \cup Ra) \cap Ra$, i.e., $a \in Ra(aR \cup Ra) = (Ra)(aR) \cup (Ra)(Ra)$. This implies that $a \in (Ra)(aR)$ or $a \in (Ra)(Ra)$. If $a \in (Ra)(aR)$, then

$$\begin{aligned}(Ra)(aR) &= (Ra)((ea)(RR)) = (Ra)((RR)(ae)) \\ &= (Ra)(((ae)R)R) = (Ra)((aR)R) \\ &= (Ra)((RR)a) = (Ra)(Ra) = ((Ra)a)R \\ &= ((Ra)(ea))R = ((\mathrm{Re})(aa))R = (Ra^2)R.\end{aligned}$$

So $a \in (Ra^2)R$. If $a \in (Ra)(Ra)$, then obvious $a \in (Ra^2)R$. This implies that a is an intra-regular. Hence R is an intra-regular, i.e., $(2) \Rightarrow (1)$.

Theorem 4.6: Let R be an LA-ring with left identity e, such that $(xe)R = xR$ for all $x \in R$. Then the following conditions are equivalent.

(1) R is an intra-regular.

(2) $A \cap I = (A \circ I) \circ A$ for every intuitionistic fuzzy quasi-ideal A and every intuitionistic fuzzy ideal I of R.

(3) $B \cap I = (B \circ I) \circ B$ for every intuitionistic fuzzy bi-ideal B and every intuitionistic fuzzy ideal I of R.

(4) $C \cap I = (C \circ I) \circ C$ for every intuitionistic fuzzy generalized bi-ideal C and every intuitionistic fuzzy ideal I of R.

Proof: Suppose that (1) holds. Let $C = (\mu_C, \gamma_C)$ be an intuitionistic fuzzy generalized bi-ideal and $I = (\mu_I, \gamma_I)$ be an intuitionistic fuzzy ideal of R. Now $(C \circ I) \circ C \subseteq (R \circ I) \circ R \subseteq I \circ R \subseteq I$ and $(C \circ I) \circ C \subseteq (C \circ R) \circ C \subseteq C$, thus $(C \circ I) \circ C \subseteq C \cap I$. Let $x \in R$, this implies that there exist elements $a_i, b_i \in R$ such that $x = \sum_{i=1}^{n} (a_i x^2) b_i$. Now

$$\begin{aligned}x &= (a_i x^2) b_i = (a_i (xx)) b_i = (x(a_i x)) b = (b_i (a_i x)) x. \\ b_i (a_i x) &= b_i (a_i ((a_i x^2) b_i)) = b_i ((a_i x^2)(a_i b_i)) = b_i ((a_i x^2) c_i) \\ &= (a_i x^2)(b_i c_i) = (a_i x^2) d_i = (a_i x^2)(e d_i) = (d_i e)(x^2 a_i) \\ &= m_i (x^2 a_i) = x^2 (m_i a_i) = (xx) l_i = (l_i x) x = (l_i x)(ex) \\ &= (xe)(x l_i) = x((xe) l_i).\end{aligned}$$

Thus

$$\begin{aligned}((\mu_C \circ \mu_I) \circ \mu_C)(x) &= \vee_{x = \sum_{i=1}^{n} p_i q_i} \{\wedge_{i=1}^{n} \{(\mu_C \circ \mu_I)(p_i) \wedge \mu_C(q_i)\}\} \\ &\geq (\mu_C \circ \mu_I)(b_i (a_i x)) \wedge \mu_C(x) \\ &= \vee_{b_i (a_i x) = \sum_{i=1}^{n} m_i n_i} \{\wedge_{i=1}^{n} \{\mu_C(m_i) \wedge \mu_I(n_i)\}\} \wedge \mu_C(x) \\ &\geq \mu_C(x) \wedge \mu_I((xe) l_i) \wedge \mu_C(x) \\ &\geq \mu_C(x) \wedge \mu_I(x) = (\mu_C \cap \mu_I)(x). \\ &\Rightarrow \mu_C \cap \mu_I \subseteq (\mu_C \circ \mu_I) \circ \mu_C.\end{aligned}$$

Similarly, we have $\gamma_C \cup \gamma_I \supseteq (\gamma_C \circ \gamma_I) \circ \gamma_C$. Hence $C \cap I = (C \circ I) \circ C$, i.e. $(1) \Rightarrow (4)$. It is clear that $(4) \Rightarrow (3)$ and $(3) \Rightarrow (2)$. Assume that (2) is true. Let A be an intuitionistic fuzzy right ideal and I be an intuitionistic fuzzy two-sided ideal of R. Since every intuitionistic fuzzy right ideal of R is an intuitionistic fuzzy quasi-ideal of R by the Lemma 2.23, so A is an intuitionistic fuzzy quasi-ideal of R. By our assumption $A \cap I = (A \circ I) \circ A \subseteq (R \circ I) \circ A \subseteq I \circ A$, i.e., $A \cap I \subseteq I \circ A$. Hence R is an intra-regular by the Theorem 4.5, i.e., $(2) \Rightarrow (1)$.

Theorem 4.7: Let R be an LA-ring with left identity e, such that $(xe)R = xR$ for all $x \in R$. Then the following conditions are equivalent.

(1) R is an intra-regular.

(2) $A \cap L \subseteq L \circ A$ for every intuitionistic fuzzy quasi-ideal A and every intuitionistic fuzzy left ideal L of R.

(3) $B \cap L \subseteq L \circ B$ for every intuitionistic fuzzy bi-ideal B and every intuitionistic fuzzy left ideal L of R.

(4) $C \cap L \subseteq L \circ C$ for every intuitionistic fuzzy generalized bi-ideal C and every intuitionistic fuzzy left ideal L of R.

Proof: Assume that (1) holds. Let $C = (\mu_C, \gamma_C)$ be an intuitionistic fuzzy generalized bi-ideal and $L = (\mu_L, \gamma_L)$ be an intuitionistic fuzzy left ideal of R. Let $x \in R$, this means that there exist elements $a_i, b_i \in R$ such that $x = \sum_{i=1}^{n} (a_i x^2) b_i$. Now $x = (a_i(xx))b_i = (x(a_i x))b_i = (b_i(a_i x))x$. Thus

$$\begin{aligned}
(\mu_L \circ \mu_C)(x) &= \vee_{x = \sum_{i=1}^{n} p_i q_i} \{\wedge_{i=1}^{n} \{\mu_L(p_i) \wedge \mu_C(q_i)\}\} \\
&\geq \mu_L(b_i(a_i x)) \wedge \mu_C(x) \geq \mu_L(x) \wedge \mu_C(x) \\
&= \mu_C(x) \wedge \mu_L(x) = (\mu_C \cap \mu_L)(x). \\
&\Rightarrow \mu_C \cap \mu_L \subseteq \mu_L \circ \mu_C.
\end{aligned}$$

Similarly, we have $\gamma_C \cup \gamma_L \supseteq \gamma_L \circ \gamma_C$. Hence $C \cap L \subseteq L \circ C$, i.e., $(1) \Rightarrow (4)$. It is clear that $(4) \Rightarrow (3)$ and $(3) \Rightarrow (2)$. Suppose that (2) holds. Let A be an intuitionistic fuzzy right ideal and L be an intuitionistic fuzzy left ideal of R. Since every intuitionistic fuzzy right ideal of R is an intuitionistic fuzzy quasi-ideal of R, this implies that A is an intuitionistic fuzzy quasi-ideal of R. By our supposition $, A \cap L \subseteq L \circ A$. Thus R is an intra-regular by the Theorem 4.5, i.e., $(2) \Rightarrow (1)$.

5 CONCLUSION

Our ambition is to inspire the study and maturity of non associative algebraic structure (LA-ring). The objective is to explain original methodological developments on ordered LA-rings, which will be very helpful for upcoming theory of algebraic structure. The ideal of fuzzy set to the characterizations of LA-rings are captivating a great attention of algebraist.

The aim of this paper is to investigate, the study of (regular, intra-regular) LA-rings by using of fuzzy left (right, interior, quasi-, bi-, generalized bi-) ideals.

Acknowledgements:

The authors thank the anonymous referees for their careful reading and insightful suggestions**.**

REFERENCES

[1] K. T. Atanassov, "New operations defined over the intuitionistic fuzzy sets", *Fuzzy Sets and Systems*, **61**, 137-142 (1994).

[2] B. Banerjee, D. K. Basnet, "Intuitionistic fuzzy subrings and ideals", *J. Fuzzy Math.*, **11**, 139-155 (2003).

[3] K. A. Dib, N. Galhum, A. A. M. Hassan, "Fuzzy rings and fuzzy ideals", *Fuzzy Math.*, **4**, 245-261 (1996).

[4] K. C. Gupta, M. K. Kantroo, "The intrinsic product of fuzzy subsets of a ring", *Fuzzy Sets and Systems*, **57**, 103-110 (1993).

[5] K. Hur, S. Y. Jang, H. W. Kang, "Intuitionistic fuzzy ideals of a ring", *J.Korea Soc. Math. Educ. Ser. B: Pure Appl. Math.*, **12**, 193-209, (2005).

[6] K. Hur, H. W. Kang, H. k. Song, "Intuitionistic fuzzy subgroups and subrings", *Honam Math. J.*, **25**, 19-41 (2003).

[7] M. A. Kazim, M. Naseerudin, "On almost semigroups", *Alig. Bull. Math.*, **2**, 1-7, (1972).

[8] N. Kausar, M. Waqar, "Characterizations of non-associative rings by their intuitionistic fuzzy bi-ideals", *European Journal of Pure and Applied Mathematics*, **1**, 226-250 (2019).

[9] N. Kausar, "Characterizations of non-associative ordered semi-groups by the properties of their fuzzy ideals with thresholds (α,β]", *Prikladnaya Diskretnaya Matematika*, **4**, 337-59 (2019).

[10] N. Kausar, "Direct product of finite intuitionistic fuzzy normal subrings over non-associative rings", *European Journal of Pure and Applied Mathematics*, **12**(2), 622-648 (2019).

[11] N. Kausar, B. Islam, M. Javaid, S, Amjad, U. Ijaz, "Characterizations of non-associative rings by the properties of their fuzzy ideals", *Journal of Taibah University for Science*, **13** (1), 820-833 (2019).

[12] N. Kausar, B. Islam, S. Amjad, M. Waqar, "Intuitionistics fuzzy ideals with thresholds(α,β] in LA-rings", *European Journal of Pure and Applied Mathematics*, **12** (3), 906-943 (2019).

[13] N. Kausar, M. Waqar, "Direct product of finite fuzzy normal subrings over non-associative rings"' *International Journal of Analysis and Applications*, **17**(5), 752-770 (2019).

[14] N. Kausar, M. Alesemi, Salahuddin, "Anti-fuzzy interior ideals on Ordered AG-groupoids", *European Journal of Pure and Applied Mathematics*, **13** (1), 113-129 (2020).

[15] N. Kausar, Meshari Alesemi, Salahuddin, Mohammad Munir, "Characterizations of non-associative ordered semigroups by their intuitionistic fuzzy bi-ideals", *Discontinuity, Nonlinearity, and Complexity*, **9** (2), 257-275 (2020).

[16] M. S. Kamran, *Conditions for LA-semigroups to resemble associative structures*, Ph.D. Thesis, Quaid-i-Azam University, Islamabad, (1993).

[17] N. Kuroki, "Regular fuzzy duo rings", *Inform. Sci.*, **94**, 119-139 (1996).

[18] W. J. Liu, "Fuzzy invariant subgroups and ideals", *Fuzzy Sets and Systems*, **8**, 133-139 (1982).

[19] T. K. Mukherjee, M. K. Sen, "On fuzzy ideals of a ring 1", *Fuzzy Sets and Systems*, **21**, 99-104 (1987).

[20] T. K. Mukherjee, M. K. Sen, "Prime fuzzy ideals in rings", *Fuzzy Sets and Systems*, **32,** 337-341 (1989).

[21] T. Shah, N. Kausar, "Characterizations of non-associative ordered semigroups by their fuzzy bi-ideals", *Theoretical Computer Science*, **529**, 96-110 (2014).

[22] T. Shah, N. Kausar, I. Rehman, "Intuitionistic fuzzy normal subrings over a non-associative ring", *An. St. Univ. Ovidius Constanta*, **1,** 369-386 (2012).
[23] U. M. Swamy, K. L. N. Swamy, "Fuzzy prime ideals of rings", *J. Math. Anal. Appl.*, **134**, 94-103 (1988).
[24] L. A. Zadeh, "Fuzzy sets", *Information and control*, **8,** 338-363 (1965).

Received July 27, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

THE GLOBAL DYNAMICS OF A QUARTIC DIFFERENCE EQUATION

JASMIN BEKTEŠEVIĆ[1], FATIH DESTOVIĆ[2*], VAHIDIN HADŽIABDIĆ[1], MIDHAT MEHULJIĆ[1]

[1] University of Sarajevo, Faculty of Mechanical Engineering, Division of Mathematics, Sarajevo, Bosnia and Herzegovina

[2] University of Sarajevo, Faculty of Educational Sciences, Sarajevo, Bosnia and Herzegovina

*Corresponding author, e-mail: fatih_d@msn.com

DOI: 10.20948/mathmontis-2020-47-4

Summary. In this paper we will present the global dynamic and the Julia set of a quartic second order difference equation with nonnegative parameters and the initial conditions are arbitrary nonnegative real numbers.

1 INTRODUCTION

In general, polynomial difference equations and polynomial maps in the plane have been studied in both the real and complex domains (see [8, 9]). First results on quadratic polynomial difference equation have been obtained in [1, 2] but these results gave us only a part of the basins of attraction of equilibrium points and period-two solutions. In [4], the general second order difference equation is completely investigated and described the regions of initial conditions in the first quadrant for which all solutions tend to equilibrium points, period-two solutions, or the point at infinity, except for the case of infinitely many period-two solutions. In [3], case of infinitely many period-two solutions is completely investigated. Our results are based on the theorems which hold for monotone difference equations. Our principal tool is the theory of monotone maps, and in particular cooperative maps, which guarantee the existence and uniqueness of the stable and unstable invariant manifolds for the fixed points and periodic points (see [5]). Consider the difference equation

$$x_{n+1} = f(x_n, x_{n-1}); n = 0,1, \dots \tag{1}$$

where f is a continuous and increasing function in both variables. The following result has been obtained in [1]:

Theorem 1 Let $I \subseteq R$ and let $f \in C[I \times I, I]$ be a function which increases in both variables. Then for every solution of Eq. (1) the sub sequences $\{x_{2n}\}_{n=0}^{\infty}$ and $\{x_{2n+1}\}_{n=-1}^{\infty}$ of even and odd terms of the solution do exactly one of the following:

(i) Eventually they are both monotonically increasing.

(ii) Eventually they are both monotonically decreasing.

(iii) One of them is monotonically increasing and the other is monotonically decreasing

2010 Mathematics Subject Classification: 39A05, 39A10, 39A23
Keywords and Phrases: Basin of Attraction, Period-two solutions, Julia set, Difference equation, stable and unstable manifold, invariant manifolds

As a consequence of Theorem 1 every bounded solution of Eq. (1) approaches either an equilibrium solution or period-two solution and every unbounded solution is asymptotic to the point at infinity in a monotonic way. Thus, the major problem in dynamics of Eq. (1) is the problem of determining the basins of attraction of three different types of attractors: the equilibrium solutions, period-two solution(s) and the point(s) at infinity. The following result can be proved by using the techniques of proof of Theorem 11 in [5].

Theorem 2 Consider Eq. (1) where f is increasing function in its arguments and assume that there is no minimal period-two solution. Assume that $E_1(x_1, y_1)$ and $E_2(x_2, y_2)$ are two consecutive equilibrium points in North-East ordering that satisfy

$$(x_1, y_1) \preccurlyeq_{ne} (x_2, y_2)$$

and that E_1 is a local attractor and E_2 is a saddle point or a non-hyperbolic point with second characteristic root in interval $(-1,1)$, with the neighborhoods where f is strictly increasing. Then the basin of attraction $B(E_1)$ of E_1 is the region below the global stable manifold $W^s(E_2)$. More precisely

$$B(E_1) = \{(x, y): \exists y_u: y < y_u, (x, y_u) \in W^s(E_2)\}.$$

The basin of attraction $B(E_1) = W^s(E_2)$ is exactly the global stable manifold of E_2. The global stable manifold extends to the boundary of the domain of Eq. (1). If there exists a period-two solution, then the end points of the global stable manifold are exactly the period two solution.

Now, the theorems that are applied in [5] provided the two continuous curves $W^s(E_2)$ (sta-ble manifold) i $W^u(E_2)$ (unstable manifold), both passing through the point $E_2(x_2, y_2)$ from Theorem 2, such that $W^s(E_2)$ is a graph of decreasing function and $W^u(E_2)$ is a graph of an increasing function. The curve $W^s(E_2)$ splits the first quadrant of initial conditions into two disjoint regions, but we do not know the explicit form of the curve $W^u(E_2)$. In this paper we investigate the following difference equation

$$x_{n+1} = ax_n^3x_{n-1} + bx_n^2x_{n-1}^2 + bx_nx_{n-1}^3 + ax_{n-1}^4 + cx_n^2x_{n-1} + dx_nx_{n-1}^2 + \\ +cx_{n-1}^3 + ex_nx_{n-1} + ex_{n-1}^2 + fx_{n-1}, n = 0,1, \ldots \tag{2}$$

We expose the explicit form of the curve that separates the first quadrant into two basins of attraction of a locally stable equilibrium point and of the point at infinity. One of the major problems in the dynamics of polynomial maps is determining the basin of attraction of the point at infinity and in particular the boundary of the that basin known as the Julia set. We precisely determined the Julia set of Eq. (2) and we obtained the global dynamics in the interior of the Julia set, which includes all the points for which solutions are not asymptotic to the point at infinity. It turned out that the Julia set for Eq. (2) is the union of the stable manifolds of some saddle equilibrium points, nonhyperbolic equilibrium points or period-two points. We first list some results needed for the proofs of our theorems. The main result for studying local stability of equilibrium is linearized stability theorem (see Theorem 1.1 in [7]).

Theorem 3 (linearized stability): Consider the difference equation

$$x_{n+1} = f(x_n, x_{n-1}); n = 0,1, \ldots \tag{3}$$

and let $\bar{x}$ be an equilibrium point of difference equation (3). Let $p = \frac{\partial f(\bar{x},\bar{x})}{\partial u}$ and $q = \frac{\partial f(\bar{x},\bar{x})}{\partial v}$ denote the partial derivatives of $f(u,v)$ evaluated at equilibrium $\bar{x}$. Let λ_1 and λ_2 be roots of quadratic equation $\lambda^2 - p\lambda - q = 0$.

a) If $|\lambda_1| < 1$ and $|\lambda_2| < 1$, then the equilibrium $\bar{x}$ is locally asymptotically stable (sink).

b) If $|\lambda_1| > 1$ or $|\lambda_2| < 1$, then the equilibrium $\bar{x}$ is unstable.

c) $|\lambda_1| < 1 \; and \; |\lambda_2| < 1 \Leftrightarrow |p| < 1 - q < 2$. Equilibrium $\bar{x}$ is a sink.

d) $|\lambda_1| > 1 \; and \; |\lambda_2| > 1 \Leftrightarrow |q| > 1 \; and \; |p| < |1 - q|$. Equilibrium $\bar{x}$ is a repeller.

e) $|\lambda_1| > 1 \; and \; |\lambda_2| < 1 \Leftrightarrow |p| > |1 - q|$. Equilibrium $\bar{x}$ is a saddle point.

f) $|\lambda_1| = 1 \; or \; |\lambda_2| = 1 \Leftrightarrow |p| = |1 - q| \; or \; q = -1 \; and \; |p| \leq 2$. Equilibrium $\bar{x}$ is called a non-hyperbolic point.

The next theorem (Theorem 1.4.1. in [6]) is a very useful tool in establishing bounds for the solutions of nonlinear equations in terms of the solutions of equations with known behaviour.

Theorem 4 Let I be an interval of real numbers, let k be a positive integer, and let $F: I^{k+1} \to I$ I be a function which is increasing in all its arguments. Assume that $\{x_n\}_{n=-\mathrm{k}}^{\infty}$, $\{y_n\}_{n=-\mathrm{k}}^{\infty}$ and $\{z_n\}_{n=-\mathrm{k}}^{\infty}$ are sequences of real numbers such that

$$x_{n+1} \leq F(x_n, \dots, x_{n-k}), n = 0,1, \dots$$
$$y_{n+1} = F(y_n, \dots, y_{n-k}), n = 0,1, \dots$$
$$z_{n+1} \geq F(z_n, \dots, z_{n-k}), n = 0,1, \dots$$

and

$$x_n \leq y_n \leq z_n, \qquad for\ all\ -k \leq n \leq 0.$$

Then

$$x_n \leq y_n \leq z_n, \qquad for\ all\ n > 0.$$

The next well-known theorem gives us the number of positive zeros of a polynomial $P(x)$.

Theorem 5 Let $P(x) = a_0 x^{b_0} + a_1 x^{b_1} + \cdots + a_n x^{b_n}$ where a_i, $i = 0,1, \dots, n$ are real numbers and $0 \leq b_0 < b_1 < \cdots < b_n$ are integers. The number of positive zeros of $P(x) = 0$, counting multiplicities, is either equal to $v(P)$ or less than that by an even number, where $v(P)$ denotes the number of sign changes in the sequence $a_0, a_1, \dots, a_n$.

2 MAIN RESULTS

By using the Theorem 3, we obtained the following result on local stability of the zero equilibrium of Eq. (2):

Proposition 1 The zero equilibrium of Eq. (2) is one of the following:

a) locally asymptotically stable if $\mathrm{f} < 1$,

b) non-hyperbolic and locally stable if $\mathrm{f} = 1$,

c) unstable if $\mathrm{f} > 1$.

Set $f(x,y) = ax^3y + bx^2y^2 + bxy^3 + ay^4 + cx^2y + dxy^2 + cy^3 + exy + ey^2 + fy$ and let $p = \frac{\partial f(\bar{x},\bar{x})}{\partial u}$ and $q = \frac{\partial f(\bar{x},\bar{x})}{\partial v}$ denote the partial derivatives of $f(x,y)$ evaluated at the equilibrium $\bar{x}$. The linearized equation at the positive equilibrium $\bar{x}$ is

$$z_{n+1} = pz_n + qz_{n-1},$$
$$p = 3(a+b)\bar{x}^3 + (2c+d)\bar{x}^2 + e\bar{x},$$
$$p = 5(a+b)\bar{x}^3 + (4c+2d)\bar{x}^2 + 3e\bar{x} + f.$$

Now, in view of Theorem 3 we obtain the following results on local stability of the positive equilibrium of Eq. (2):

Proposition 2 The positive equilibrium of Eq. (2) is one of the following:

a) locally asymptotically stable if $\mathrm{p} + \mathrm{q} < 1$,

b) non-hyperbolic and locally stable if $\mathrm{p} + \mathrm{q} = 1$,

c) unstable if $\mathrm{p} + \mathrm{q} > 1$,

d) saddle point if $p > |q - 1|$,

e) repeller if $1 - \mathrm{q} < p < q - 1$.

Theorem 6 If $f \geq 1$ then every solution $\{x_n\}$ of Eq. (2) satisfies$\lim_{n\to\infty} x_n = \infty$.

Proof. If $\{x_n\}$ is a solution of Eq. (2) then $\{x_n\}$ satisfies the inequality

$$x_{n+1} \geq f x_{n-1}, n = 0,1, \ldots$$

which in view of the result on difference inequalities, see Theorem 4, implies that $x_n \geq y_n, n \geq 1$ where $\{x_n\}$ is a solution of the initial value problem

$$y_{n+1} f y_{n-1}, y_{-1} = x_{-1} \text{ and } y_0 = x_0 \;\; n = 0,1, \ldots$$

Consequently, $x_0, x_{-1} > 0$ then $y_0, y_{-1} > 0, y_n \geq 0$ $for\ all\ n$, and

$$x_n \geq y_n = \lambda_1 \sqrt{f}^n + \lambda_2 \left(-\sqrt{f}\right)^n, n = 1,2, \ldots$$

where $\lambda_1, \lambda_2 \in \boldsymbol{R}$ such that $y_n \geq 0$ for all n, which implies $\lim_{n\to\infty} x_n = \infty$.

Theorem 7 Consider the difference equation (2) in the first quadrant of initial conditions, where $a > 0, b, c, d, e, f \geq 0$ and $2e + f < 1$. Then Eq. (2) has a zero equilibrium and a unique positive equilibrium $\bar{x}_+$. The line $ax^3 + bx^2y + bxy^2 + ay^3 + cx^2 + dxy + cy^2 + ex + ey + f = 1$ is the Julia set and separates the first quadrant into two regions: the region below the line is the basin of attraction of point $E_0(0,0)$ the region above the line is the basin of attraction of the point at infinity and every point on the line except $E_+(x_+, x_+)$ is a period-two solution of Eq.(2)

Proof. The equilibrium points of Eq. (2) are the solutions of equation

$$2\,(a + b)\,x^4 + (2c + d)\,x^3 + 2ex^2 + f\,x = x$$

that is equivalent to

$$(2\,(a + b)\,x^3 + (2c + d)\,x^2 + 2ex + f - 1)\,x = 0 \qquad (4)$$

Since the number of sign changes in the sequence $2\,(a + b), 2c + d, 2e, f - 1$ is one, then by applying Theorem 5 implies Eq. (4) has two equilibria: zero equilibrium and unique positive equilibrium x_+. Since $f \geq 0$ and $f < 1$, then by applying Proposition (1) the zero equilibrium is locally asymptotically stable. Denote by $h(x;,y) = ax^3y + bx^2y^2 + bxy^3 + ay^4 + cx^2y + dxy^2 + cy^3 + exy + ey^2 + fy$ and let p and q denote the partial derivatives of function $h(x, y)$ at point E_+. By straightforward calculation we obtain that the following hold:

$$p + q = (3(a+b)\bar{x}^3 + (2c+d)\bar{x}^2 + e\bar{x}) + (5(a+b)\bar{x}^3 + (4c+2d)\bar{x}^2 + 3e\bar{x} + f)$$
$$= 8(a+b)\bar{x}^3 + (6c+3d)\bar{x}^2 + 4e\bar{x} + f$$
$$= 2\bar{x}(3(a+b)\bar{x}^2 + (2c+d)\bar{x} + e) + (2(a+b)\bar{x}^3 + (2c+d)\bar{x}^2 + 2e\bar{x} + f)$$
$$= 2\bar{x}(3(a+b)\bar{x}^2 + (2c+d)\bar{x} + e) + 1 > 1.$$
$$q - p = (5(a+b)\bar{x}^3 + (4c+2d)\bar{x}^2 + 3e\bar{x} + f) - (3(a+b)\bar{x}^3 + (2c+d)\bar{x}^2 + e\bar{x})$$
$$= (2(a+b)\bar{x}^3 + (2c+d)\bar{x}^2 + 2e\bar{x} + f - 1) + 1 = 1.$$

Hence, by applying Proposition (2) the positive equilibrium is an unstable non-hyperbolic point. Period-two solution u, v satisfies the system

$$u = (av^3 + buv^2 + bu^2v + au^3 + cv^2 + duv + cu^2 + ev + eu + f)u$$
$$v = (au^3 + bu^2v + buv^2 + av^3 + cu^2 + duv + cv^2 + eu + ev + f)v.$$

Obviously, the point $(0,0)$ is solution of the system above, but it is not period two solution. Hence, it has to be $v > 0$ which implies

$$au^3 + bu^2v + buv^2 + av^3 + cu^2 + duv + cv^2 + eu + ev + f = 1.$$

Therefore every point of the set

$$\{(x,y):\ ax^3 + bx^2y + bxy^2 + ay^3 + cx^2 + dxy + cy^2 + ex + ey + f = 1\}$$

is a period-two solution of Eq. (2) except point E_+. Now, we have to show that line

$$g\,(x,y) = ax^3 + bx^2y + bxy^2 + ay^3 + cx^2 + dxy + cy^2 + ex + ey + f = 1$$

is a graph of the decreasing function in the first quadrant. Let for some $x > 0$ there are y_1 and y_2 $(0 < y_1 < y_2)$ such that $g(x,y_1) = g(x,y_2) = 1$. As $g(x,y)$ is increasing in both variables then

$$1 = g(x,y_1) < g(x,y_2) = 1,$$

which is impossible. Thus the curve $g(x,y) = 1$ is the graph of function in the first quadrant. Further over $g(x,y) = 1$ then

$$\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} y' = 0.$$

By applying the fact that is $g(x,y)$ is increasing in both variables we obtain $y' < 0$ in the first quadrant. Hence, $g(x,y) = 1$ is the graph of the decreasing function in the first quadrant. Let $\{x_n\}$ be a solution of Eq. (2) for initial condition (x_0, x_{-1}) which lies below the line

$$ax^3 + bx^2y + bxy^2 + ay^3 + cx^2 + dxy + cy^2 + ex + ey + f = 1.$$

Then

$$g(x_0,x_{-1}) = ax_0^3 + bx_0^2x_{-1} + bx_0x_{-1}^2 + ax_{-1}^3 + cx_0^2 + dx_0x_{-1} + cx_{-1}^2 + ex_0 + ex_{-1} + f < 1,$$

$$x_{n+1} = g(x_n, x_{n-1})x_{n-1},$$

and

$$x_1 = g(x_0,x_{-1})x_{-1} < x_{-1},$$
$$x_2 = g(x_1,x_0)x_0 < g(x_{-1},x_0)x_0 = g(x_0,x_{-1})x_0 < x_0.$$

Thus (x_2,x_1) and (x_0,x_{-1}) are two points in North-East ordering $(x_2,x_1) \leq_{ne} (x_0,x_{-1})$ which means that the point (x_2,x_1) is also below the line $g(x,y) = 1$ and also holds

$$g(x_2,x_1) < 1.$$

Similarly, we find

$$x_3 = g(x_2,x_1)x_1 < x_{-1},$$
$$x_4 = g(x_3,x_2)x_2 < g(x_1,x_2)x_2 = g(x_2,x_1)x_2 < x_2.$$

Continuing on this way we get

$$(0,0) \leq_{ne} \dots \leq_{ne} (x_4,x_3) \leq_{ne} (x_2,x_1) \leq_{ne} (x_0,x_{-1})$$

which implies that both sub sequences $\{x_{2n}\}$ and $\{x_{2n+1}\}$ are monotonically decreasing and bounded below by 0. Since below the line $g(x,y) = 1$ there are no period-two solutions it must be $x_{2n} \to 0$ and $x_{2n+1} \to 0$. On the other hand, if we consider solution $\{x_{2n}\}$ of Eq.(2)

for initial condition (x_0, x_{-1}) which lies above the line $g(x, y) = 1$ then $g(x_0, x_{-1}) > 1$ and by applying the method shown above we obtain the following condition:

$$(x_{-1}, x_0) \leq_{ne} (x_1, x_2) \leq_{ne} (x_3, x_4) \leq_{ne} \ldots$$

Therefore, both sub sequences $\{x_{2n}\}$ and $\{x_{2n+1}\}$ are monotonically increasing, hence $x_{2n} \to \infty$ and $x_{2n+1} \to \infty$ as n $\to \infty$.

The next figure is visual illustration of Theorem 7 obtained by using Mathematica 9.0, with the boundaries of the basins of attraction obtained by using the software package Dynamica [6].

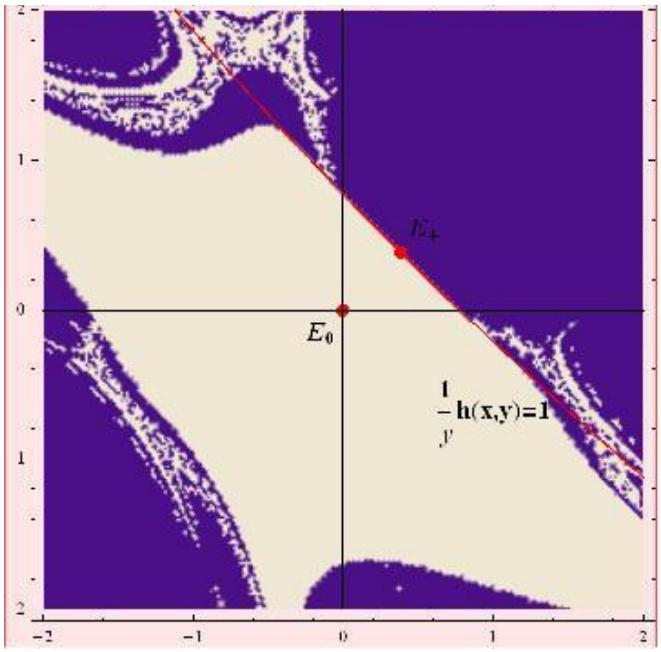


Figure 1. Illustration of Theorem 7
a = 0.3, b = 1, c = 0.5, d = 1, e = 0.4 and f = 0.25

In view of Theorem 4 which implies results on difference inequalities we get the following:

Proposition 3 Consider the difference equation of type

$$x_{n+1} = Ax_n^3 x_{n-1} + Bx_n^2 x_{n-1}^2 + Cx_n x_{n-1}^3 + Dx_{n-1}^4 + Ex_n^2 x_{n-1} + Fx_n x_{n-1}^2 + \tag{5}$$

$$+Gx_{n-1}^3 + Hx_n x_{n-1} + Ix_{n-1}^2 + Jx_{n-1},$$

where the given parameters satisfy conditions $A, B, C, D, E, F, G, H, I > 0$ and $0 < J < 1$. Then the global stable manifold of the positive equilibrium is between two lines

$$p_1: min\{A, D\}(x^3 + y^3) + min\{B, C\}(x^2 y + xy^2) + min\{E, G\}(x^2 + y^2) \tag{6}$$

$$+ min\{H, I\}(x + y) + Fxy + J = 1.$$

and

$$p_2: max\{A,D\}\,(x^3+y^3)+max\{B,C\}\,(x^2y+xy^2)+max\{E,G\}\,(x^2+y^2) \\ +max\{H,I\}\,(x+y)+Fxy+J=1. \tag{7}$$

Proof. Since the number of sign changes in the sequence $A+B+C+D, E+F+G, H+I, J-1$ is one, then by applying Theorem 5 implies Eq. (5) has two equilibria: zero equilibrium and unique positive equilibrium $\bar{x}_+$. Since $J<1$ the zero equilibrium is always locally asymptotically stable thus the positive equilibrium must be unstable equilibrium point. The theorems applied in [5] provided the following global behavior. More precisely, if the positive equilibrium is a saddle point or a non-hyperbolic point then there exists a global stable manifold which contains point $E_+(\bar{x},\bar{x})$; where $\bar{x}$ is the positive equilibrium. In this case global behavior of Eq. (5) is described by Theorem 9 in [4]. If the positive equilibrium is a repeller then there exists a period-two solution and we obtain that the period-two solution is a saddle point and there are two global stable manifolds which contain points $P_1(u,v)$ and $P_2(v,u)$ whre (u,v) is unique period-two solution of Eq.(5). In this case the global behavior of Eq. (5) is described by Theorem 10 in [4]. Although the Theorems 9 and 10 in [4] have been applied on a polynomial second order difference equation they are special cases of general Theorems in [5] applied on function f, where f is increasing function in its arguments. So, the global dynamics of Eq. (5) is exactly the same as the global dynamics of equations described by Theorems 9 and 10 in [4]. Furthermore

$$x_{n+1}=Ax_n^3x_{n-1}+Bx_n^2x_{n-1}^2+Cx_nx_{n-1}^3+Dx_{n-1}^4+Ex_n^2x_{n-1}+Fx_nx_{n-1}^2+ \\ +Gx_{n-1}^3+Hx_nx_{n-1}+Ix_{n-1}^2+Jx_{n-1} \\ \geq \big(min\{A,D\}\,(x_n^3+x_{n-1}^3)+min\{B,C\}\,(x_n^2x_{n-1}+x_nx_{n-1}^2)\big)x_{n-1}+ \\ +(min\{E,G\}(x_n^2+x_{n-1}^2)+min\{H,I\}(x_n+x_{n-1})+Fx_nx_{n-1}+J)x_{n-1},$$

and

$$x_{n+1}=Ax_n^3x_{n-1}+Bx_n^2x_{n-1}^2+Cx_nx_{n-1}^3+Dx_{n-1}^4+Ex_n^2x_{n-1}+Fx_nx_{n-1}^2+ \\ +Gx_{n-1}^3+Hx_nx_{n-1}+Ix_{n-1}^2+Jx_{n-1} \\ \leq \big(max\{A,D\}(x_n^3+x_{n-1}^3)+max(x_n^2x_{n-1}+x_nx_{n-1}^2)\big)x_{n-1}+ \\ +(max\{E,G\}(x_n^2+x_{n-1}^2)+max\{H,I\}(x_n+x_{n-1})+Fx_nx_{n-1}+J)x_{n-1}.$$

for all n, by applying Theorem 4 for solution $\{x_n\}$ of Eq. (5) the following inequality holds

$$y_n \leq x_n \leq z_n,$$

for all n, where $\{y_n\}$ is a solution of the difference equation

$$y_{n+1}=(min\{A,D\}\,(y_n^3+y_{n-1}^3)+min\{B,C\}\,(y_n^2y_{n-1}+y_ny_{n-1}^2))x_{n-1} \\ +(min\{E,G\}\,(y_n^2+y_{n-1}^2)+min\{H,I\}\,(y_n+y_{n-1})+Fy_ny_{n-1}+J)x_{n-1}, \tag{8}$$

and $\{z_n\}$ is a solution of the difference equation

$$z_{n+1}=(max\{A,D\}\,(z_n^3+z_{n-1}^3)+max\{B,C\}\,(z_n^2z_{n-1}+z_nz_{n-1}^2))x_{n-1} \tag{9}$$

$$+ (max\{E, G\}\,(z_n^2 + z_{n-1}^2) + max\{H, I\}\,(z_n + z_{n-1}) + F z_n z_{n-1} + J) x_{n-1}.$$

Since Eq. (8) and Eq. (9) satisfy all conditions of Theorem 7 this implies that the statement of Proposition 3 holds.

3 CONCLUSION

In this paper we restrict our attention to certain polynomial quartic second order difference equation Eq. (2). It is important to mention that we have accurately determined the Julia set of Eq. (2) and the basins of attractions for the zero equilibrium and the positive equilibrium point. In general, all theoretical concepts which are very useful in proving the results of global attractivity of equilibrium points and period-two solutions only give us existence of global stable manifold(s) whose computation leads to very uncomfortable calculus.

Acknowledgements: The authors would like to express their sincere thanks to reviewers on their valuable and constructive suggestions that contributed to this paper in clarity and form.

REFERENCES

[1] Amleh A. M., Camouzis E., Ladas G., "On the dynamics of rational difference equation, Part I", *Int. J. Difference Equ*., **3**, 1-35 (2008).

[2] Amleh A. M., Camouzis E., Ladas G., "On the dynamics of rational difference equation, Part II", *Int. J. Difference Equ*., **3**, 195-225 (2008).

[3] Bektešević J., Hadžiabdić V., Mehuljić M., Mujić N., "The Global Behavior of a Quadratic Difference Equation", *Filomat*, **32** (18), 6203-6210 (2018).

[4] Bektešević J., Kulenović M. R. S., Pilav E., "Global Dynamics of Quadratic Second Order Difference Equation in the First Quadrant", *App. Math. Comp*., **227**, 50-65 (2014).

[5] Brett A., Kulenović M. R. S., "Basins of attraction of equilibrium points of monotone difference equations" *Sarajevo J. Math*., **5**(18), 211–233 (2009)

[6] Camouzis E. Ladas G., *Dynamics of Third Order Rational Difference Equations with Open Problems and Conjectures*, Chapman and Hall/CRC Boca Raton London (2008).

[7] Kulenović M. R. S., Ladas G., *Dynamics of Second Order Rational Difference Equations with Open Problems and Conjectures*, Chapman and Hall/CRC Boca Raton London (2001).

[8] Milnor J., *Dynamics in One Complex Variable*, Stony Brook New York (2000).

[9] Morosawa S., Nishimura Y., Taniguchi M., Ueda T., *Holomorphic dynamics*, Cambridge University Press UK (2000).

Received January 13, 2020

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

RUNGE-KUTTA DISCONTINUOUS GALERKIN METHOD FOR HYPERBOLIC HYPERELASTICITY EQUATIONS FOR INHOMOGENEOUS MEDIUM

M.V. ALEKSEEV*, E.B. SAVENKOV

Keldysh Institute of Applied Mathematics, Russian Academy of Sciences. Moscow, Russia

*Corresponding author. E-mail: mikhail.alekseev@phystech.edu

DOI: 10.20948/mathmontis-2020-47-5

Summary. The work considers application of Runge-Kutta Discontinuous Galerkin method for solution of Godunov-Romenskii type hyperbolic model for hyperelastic medium. The medium is considered inhomogeneous with piecewise uniform distributed properties. To describe evolution of medium Godunov-Romenskii model is used supplemented with transport equation that describes evolution of properties distribution. The numerical approach is based on application of Runge-Kutta Discontinuous Galerkin method with Godunov type fluxes both for conservative and non-conservative terms. We describe mathematical model and corresponding numerical algorithm briefly. Results of numerical simulations are presented.

1 INTRODUCTION

The present paper is devoted to numerical study of Runge-Kutta Discontinuous Galerkin (RK/DG) method of high order of accuracy for solution of first order hyperbolic system of equations of hyperelasticity.The model describes dynamics of continuous media (deformation and strain fields, velocity, temperature and entropy) in Eulerian reference frame. The model was originally proposed[1] by S.Godunov and E.Romenskii. Recently there has been a significant increase in interest in such type of the models since it is assumed that in some cases they are more suitable (comparably to traditional models based on arbitrary Lagrangian-Eulerian description) to simulation of physical phenomena involving extremely large deformations of the media[2, 3]. Such type of problems are often arise in numerical simulation of shock wave phenomena in solids induced by rapid mechanical, thermal or radiation loads[24, 25].

Currently, a number of papers is devoted to numerical solution of hyperelasticity model[4, 5]. However, the most of them considers WENO-based approaches[6, 7]. In present work Runge-Kutta Discontinuous Galerkin method[8] is considered. The general motivation for such a choice is its universality and possibility of generalization to higher-order equations, that may occur in the multiphase problems. The second reason is to estimate efficiency of the RK/DG method when simulating Godunov-Romenskii model for particular cases of more simple (gas and fluid dynamics) and more complex (inhomogenious hyperelastic medium) settings. Both issues can be considered as a preliminary tests for further development of RK/DG numerical techniques for complex multiphase and multicomponent models developed in, e.g., Baer-Nunziato framework[9].

The main features of present paper are:

2010 Mathematics Subject Classification: 74H05, 65K05, 65Q10
Key words and Phrases: Runge-Kutta discontinuous Galerkin methods, hyperelasticity

- first order hyperbolic hyperelastic model is considered as a unified framework to describe solid/liquid/gaseous media;
- heterogeneous elastic medium case which assumes piecewise uniform distribution of medium properties is considered .

The structure of the paper is as follows. The basic Godunov-Romenskii hyperelastic model is described at the beginning of the section 2. In the subsection 2.1 its simple generalization to the case of piecewise homogeneous case is considered. Section 3 is devoted to the description of the RK/DG numerical algorithm for both conservative and non-conservative hyperbolic equations. In section 4 the implementation detalis and results of numerical experiments are presented.

2 MATHEMATICAL MODEL

To describe dynamics of the continuous hyperelastic medium in the Euler reference frame the Godunov-Romenskii model[10] is used. The corresponding system of equations is hyperbolic and consists of conservation law of momentum (1), dynamic equations for distortion tensor components (2) and conservation law of energy (3):

$$\frac{\partial(\rho\mathbf{u})}{\partial t}+\nabla\cdot\left(\rho\mathbf{u}\otimes\mathbf{u}-\mathbf{T}\right)=0, \tag{1}$$

$$\frac{\partial(\rho\mathbf{F})}{\partial t}+\nabla\cdot\left(\rho\mathbf{F}\otimes\mathbf{u}-\rho\mathbf{u}\otimes\mathbf{F}^{T}\right)=-\mathbf{u}\otimes\nabla\cdot\left(\rho\mathbf{F}\right), \tag{2}$$

$$\frac{\partial(\rho E)}{\partial t}+\nabla\cdot\left(\rho\mathbf{u}E-\mathbf{u}\otimes\mathbf{T}\right)=0. \tag{3}$$

Here $\mathbf{T}$ is Cauchy stress tensor, $E=\mathrm{U}+|\mathbf{u}|^2/2$ – total energy, $\mathrm{U}=\mathrm{U}(\mathbf{F},\mathrm{S})$ – internal energy. The primary variables are components of distortion tensor $\mathbf{F}=\nabla_{\mathbf{X}}\mathbf{x}$ ($\mathbf{x}$ and $\mathbf{X}$ are Euler and Lagrange coordinates of medium points, respectively), velocity $\mathbf{u}$ and entropy S. Symbol "$\otimes$" denotes the tensor product. The medium density ρ is defined as

$$\rho=\rho_0/\det(\mathbf{F}), \tag{4}$$

where $\rho_0=\rho_0(\mathbf{x})$ denotes the density of undeformed medium. The combination of equations (2) and (4) recovers the continuity equation:

$$\frac{\partial\rho}{\partial t}+\nabla\cdot\left(\rho\mathbf{u}\right)=0.$$

This equation can be used instead of one of the equations in (2) for the distortion tensor components.

System (1)-(3) has to be closed by the specific internal energy (equation of state, EOS) in its canonical form, $\mathrm{U}=\mathrm{U}(\mathbf{F},\mathrm{S})$. To provide the frame indifference of internal energy it must be expressed in terms of some symmetric strain tensor $\mathbf{G}$ [10, 5]:

$$\mathrm{U} = \mathrm{U}(\mathbf{F}, \mathrm{S}) = \hat{\mathrm{U}}(\mathbf{G}, \mathrm{S}). \tag{5}$$

Here and further $\mathrm{U}, \hat{\mathrm{U}}, \mathrm{U}^{+}, \ldots$ denote functional dependencies of the same variable on particular set of arguments.

A number of strain tensors[10, 11] can be used in (5). In the present work the Finger tensor $\mathbf{G} = \mathbf{F}^{-T}\mathbf{F}^{-1}$ is considered[4]. In this case the Cauchy stress tensor $\mathbf{T}$ is expressed by Murnaghan formula[10]:

$$\mathbf{T} = -2\rho\mathbf{G} \cdot \frac{\partial \hat{\mathrm{U}}}{\partial \mathbf{G}}. \tag{6}$$

Since $\hat{\mathrm{U}}$ is a function of Finger tensor components it can be expressed, due to objectivity arguments, as a function of its invariants $I_{1,2,3}$:

$$\overline{\mathrm{U}} = \mathrm{U}^{+}(I_1, I_2, I_3, \mathrm{S}), \tag{7}$$

$$I_1 = tr(\mathbf{G}),\ I_2 = \left[tr(\mathbf{G})^2 - tr(\mathbf{G}^2)\right]/2,\ I_3 = \det(\mathbf{G}).$$

The internal energy $\overline{\mathrm{U}}$ can be considered as the sum of two terms. The first one, U^{h}, is "hydrodynamical" part that depends only on bulk deformation and the second one, U^{sh}, describes dependency on shear deformation:

$$\mathrm{U}^{+} = \mathrm{U}^{\mathrm{h}}(I_3, \mathrm{S}) + \mathrm{U}^{\mathrm{sh}}(I_1, I_2, I_3, \mathrm{S}). \tag{8}$$

Hereafter the isotropic hyperelastic EOS[4] is used:

$$\mathrm{U}^{\mathrm{h}}(I_3, \mathrm{S}) = \frac{K_0}{2\alpha^2}\left(I_3^{\alpha/2} - 1\right)^2 + c_V T_0 I_3^{\gamma/2}\left(\exp\left[\mathrm{S}/c_V\right] - 1\right), \tag{9}$$

$$\mathrm{U}^{\mathrm{sh}}(I_1, I_2, I_3) = B_0 I_3^{\beta/2}\left(I_1^2/3 - I_2\right)/2. \tag{10}$$

Here $K_0 = c_0^2 - (4/3)b_0^2$ is the squared bulk sound velocity, c_0 is sound velocity, b_0 is shear elastic wave velocity, $B_0 = b_0^2$, c_V is the volumetric heat capacity, T_0 is reference temperature, α, β, γ are constant parameters.

The considered model can describe both solids ($\partial \mathrm{U}^{+}/\partial I_{1,2,3} \not\equiv 0$) and liquids/gases ($\partial \mathrm{U}^{+}/\partial I_{1,2} \equiv 0, \partial \mathrm{U}^{+}/\partial I_3 \not\equiv 0$). In the latter case the system of equations (1)-(3) can be reduced to classical gas dynamics equations with only bulk deformation accounted. That can be done by replacing the equations for distortion tensor components (2) with mass conservation law.

2.1 Piecewise homogeneous model

Consider now spatial domain Ω occupied by piecewise homogeneous medium. The latter means that parameters of EOS (9), (10) are different in different subdomains (phases) of Ω.

Let Ω_k, $k=\overline{1,N_f}$ where N_f is number of phases, be such subdomains, $\Omega=\bigcup\Omega_k$, $k=\overline{1,N_f}$. Let $\phi_k=\phi_k(\mathbf{X})$ be characteristic functions of Ω_k:

$$\phi_k(\mathbf{X})=\begin{cases}1, & \mathbf{X}\in\Omega_k,\\ 0, & \mathbf{X}\notin\Omega_k,\end{cases}\quad k=\overline{1,N_f};\quad \sum_{k=1}^{N_f}\phi_k(\mathbf{X})=1. \tag{11}$$

Set of EOS parameters specific for subdomain Ω_k is defined as $\mathbf{a}^{(k)}=\{\alpha,\beta,\gamma,K_0,\ldots\}$, $k=\overline{1,N_f}$. The EOS in that case has the form $\mathrm{U}_k=\mathrm{U}_k\left(\mathbf{G},\mathrm{S};\mathbf{a}^{(k)}\right)$ in domain Ω_k. Then the distribution of medium property in the Lagrangian reference frame is defined as

$$\mathbf{a}(\mathbf{X})=\sum_{k=1}^{N_f}\mathbf{a}^{(k)}\phi_k(\mathbf{X}). \tag{12}$$

During deformation of the medium, the values of characteristic functions $\hat{\phi}_k=\hat{\phi}_k(\mathbf{x},t)$ in Eulerian reference frame satisfy the following equation:

$$\frac{\partial\hat{\phi}_k}{\partial t}+\mathbf{u}\frac{\partial\hat{\phi}_k}{\partial\mathbf{x}}=0, \tag{13}$$

where $\mathbf{u}(\mathbf{x},t)$ is velocity defined as a function of Eulerian coordinates. Equation (13) should be supplemented by appropriate initial conditions.

Further we do not use characteristic functions $\hat{\phi}_k$ but rather their "smoothed" version. The smoothed zone width is a parameter of the model and is resolved by the computational mesh used in simulations. Equation (13) is nonconservative and is solved together with hyperelasticity model equations (1)-(3).

Considered above inhomogeneous model assumes that only one EOS is used to describe behavior of all phases, – that is, it is not "real" multiphase model. However, it has a number of features of multiphase models: e.g., it consists of two groups of equations (conservative and nonconservative)[12, 9]. In present work this model is considered as the simplest one to test and verify algorithmic techniques for numerical solution of more complicated models.

3 NUMERICAL ALGORITHMS

The considered class of problems is described by the hyperbolic system of equations of the first order. The total number of equations is large (13 equations of hyperelastic model plus N_f equations for ϕ_k), and its solution has a rich wave structure. The system consists of conservative and nonconservative equations. Its possible generalizations include spatial operators of higher order (for example, when considering surface tension). This motivates the following requirements for numerical methods:

- The method must be capable for calculations with high approximation accuracy.
- It has to provide possibility to construct numerical approximations of the hyperbolic operators as well as of diffusion ones.
- It can be applied in conservative and nonconservative settings.

- It should provide unified framework when considering both theoretical and software implementation issues.

We consider RK/DG as a candidate for such a framework. Further we briefly describe it for both conservative and nonconservative cases.

Conservative case. Consider one-dimensional conservation law in spatial domain $\Omega = [0, L] \subset \mathrm{R}$:

$$\frac{\partial g(x,t)}{\partial t} + \frac{\partial \mathrm{F}\left(g(x,t)\right)}{\partial x} = 0, \tag{14}$$

where $g(x,t)$ is conserved quantity and $\mathrm{F}\left(g\right)$ is corresponding physical flux.

Let $\{\omega_i\}_{i=0}^{i=N}$ be a partition of Ω into computational cells, and $\omega_i = \left[x_{i-1/2}, x_{i+1/2}\right], 1,, i,, N$. We shall denote by $\mathrm{V}_h^k(\Omega)$ the space of elements of $\mathrm{L}^\infty(\Omega)$ whose restriction to ω_i belongs to a vector space $\mathrm{P}^k(\omega_i)$ of polynomials of degree k :

$$\mathrm{V}_h^k = \left\{ v : v|_{\omega_i} \in \mathrm{P}^k(\omega_i); 1,, i,, N \right\}.$$

Define the elements of $\mathrm{P}^k(\omega_i)$ by linearly independent orthogonal set of Legendre polynomials $\left\{\psi_i^{(l)}\right\}_{l=0}^{l=k}$ and replace exact solution $g(x,t)$ in ω_i by its approximation

$$g_h(x,t)|_{\omega_i} = \sum_{l=0}^{k} \psi_i^{(l)}(x) g_i^{(l)}(t), \tag{15}$$

In order to obtain the semidiscrete equation for function $g_h(x,t)$ we multiply equation (14) by test function $v_h \in \mathrm{V}_h^k$, integrate over ω_i and apply Green's formula:

$$\int_{\omega_i} \frac{\partial g_h(x,t)}{\partial t} v_h(x) dx + \int_{\omega_i} \mathrm{F}\left(g_h(x,t)\right) \frac{\partial v_h(x)}{\partial x} dx + \sum_{e \in \partial\omega_i} \int_e \hat{\mathrm{F}}\left(g_h(x,t)\right) v_h(x) d\Gamma = 0. \tag{16}$$

In equation (16) physical flux $\mathrm{F}\left(g_h(x,t)\right)$ is replaced by numerical flux $\hat{\mathrm{F}}\left(g_h(x,t)\right)$ in surface integral. In one-dimensional case one can obtain:

$$\sum_{e \in \partial\omega_i} \int_e \hat{\mathrm{F}} v_h d\Gamma = \hat{\mathrm{F}}_{i+1/2} v_h\left(x_{i+1/2}^+\right) - \hat{\mathrm{F}}_{i-1/2} v_h\left(x_{i-1/2}^-\right).$$

Here $\hat{\mathrm{F}}_{i\pm1/2}$ is numerical flux at $x \in \partial\omega_i$.

$$\hat{\mathrm{F}}_{i\pm1/2} = \hat{\mathrm{F}}\left(g_{i\pm1/2}^+, g_{i\pm1/2}^-\right),$$

where $g_{i\pm1/2}^+$ and $g_{i\pm1/2}^-$ are left and right hand side limiting values of g_h at $x \in \partial\omega_i$. Different numerical fluxes for hyperbolic hyperelastic models are known (including HLLC, HLL, etc.)[4]. We consider here only two options:

1. Lax-Friedrichs flux:

$$\hat{\mathrm{F}}^{\mathrm{LF}}_{i\pm1/2} = \frac{1}{2}\left[\mathrm{F}\left(g^{+}_{i\pm1/2}\right) + \mathrm{F}\left(g^{-}_{i\pm1/2}\right)\right] - \frac{1}{2}\frac{\Delta x}{\Delta t}\left(g^{+}_{i\pm1/2} - g^{-}_{i\pm1/2}\right);$$

2. Rusanov flux:

$$\hat{\mathrm{F}}^{\mathrm{RS}}_{i\pm1/2} = \frac{1}{2}\left[\mathrm{F}\left(g^{+}_{i\pm1/2}\right) + \mathrm{F}\left(g^{-}_{i\pm1/2}\right)\right] - \frac{1}{2}\Lambda\left(g^{+}_{i\pm1/2} - g^{-}_{i\pm1/2}\right),$$

here

$$\Lambda = \Lambda\left(g^{+}_{i\pm1/2}, g^{-}_{i\pm1/2}\right) = \max\left(|\lambda^{+}_{i\pm1/2}|, |\lambda^{-}_{i\pm1/2}|\right),$$

where $\lambda^{+}_{i\pm1/2}$, $\lambda^{-}_{i\pm1/2}$ – eigenvalues of Jacobian matrices $J(g^{+}_{i\pm1/2})$ and $J(g^{-}_{i\pm1/2})$, $J(g) = \partial\mathrm{F}\left(g\right)/\partial g$.

Considering test functions in the form $v_h\left(x\right) = \left\{\psi_i^{(l)}\right\}_{l=0}^{l=k}$ the following system of ordinal differential equation is obtained for vector of coefficients $\hat{\mathbf{g}} = \left\{g_i^{(l)}\right\}_{l=0}^{l=k}$:

$$\frac{d\hat{\mathbf{g}}}{dt} = \mathbf{M}(\hat{\mathbf{g}}). \tag{17}$$

For time discretization of (17) a strong stability preserving TVD/RK3 method[8] is used. The appropriate limiting procedure (see below) is applied at each Runge–Kutta stage.

Nonconservative case. Let us describe now the RK/DG method applied to the nonconservative equation (13) in spatial domain $\Omega = [0, L] \subset \mathrm{R}$.

For (13) the traditional approach can not be applied in the same way as it was done previously for conservative case. The main cause is the difficulty of the definition of the solution $g\left(x,t\right)$ in terms of distributions. The correct formulation of the Riemann problem and corresponding generalized Hugoniot conditions can not be set in traditional way. The constructive solution to this problem is provided by DLM (DalMaso–LeFloch–Murat) theoretical framework[13].

Consider the discontinuous function

$$g(x) = g_{\mathrm{l}} + H(x - x_{\mathrm{d}})g_{\mathrm{r}},$$

where $H = H(x)$ is Heaviside function, x_{d} is the discontinuity coordinate, $g_{r,l} = g\left(x_d \pm 0\right)$ are the right and left-hand solution limits, respectively. For this case the expression

$$a\left(x, g\right)\frac{\partial g}{\partial x}, \tag{18}$$

where function $a(x,g)$ has the discontinuity at the same point x_{d}, can not be defined correctly as distirbution[14] and the special treatment is needed as developed in DLM theory. To proceed, replace g by its smooth regularization g_ε:

$$g_\varepsilon(x) = \begin{cases} g_{\mathrm{l}}, & x < x_{\mathrm{d}} - \varepsilon, \\ \Phi\left(\dfrac{x - x_{\mathrm{d}} + \varepsilon}{2\varepsilon}\right), & x_{\mathrm{d}} - \varepsilon,, x,, x_{\mathrm{d}} + \varepsilon, \\ g_{\mathrm{r}}, & x > x_{\mathrm{d}} + \varepsilon, \end{cases}$$

where $\varepsilon > 0$, mapping $\Phi : [0,1] \mapsto [g_{\mathrm{l}}, g_{\mathrm{r}}]$ is Liepshitz continuous and is called path[13]. Now define non-conservative product (18) as

$$a(x,g)\frac{\partial g}{\partial x} = \lim_{x \to x_{\mathrm{d}}} a(x, g_\varepsilon)\frac{\partial g_\varepsilon}{\partial x},$$

In such a way at $\varepsilon \to 0$ the product (18) can be defined as bounded Borel measure, converging to (18) in $*$-weak topology:

$$a(g_\varepsilon)\frac{\partial g_\varepsilon}{\partial x} \to C\delta(x - x_{\mathrm{d}}), \quad C = \int_0^1 a(\Phi(\tau))\frac{\partial \Phi(\tau)}{\partial \tau} d\tau, \tag{19}$$

where $\delta(x)$ is Dirac delta-function.

Consider again spatial domain $\Omega = [0, L] \subset \mathrm{R}$ with given partition $\{\omega_i\}_{i=0}^{i=N}$. Define the space of boundary points $\Gamma = \{x : x \in \partial\omega_i\}, i = \overline{1, N}$. Introduce the piecewise polynomials space V_h^k as was done previously. Multiply (13) by test function $v_h \in \mathrm{V}_h^k$ and integrate it over Ω taking (19) into account. This leads to the following semidiscrete equation for $g_h \in \mathrm{V}_h^k$:

$$\int_\Omega \frac{\partial g_h}{\partial t} v_h\, dx + \sum_{\omega_i \in \Omega} \int_{\omega_i} a(x, g_h)\frac{\partial g_h}{\partial x} v_h\, dx + \sum_{x \in \Gamma} \langle v_h \rangle_x \int_0^1 \frac{\partial \Phi}{\partial \tau}(\tau)\, a(\Phi(\tau))\, d\tau = 0,$$

where $\langle v_h \rangle = (v_h^{\mathrm{r}} + v_h^{\mathrm{l}})/2$. Borel measure (19) depends on the choice of the path Φ. In the present work the linear path $\Phi(\tau) = (g_{\mathrm{r}} - g_{\mathrm{l}})\tau + g_{\mathrm{l}}$ is chosen[13]. Further one can proceed as in the conservative setting.

Limiting procedure. For considered method the numerical solution will not be monotonic in case of discontinuous solution. To avoid non-physical oscillations in numerical solution an artificial dissipation has to be introduced. It can be done in various ways, among which methods based on geometric limiting, explicit introduction of additional dissipative terms and algorithms based on high-pass filtering component of the solution are known [15]. In [16] method for monotonizing the solution by explicitly introducing von Neumann-Richtmeier type artificial viscosity is described. The most popular technique is to use geometrical and moment-based limiters such as maximum preserving limiter[17], minmod limiter[8], or

Krivodonova limiter[18]. In the present work we use maximum preserving limiter for concentration function in nonconservative transport equation (13) and Krivodonova moment limiter for hyperelastic model (1)-(3).

4 NUMERICAL RESULTS

The described algorithm was implemented as program code using C++ language. The feature of software implementation is usage thermodynamical potential with its natural variables as EOS. For appropriate medium properties calculation the automatic differentiation technique is used (STAN[19]). Thermodynamical parameters (stress tensor, acoustic tensor, temperature, entropy, etc) are obtained directly from thermodynamical potential without numerical approximation of its derivatives. The developed program also uses libraries BOOST[20] and EIGEN[21].

In present section the numerical results for Godunov-Romenskii model are given for homogeneous and heterogeneous medium testcases. The well-known model tests for solid and gaseous phases are considered. In the examples below, initial value problems are solved in a computational domain $\Omega = [0,1]$ cm. The position of the discontinuity in the initial data is $x = 0.5$ cm.

4.1 Homogeneous case

Gas dynamics. As mentioned above, Godunov-Romenskii model can describe gas flow assuming that EOS is chosen in a proper way. This approach is used here to solve the well-known Sod shock tube problem[22], adapted for hyperelastic model setting. Complete hyperelastic model with 13 equations for variables $\left(u_k, F_{ij}, \mathrm{S}\right)$ is considered instead Euler ideal gas dynamics system with 5 equations[23].

Mesh step is 0.001 cm. Time step is 0.01 sec. The piecewise polynomials inside each cell are up to third order. As it is mentioned above, EOS consists of only hydrodynamical term (9)

$$\mathrm{U} = \mathrm{U}^{\mathrm{h}}\left(I_3, \mathrm{S}\right) = c_V T_0 I_3^{\gamma/2}\left(\exp\left[\mathrm{S} / c_V\right] - 1\right).$$

with $\rho_0 = 1.0\ \mathrm{g/cm^3}$ being initial density, $c_V = 1.0 \cdot 10^{-6}\,\mathrm{kJ/(gK)}$ – heat capacity, $T_0 = 100\mathrm{K}$ – reference temperature, $b_0 = 0.0\mathrm{km/s}$ – shear wave speed and $\gamma = 0.4$ – constant parameters.

Distortion tensor coefficients and entropy values are chosen in such a way that they correspond to the parameters of the Sod problem for Euler equations. The initial state corresponds to Riemann problem with two constant states:

$$\mathbf{u}_{\mathrm{l}} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \frac{\mathrm{km}}{\mathrm{s}}, \quad \mathbf{F}_{\mathrm{l}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathrm{S}_{\mathrm{l}} = 4.0 \cdot 10^{-6} \frac{\mathrm{kJ}}{\mathrm{gK}},$$

$$\mathbf{u}_{\mathrm{r}} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \frac{\mathrm{km}}{\mathrm{s}}, \quad \mathbf{F}_{\mathrm{r}} = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathrm{S}_{\mathrm{r}} = 1.7 \cdot 10^{-6} \frac{\mathrm{kJ}}{\mathrm{gK}},$$

where indices “l” and “r” denotes left and right states, respectively.

Figure 1 shows various state dimensionless profiles at time $t = 0.6$ sec. The results are fully identical to gas dynamics ones[23].

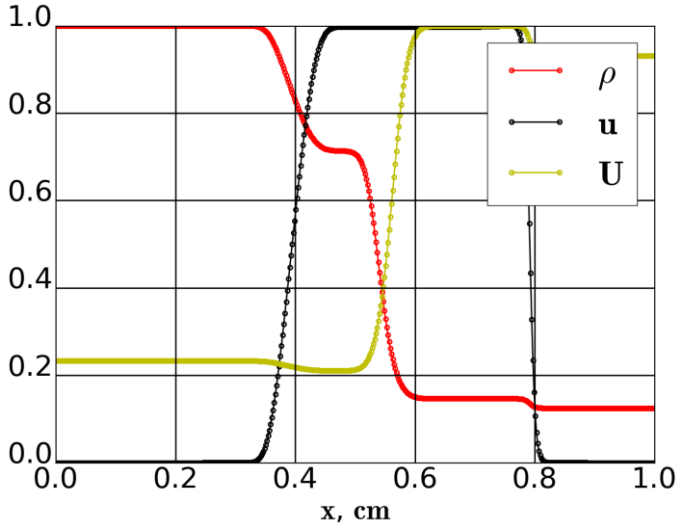


Figure 1: Dimensionless density (ρ), velocity (u) and internal energy (U) profiles at a time $t = 0.6$ s.

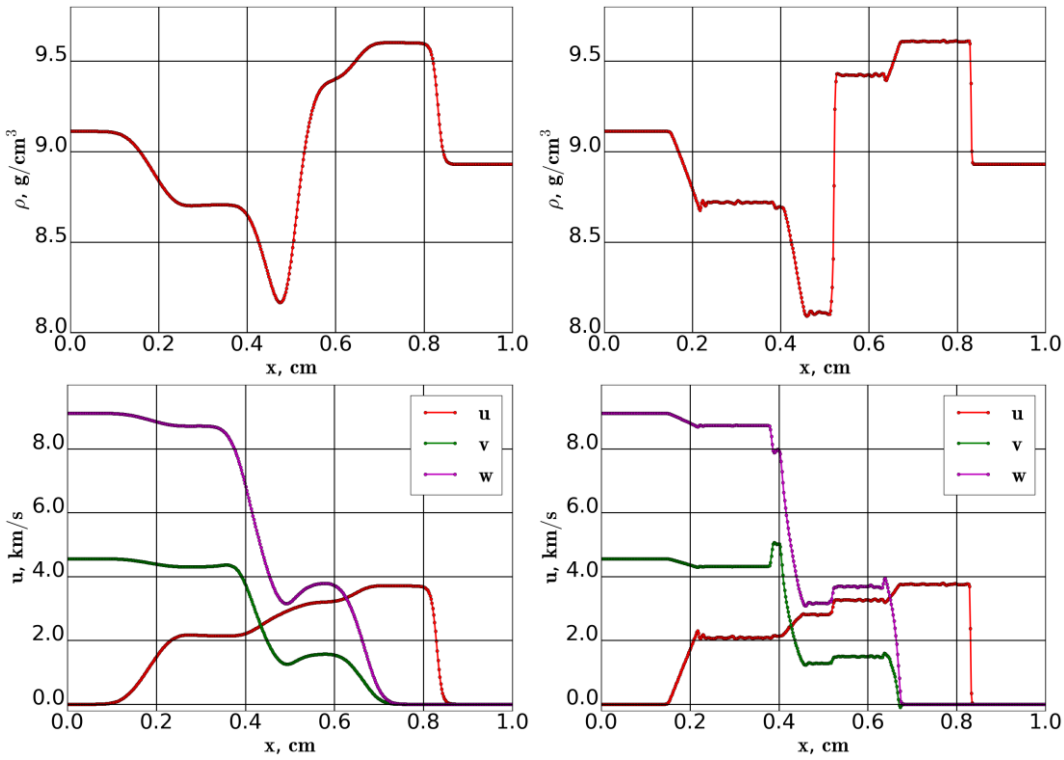


Figure 2: Density ρ (up) and velocities u, v, w (down) distributions for finite volume (left) and RK/DG (right) methods at time 0.5 μ s.

Nonlinear hyperelasticity. This test is from [4] with nondiagonal distortion tensor. The RK/DG and finite volume methods are considered. Mesh step is 0.002 cm. Time step is 0.005

μ s. Piecewise polynomials in each cell are considered up to third order. The initial state is given by:

$$\mathbf{u}_{\mathrm{l}} = \begin{pmatrix} 0 \\ 0.5 \\ 1 \end{pmatrix} \frac{\mathrm{km}}{\mathrm{s}}, \quad \mathbf{F}_{\mathrm{l}} = \begin{pmatrix} 0.98 & 0 & 0 \\ 0.02 & 1 & 0.1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathrm{S}_{\mathrm{l}} = 10^{-3} \frac{\mathrm{kJ}}{\mathrm{gK}},$$

$$\mathbf{u}_{\mathrm{r}} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \frac{\mathrm{km}}{\mathrm{s}}, \quad \mathbf{F}_{\mathrm{r}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathrm{S}_{\mathrm{r}} = 0 \frac{\mathrm{kJ}}{\mathrm{gK}}.$$

The material is assumed to be copper with EOS parameters defined in [4]: $\rho_0 =$ 8.9 g/cm^3, $c_0 = 4.6$ km/s, $c_V = 3.9 \cdot 10^{-4}$ kJ/(g·K), $T_0 = 300$ K, $b_0 = 2.1$ km/s, $\alpha = 1.0$, $\beta = 3.0$, $\gamma = 2.0$.

Results are shown in Figure 2 at time 0.5 μ s in the comparison with finite volume method results. RK/DG method has better resolution of waves. The results are fully identical to the published ones[4].

4.2 Heterogeneous case

In this testcase mesh step is $h = 5 \cdot 10^{-4}$ cm and time step is 0.005 μ s. The polynomials inside each cell are considered up to third order.

Consider the heterogeneous medium model described in subsection 2.1. Homogeneous domains correspond to $\Omega_{\mathrm{l}} = (0, 1/2]$ and $\Omega_{\mathrm{r}} = (1/2, 1)$. The smoothed characteristic functions for $\Omega_{\mathrm{l,r}}$ are chosen as $\phi_{\mathrm{l}} = 1 - \phi_{\mathrm{r}}$,

$$\phi_{\mathrm{l}} = \begin{cases} 0, & x \in (0, 1/2 - \varepsilon) \\ \dfrac{1}{2} - \dfrac{1}{2\varepsilon}\left(x - \dfrac{1}{2}\right) - \dfrac{1}{2\pi} \sin\left(\dfrac{\pi}{\varepsilon}\left(x - \dfrac{1}{2}\right)\right), & x \in [1/2 - \varepsilon, 1/2 + \varepsilon] \\ 1, & x \in (1/2 + \varepsilon, 1), \end{cases}$$

where $\varepsilon = 10h$. Initial conditions are given by[4]:

$$\mathbf{u}_{\mathrm{l}} = \begin{pmatrix} 2 \\ 0 \\ 0.1 \end{pmatrix} \frac{\mathrm{km}}{\mathrm{s}}, \quad \mathbf{F}_{\mathrm{l}} = \begin{pmatrix} 1 & 0 & 0 \\ -0.01 & 0.95 & 0.02 \\ -0.015 & 0 & 0.9 \end{pmatrix}, \quad \mathrm{S}_{\mathrm{l}} = 0 \frac{\mathrm{kJ}}{\mathrm{gK}},$$

$$\mathbf{u}_{\mathrm{r}} = \begin{pmatrix} 0 \\ -0.03 \\ -0.01 \end{pmatrix} \frac{\mathrm{km}}{\mathrm{s}}, \quad \mathbf{F}_{\mathrm{r}} = \begin{pmatrix} 1 & 0 & 0 \\ 0.015 & 0.95 & 0 \\ -0.01 & 0 & 0.9 \end{pmatrix}, \quad \mathrm{S}_{\mathrm{r}} = 0 \frac{\mathrm{kJ}}{\mathrm{gK}}.$$

Left material EOS parameters: $\rho_0 =$ 8.93 g/cm^3, $c_0 = 4.6$ km/s, $c_V = 3.9 \cdot 10^{-4}$ kJ/(g·K), $T_0 = 300$ K, $b_0 = 2.1$ km/s, $\alpha = 1.0$, $\beta = 3.0$, $\gamma = 2.0$.

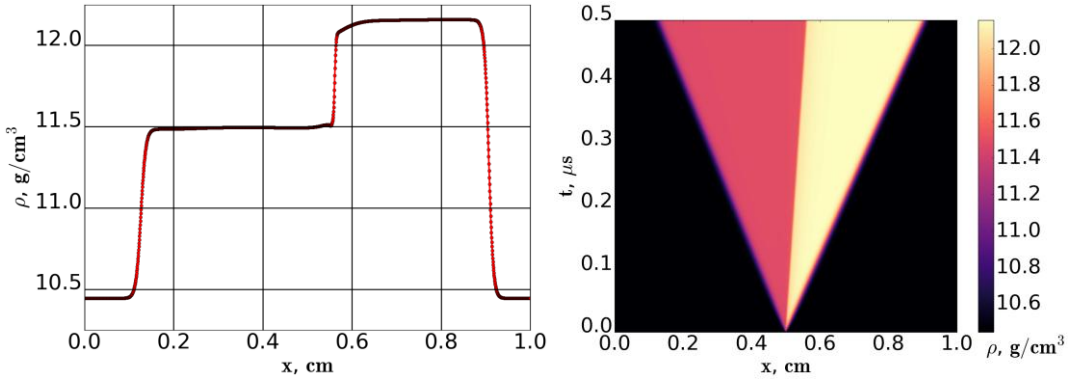


Figure 3: Denstiy (ρ) distribution at $t =$ 0.5 μ s (left) and x - t diagram (right).

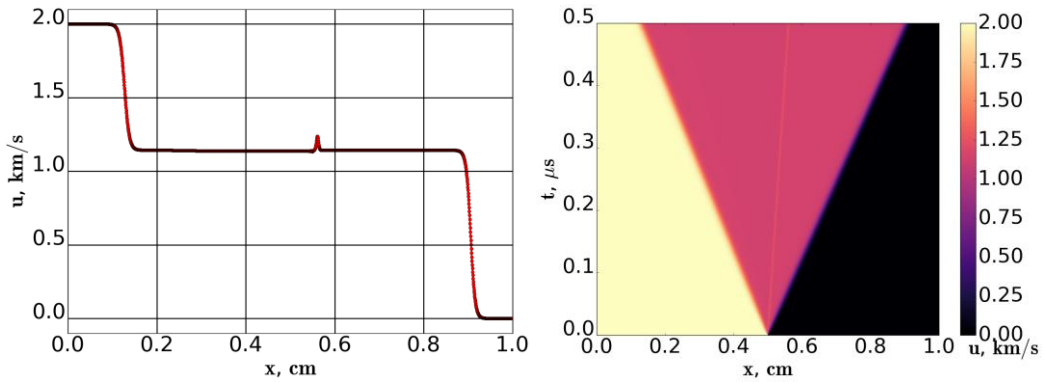


Figure 4: Velocity (u) distribution at $t =$ 0.5 μ s (left) and x - t diagram (right).

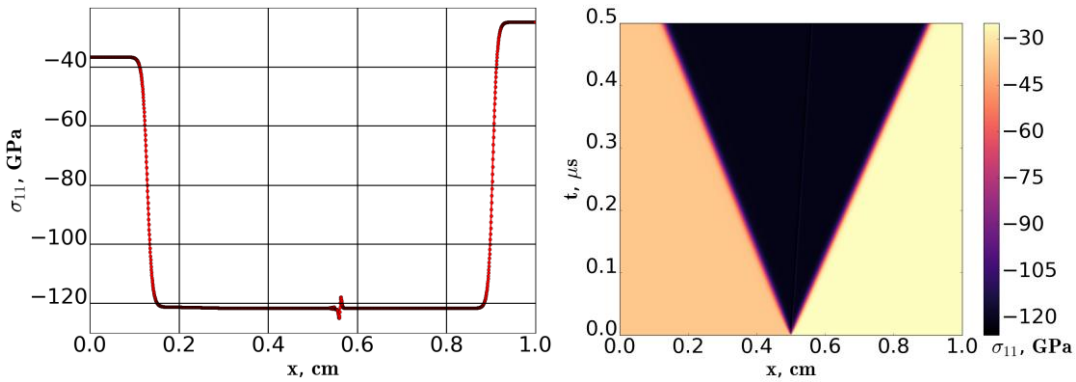


Figure 5: Stress tensor component (σ_{11}) distribution at $t =$ 0.5 μ s (left) and x - t diagram (right).

Right material EOS parameters: $\rho_0 =$ 8.93 g/cm^3, $c_0 = 6.22$ km/s, $c_V = 9.0 \cdot 10^{-4}$ kJ/(g·K), $T_0 = 300$ K, $b_0 = 3.16$ km/s, $\alpha = 1.0$, $\beta = 3.577$, $\gamma = 2.088$.

Solution at time $t =$ 0.5 μ s is shown in Figures 3-5. Interphase boundary is moving from left to right. Left figures correspond to variables profiles at given time, right ones are corresponding x - t diagrams.

In Figure 3 the density profile and corresponding x - t diagram is shown at time $t =$ 0.5 μ s.

In Figures 4 and 5 velocity u and stress tensor component σ_{11} and corresponding x - t diagrams are shown.

5 CONCLUSION

The paper discusses the application of the RK/DG method for solving problems of hyperelasticity in an inhomogeneous medium. Both models, the homogeneous and heterogeneous one, admitting piecewise-constant distribution of medium properties, are investigated. As a result of a series of calculations, it was shown that the Godunov-Romenskii hyperelastic model can be practically applied to solve gas dynamics problems, when the internal energy of a medium depends only on its bulk deformations and entropy. The application of the RK/DG method demonstrates sharp resolution of wavefronts, comparable to the use of methods of the WENO type.

REFERENCES

[1] Godunov S.K., Romenskii E.I., “Nonstationary equations of nonlinear elasticity theory in Eulerian coordinates”, *J. Appl. Mech. Tech. Phys*., **13**(6), 868-884 (1972).

[2] Favrie N., Gavrilyuk S.L., Saurel R., “Solid–fluid diffuse interface model in cases of extreme deformations”, *J. Comput. Phys*., **228**(16), 6037-6077 (2009).

[3] Saurel R., Petitpas F., Berry R.A., “Simple and efficient relaxation methods for interfaces separating compressible fluids, cavitating flows and shocks in multiphase mixtures”, *J. Comput. Phys.*, **228**(5), 1678-1712 (2009).

[4] Barton P.T. et al., “Exact and approximate solutions of Riemann problems in non–linear elasticity”, *J. Comput. Phys.*, **228**(18), 7046-7068 (2009).

[5] Miller G.H., “Minimal rotationally invariant bases for hyperelasticity”, *SIAM J APPL MATH*, **64**(6), 2050-2075 (2004).

[6] Liu X.D., Osher S., Chan T., “Weighted essentially non–oscillatory schemess”, *J. Comput. Phys.*, **115**(1), 200-212 (1994).

[7] Jiang G.S., Shu C.W., “Efficient implementation of weighted ENO schemes”, *J. Comput. Phys.*, **126**(1), 202-228 (1996).

[8] Cockburn B., Shu C.W., “The Runge–Kutta local projection discontinuous–Galerkin finite element method for scalar conservation laws”, *ESAIM-MATH MODEL NUM*, **25**(3), 337-361 (1991).

[9] Baer, M.R., Nunziato, J.W., “A two-phase mixture theory for the deflagration-to-detonation transition (DDT) in reactive granular materials”, *Int. J. Muhiphase Flow*, **12**(6), 861-889 (1986).

[10] Godunov S. K., Romenskii E., *Elements of continuum mechanics and conservation laws*, Springer Science and Business Media (2003).

[11] Dimitrienko Y.I., *Nonlinear continuum mechanics and large inelastic deformations*, Springer Science and Business Media, Vol.174., (2010).

[12] Kapilla, A.K., Menikoff, R., Bdzil, J.B. et al., “Two-phase modelling of deflagration to detonation transition in granular materials: reduced equations”, *PHYS FLUIDS*, **13**(10), 3002-3024 (2001).

[13] Dal Maso G., Le Floch P., Murat F., “Definition and weak stability of nonconservative products”, *J. Math. Pures Appl*, **74**(6), 483–548 (1995).

[14] Kolmogorov A. N., Fomin S. V., *Elements of the theory of functions and functional analysis*, Vol. 1., (1957).

[15] Michoski C. et al., “A comparison of artificial viscosity, limiters, and filters, for high order discontinuous Galerkin solutions in nonlinear settings”, *J. Sci. Comput.*, **66**(1), 406-434 (2016).

[16] Persson P.O., Peraire J., “Sub–cell shock capturing for discontinuous Galerkin methods”, *44th*

AIAA Aerospace Sciences Meeting and Exhibit, 112 (2006).
[17] X. Zhang, Y. Xia, C.-W. Shu., "Maximum–principle–satisfying and positivity–preserving high order discontinuous Galerkin schemes for conservation laws on triangular meshes", *J. Sci. Comput.*, **50**(1), 29–62 (2012).
[18] Krivodonova L., "Limiters for high–order discontinuous Galerkin methods", *J. Comput. Phys.*, **226**(1), 879-896 (2007).
[19] The Stan Math Library. https://github.com/stan-dev/math (Accessed January 19, 2020).
[20] BOOST C++ libraries. https://www.boost.org/ (Accessed January 10, 2020).
[21] EIGEN. http://eigen.tuxfamily.org/index.php?title=Main_Page (Accessed January 5, 2020).
[22] Sod G.A., "A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws", *J. Comput. Phys.*, **27**(1), 1-31 (1978).
[23] Toro E.F., *Riemann solvers and numerical methods for fluid dynamics: a practical introduction*, Springer Science and Business Media (2013).
[24] Egorova V.A., Uskov R.V., Voronin F.N., Zhukovskiy M.E., "On the modeling of the thermomechanical fields in porous material being under radiation", *Mathematica Montisnigri*, **43**, 58-69 (2018).
[25] Markov M.B., Uskov R.V., Zhukovskiy M.E., "Monte–Carlo modeling of the photon-electron cascade In heterogeneous matter", *Mathematica Montisnigri*, **46**, 49-60 (2019).

Received January 10, 2020

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

MODELING OF RADIATION-INDUCED ELECTRIC CURRENT IN THE MATERIALS OF FINELY DISPERSED STRUCTURE

M.E. ZHUKOVSKIY[1*], M.B. MARKOV, R.V. USKOV[1] AND L.V. KUZNETSOVA[2]

[1]Keldysh Institute of Applied Mathematics of RAS (KIAM), Moscow, Russia

[2] 12th Central research Institute of the Ministry of defense of the Russian Federation, Moscow, Russia

* Corresponding author. E-mail: usermath@mail.ru

DOI: 10.20948/mathmontis-2020-47-6

Summary. The technique of Monte Carlo modeling of radiation-induced electric currents in heterogeneous finely dispersed medium with direct consideration of their microstructure is worked out. The main attention is paid to developing the method of the construction of a geometric model of the polydisperse structures. The method based on random tracing algorithm is intended for implementation on heterogeneous computing clusters with using the graphical processors and the CUDA parallelization of calculations. The geometric model includes a detecting system for statistical evaluation of the desired physical quantities (electric current density). A computational experiment was performed to study the basic regularities of generation of electrical currents arising in polydispersed mfterial being under X-radiation. The results of the experiment showed the irradiation of the object under study produces electric currents with a sharply inhomogeneous spatial structure. Inhomogeneities occur near the boundary surfaces between the binder and inclusions.

1 INTRODUCTION

Heterogeneous materials of finely dispersed structures are widely used in mechanical engineering, heat power engineering, rocket, aviation, chemical and other industries. This is because these materials provide the required strength, thermal, hydraulic, technological properties and can operate at high temperatures and pressures. Such materials are used, for instance, in protection of structures from intensive energy flows [1], creation of solid propellants [2, 3].

Investigation of radiation-induced electrical effects in heterogeneous finely dispersed media is very actual for researching the protective and functional properties of such media being under radiation [4-9]. Heterogeneous dispersed structures are the materials having huge number of inner boundaries between homogeneous components. The presence of these boundaries leads to generation of electrical phenomenon due to the lack of electronic equilibrium near them [10-12].

Mathematical modeling of radiation-induced electrical effects in finely dispersed media involves the development of radiation transport simulation algorithms as well as the construction of a geometric model of a substance with a direct resolution of its microstructure.

The transport of radiation in heterogeneous materials of complex geometric structure is cascade process and characterized by the fact that the particle pathways are comparable with the size of the inhomogeneities of the medium. In this situation, a detailed simulation of each collision of radiation particles with atoms of the medium is required [13]. Statistical

2010 Mathematics Subject Classification: 97M50, 97N50, 93A30.
Key words and Phrases: Finely dispersed medium, Radiation induced current, Monte Carlo simulation.

algorithms of radiation cascade transport modelling are worked out considering the features of radiation transport in finely dispersed media [14, 15].

The main attention in this work is paid to the construction of a geometrical model of a dispersed structure. The model includes the detector (registration) system for statistical estimation of the electrical current in an irradiated object.

Various algorithms can be used to construct a geometric model of the material (placement of microstructure particles with specified geometric properties inside the sample). The most popular of these is the algorithm of Lubachevsky-Stillinger [16-18]. The algorithm simulates the process of mechanical compression of a set of solid particles. There are other techniques for construction of geometrical model of materials in question [19, 20].

These algorithms are poorly parallelized on GPUs due to their complex internal logic and therefore cannot be integrated into common code designed for heterogeneous computing clusters (HCC).

A method of creating the geometrical model of the irradiated object based on ray-tracing algorithm [21] is worked out in this work. The method has almost unlimited scalability and is easily implemented on the graphics subsystem of the HCC.

The developed code for supercomputer simulation of radiation-induced electric currents in heterogeneous dispersed materials with direct consideration of their microstructure is implemented on heterogeneous computing clusters.

The results of a computational experiment to calculate the current density in a fragment of finely dispersed material show that electric currents with a sharply inhomogeneous spatial structure are formed during irradiation of the object under study. Inhomogeneities occur near the boundary surfaces between the binder and inclusions.

2 GEOMETRICAL MODEL OF THE FINELY DISPERSED STRUCTURES

Base characteristics of the dispersed medium are the size of suspended particles in dispersed systems and dispersity (relative volume fraction of suspended particles of every type). It is assumed in this paper that all particles of given type are of the same size.

The geometric model also includes a model of the detector system for the statistical evaluation of the required physical quantities (electrical current). The detector system intended for the statistical estimation of functionals on the space of solutions of the transport equation includes a set of “detectors”, spheres of a specified size and location within which the events of the interaction of the radiation quanta and the secondary particles with the material are recorded.

The detectors must be isolated from each other (should not intersect) and the entire volume of the detector should be inside the given matter (in the context of the considered media, they should not “capture” the boundaries between homogeneous components).

It can consider the model of the dispersed matter with the detectors as a polydisperse medium consisting of some types of solid nonoverlapping objects (inclusions): suspended particles and detectors. However, several detectors may be inside a single particle.

Let some object be a polydisperse medium consisting of a binder and N types of suspended spherical particles of radius r_n ($n=1,\ldots,N$).

The developed algorithm for creating a geometric model has the following structure.

Initially, the placement of inclusions of the 1st type is constructed.

1. The coordinates (X_{min}, Y_{min}, Z_{min}) and (X_{max}, Y_{max}, Z_{max}) of the object are determined;

2. $M_x^{n=1} = (Z_{\max} - Z_{\min}) \cdot (Y_{\max} - Y_{\min}) / \pi r_{n=1}^2$ beams are drawn from a random point on the plane $x=X_{min}$ along the X-axis; $M_y^{n=1} = (Z_{\max} - Z_{\min}) \cdot (X_{\max} - X_{\min}) / \pi r_{n=1}^2$ beams are drawn from a random point on the plane $y=Y_{min}$ along the Y-axis and $M_z^{n=1} = (Y_{\max} - Y_{\min}) \cdot (X_{\max} - X_{\min}) / \pi r_{n=1}^2$ beams are drawn from a random point on the plane $z=Z_{min}$ along the Z-axis;
3. The intersection points of the beams with boundary surfaces of homogeneous parts of the object are calculated (fig. 1).
4. A random point (center of a particle of 1st type) on every interval between two consecutive intersection points (segments 1-2, 3-4 and 5-6 in fig. 1) is played.

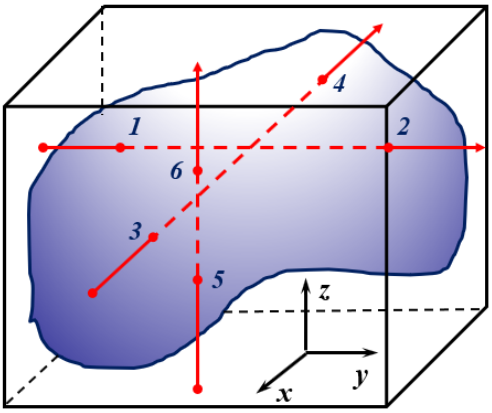


Fig. 1. Scheme of particle placement algorithm

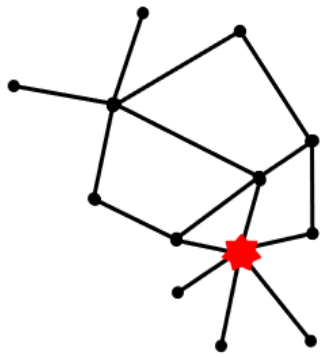


Fig. 2. Scheme of particle excluding

The next stage of the algorithm is to filter (exclude) particles according to the following criteria.

- The particles should not intersect the boundary surfaces of the homogeneous parts of the object (if intersection takes place, particle is excluded from corresponding set);
- The particles must be isolated from each other.

Elimination of mutual intersections of detectors is carried out by using the following method.

- It is built a graph on the set of constructed points (centers of the inclusions). The nodes of the graph are the centers of the particles. The edges of the graph are constructed between two nodes for which the distance between the centers of the particles is less than $2r_{n=1} + S_{n=1}$ ($S_{n=1}$ is special value that restricts the minimum distance between the particles);
- The node of the graph having the maximum number of edges is defined. It is excluded from the graph (fig. 2).

Last step is repeated until there are no edges left in the graph.

The algorithm is repeated with additional beams if a number of inclusions is less then requiered.

The set of detectors is constructed after the geometrical model is built with inclusions of all types. The developed algorithm is used for the construction of the detector set but there is one

exception. Some of the detectors may be located entirely inside the inclusions to estimate the desired value in suspended particles.

A fragment of geometrical model consisting of epoxy binder, metal or dielectric inclusions (blue spheres) and a set of detectors (magenta spheres) is shown in fig. 3 (the image is enlarged for clarity).

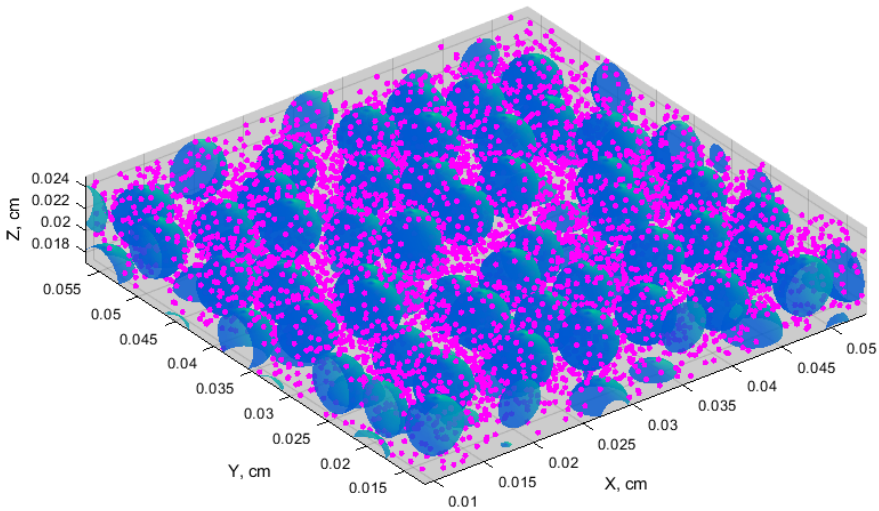


Fig. 3. A fragment of geometrical model constructed by use of developed algorithm

All inclusions are closed, but appear cropped due to image magnification.

3 MODELING OF THE RADIATION TRANSPORT

The complicated process of particle transport through the matter can be represented by a sequence of elementary processes of the interaction between the particle and the atoms of matter (particle trajectory). These processes include the scattering, braking or disappearance of the particle due to absorption or escape from the considered system (from the object). This representation is convenient for modelling the radiation transport by the Monte-Carlo method.

The transport of the particles accompanied by the birth of secondary particles in cascade processes of the interaction of the radiation with matter is described by a system of integral equations.

$$Q = Q_1 + \int k(x', x) Q(x') dx' = Q_1 + KQ . \tag{1}$$

Here $x = (\mathrm{r}, \Omega, E)$, where r, Ω, E are coordinates, direction of motion and energy, respectively; $Q(x)$ is the density of collisions and $Q_1(x)$ is the density of the first collisions; $k = (x, x')$ is the kernel of the integral operator and has the meaning of probability density of $x' \to x$ transition.

Equation (1) is true for every type of particles of cascade. The previous generation particle flux is the next generation particle source.

The objective of the radiation transport theory is to compute the readings of detector located in the field of radiation. The desired (measured) values are presented as the readings of some detector and are written as functional on the space of the transport equation solutions. We consider such registering facilities (detectors) whose readings are equal to the sum of the contributions of some particle's collisions in a sensitive volume of the detector (additive detectors). To evaluate the desired measured value by the Monte Carlo method, the random trajectories of the particles are simulated (fig. 4).

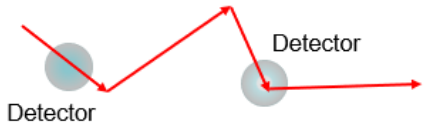


Fig. 4. Particle trajectory

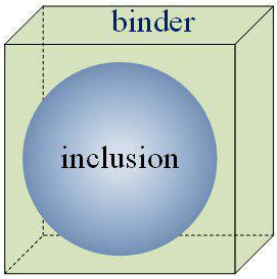


Fig. 5 A fragment of dispersed material

The contributions of these trajectories to the detector's measurable value are summed up. The particle trajectory construction is performed according to the chosen physical model of the interaction between the radiation and matter.

Trajectories are simulated using the individual computational algorithms for each type of particle considering their physical properties [13].

The developed algorithm is described in detail in [15].

4 RESULTS OF THE COMPUTATIONAL EXPERIMENT

This section presents the results of computational experiments on simulation of radiation-induced electric currents in heterogeneous materials of finely dispersed structure.

A fragment of dispersed structure (fig. 5) is considered for researching the basic features of the current generation process in an object being under radiation. The cubic fragment of 0.003 cm size consists of binder (epoxy resin, density is about 1 g/cm^3) and one spherical inclusion (ammonium perchlorate, $NH4ClO_4$, density is about 2 g/cm^3) of 0.002 cm diameter. The studied fragment is irradiated by photons of 20 KeV energy in the direction of the Z axis.

4.1 The main regularities of the generation of radiation-induced effects

The distribution of fields of radiation-induced effects (heating, charge effects, electric currents) is determined mainly by the number of electrons born and their penetrating power. The first value is proportional to the macroscopic cross section of the interaction of photons with matter, and the second is proportional to the braking path of electrons. The dependence of these values on the energy of the radiation particles is shown in fig. 6, 7.

These figures show that in inclusion, the macroscopic cross-section of the interaction of photons significantly (up to two orders of magnitude) exceeds this value for the binder. Therefore, much more electrons are born in the inclusion than in the binder. The penetrating

power (braking distance) of electrons, on the contrary, is noticeably greater in the binder.

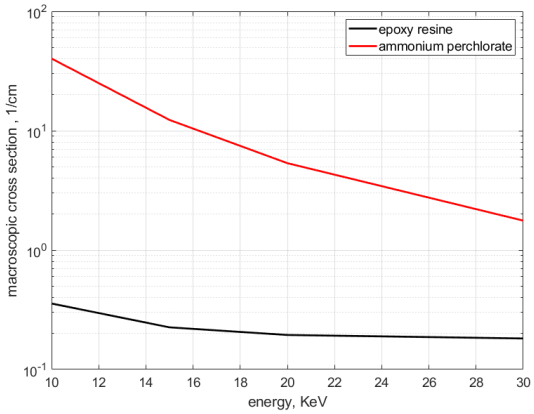


Fig. 6. Microscopic cross sections

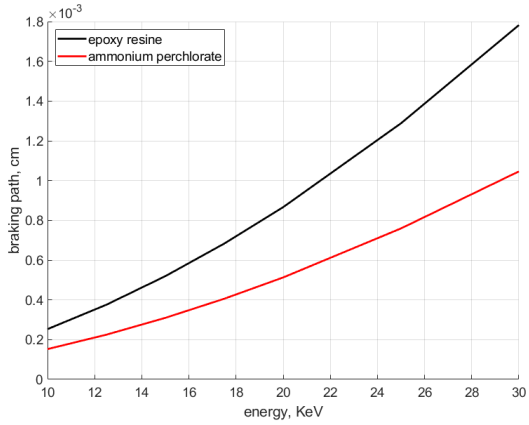


Fig. 7. Braking paths

The generation of electric currents in a substance being under gamma-or x-ray radiation is caused by the fluxes of photo and Compton electrons generated as a result of the photo-electron cascade processes in the material under study.

Current components for which there is no electronic equilibrium, that is, electron flows along the direction of this current component and in the opposite direction do not compensate for each other, will be different from zero at a given spatial point. The photon flux propagates along the *z* axis in the computational experiment under consideration, so the transverse (*x*, *y*) components of the current will obviously be negligible in a homogeneous medium.

The electron braking distance does not exceed 4 microns, and the number of collisions reaches tens and hundreds in the studied fragment of heterogeneous material. Therefore, at the periphery of the fragment (at 2-4 microns from the inclusion boundaries) in the binder, the absence of transverse current components should be expected.

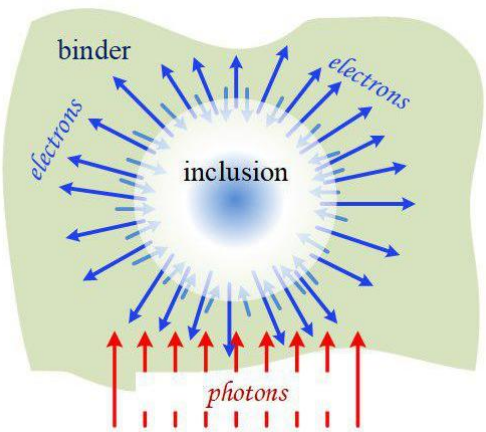


Fig. 8. Electron fluxes from and into the inclusion

Another situation is realized near the interface of two media with different physical properties (density, cross sections, braking paths) on spatial scales of the order of the electron path. The concentration of electrons born in the inclusion is much greater than in the binder.

In addition, electrons emitted from the inclusion into the binder have a much greater penetrating power than electrons moving in the opposite direction (fig. 8).

Therefore, uncompensated electron fluxes arise near the boundaries of two media and the direction of the electron fluxes is from the inclusion into the binder. This direction is due to the predominance of electron emission from inclusion in the binder over emission in the opposite direction [23].

4.2 Results of the modeling of radiation-induced electrical current

The fig. 9-12 below show the spatial distributions of the amplitude of the transverse components J_x and J_y of the current in the irradiated fragment. These figures show the amplitudes of the electric current density in CGSE units per 1 photon/cm^2.

The fig. 9, 10 show graphs of the transverse components along straight lines {z=0.0015 cm, y=0.0015 cm} and {z=0.0015 cm, x=0.0015 cm} respectively. Dotted lines mark the boundaries of the inclusion.

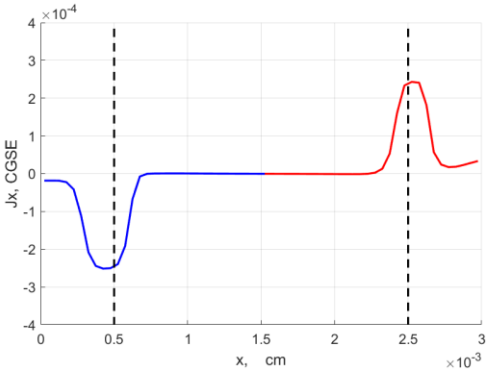


Fig. 9. The transverse component J_x

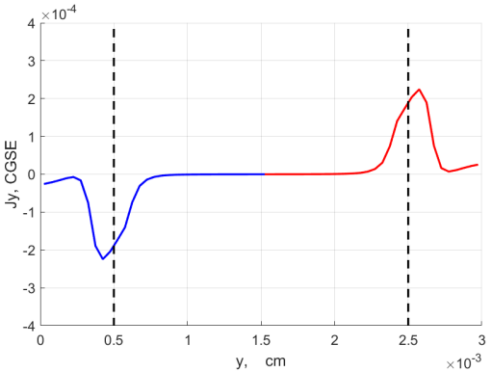


Fig. 10. The transverse component J_y

The transverse components of the current are negligible along the longitudinal axis {x=0.0015 cm, y=0.0015 cm} passing through the "poles" of the inclusion.

Spatial distributions of the transverse components J_x and J_y in the plane z=0.0015 cm are shown in Fig. 11, 12 in the form of the surfaces.

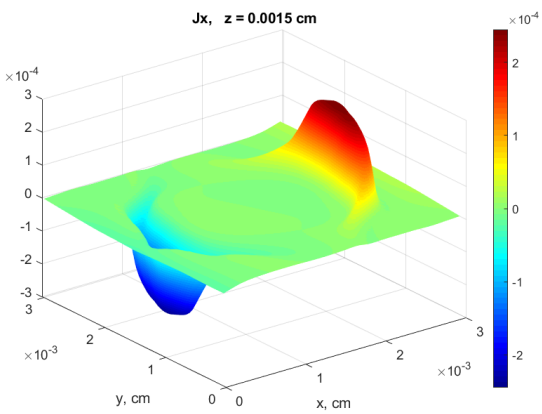


Fig. 11. 2D image of the component J_x

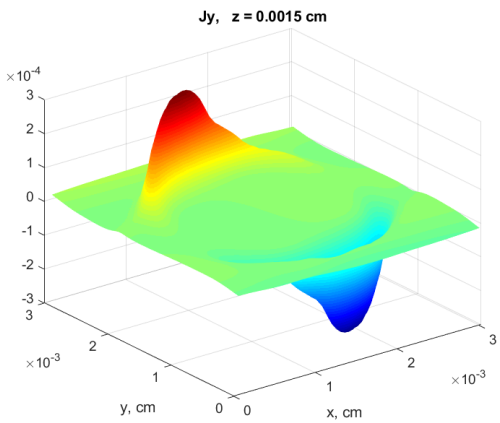


Fig. 12. 2D image of the component J_y

In General, these figures demonstrate the expected symmetry of the distribution of the transverse components of the current relative to the corresponding coordinate axes.

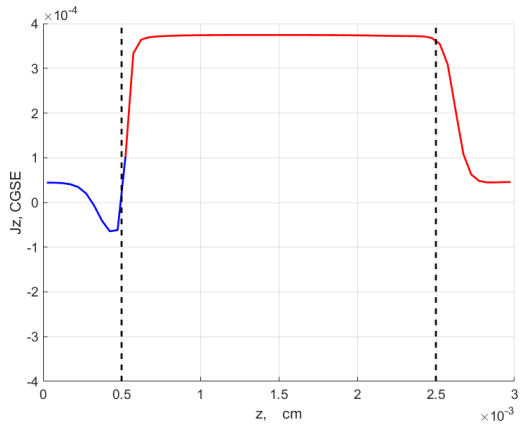


Fig. 13. J_z component along the longitudinal axis {x=0.0015 cm, y=0.0015 cm}

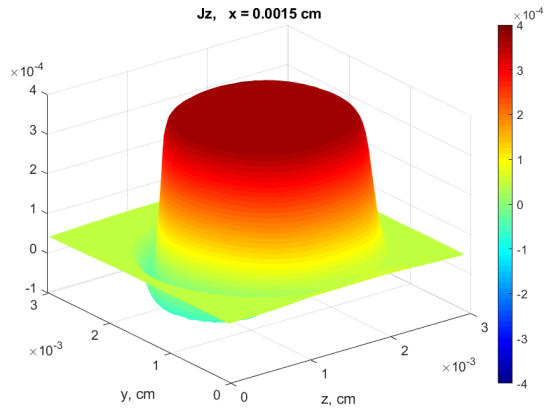


Fig. 14. 2D image of the component J_z

Fig. 13, 14 show the spatial distributions of the longitudinal (along the direction of the photon flow) component of the current J_z. The component J_z in the plane x=0.0015 cm is shown in Fig. 14 in the form of the surface.

The component J_z reaches the maximum value at the boundary of the two media because the electron emission from the inclusion into the binder is much more intense than in the opposite direction (fig. 8). The background longitudinal component of the current is generated at the periphery of the binder (at a distance from the boundary surfaces exceeding the braking path of the electron). Its value is significantly less than one of the inclusion-binder boundaries.

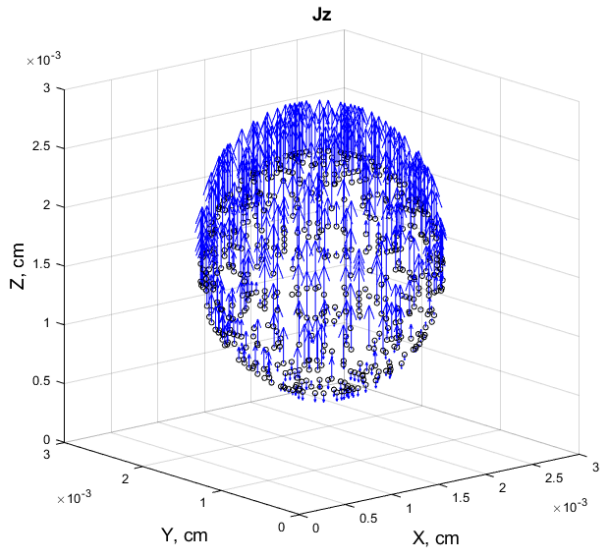


Fig. 15. $\mathbf{J}_z$

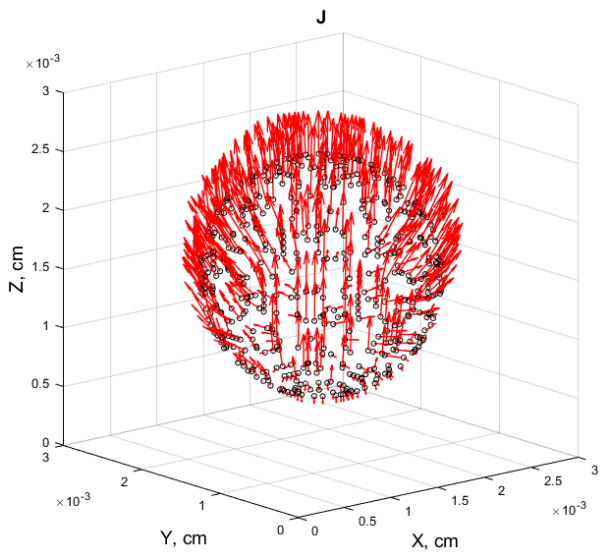


Fig. 16. $\mathbf{J}=\mathbf{J}_x+\mathbf{J}_y+\mathbf{J}_z$

Vector field of the current in the binder near the boundary surface inclusion-binder is depicted in fig. 15 ($\mathbf{J}_z$) and fig. 16 ($\mathbf{J}=\mathbf{J}_x+\mathbf{J}_y+\mathbf{J}_z$).

These figures demonstrate the expected asymmetry of the electric current distribution with respect to the z=const plane passing through the center of the inclusion.

5 CONCLUSION

The technology of supercomputer simulation of radiation-induced electric currents in heterogeneous dispersed materials with direct consideration of their microstructure is developed. The main attention is paid to the creation of an algorithm for constructing a geometric model of a polydisperse medium, which is intended for implementation on heterogeneous computing clusters. The geometric model includes a detecting system for statistical evaluation of the desired physical quantities (electric current density).

The results of a computational experiment to calculate the current density in a fragment of finely dispersed material show that electric currents with a sharply inhomogeneous spatial structure are formed during irradiation of the object under study. Inhomogeneities occur near the boundary surfaces between the binder and inclusions. The generation of a current at the boundaries between two media is caused by the predominance of electron emission from the inclusion (a material with a large macroscopic cross-section of photons) in the binder (a material with a greater penetration of electrons) over the emission in the opposite direction (Fig. 8). Only the longitudinal component of the current (along the direction of the photon flow) is present in the binder at more than the length of the braking distance from the outer surface of the inclusion. Its value is much smaller than the amplitude of the current near the binder-inclusion boundary.

Acknowledgements: The work was financial supported by the Russian Scientific Foundation (project No. 17-71-30014).

REFERENCES

[1] A.V. Ostrik, A.I. Potapenko, “Geterogennye materialy dlia zashchity konstruktsii ot intensivnykh potokov energii”, *Konstruktsii iz kompozitsionnykh materialov*, 1, 48-53 (2001).

[2] *Kondensirovannye energeticheskie sistemy*, B.P. Zhukov (Ed.), Yanus-K (1999).

[3] A.M. Lipanov, A.V. Aliev, *Proektirovanie RDTT*, M.: Mashinostroenie (1995).

[4] A.R. Frederickson, L. Levy, C.L. Enloe, “Radiation-induced electrical discharges in complex structures”, *IEEE Trans. Electr. Insulation*, **27** (6), 1166–1178 (1992).

[5] V.V. Gromov, *Elektricheskii zariad v obluchennykh materialakh*, M.: Energoizdat (1982).

[6] D.N. Sadovnichij, E.A. Golub, A.P. Tyutnev, A.V. Yushkov, “Raschet elektricheskikh polei v ploskikh geterogennykh strukturakh”, *Khimiia vysokikh energii*, 5, 3-8 (1993).

[7] *Dielektriki i radiatsiia*, N.S. Kostyukov (Ed.), **3**, M.: Nauka (2003).

[8] Yu.M. Milyoxin, D.N. Sadovnichij, S.A. Gusev, “Elektrizatsiia tverdotoplivnykh energeticheskikh ustanovok pri vozdeistvii ioniziruiushchego izlucheniia”, *Khimicheskaia fizika i mezoskopiia*, **8** (1), 2006.

[9] A.V. Ostrik, “Prognozirovanie zashchitnykh svoistv pokrytii iz sferoplastikov pri vozdeistvii vysokointensivnykh potokov ioniziruiushchego izlucheniia”, *Deformatsiia i razrushenie materialov*, 2, 36-43 (2010).

[10] M.V. Alekseev and others, “O modelirovanii radiatsionno-indutsirovannykh zariadovykh effektov v melkodispersnykh materialakh zamknuto-iacheistoi struktury” Preprint IPM, 80

(Moscow: KIAM), (2019). doi:10.20948/prepr-2019-80 URL: http://library.keldysh.ru/preprint.asp?id=2019-80.

[11] V.I. Mazhukin, A.V. Shapranov, A.V. Mazhukin, "The structure of the electric double layer at the metal-vacuum interface", *Mathematica Montisnigri*, **44**, 110-121 (2019).

[12] K.K. Inozemtseva, M.B. Markov, F.N. Voronin, "The electromagnetic and thermomechanical effects of electron beam on the solid barrier", *Mathematica Montisnigri*, **39**, 79-100 (2017).

[13] M. E. Zhukovskiy, S. V. Podolyako, and R. V. Uskov, "Model of individual collisions for description of electron transport in matter," *Mathematical Models and Computer Simulations*, **4** (1), 101–109 (2012).

[14] M. E. Zhukovskiy and R. V. Uskov, "Hybrid Parallelization of the Algorithms of Radiation Cascade Transport Modelling", *Mathematical Models and Computer Simulations*, **7** (6), 601–610 (2015).

[15] M. B. Markov, R. V. Uskov, M. E. Zhukovskiy, "Monte Carlo modelling of the photon-electron cascade in heterogeneous matter", *Mathematica Montisnigri*, **46**, 49-60, (2019).

[16] F.H. Stillinger, B.D. Lubachevsky, "Crystalline-Amorphous Interface Packings for Disks and Spheres", *J. Stat. PhysicsB*, **73 (**3-4), 497-514 (1993).

[17] B.D. Lubachevsky, F.H. Stillinger, "Geometric properties of random disk packings", *J. Stat. Physics*, **60** (5-6), 561-583 (1990).

[18] V. A. Egorova, R. V. Uskov, F. N. Voronin and M. E. Zhukovskiy, "On the modeling of the thermomechanical fields in porous material being under radiation", *Mathematica Montisnigri*, **43**, 58-69 (2018).

[19] M.D. Webb, I.L. Davis, "Random particle packing with large particle size variations using reduced-dimension algorithms", *Powder Technology*, **167** (1), 10–19 (2006).

[20] Y. Shi, Y. Zhang, "Simulation of random packing of spherical particles with different size distributions", *Applied Physics A: Materials Science & Processing*, **92** (3), 621–626 (2008).

[21] A.A. Bolkisev, "O primenenii metoda sluchainogo poiska k zadache sluchainoy upakovki tverdykh chastic dlya modelirovaniya struktury smesevogo tverdogo topliva", *Vesti Udmurtskogo Universiteta, Matem. Mekh. Kompyut. Nauki*, 2, 106-113 (2012).

[22] M. Zhukovskiy, S. Podoliako, G.-R. Jaenisch, C. Bellon, U. Samadurau, "Monte-Carlo radiographic model with CAD-based geometry description". *Insight*, **48**, (10), 618-623 (2006).

[23] M. Zhukovskiy, M. Markov, S. Podoliako and R. Uskov, "Hybrid parallelization of computing the electron fluxes produced by photon radiation", *Mathematica Montisnigri*, **31**, 43-53 (2014).

Received November 10, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

NUMERICAL MODELLING VIA INMOST SOFTWARE PLATFORM

I.N. KONSHIN[1,2,3,4], K.M. TEREKHOV[1*], AND YU.V. VASSILEVSKI[1,2,3]

[1] Marchuk Institute of Numerical Mathematics, RAS. Moscow, Russia.
[2] Moscow Institute of Physics and Technology. Dolgoprudny, Moscow Region, Russia
[3] Sechenov University. Moscow, Russia
[4] Dorodnicyn Computing Centre, FRC CSC RAS. Moscow, Russia
*Corresponding author. E-mail: kirill.terehov@gmail.com

DOI: 10.20948/mathmontis-2020-47-7

Summary. INMOST is a software platform for the development of parallel numerical models on general polyhedral grids. In this paper we present the INMOST platform as the powerful tool for numerical modelling. The place of INMOST platform among other modern widespread libraries and numerical modelling packages is shown. A brief overview of tools that help in implementation of each stage of mathematical modelling is presented. Examples of INMOST application demonstrate appealing features of INMOST-based numerical modelling.

1 INTRODUCTION

The amount of software for unstructured mesh generation, numerical modelling and graphic visualization is huge. Along with the development of modern parallel computer systems, there is a need to use parallel algorithms with distributed mesh data. All these applications undoubtedly have a common set of needs for representing and manipulating distributed unstructured meshes. However, a large number of mesh representations are in use in the computational community each tailored to a specific application. Therefore, to gain widespread acceptance it is important to have a full mesh framework which allows applications to operate with all types of mesh data including the general polyhedral grids. At the same time, the infrastructure should be lightweight and efficient to have sufficient utilities for real-world numerical modelling applications. In addition, such an infrastructure should provide an opportunity for convenient assembling of systems of linear and nonlinear equations, their solving, as well as analysis and visualization of obtained solutions.

INMOST is a software platform for the development of parallel numerical models on general polyhedral grids [1, 2, 7, 9]. In this paper we present the INMOST platform as the powerful tool for numerical modelling. The examples of INMOST application demonstrate appealing features of INMOST-based numerical modelling.

In fact, the INMOST software platform does not include ready-made numerical models or even ready-made discretization schemes (such as finite volumes and finite elements). INMOST is just a software “platform” on the basis of which scientific researchers or developers of industrial codes can build their general-purpose grids distributed across processors, apply their discretization schemes and perform calculations for their numerical models. The presented examples of INMOST application demonstrate appealing features of INMOST-based numerical modelling. However, in view of the above reasoning, in this paper there is no direct comparison with the results of calculations using other packages.

2010 Mathematics Subject Classification: 86-08, 97N80, 65M50, 65N22, 65Y05.
Key words and Phrases: INMOST, Numerical Modelling, Computational Fluid Dynamics.

It is worth noting that at present there is a development (see, for example, [6]) presenting the extension of parallel platform INMOST using the Ani3D package [16], which allows the construction of adapted tetrahedral meshes along with the use of a large number of finite element discretizations (see INMOST Ani_Inmost examples [6]).

The present paper is organized as follows. Section 2 contains a brief description of INMOST platform functionality and shows the place of INMOST among other modern widespread libraries and numerical modelling packages. A detailed review of specific approaches to construction of numerical models is given in Section 3. In Section 4 we present several INMOST-based numerical models and their computational performance. The conclusion summarize the research.

2 INMOST SOFTWARE PLATFORM

The main purpose of INMOST platform is to provide to the user all the necessary tools for development and exploration of various numerical models. This involves wide functionality, i.e. operations for general distributed mesh data, convenient interface for assembling and solving systems of nonlinear and linear equations, with built-in or plugged-in linear algebra packages. INMOST software platform was developed to meet the following criteria:

- wide functionality;
- efficiency;
- reliability;
- universality;
- ease of use;
- portability;
- open source code.

At present, it is very difficult to find software packages that satisfy all of the above requirements. There are many alternative solutions, such as the FMDB library (Flexible distributed Mesh DataBase) [36], the MOAB library (Mesh-Oriented datABase) [37], the MSTK library (MeSh ToolKit) [38], the STK library (Sierra ToolKit) [39], Salome package [40], OpenFOAM package (Open Source Field Operation And Manipulation CFD ToolBox) [41] and others that do not fully meet the stated criteria. Existing solutions do not always have easy portability between different platforms (Windows, Linux), existing implementations are not always reliable, it is impossible (in some packages) or difficult to implement user's discretization schemes. However most of publicly available packages are the best choice for the solution of a particular problem. For instance, packages ParMETIS [42] and Zoltan [43] distribute and redistribute a general mesh and graph data across processors, while libraries PETSc [21] and Trilinos [22] solve distributed systems of linear equations.

These considerations motivated a group from INM RAS to develop a comprehensive set of software tools that provides the necessary functional capabilities and allow the use of ParMETIS, Zoltan, PETSc, and Trilinos packages. These tools form the INMOST (Integrated Numerical Modelling and Object-oriented Supercomputing Technologies) software platform [9].

The main modules of the platform are:

- mesh operations module (INMOST Mesh);
- mesh data balancing module (INMOST Partitioner);

- automatic differentiation module (INMOST Automatizator);
- module for assembling and solving linear systems (INMOST Solver).

In the near future, visualization (INMOST DrawGrid) and parameter optimization tools (INMOST TTSP (Tool for Tuning Solver Parameters), INMOST OptimizerSolve example) will also be added as separate modules.

The main objective of INMOST is to support data structure and distributed mesh operations (INMOST Mesh) demanded by grid generators and numerical implementations of physical models. Consistent computational grids may contain cells with arbitrary numbers of faces, each face may be formed by an arbitrary number of edges. Thus, INMOST supports polyhedral cells of arbitrary configuration.

The grid balancing module (INMOST Partitioner) is responsible for automatic and efficient distribution of a computational grid among processors. External ParMETIS and Zoltan partitioners can be exploited as well as internal parallel paritioner based on K-means clustering. Upon grid distribution, the user can determine ghost cells along the interfaces of mesh subdomains where data of neighboring processors are synchronized. The number of layers of ghost cells is defined by the user who can also assign ghost cells explicitly. The last option is useful for complex discretization stencils. Importantly, the MPI exchange library is hidden from the user and compilation of the sequential single-processor version of INMOST is easy.

The automatic differentiation module (INMOST Automatizator) is developed to help in implementation of nonlinear numerical models, automatic generation and assembling of Jacobian matrices and nonlinear residuals appearing in Newton type linearizations. This module simplifies discretizations of new mathematical models.

The module (INMOST Solver) for assembling and solving systems of linear equations has the following appealing features:

- convenient and efficient assembling of a linear system matrix via addressing by the global row and column indices (INMOST MatSolve example);
- variety of built-in linear solvers based on threshould incomplete triangular factorizations as well as linear solvers from external packages, such as PETSc and Trilinos;
- the user can switch between linear solvers independently of the matrix assembling procedure.

3 NUMERICAL MODELLING STAGES AND ASSOCIATED TECHNOLOGIES

Numerical modelling is used both in academic and industrial purposes. The design of a numerical model consists of the following stages:

- Physical model;
- Mathematical model;
- Discretization;
- Solution methods;
- Computer program;
- Post-processing and analysis of the results.

Physical model. Understanding of underlying physical phenomena is crucial: advection, diffusion, reaction or decay processes; heat or density transfer; elasticity or rheology effects,

etc. Chosen physics implies a list of primary physical quantities to be measured and/or simulated: pressure, velocity, concentration, temperature, probabilities, etc.

Mathematical model is represented by a system of differential or integral equations describing the chosen phenomena and involving the chosen physical quantities. Often, formulation of boundary and initial conditions is required. Apart of problem formulation, analysis of solution existence and uniqueness is important. In some cases the analysis is not available though this does not inhibit the numerical solution.

Mathematical description involves governing equations and domain geometry. For academic models in parallelepipedal domains it is sufficient to specify the length-width-height parameters only. In more realistic cases the domain geometry can be presented as a set of unions or intersections of primitives: spheres, cylinders, cones, half spaces, parallelepipeds, etc. The most complex geometries for industrial applications can be constructed with one of the following tools: AutoCAD [10], 3DS Max [11], Sketchup [12], OpenCascade [13], etc.

Discretization of a mathematical model includes discretization of the domain and the equations. Discretization of the domain implies generation of a computational mesh. *A priori* or runtime local mesh refinement near solution gradients crucially affects the solution accuracy. Accurate approximation of domain boundaries is also important. A lot of mesh generation and refinement tools (commercial or open source) are available: Tetgen [14], GMSH [15], ANI3D [16], INMOST OctreeCutcell [9], etc.

The type of the computational grid correlates with the method for discretization of differential equations: finite differences (FD) imply structured grids, finite elements (FE) imply tetrahedral or hexahedral meshes, while finite volumes (FV) fit to general polyhedral meshes.

Discretization results in a system of linear or nonlinear algebraic equations. The total number of equations is proportional to the number of mesh elements (cells, faces, edges, nodes) and ranges from thousands to billions.

Solution methods. For the solution of systems of nonlinear algebraic equations, the following approaches can be used: Newton method, Picard method, line search method, trust interval method and others. Examples of nonlinear solver packages are: SUNDIALS [17], Trilinos NOX [18]. Automatic differentiation capabilities of INMOST allow for easy assembly and solution of the nonlinear system with the Newton method.

The solution of a nonlinear system is based on its linearization which results in a large system of sparse linear equations. For solution of the system the following methods are applicable: direct factorization, Krylov's iterative methods with different preconditioners such as incomplete factorization, algebraic multigrid methods, domain decomposition methods, etc. A variety of advanced software packages can be used for this purpose: SuperLU [19], MUMPS [20], PETSc [21], Trilinos [22], Hypre [23], ILUPACK [24], INMOST Solve [9], etc.

Computer program. To design the computer program and connect the required software packages, one exploits computer languages C, C++, Fortran, Python, etc. To utilize parallelism of modern computers including various accelerators, one may use computer libraries and languages: OpenMP [25], MPI [26], OpenCL [27], CUDA [28], etc. In some cases the number of computer languages within a numerical model may reach 4 or 5.

For computer program debugging the following tools are available: gdb [29], valgrind [30], drmemory [31], etc.

Post-processing and analysis of results. Computation of required metrics, plotting of

graphs and diagrams, generation of pictures and videos facilitate analysis of the computed results. Useful visualization tools are ParaView [32], VisIt [33], General Mesh Viewer [34], gnuplot [35], INMOST DrawGrid [9], etc.

4 INMOST-BASED NUMERICAL MODELS

In this section we review several applications of INMOST platform in numerical modelling.

4.1 Incompressible fluid flow

We first consider the incompressible fluid flow. The problem is described by the Navier–Stokes equations:

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \operatorname{div}\left(\rho \mathbf{u}\mathbf{u}^T - \mu \nabla \mathbf{u} + p\mathbf{I}\right) = \mathbf{0}, \qquad \operatorname{div}\left(\mathbf{u}\right) = 0. \tag{1}$$

Here $\mathbf{u}=\{u,v,w\}^T$ is the unknown fluid flow velocity, p is the unknown pressure, ρ is the constant density and μ is the dynamic viscosity. The system of Navier–Stokes equations is augmented by the boundary conditions. On the boundary we can impose no-slip or slip condition, Maxwell–Navier friction, do-nothing condition, condition for the normal stress (fixed pressure).

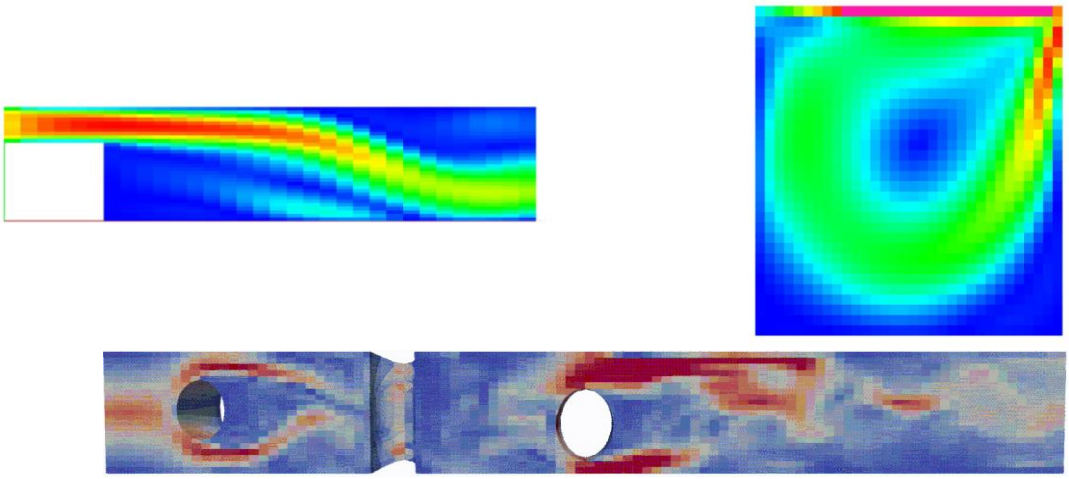


Figure 1. FV solutions: flow over the backward step (top-left), flow in the lid-driven cavity (top-right), flow in a channel with obstacles (bottom).

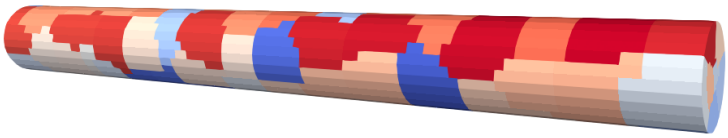


Figure 2. Decomposition of a computational domain for 92 processors

Both the momentum and the continuity equations are discretized by the finite volume method. The unknowns are collocated at the cell centers. To avoid checkerboard patterns in the discrete pressure due to the Ladyzhenskaya–Babushka–Brezzi instability, we stabilize the discretization by eigen-splitting of the part of the flux that corresponds to the pressure and the incompressibility condition, by analogy with the approach [44]. The FV system is nonlinear due to the convective term and the numerical scheme is stable even for large time steps.

In Figure 1 (top row of pictures) visualization of the flow for the solutions to two benchmark problems are shown: the steady flow over the backward step and the steady flow in the lid-driven cavity. In the bottom picture of Figure 1 we present the unsteady flow in a channel with cylindrical obstacles in the case of high Reynolds number (Re~1000). Figure 1 demonstrates physically adequate absolute velocity solution for a relatively small number of cells using finite volume discretization scheme. The presented examples make the full use of the INMOST programming platform, i.e. they exploit the general mesh data structure, the sparse matrix structure and automatic differentiation for the Jacobian system assembly and built-in sparse parallel linear solvers. Thanks to INMOST the flow model may be run in parallel. Decomposition of a computational mesh in a channel for 92 processors is demonstrated in Figure 2. The solution times of steady Pousielle flow problem for 36 and 92 MPI processes were 2.51 and 1.25 seconds, respectively, which implies feasible parallel efficiency.

The presented incompressible fluid flow model was complemented with the blood coagulation model [7] using INMOST multiphysics extension. This extension is currently under active development. It allows to couple the flow model with a reaction-advection-diffusion model triggering coagulation into a joint nonlinear system which is solved on each time step.

4.2 Free surface fluid flow

We solve numerically the problem of fluid flow with a free surface on dynamically adapted octree grids. The problem is guided by the coupled solution of the system of incompressible Navier–Stokes equations and the level set equation:

$$\begin{gathered}\rho\frac{\partial\mathbf{u}}{\partial t}+\operatorname{div}\left(\rho\mathbf{u}\mathbf{u}^T-\mu\nabla\mathbf{u}+p\mathbf{I}\right)=\rho g\nabla z,\\ \operatorname{div}\left(\mathbf{u}\right)=0,\\ \frac{\partial\varphi}{\partial t}+\operatorname{div}\left(\varphi\mathbf{u}\right)=0,\\ |\nabla\varphi|=1,\end{gathered}\tag{2}$$

where $\mathbf{u}=\{u,v,w\}^T$ and p are unknown velocity and pressure, ρ is the density, μ is the dynamic viscosity, g is the gravitational constant. In this problem the velocity is staggered on cell faces and the pressure is collocated at cell centers. The Navier–Stokes equations are augmented with the boundary conditions listed in Section 4.1. The unknown level-set function φ, satisfying the Neumann boundary condition, is passively advected with fluid and reinitialized by the solution of the Eikonal equation. The level-set function is prescribed at mesh nodes, its sign separates the domain into two parts: the fluid domain and the empty domain, see Figure 3 (left). The boundary condition at the interface accounts the surface

tension:

$$-\mu \frac{\nabla \mathbf{u} + (\nabla \mathbf{u})^T}{2} \mathbf{n} + p\mathbf{n} = \sigma \kappa \mathbf{n}. \tag{3}$$

Here σ is the surface tension coefficient, κ is the surface curvature computed from the level-set function, **n** is the surface normal. More implementation details can be found in [45] and references therein. Figure 3 (right) shows the surface tension induced water crown as a splash from falling drops into a pool filled with the fluid.

Currently, the numerical model relies on the sparse matrix structure, automatic differentiation and linear solvers from INMOST tackling the diffusion problem and pressure projection problem. However, the octree mesh data structure of the model is not based on INMOST tools limiting parallelization of the model to OpenMP technology. The parallel general mesh adaptivity functionality is already developed [8, 46], and the transition of the code to the INMOST mesh data structure is underway.

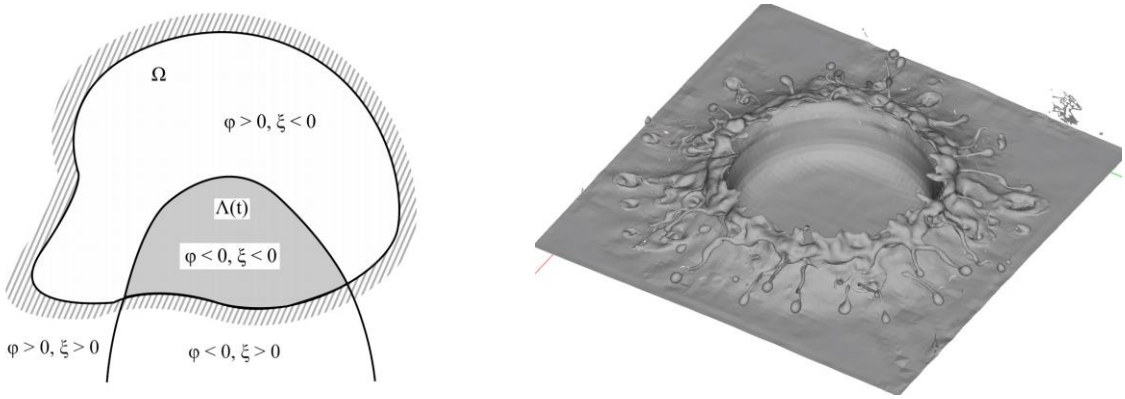


Figure 3. Computational domain (left) and the computed water crown (right)

4.3 Oil and gas modelling

The flow of the mixture of fluids subject to the Darcy law is used for simulation of primary or secondary oil and gas recovery from a heterogeneous anisotropic fractured reservoir. The black oil model equations for unknown pressure and saturations are [50]:

$$\begin{aligned}
&\frac{\partial \rho_w \theta S_w}{\partial t} - \operatorname{div}\left(\lambda_w \mathbf{K}\left(\nabla p - \rho_w g \nabla z\right)\right) = q_w, \\
&\frac{\partial \rho_o \theta S_o}{\partial t} - \operatorname{div}\left(\lambda_o \mathbf{K}\left(\nabla p - \nabla Pc_o - \rho_o g \nabla z\right)\right) = q_o, \\
&\frac{\partial \rho_g \theta \left(R S_o + S_g\right)}{\partial t} - \operatorname{div}\left(\lambda_g \mathbf{K}\left(\nabla p - \nabla Pc_g - \rho_g g \nabla z\right)\right) \\
&\quad - \operatorname{div}\left(\lambda_{go} \mathbf{K}\left(\nabla p - \nabla Pc_o - \rho_o g \nabla z\right)\right) = q_g.
\end{aligned} \tag{4}$$

Here **K** is the permeability of the rock, R is the gas solubility, θ is the porosity, ρ is the density dependent on pressure, $\lambda = \rho k_r / \mu$ are phase mobilities dependent on pressure and saturations for water, oil and gas, q are sources and sinks representing the wells guided by the

Peaceman formula [49]. The no-flow boundary condition is imposed on all the boundaries. Various FV discretization methods for the Darcy problem may be used for the numerical solution of system (4) [51]. In Figure 4 (left) we present an example of water saturation field in a reservoir with complex geology defined through a general mesh adapted to geological structures. In Figure 4 (right) we demonstrate water saturation in a network of fractures.

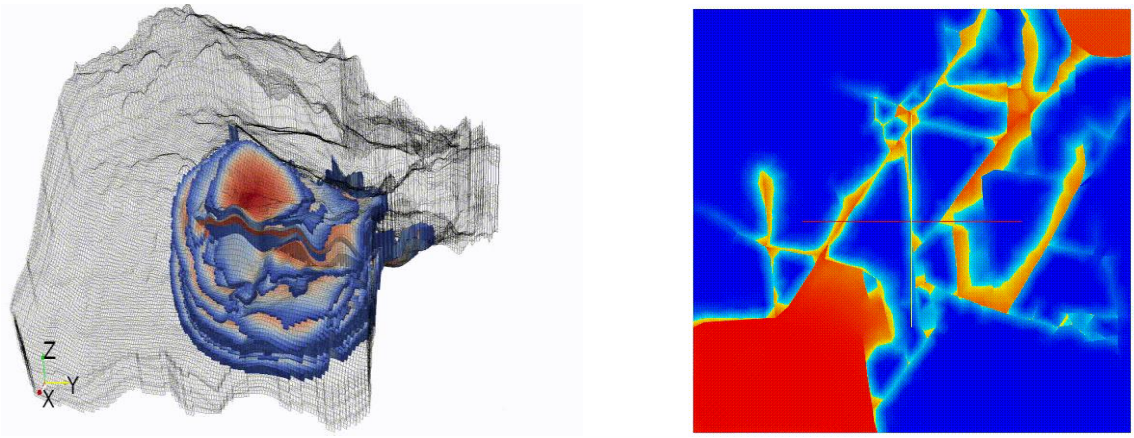


Figure 4. Water saturation field and general mesh in a reservoir (left); water saturation in a fracture network (right)

The numerical model uses the mesh data structure, sparse matrix structure, automatic differentiation and linear solvers from INMOST.

4.4 Mechanics of deformable bodies

The model describes the elastic deformation of bodies from heterogeneous anisotropic material. The model is given by the elasticity equations $\sigma = \mathbf{C}:(\nabla\mathbf{u} + (\nabla\mathbf{u})^T)/2,\ \ \mathrm{div}(\sigma) = 0.$ Here $\mathbf{u}=\{u,v,w\}^T$ is the displacement field, σ is the 3x3 stress tensor, $\mathbf{C}$ is the 4-th rank material stiffness tensor, ":" is the contraction operator. The system is augmented by Diriclet, Neumann or roller type boundary conditions.

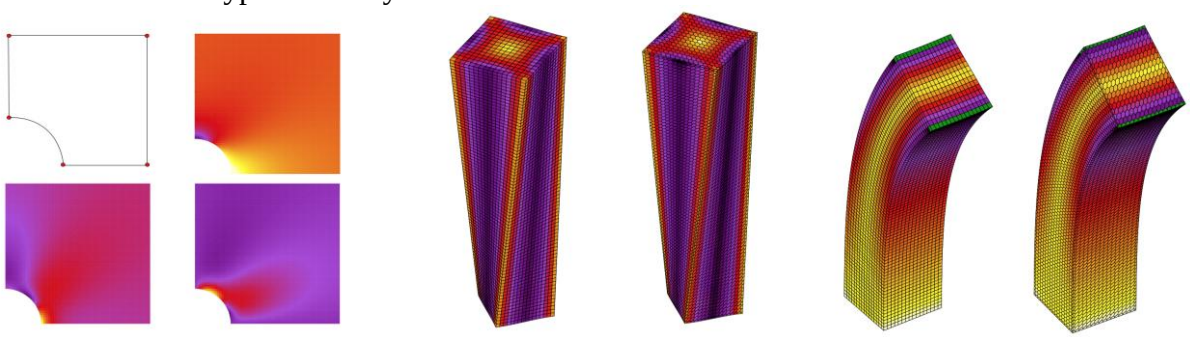


Figure 5. Stress components under load (left), stress magnitude under twisting (middle), and bending (right)

The numerical implementation is based on the FV method described in [47]. Figure 5 demonstrates the application of the model to benchmark problems: three components of stress in perforated infinite strip under load (left), magnitude of stress in a beam under twisting (middle) and bending (right) on structured and unstructured grids.

The numerical model uses the mesh data strucutre, sparse matrix structure, automatic differentiation and linear solvers from INMOST.

4.5 Poromechanics

The last example describes a solid body saturated with the fluid. The model is used for analysis of land subsidence and Earth fissuring due to water pumping into or out of the ground, as well as bed failure under constructions. It couples the solid mechanics problem (Section 4.4) with the fluid filtration problem guided by the Darcy problem (Section 4.3), the interaction between solid and fluid being described by Biot coupling terms [52]. For the single phase flow with unknowns u, v, w, and p it reads as:

$$\begin{aligned} &\frac{1}{M}\frac{\partial p}{\partial t} - \operatorname{div}\left(\mathbf{K}\left(\nabla p - \rho g \nabla z\right) - \mathbf{B}\frac{\partial \mathbf{u}}{\partial t}\right) = q, \\ &-\operatorname{div}\left(\mathbf{S}:\frac{\nabla \mathbf{u} + \left(\nabla \mathbf{u}\right)^T}{2} - \mathbf{B}p\right) = \rho g \nabla z. \end{aligned} \tag{5}$$

Here p is the fluid pressure, $\mathbf{u}=\{u,v,w\}^T$ is the displacement of solid, $\mathbf{S}$ is the material compliance tensor, inverse of the stiffness tensor, ρ is the density, $\mathbf{B}$ is the Biot coefficient tensor, M is the Biot modulus, $1/M$ is the specific storage coefficient. The boundary conditions are similar to those discussed in Sections 4.3 and 4.4.

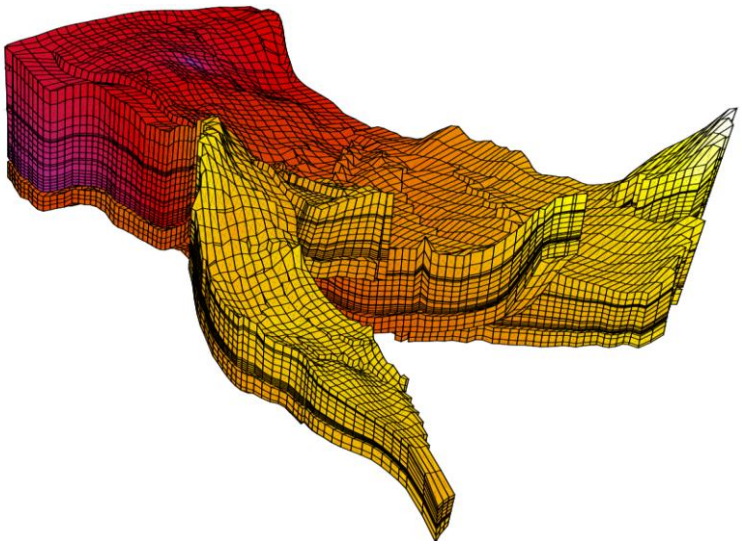


Figure 6. Computational mesh and the pressure field for poromechanics problem

Figure 6 demonstrates the pressure field for the poromechanics problem due to oil recovery from the Norne oil field, the elastic properties are synthetic. The grid is unstructured with faults, the system (5) is discretized by the FV method [48].

The numerical model uses the mesh data structure, sparse matrix structure, automatic differentiation and linear solvers from INMOST.

CONCLUSION

The present paper presents the INMOST platform as a powerful tool for numerical modelling. INMOST provide a wide functionality of operations for general distributed mesh data, convenient interface for assembling and solving systems of linear and nonlinear equations, as well as analysis and visualization of obtained solutions. The description of the most important modules of INMOST is given. An overview of numerical modelling stages along with the tools that help in their implementation is presented. The place of INMOST platform among other modern widespread libraries and numerical modelling packages is shown.

For a specific application in numerical modeling, it may turn out to be the most optimal choice of a special grid generator, as well as the sampling method for this type of grid. This is the main reason that these modules are not included in the INMOST software platform. For the same reason, it is difficult to make a full comparison of INMOST with other popular numerical modelling packages, which mainly use the simplest semi-regular types of grids, allowing the direct use of certain types of discretization, for example, two-point finite-volume ones. Conversely, INMOST software platform focuses on support for operation with general type grids. Despite this, the presented examples of INMOST application demonstrate appealing features of INMOST-based numerical modelling. Especially, this includes a wide range of its application for solving problems from gas hydrodynamics and problems with a free surface to problems of mechanics of a solid and deformable bodies, as well as problems of subsurface flow, including problems of poromechanics.

Acknowledgements: The work was supported in part by the world-class research center “Moscow Center for Fundamental and Applied Mathematics” and by the RAS Research program No. 26 “Basics of algorithms and software for high performance computing”.

REFERENCES

[1] Yu. Vassilevski, I. Konshin, G. Kopytov, K. Terekhov, *INMOST – a software platform and graphical environment for development of parallel numerical models on general meshes.* Lomonosov Moscow State Univ. Publ., Moscow, (2013), 144 p. (in Russian)

[2] A.A. Danilov, K.M. Terekhov, I.N. Konshin, Yu.V. Vassilevski, “INMOST parallel platform: framework for numerical modeling”, *Supercomp. Frontiers and Innovations*, **2**(4), 55–66 (2015).

[3] I. Konshin, I. Kapyrin, K. Nikitin, K. Terekhov, “Application of the parallel INMOST platform to subsurface flow and transport modelling”, *Lecture Notes in Computer Science*, **9574**, 277–286 (2016).

[4] F.V. Grigoriev, I.V. Kapyrin, I.N. Konshin, “Software platform INMOST in the GeRa code to operate with the distributed mesh data”. *Mathematica Montisnigri*, **XXXVI**, 27–44 (2016) (in Russian)

[5] I. Konshin, I. Kapyrin, “Scalable computations of GeRa code on the base of software platform INMOST”, *Lecture Notes in Computer Science*, **10421**, 433–445 (2017).

[6] V. Kramarenko, I. Konshin, Y. Vassilevski, “Ani3D-extension of parallel platform INMOST and hydrodynamic applications”, *Communications in Computer and Information Science*, **793**, 219–228 (2017).

[7] K. Terekhov, Yu. Vassilevski, INMOST parallel platform for mathematical modeling and applications. *Communications in Computer and Information Science*, **965**, 230–241 (2018).

[8] K. Terekhov, Yu. Vassilevski, Mesh modification and adaptation within INMOST programming

platform. *Lecture Notes in Computer Science and Engineering*, **131**, 243–255 (2019).
[9] INMOST – a toolkit for distributed mathematical modelling. http://www.inmost.org (Accessed November 29, 2019).
[10] AutoCAD – a computer-aided design (CAD) software. https://www.autodesk.com/products/autocad/overview (Accessed November 29, 2019).
[11] 3DS Max – 3D modeling and rendering software for design visualization. http://www.autodesk.com/products/3ds-max/overview (Accessed November 29, 2019).
[12] Sketchup – 3D modeling for professionals. https://www.sketchup.com/ (Accessed November 29, 2019).
[13] OpenCascade – Advanced 3D Technology. https://www.opencascade.com/ (Accessed November 29, 2019).
[14] Tetgen – a quality tetrahedral mesh generator and a 3D Delaunay triangulator. http://wias-berlin.de/software/index.jsp?id=TetGen (Accessed November 29, 2019).
[15] GMSH – a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities. http://gmsh.info/ (Accessed November 29, 2019).
[16] ANI3D – Advanced Numerical Instruments 3D. https://sourceforge.net/projects/ani3d/ (Accessed November 29, 2019).
[17] SUNDIALS: SUite of Nonlinear and DIfferential/ALgebraic equation Solvers. URL: https://computing.llnl.gov/projects/sundials (Accessed November 29, 2019).
[18] Trilinos NOX – nonlinear analysis tools. https://trilinos.github.io/nox_and_loca.html (Accessed November 29, 2019).
[19] SuperLU – Supernodal sparse direct solver. https://portal.nersc.gov/project/sparse/superlu/ (Accessed November 29, 2019).
[20] MUMPS: MUltifrontal Massively Parallel sparse direct Solver. http://mumps.enseeiht.fr/ (Accessed November 29, 2019).
[21] PETSc – Portable Extensible Toolkit for Scientific Computation. https://www.mcs.anl.gov/petsc/ (Accessed November 29, 2019).
[22] Trilinos linear solver. https://trilinos.github.io/linear_solver.html (Accessed November 29, 2019).
[23] Hypre – scalable linear solvers and multigrid methods. https://computing.llnl.gov/projects/hypre-scalable-linear-solvers-multigrid-methods (Accessed November 29, 2019).
[24] ILUPACK – multilevel ILU preconditioners for general real and complex matrices. http://ilupack.tu-bs.de/ (Accessed November 29, 2019).
[25] OpenMP – multi-platform shared-memory parallel programming in C/C++ and Fortran. https://www.openmp.org/ (Accessed November 29, 2019).
[26] MPI – Message Passing Interface. https://www.mpi-forum.org/ (Accessed November 29, 2019).
[27] OpenCL (Open Computing Language) – The open standard for parallel programming of heterogeneous systems. https://www.khronos.org/opencl/ (Accessed November 29, 2019).
[28] CUDA (Compute Unified Device Architecture) – a parallel computing platform and programming model for general computing on GPUs. https://developer.nvidia.com/cuda-zone (Accessed November 29, 2019).
[29] gdb – The GNU project debugger. https://www.gnu.org/software/gdb/ (Accessed November 29, 2019).
[30] valgrind – instrumentation framework for building dynamic analysis tools. http://valgrind.org/ (Accessed November 29, 2019).
[31] drmemory – Memory Debugger for Windows, Linux, and Mac. https://drmemory.org/ (Accessed November 29, 2019).
[32] ParaView – an open-source, multi-platform data analysis and visualization application. https://www.paraview.org/ (Accessed November 29, 2019).
[33] VisIt – an open-source interactive parallel visualization and graphical analysis tool for viewing scientific data. https://visit.llnl.gov/ (Accessed November 29, 2019).

[34] GMV – The General Mesh Viewer. http://www.gmv-barracuda.com/index.html (Accessed November 29, 2019).
[35] gnuplot – a portable command-line driven graphing utility. http://www.gnuplot.info/ (Accessed November 29, 2019).
[36] E. S. Seol, FMDB: Flexible Distributed Mesh DataBase for parallel automated adaptive analysis. *Ph.D. Thesis*, Rensselaer Polytechnic Institute, New York, 2005, 151 p. Available at URL: https://www.scorec.rpi.edu/REPORTS/2005-9.pdf (Accessed November 29, 2019).
[37] MOAB – a Mesh-Oriented datABase. URL: https://press3.mcs.anl.gov/sigma/moab-library/ (Accessed November 29, 2019).
[38] MSTK (MeSh ToolKit) – Infrastructure library for reading, writing, querying and manipulating distributed, general unstructured polyhedral meshes. https://github.com/MeshToolkit/MSTK (Accessed November 29, 2019).
[39] STK (Sierra ToolKit mesh) – an unstructured mesh database. https://trilinos.github.io/stk.html (Accessed November 29, 2019).
[40] Salome – an open-source software that provides a generic pre- and post-processing platform for numerical simulation. URL: https://www.salome-platform.org/ (Accessed November 29, 2019).
[41] OpenFOAM – Open Source Field Operation And Manipulation CFD ToolBox. URL: https://www.openfoam.com/ (Accessed November 29, 2019).
[42] ParMETIS: parallel graph partitioning and fill-reducing matrix ordering. http://glaros.dtc.umn.edu/gkhome/metis/parmetis/overview (Accessed November 29, 2019).
[43] Zoltan – a toolkit of parallel services for dynamic, unstructured, and/or adaptive simulations. https://trilinos.github.io/zoltan.html (Accessed November 29, 2019).
[44] K.M. Terekhov, Yu.V. Vassilevski, "Finite volume method for coupled subsurface flow problems, I: Darcy problem", *J. of Computational Physics*, **395**, 298–306 (2019).
[45] K.D. Nikitin, K.M. Terekhov, Yu.V. Vassilevski, "Two methods of surface tension treatment in free surface flow simulations", *Applied Mathematics Letters*, **86**, 236–242 (2018).
[46] K.M. Terekhov, "Parallel dynamic mesh adaptation within INMOST platform", *Communications in Computer and Information Science*, **1129**, 313–326 (2019).
[47] K.M. Terekhov, H.A. Tchelepi, "Cell-centered finite-volume method for elastic deformation of heterogeneous media with full-tensor properties", *J. of Computational and Applied Mathematics*, **364**, 112331, (2020).
[48] K.M. Terekhov, "Cell-centered finite-volume method for heterogeneous anisotropic poromechanics problem", *J. of Computational and Applied Mathematics*, **365**, 112357 (2020).
[49] D.W. Peaceman, "Representation of a horizontal well in numerical reservoir simulation", *SPE Advanced Technology Series*, **1**(1), 7–16 (1993).
[50] Zh. Chen, "Reservoir simulation: mathematical techniques in oil recovery", *SIAM Series in Applied Mathematics*, Vol. 77, 2007.
[51] K.M. Terekhov, B.T. Mallison, H.A. Tchelepi, "Cell-centered nonlinear finite-volume methods for the heterogeneous anisotropic diffusion problem", *J. Comput. Physics*, **330**, 245–267 (2017).
[52] M.A. Biot, "General theory of three-dimensional consolidation", *J. of Applied Physics*, **12**(2), 155–164 (1941).

Received November 20, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

ATOMISTIC MODELING OF THE DYNAMICS OF THE SOLID/LIQUID INTERFACE OF SI MELTING AND CRYSTALLIZATION TAKING INTO ACCOUNT DEEPLY SUPERHEATED/SUPERCOOLED STATES

V.I. MAZHUKIN*, A.V. SHAPRANOV, A.V. MAZHUKIN, P.V. BRESLAVSKY

Keldysh Institute of Applied Mathematics, Russian Academy of Science.

*Corresponding author. E-mail: vim@modhef.ru

DOI: 10.20948/mathmontis-2020-47-8

Summary. The velocity of the solid/liquid interface (SLI) $\upsilon_{s\ell}$ plays an important role in the processes of crystallization and melting. It is one of the fundamental concepts of materials science. Based on the analysis of kinetic models of melting/crystallization with diffusion and collisional-thermal constraints, a modification of the transition state theory is performed. Using two interaction potentials (KIHS and SW), molecular dynamics modeling of Si melting/crystallization under deep overheating/undercooling was performed. From comparing the simulation results with the data of the modified kinetic model, we constructed the response function of interface $\upsilon_{s\ell}$ in the region of the maximum allowable values of superheating/undercooling of Si. The temperature dependence of the velocity of the solid/liquid interface $\upsilon_{s\ell}$ is diffusion-limited and is described by the same equation in the entire temperature range.

1 INTRODUCTION

Melting of solid body and liquid solidification are among the widely used and actively studied [1] phenomena. Two mechanisms of melting/crystallization of solids/liquids are known: heterogeneous (surface or frontal) and homogeneous (volume). In the first case, in the framework of classical thermodynamics [2], melting of solids and solidification of liquids belong to the first-order phase transformations that occur at a certain (equilibrium) temperature T_m, which corresponds to the equality of the Gibbs free energies of the solid and liquid states. The phenomena of heterogeneous melting and solidification are always non-uniform. They correspond to the motion of a continuous medium with a strong discontinuity surface, on which the mechanical, thermodynamic, thermophysical, and optical characteristics of a substance abruptly change.

The velocity $\upsilon_{s\ell}(\Delta T) = \upsilon_{s\ell}(T_{s\ell})$ of the solid/liquid interface (SLI) is the function of the deviation $\Delta T = T_{s\ell} - T_m$ from the equilibrium melting temperature T_m and is called the response function of the interface to overheating or undercooling, respectively, of the solid/liquid phase. The interface velocity $\upsilon_{s\ell}(T_{s\ell})$ is a fundamental quantity that describes crystallization and melting processes and plays a fundamental role in materials science [3-5].

A significant part of the dynamics of melting and crystal growth from the melt is determined by the heat transfer from the moving interface. However, there are limiting circumstances in which the growth rate is regulated not only by the macroscopic heat flux. One of such circumstances is the emergence of fast phase transitions of the first kind, which

2010 Mathematics Subject Classification: 82C24, 82C26, 82D37, 81T80.
Key words and Phrases: Interface problems; diffusion-limited aggregation, Atomistic modeling, Dynamic and nonequilibrium phase transitions, Silicon.

are typical to the action of concentrated energy fluxes on materials. These processes have their own specifics [6]. Fast phase transitions are accompanied by the appearance of metastable, strongly superheated/undercooled states. Reaching the maximum permissible values of overheating/undercooling [7,8] leads to a number of interface instabilities.

Initially, experimental and theoretical studies were carried out more intensively in the field of crystallization/solidification of melts [9–11]. Along with the fundamental aspects [12], this is due to a large number of technological applications related to the production of metal glasses [13, 14] amorphous semiconductors [15], nanomaterials [16], etc.

However, the widespread use in the last two decades of ultrashort pulsed (pico- and femtosecond) laser irradiation on various materials causes increased interest in fast first-order phase transitions. An analysis of the processes caused by pulsed laser irradiation leads to the consideration of a number of important fundamental problems, which, at high heating rates, include the features of homogeneous and heterogeneous melting/solidification and evaporation mechanisms and the associated extreme overheating and supercooling of matter. An understanding of the melting/solidification processes is also of great interest for the applied problems of photonics [17], ultrafast laser microprocessing of materials [18, 19], generation of nanoparticles and nanostructures [20,21], etc.

Homogeneous melting/crystallization mechanisms are characterized by the nucleation of a new phase (liquid/crystal) in a certain volume of respectively superheated crystal and undercooled melt and will not be considered in this paper.

In the theoretical studies of the mobility of the solid/liquid interface and the kinetics of crystal and melt growth associated with it, kinetic models [10, 22–24] and atomistic modeling [25–28] are widely used. In most works, the studies of the temperature dependence of the stationary velocity $\upsilon_{s\ell}(T_{s\ell})$, as before, are carried out mainly in the temperature range of crystallization. Nevertheless, none of the models discussed describes the solidification process in the entire range of undercooling: from the minimum near the melting point T_m to the maximum in the region of amorphization (glass transition). In addition, the important question remains open about the possibility of using the analyzed kinetic models to determine the temperature dependence of the solid/liquid interface velocity for the melting process in the entire overheating region: from the minimum near the melting point T_m to the maximum in the spinodal region.

The main tool for studying the kinetic rate of melting/crystallization in the region of the maximum allowable values of superheating/supercooling is atomistic modeling, the results of which are compared with the data of kinetic models. An acceptable match is achieved by introducing appropriate correction parameters in the model [25 - 30].

The main goal of the work is to construct a modified kinetic model with diffusion constraint that describes the mobility of the solid/liquid interface in a wide temperature range, including the region of maximum permissible values of superheat/supercooling. As the studied material, crystalline silicon (Si) was chosen, for which a series of MD calculations of the temperature dependence of the stationary melting/crystallization rate $\upsilon_{s\ell}(T_{s\ell})$ were performed. MD results are compared with the data obtained from the modified kinetic model.

2 KINETICS THEORY OF THE SOLID/LIQUID INTERFACE

Of the many solid/liquid interface kinetics theories that display various crystal growth mechanisms, two of the most commonly used and cited directions can be distinguished. The

first is described by original Wilson – Frenkel (WF) growth models [31, 32], in which the interface velocity is associated with the diffusion of atoms in the liquid phase. This theory is often called the transition state theory, since it assumes that melting or solidification occurs through some intermediate or transition state. In this theory, a diffusion limitation mechanism is used to control the speed of the crystallization-melting front. This mechanism is based on the assumption that atoms (molecules) must overcome the diffusion barrier upon transition from a liquid to a solid phase [31, 32]. The transition is accompanied by a significant restructuring of the interface. The rate of the crystallization process was assumed to be proportional to the diffusion coefficient, which is usually presented in the form of the Arrhenius equation

$$D = D_0 \exp\left(-\frac{Q}{k_B T_{s\ell}}\right) \tag{1}$$

where Q is the activation energy for the diffusive motion in liquid, k_B is the Boltzmann constant, $k_B T$ is the average thermal energy of one atom, D_0 is the prefactor, controlling the rate of the process.

In the final form, the velocity of the meting/crystallization front $\upsilon_{s\ell}(\Delta T) = \upsilon_{s\ell}(T_{s\ell})$ with the diffusive limitation is formulated as [33]:

$$\upsilon(T_{s\ell}) = \frac{a\,f}{\lambda^2} D\left[1-\exp\left(-\frac{\Delta G}{k_B T_{s\ell}}\right)\right] = C^{\mathrm{WF}}\frac{D}{a}\left[1-\exp\left(-\frac{L_m}{k_B T_m}\frac{\Delta T}{T_{s\ell}}\right)\right] \tag{2}$$

where $C^{\mathrm{WF}} = \frac{a^2}{\lambda^2} f$, a is the interatomic distance, λ is the mean free path of atoms for this process usually assumed being proportional to the lattice constant, a: $\lambda < a$, f is the efficiency coefficient (a constant of the order of unity, $f \leq 1$), characterizing the fraction of collisions of liquid atoms with the solid leading to crystallization. The values of λ, f, D_0, Q do not have a strict definition and are difficult to measure. Moreover, they depend on the crystallographic orientation of the interface.

The second direction is based on the kinetic model originally proposed by Broughton, Gilmer and Jackson (BGJ) [34] as an improvement on the earlier model (WF). The BGJ model uses, as a limitation, the frequency of thermal collisions of atoms with an interphase boundary [35]. A modification of the transition state theory [31, 32] was motivated by the results of MD modeling with the Lenard – Jones interatomic potential [36], which showed that the growth of crystals of monatomic systems may not in all cases be limited by diffusion. In particular, far from the melting temperature in the region of very low temperatures, diffusion tends to zero, while, according to the simulation results, the SLI speed is still finite. On this basis, a conclusion was formulated on the unacceptability of the model with a diffusion restriction for crystallization in the entire temperature range.

The BGJ model [34] was based on the hypothesis formulated earlier that solidification of monatomic metals is limited only by the frequency of collisions of the melt atoms with the crystal surface [35]. Following this hypothesis, an assumption was made that the kinetic model with the collision-thermal restriction mechanism best fits the obtained simulation results. According to this limiting mechanism, the maximum crystallization rate is controlled by the average thermal velocity of the atoms in the melt. The result of such reasoning was the

replacement by the authors of the BGJ model [34] of the diffusion term in (2) with the average thermal velocity of atoms $\upsilon_T = \sqrt{3k_B T_{s\ell} / m}$.

$$\upsilon_{s\ell}(T_{s\ell}) = \frac{a}{\lambda} f_0 \upsilon_T \left[1 - \exp\left(-\frac{L_m}{k_B T_m}\frac{\Delta T}{T_{s\ell}}\right)\right] = C_{BGJ}\sqrt{\frac{3k_B T_{s\ell}}{m}}\left[1 - \exp\left(-\frac{L_m}{k_B T_m}\frac{\Delta T}{T_{s\ell}}\right)\right] \quad (3)$$

where $C_{BGJ} = \frac{a}{\lambda} f_0$ is a dimensionless coefficient, m is the atomic mass.

However, in later works [25, 29, 33], atomistic modeling showed that in the range of values close to the melting temperature, the crystallization process can be displayed with acceptable accuracy by the kinetic models with the diffusion (2) and collision-thermal constraints (3), as well as models of the density functional theory [24, 37].

A much smaller number of works is devoted to analysis of the possibility of using the analyzed kinetic models (2), (3) to determine the speed of movement of the solid/liquid interface $\upsilon_{s\ell}(T_{s\ell})$ with acceptable agreement in the temperature region not only of crystallization, but also of melting with strong overheating of the solid phase.

In one of the first works [29], the results of molecular dynamics simulation were presented, in which the stationary rate of the silicon crystallization/melting interface was determined as a function of temperature in the conditions of strong undercooling/superheating. Particle interaction in the atomistic model was determined by the Stillinger – Weber potential [38]. The simulation results were compared with the data of the kinetic model with diffusion constraint (2). The model was adjusted to the simulation results by selecting the C^{WF} parameter. An analysis of the results showed that the transition state theory provides a reasonable qualitative description of heterogeneous crystallization and melting of silicon in the temperature range $0.64\ T_m \le T_{s\ell} \le 1.18 T_m$. Close to the melting temperature, good agreement with experimental data was observed. In the region of deep undercooling, there was a strong discrepancy between the results of atomistic modeling and the crystallization rate data obtained from equation (2).

Similar studies using the kinetic model with the collision-thermal constraint mechanism (3) for metals were performed in [23, 26, 28, 30]. In an early work [30], the results of comparing the molecular dynamics simulation of crystallization and melting of sodium with the data of the kinetic model (3) in the temperature range $0.2\ T_m \le T_{s\ell} \le 1.26\ T_m$ are presented. The comparison of the temperature dependence of the stationary velocity of motion solid/liquid interface $\upsilon_{s\ell}(T_{s\ell})$ showed that the kinetic model data are in good agreement with the results of atomistic modeling in the entire crystallization region. However, an agreement was not obtained for the melting branch. A similar agreement was obtained with high accuracy for the model with a frequency-thermal limitation (3) in [23, 26, 28] for atomistic modeling with EAM potentials for aluminum (Al), copper (Cu), and iron (Fe) in the following temperature ranges:

Al: $0.3 \cdot T_m \le T_{s\ell} \le 1.26 \cdot T_m$; Cu: $0.61 \cdot T_m \le T_{s\ell} \le 1.2 \cdot T_m$; Fe: $0.6 \cdot T_m \le T_{s\ell} \le 1.15 \cdot T_m$.

Nevertheless, the important question remains unanswered to date: which of the kinetic models and in which temperature range can be used to determine the stationary SLI motion velocity in the processes of metal melting/crystallization at high supercooling/overheating.

This problem is also crucial when using the continuum models describing the heterogeneous mechanisms of melting and crystallization.

3 MOLECULAR DYNAMICS STUDY

The molecular dynamics method was used to simulate the process of heterogeneous melting/crystallization of silicon under conditions when the melting/crystallization front propagates over the superheated/undercooled phase. Two series of numerical experiments were performed for two selected interaction potentials: KIHS [39] and Stillinger-Weber (SW) [38].

The calculated area of 5.5×5.5×42.3 nm in the form of a parallelepiped was filled with 64,000 particles interacting by means of the corresponding potential. In all three spatial directions, periodic boundary conditions are imposed on the boundaries of the computational domain. The particles form a layered structure containing two phases in contact: crystalline and liquid. The crystalline phase is formed by a diamond-like cubic lattice. The orientation of the lattice cells is such that the crystallographic direction [100] coincides with the large edge of the parallelepiped of the computational domain, and the melting/crystallization fronts propagate in this direction. At the initial stage, to study the melting process, the liquid phase occupies approximately 20% of the volume of the computational domain, and 80% in the study the crystallization process.

Using a thermostat in the entire calculation area, a fixed temperature value is set and maintained during the entire numerical experiment. At the same time, a constant value of external pressure is held by the barostat. The entire computational domain is divided into 30 layers of equal thickness along the long edge of the parallelepiped. Separately, in each layer, the thermostat controls the local temperature and keeps it equal to the target one. Thus, the inverse effect on the local temperature of the absorption/release of the latent heat of fusion L_m at the melting/crystallization fronts is leveled out.

As a result, the process of heterogeneous melting/crystallization quickly goes to the stationary mode, and the change in the amount of the new phase occurs almost linearly.

The integration of the equations of motion was carried out with a time step of 1 fs. Depending on the values of overheating/supercooling, from 400 thousand to 1 million steps were required to obtain each value of the melting/crystallization rate.

The values of the target temperature of the calculation region in each series were selected from the range of approximately – 40% ... + 30% of the equilibrium melting temperature. An exit to the right outside the specified temperature range ended in failure, due to the beginning of the volume nucleation process, in which the solid / liquid interface became unstable. On the left outside this range, the crystallization process did not occur, the liquid “froze”, and an amorphous phase formed.

The control of the order parameter made it possible to automatically track the positions of the melting/crystallization fronts and to calculate the speed of their motion.

The discrete set of solid/liquid interface speed values thus obtained are presented in Table 1 and with markers on the curves in Figs. 1, 2. Subsequently, they were used to construct the analytical dependences $\upsilon_{s\ell}(T_{s\ell})$ for each of the considered interaction potentials of model silicon.

KIHS										
$T_{s\ell}$	1007.4	1175.4	1343.7	1428	1512	1595.8	1680	1844.9	2009.2	2171.9
$\upsilon_{s\ell}$	-0.04	-0.41	-8	-13.7	-14	-10.03	0	30	81.75	160.22
SW										
$T_{s\ell}$	839.5	1007.5	1175.6	1344.2	1512	1680	1845.2	2010	2174	2256
$\upsilon_{s\ell}$	-0.07	-1.5	-7.45	-17.9	-15.4	0	27.7	77.5	141.8	177

Table 1. The values of velocity $\upsilon_{s\ell}(T_{s\ell})$, obtained from MD calculations.

4 MODIFIED KINETIC MODEL

In the theory of the transition state of Jackson and Chalmers [40], it is assumed that crystallization and melting proceed through an indefinite transition state, which is characterized by the presence of processes with two rates: one describes the melting rate $R_{s\to\ell}$, and another - solidification rate $R_{\ell\to s}$. The difference of these two gives the velocity of the solid/liquid interface:

$$\upsilon_{s\ell} = R_{s\to\ell} - R_{\ell\to s} \tag{4}$$

The intermediate state (intermediate phase) through which direct and reverse transitions take place has some Gibbs energy G^*. The driving force of these transitions is the difference between the Gibbs energy G^* and the one of the corresponding phase G_s, G_ℓ, being in the Arrhenius exponent:

$$\begin{aligned} R_{s\to\ell} &= \chi_s \exp[-(G^* - G_s)/kT] \\ R_{\ell\to s} &= \chi_\ell \exp[-(G^* - G_\ell)/kT] \end{aligned} \tag{5}$$

Here k is the Boltzmann constant, T is the temperature, χ_s, χ_ℓ are the proportionality coefficients, which in classical theory are assumed to be equal to each other $\chi_s = \chi_\ell = \chi$.

Then the velocity is written as:

$$\upsilon_{s\ell} = \chi \exp\left[-\frac{(G^* - G_\ell)}{kT}\right]\left\{\exp\left[\frac{(G_s - G_\ell)}{kT}\right] - 1\right\} \tag{6}$$

The energy barrier in front of the curly bracket in the exponent can be interpreted as the activation energy of the process that limits the rate of melting/crystallization. In the Wilson-Frenkel phenomenological theory, this limiting process is associated with the diffusion of atoms in the liquid delivering the atoms to the crystallization front:

$$\upsilon_{s\ell} = \chi \exp\left[-\frac{Q}{kT}\right]\left\{\exp\left[\frac{(G_s - G_\ell)}{kT}\right] - 1\right\} \tag{7}$$

After performing thermodynamic transformations of the Gibbs energy difference between the solid and liquid phases, the final expression is obtained. It is the main conclusion of the transition state theory as applied to melting/crystallization processes with diffusion constraint:

$$\upsilon_{s\ell}(T_{s\ell}) = \chi \exp\left[-\frac{Q}{kT_{sl}}\right]\left\{\exp\left[\frac{L_m(T_{s\ell} - T_m)}{T_{s\ell}T_m}\right] - 1\right\} \tag{8}$$

Here T_m is the equilibrium melting temperature, $T_{s\ell}$ is the temperature of the melting front, L_m is the latent heat of melting at the melting temperature T_m.

The equation (8) showed an acceptable coincidence of the rate in the crystallization and melting region in a small vicinity of the equilibrium temperature T_m with the results of molecular dynamics modeling (MDM) and experimental data. However, in a wide temperature range in which the melting/crystallization processes can proceed with the maximum allowable values of superheating/undercooling, an acceptable agreement could not be obtained, since the kinetics of melting/crystallization far from the temperature T_m differs significantly from the kinetics in the vicinity of T_m.

To overcome this difficulty, it is necessary to modify the Wilson-Frenkel kinetic model with diffusion constraint (8). The proposed modification is based on the assumption that the processes of direct and reverse transitions (5) in the transition state are asymmetric. The simplest form of asymmetry can be represented as the absence of equality of proportionality coefficients $\chi_s \neq \chi_\ell$, and the relation χ_s/χ_ℓ as a functional dependence on the temperature $f(T_{s\ell})$. Considering the relation χ_s/χ_ℓ, the expression for the velocity (6) can be written as:

$$\upsilon_{s\ell} = \chi_\ell \exp\left[-\frac{(G^* - G_\ell)}{kT}\right]\left\{\frac{\chi_s}{\chi_\ell}\exp\left[\frac{(G_s - G_\ell)}{kT}\right] - 1\right\} \tag{9}$$

Or, repeating the above discussion,

$$\upsilon_{s\ell}(T_{s\ell}) = \chi_\ell \exp\left[-\frac{Q}{kT_{s\ell}}\right]\left\{\frac{\chi_s}{\chi_\ell}\exp\left[\frac{L_m(T_{s\ell} - T_m)}{kT_{s\ell}T_m}\right] - 1\right\} \tag{10}$$

To determine the functional dependence $\chi_s/\chi_\ell = f(T_{s\ell})$ we use the following considerations. The relation χ_s/χ_ℓ is generally different from unity. But in the state of equilibrium, when $T_{s\ell} = T_m$ and $G_s = G_\ell$, the expression in the curly brackets in (9), (10) must be equal to zero. This means that functional dependence $\chi_s/\chi_\ell = f(T_{s\ell})$ should in equilibrium take a value equal to unity.

As such a dependence satisfying the condition χ_s/χ_ℓ =1, one can use the ratio

$$\frac{\chi_s}{\chi_\ell} = \exp\left[\alpha\frac{T_{s\ell} - T_m}{T_m}\right], \tag{11}$$

where α is a dimensionless coefficient.

In view of expression (11), the modified model with diffusion constraint takes the final form:

$$\upsilon_{s\ell}(T_{s\ell}) = D_0 \exp\left[-\frac{Q}{kT_{s\ell}}\right]\left\{\exp\left[\left(\frac{L_m}{kT_{s\ell}} + \alpha\right)\frac{T_{s\ell} - T_m}{T_m}\right] - 1\right\}, \tag{12}$$

where $D_0 = \chi_\ell$ is a dimensional constant.

The modified equation (12) contains 3 constants D_0, Q, α, the values of which were determined from a comparison with MD results. To do this, we used a procedure containing

the least squares criterion σ [41], which minimizes the deviation of the values of equation (12) with the selected parameters from the MD results. The values of the calculated parameters D_0, Q, α, as well as the least squares criterion σ for each of the interaction potentials are shown in Table 2.

Interaction potential	D_0 , [m/s]	Q, [eV]	α	σ, [m/s]
KIHS	83460.42	0.9241	– 1.1308	2.30
SW	22816.30	0.7220	– 0.8994	1.76

Table 2. The values of the calculated parameters D_0, Q, α, and the least squares criterion σ for each of the interaction potentials.

The equilibrium temperatures Tm and latent heat of melting L_m corresponding to the potentials KIHS [39] and SW [38] are taken from [42, 43] and are shown in Table 3.

Interaction potential	T_m, [K]	L_m, [kJ/mole]
KIHS [42,43]	1680	35
SW [42,43]	1680	32
Reference data [44]	1688	45,3

Table 3. Equilibrium values of temperature T_m and latent heat L_m.

5 DISCUSSION OF THE RESULTS

Figs. 1a, b, present the results of the molecular dynamics simulation, in which the stationary rate of the Si crystallization/melting interface was determined as a function of temperature in the conditions of strong undercooling/superheating. The interaction of particles in the atomistic model was determined by the potentials of SW [38] and KIHS [39]. The simulation results were compared with the data of the modified kinetic model with diffusion constraint (12). The approximation of the model to the simulation results was carried out by selecting 3 parameters D_0, Q, α. In the deep undercooling region, the modified kinetic model (12) showed that the data on crystallization rate obtained from the equation (12) is almost completely identical to the results of atomistic modeling for both potentials. Thus, the introduction into the theory of the transition state of a functional temperature dependence of direct and reverse transitions makes it possible to eliminate the main drawback of the Wilson-Frenkel model, which shows that in the region of very low temperatures, diffusion tends to zero faster than modeling data indicating that the velocity solid/liquid interface is still finite.

In particular, far from the melting temperature in the region of very low temperatures, diffusion tends to zero, while, according to the simulation data, the SLI speed is still finite.

An analysis of the results showed that the modified theory of the transition state gives a reasonable qualitative description of heterogeneous crystallization and melting of silicon in the temperature range $0.596T_m \leq T_{s\ell} \leq 1.28T_m$ for the KIHS potential and $0.49T_m \leq T_{s\ell} \leq 1.35T_m$ for the SW potential.

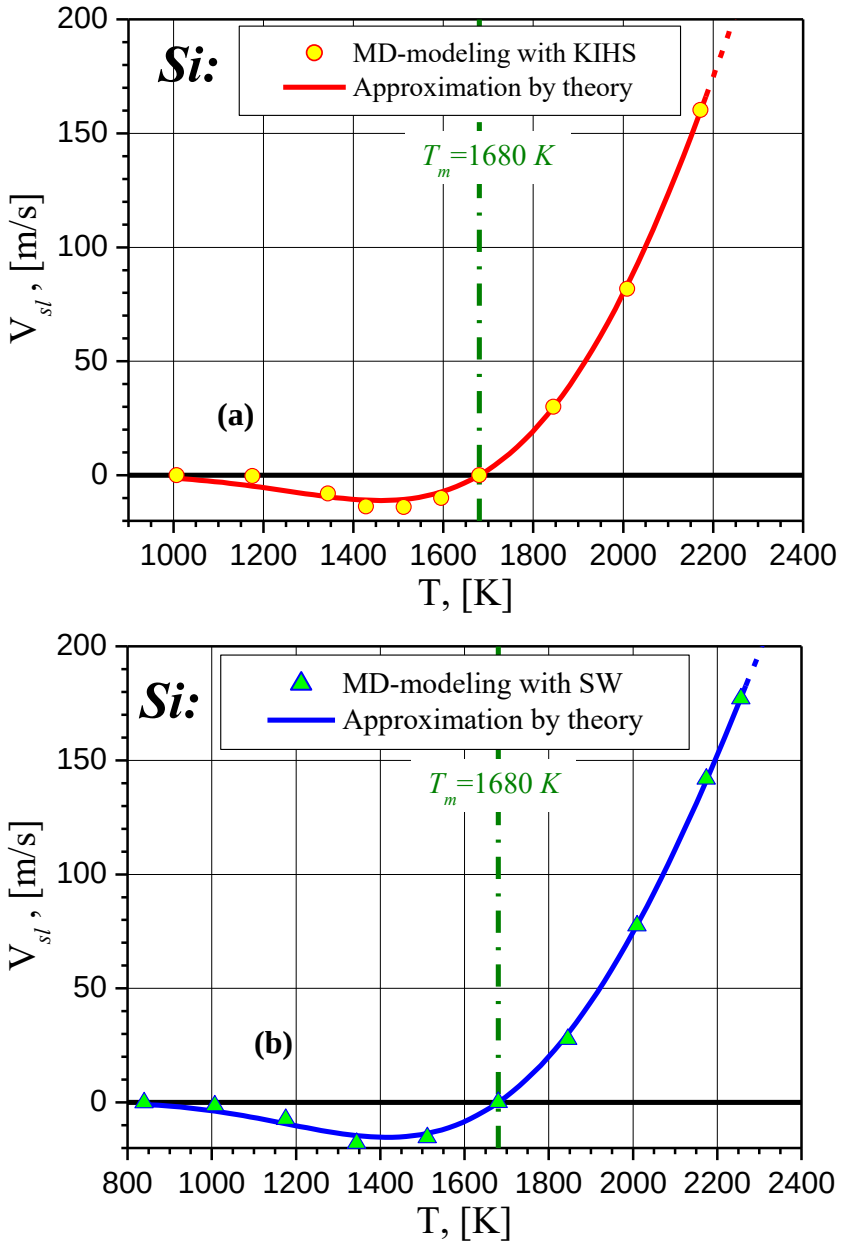


Fig.1. Temperature dependence of the melting/crystallization front for silicon: MD modeling and approximation. KIHS potential (a), SW potential (b).

In the vicinity of the melting temperature T_m, a smooth change in the temperature dependence of the velocity, determined from the simulation, is observed without breaking the slope when passing through the melting point for both potentials.

At the same time, the calculation results demonstrate a pronounced asymmetry of the curve $\upsilon_{s\ell}(T_{s\ell})$ relative to the melting point T_m, i.e. between undercooling and overheating. The velocity profiles approach large unercooling and large superheating in different ways, since the solidification kinetics far from the melting temperature differs significantly from the melting kinetics in a very superheated state. The crystal growth rates in the melt are largely

determined by the structural order parameter, which measures the local degree of crystallinity. When crystals grow, each atom on the interface must go to a specific place in the lattice. As the melt grows, the crystal melts and turns into a liquid without any structural limitations for atoms passing from the interface to the melt. Thus, structurally, it is easier for a crystal to transform into a liquid than to melt into a crystal. These structural differences in the processes lead to the asymmetry observed in the crystal/melt growth rate. The structural factor also has a great influence on the velocity profiles when approaching severe overheating and undercooling. Near the limiting superheating, the metastability state reaches its maximum and upon further heating, the crystal becomes unstable due to homogeneous nucleation, which leads to the disappearance of the solid/liquid interface. With large undercoolings, the mobility of the melt decreases significantly, reducing the value of the interface velocity.

6 CONCLUSION

1. A modification of the transition state theory for melting/crystallization processes is proposed using the Wilson-Frenkel kinetic model as an example. The modification consists in replacing the constant coefficients in the rate of direct and reverse transitions with a functional dependence on the temperature of the solid/liquid interface $T_{s\ell}$.

2. Atomistic modeling of Si melting / crystallization processes under the conditions of deep overheating/undercooling was performed using two interaction potentials KIHS and SW.

3. From comparing the simulation results with the data of the modified kinetic model, the interface response function is constructed in the region of the maximum allowable values of superheating/undercooling in Si. The crystallization part of the interface response function, as well as the second part of the interface response function for melting, are diffusion-limited and are described by the same equation over the entire temperature range.

4. The temperature dependence of the speed of the solid/liquid interface determined from the simulation results using both interaction potentials shows a clear asymmetry with respect to the melting point T_m. This is explained by the strong difference between the solidification kinetics in a highly undercooled state and the melting kinetics in a very superheated state.

5 The change in the temperature dependence of the velocity $\upsilon_{s\ell}(T_{s\ell})$ upon passing through the melting point T_m occurs smoothly without a kink of the slope.

Acknowledgements: This study was supported by the Russian Science Foundation (project no. 18-11-00318).

REFERENCES

[1] J.W. Christian, *The Theory of Transformations in Metals and Alloys (v. 1-2)*, 1st Edition, Pergamon, (2002).

[2] D.V. Sivuxin, *Obshhij kurs fiziki. T. II. Termodinamika i molekulyarnaya fizika*, M.: Nauka, (1975).

[3] M.C. Flemings, “Solidification processing”, *Metallurgical Transactions*, **5**, 2121–2134 (1974). https://doi.org/10.1007/BF02643923

[4] Guo-Quan Lu, E. Nygren, M.J. Aziz, “Pressure-enhanced crystal/ation kinetics of amorphous Si and Ge: Implications for point-defect mechanisms”, *J. Appl. Phys.*, **70** (10), 5323 - 5345 (1991)

[5] M. Li, S. Ozawa, K. Kuribayashi, "On determining the phase-selection principle in solidification from undercooled melts-competitive nucleation or competitive growth?", *Philosophical Magazine Letters*, **84**(8), 483–493(2004). doi:10.1080/0950083042000271090
[6] B.J. Siwick, J.R. Dwyer, R.E. Jordan, R.J.D. Miller, "An Atomic-Level View of Melting Using Femtosecond Electron Diffraction", *Science*, **302**(5649), 1382–1385 (2003). doi:10.1126/science.1090052
[7] Q.S. Mei, K. Lu . "Melting and superheating of crystalline solids: From bulk to nanocrystals", *Progress in Materials Science*, **52**(8), 1175–1262 (2007)
[8] A.B. Belonoshko, N.V. Skorodumova, A. Rosengren, B. Johansson, "Melting and critical superheating", *Phys. Rev. B.*, **73,** 012201(1-3) 2006. doi: 10.1103/PhysRevB.73.012201
[9] N.A. Berjeza, S.P. Velikevitch, V.I. Mazhukin, I. Smurov, G. Flamant, "Influence of temperature gradient to solidification velocity ratio on the structure transformation in pulsed- and CW-laser surface treatment", *Applied Surface Science*, **86**(1-4), 303–309 (1995). doi:10.1016/0169-4332(94)00446-3
[10] M.J. Aziz, "Interface attachment kinetics in alloy solidification". *Metallurgical and Materials Transactions A* (*MMTA)*, **27,** 671–686 (1996). https://doi.org/10.1007/BF02648954
[11] J.Lipton, W.Kurz, R.Trivedi, "Rapid dendrite growth in undercooled alloys". *Acta Metallurgica*. **35** (4), 957-964 (1987). doi.org/10.1016/0001-6160(87)90174-X
[12] M. Asta, C. Beckermann, A. Karma, W. Kurz, R. Napolitano, M. Plappf, G. Purdy, M. Rappaz, R. Trivedi, "Solidification microstructures and solid-state parallels: Recent developments, future directions. Overview № 146", *Acta Materialia*, **57**(4), 941–971 (2009). https://doi.org/10.1016/j.actamat.2008.10.020
[13] J.J. Hoyt, M. Asta, A Karma. "Atomistic and continuum modeling of dendritic solidification". Materials Science and Engineering: R:Report, 41, 121–163 (2003). DOI:10.1016/S0927-796X(03)00036-6
[14] P. K. Galenko, D. V. Alexandrov. "From atomistic interfaces to dendritic patterns". *Phil. Trans. R. Soc. A.* **376:** 20170210, 1-9 (2018). http://doi.org/10.1098/rsta.2017.0210
[15] J.Y. Tsao, M.J. Aziz, M.O. Thompson, P.S. Peercy, "Asymmetric Melting and Freezing Kinetics in Silicon", *Physical Review Letters*, **56**(25), 2712–2715 (1986). doi:10.1103/physrevlett.56.2712
[16] *Advances in the Application of Lasers in Materials Science*, P.M. Ossi (ed.), Springer Series in Materials Science 274, Springer Nature Switzerland AG, (2018). https://doi.org/10.1007/978-3-319-96845-2_5
[17] J. Yan, P. Liu, Z. Lin, H. Wang, H. Chen, C. Wang, G. Yang, "Directional Fano Resonance in a Silicon Nanosphere Dimer", *ACS Nano*, **9**(3) 2968-2980 (2015).
[18] https://doi.org/10.1021/nn507148z
[19] J. Cheng, Ch. Liu, Sh. Shang, D. Liu, W. Perrie, G. Dearden, K. Watkins, "A review of ultrafast laser materials micromachining", *Optics & Laser Technology*, **46**, 88-102 (2013).
[20] doi:10.1016/j.optlastec.2012.06.037
[21] V.I. Mazhukin, M.M. Demin, A.V. Shapranov, "High-speed laser ablation of metal with pico- and subpicosecond pulses", *Applied Surface Science*, **302**, 6-10 (2014).
[22] http://dx.doi.org/10.1016/j.apsusc.2014.01.111
[23] M. Cesaria, A.P. Caricato, M. Beccaria, A. Perrone, M. Martino, A. Taurino, M. Catalano, V. Resta, A. Klini, F. Gontad, "Physical insight in the fluence-dependent distributions of Au nanoparticles produced by sub-picosecond UV pulsed laser ablation of a solid target in vacuum environment", *Applied Surface Science*, **480,** 330-340 (2019).
[24] https://doi.org/10.1016/j.apsusc.2019.02.022
[25] A. Mene´ndez-Manjo´n, S. Barcikowski, G.A. Shafeev, V.I. Mazhukin, B.N. Chichkov, "Influence of beam intensity profile on the aerodynamic particle size distributions generated by

femtosecond laser ablation", *Laser and Particle Beams*, **28**, 45-52 (2010). doi:10.1017/S0263034609990553

[26] J.H. Perepezko, G. Wilde, "Melt undercooling and nucleation kinetics", *Current Opinion in Solid State and Materials Science*. **20**(1), 3-12 (2016). https://doi.org/10.1016/j.cossms.2015.07.001

[27] V.I. Mazhukin, A.V. Shapranov, A.V. Mazhukin, O.N. Koroleva. "Mathematical formulation of a kinetic version of Stefan problem for heterogeneous melting/crystallization of metals", *Mathematica Montisnigri*, **36**, 58 - 77 (2016).

[28] Yu Chen Shen, D.W. Oxtoby. Density functional theory of crystal growth: Lennard-Jones fluids. J. Chem. Phys., **104** (11), 4233 - 4242 (1996). doi: 10.1063/1.471234

[29] M.I. Mendelev, M.J. Rahman, J.J. Hoyt, M. Asta, "Molecular-dynamics study of solid–liquid interface migration in fcc metals", *Modeling Simul. Mater. Sci. Eng.*, **18**, 074002(1-18) (2010). doi:10.1088/0965-0393/18/7/074002

[30] V.I. Mazhukin, A.V. Shapranov, M.M. Demin, N.A. Kozlovskaya, "Temperature dependence of the kinetics rate of the melting and crystallization of aluminum", *Bulletin of the Lebedev Physics Institute*, **43**(9), 283-286 (2016).

[31] A. Karma, D. Tourret, "Atomistic to continuum modeling of solidification microstructures", *Current Opinion in Solid State and Materials Science*, **20**(1), 25-36 (2016). doi:10.1016/j.cossms.2015.09.001

[32] V.I. Mazhukin, A.V. Shapranov, V.E. Perezhigin, O.N. Koroleva, A.V. Mazhukin, "Kinetic melting and crystallization stages of strongly superheated and supercooled metals", *Mathematical Models and Computer Simulations*, **9**(4), 448-456 (2017).

[33] M.D. Kluge, J.R. Ray, "Velocity versus temperature relation for solidification and melting of silicon: A molecular-dynamics study", *Phys. Rev. B*, **39** (3), 1738 -1746 (1989).

[34] C.J. Tymczak, J.R. Ray. Asymmetric Crystallization and Melting Kinetics in Sodium: A Molecular-Dynamics Study. Phys. Rev. Let., **64** (11), 1278 - 1281(1990).

[35] H.A. Wilson, "On the velocity of solidification and viscosity of supercooled liquids", *Philos. Mag.*, **50**, 238-250 (1900).

[36] Ja.I. Frenkel, "Note on the relation between the speed of crystallization and viscosity", *Phys. Z. Sowjet Union*, **1**, 498 - 499 (1932).

[37] K. A. Jackson, "The Interface Kinetics of Crystal Growth Processes". *Interface Science*, **10**(2/3), 159–169 (2002). doi:10.1023/a:1015824230008

[38] J.Q. Broughton, G.H. Gilmer, K.A. Jackson, "Crystallization Rates of a Lennard-Jones Liquid", *Phys. Rev. Let.*, **49**, 1496 -1500 (1982).

[39] D. Turnbull, "On the relation between crystallization rate and liquid structure". J. Phys. Chem., **62(**4), 609 – 613 (1962).

[40] J.E. Jones, "On the determination of molecular fields. II. From the equation of state of a gas". *Proceedings of the Royal Society A*, **106** (738), 463 – 477 (1924).

[41] L.V. Mikheev, A.A. Chernov, "Mobility of a diffuse simple crystal—melt interface". J. *Crystal Growth*, **112** (2–3), 591-596 (1991). https://doi.org/10.1016/0022-0248(91)90340-B

[42] F.H. Stillinger, T.A. Weber, **"**Computer simulation of local order in condensed phases of silicon". *Phys. Rev.B*, **31**(8), 5262-5271 **(**1985**)**.

[43] T. Kumagai, S. Izumi, S. Hara, S. Sakai, "Development of bond-order potentials that can reproduce the elastic constants and melting point of silicon for classical molecular dynamics simulation", *Comp. Mater. Sci.*, **39**(2), 457-464 (2007).

[44] K.A. Jackson, B. Chalmers, "Kinetics of solidification", *Can. J. Phys.*, **34**, 473 – 490 (1956).

[45] A.A. Samarskii, F.I. Gulin, *Chislennye metody*, M.: Fizmatlit, (1989)

[46] O.N. Koroleva, A.V. Mazhukin, V.I. Mazhukin, "Modeling of silicon characteristics in the semiconductor-metal phase transition region", *Mathematica Montisnigri*, **41**, 73-90 (2018).

[47] A.V. Mazhukin, O.N. Koroleva, V.I. Mazhukin, A.V.Shapranov, "Continual and molecular dynamics approaches in determining thermal properties of silicon", *Proceedings of SPIE V.*, 10453(1-11) (2017).

[48] V.M. Glazov, O.D. Shchelikov, "Volume Changes during Melting and Heating of Silicon and Germanium Melts", *High Temperature*, **38** (3), 405-41 (2000).

Received November 17, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

APPLICATION OF THE CLOSEST POINT PROJECTION METHOD TO SOLUTION OF REYNOLD'S LUBRICATION EQUATIONS ON EVOLVING SURFACES

E. ZIPUNOVA, A. IVANOV, E. SAVENKOV*

Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia

*Corresponding author. E-mail: e.savenkov@googlemail.com

DOI: 10.20948/mathmontis-2020-47-9

Summary. The paper considers a computational algorithm for solution of surface PDEs defined on the evolving surfaces. The basis of the algorithm is a finite element version of the closest point projection method. The method is applied to the Reynold's lubrication equation which governs fluid flow in thin fractures. The closest point approach is used for description of fracture mid-surface evolution and for construction of the embedding equation. We describe algorithmic details of the proposed approach as well as a number of numerical experiments which demonstrates robustness of the method.

1 INTRODUCTION

Currently, hydraulic fracturing (HF) is one of the most widely used methods of oil and gas reservoir stimulation. The essence of the technology is an injection of special fluid into reservoir in order to create an artificial fracture of considerable area (length ~ 100 m, height ~ 10 m, average opening $\sim 5-10$ mm). The fracture is filled with a proppant – calibrated artificial or natural sand-like granular material. The result of HF procedure is an artificial flow channel connected to the production or injection well with large inflow area and high permeability. This provides a significant increase in inflow of reservoir fluid to the well. Engineering aspects of technology are considered, for example, in [1].

Mathematical description of the hydraulic fracture evolution during its development comes down to solution of complex coupled problem which includes (among other groups of equations) flow equations of (usually non-Newtonian) for fluid in the evolving fracture.

A number of models for HF evolution are known. Most general of them (see, e.g., [2]) assume that:

- fracture mid-surface is an arbitrary sufficiently smooth surface with boundary;
- at a fixed point of the mid-surface its opening is defined by reaction of the surrounding medium and pressure in the fracture;
- during the hydraulic fracturing procedure, the fracture evolves, and the exact way of this evolution is not known in advance (in other words, the fracture mid-surface is a part of the solution of complete problem);
- the fracture mid-surface is not flat – fracture can switch direction of propagation locally; the direction of its propagation may be different at different points on its front.

Note that common approaches to solve the problem of fracture propagation in an elastic medium are based on boundary integral equations. In this case it turns possible to solve both

2010 Mathematics Subject Classification: 74H05, 65K05, 74H35
Keywords and Phrases: Closest point projection method, Finite element method, hydraulic fracturing, Reynold's lubricantion layer equations

fracture flow equations and elasticity equations in the surrounding medium using single surface computational mesh defined on the fracture mid-surface. This simplifies overall algorithm for the complete problem. In the case when the medium is, for example, heterogeneous, these methods cannot be applied – and the mesh-based methods like, e.g. finite element method are more preferable. In that case an additional computational mesh has to be introduced for solution of surface PDE. The same situation arises when realistic fluid flow models in reservoir has to be considered, see, e.g. [3].

Another approach is to use the so called embedding methods when surface PDE is solved using the same spatial mesh as a problem in the surrounding medium. A number of such methods are known. In the works [4, 5, 6, 7, 8] an original method was proposed for solving surface PDEs based on surface representation using closet point projection operator. The essence of the method is that closest point projection operator is used to extend surface PDE into the surrounding space to obtain the so called embedding PDE. Further, this embedding PDE is approximated by a suitable difference method on a mesh which, generally, is not consistent with surface geometry. As a solution to the original problem on the surface, the trace of 3D embedding equation is considered. At the same time, the closet point projection operator is used to approximate Dirichlet (or Neumann) boundary conditions defined on the boundary of both the original surface and the three-dimensional domain, where embedding PDE is solved.

In this paper, we consider the finite element version of the closet point projection method to solve Reynold's lubrication equations which governs fluid flow inside evolving fracture.

In contrast to the works cited above, the finite element method is used a basic approximation scheme. More detailed description of the algorithm as well as numerical examples are presented in [19, 20].

2 CLOSEST POINT PROJECTION METHOD

This section briefly describes the main ideas of the closest point projection method for numerical solution of surface PDEs. The method was proposed and developed in [4, 5, 8, 6, 7]. It uses implicit representation of the surface and is based on the extension of surface PDE into the space. To construct such extension the closest point projection operator is used rather then commonly used level set method.

In this section, for simplicity, we consider model boundary value problem for a parabolic equation with Laplace–Beltrami operator (see, for example, [21]) defined on a curved surface F with boundary:

$$\frac{\partial u}{\partial t} - \Delta_{\mathrm{F}}\, u = 0, \tag{1}$$

complemented by the initial and boundary condition of the desired type.

We assume that the surface F is entirely located inside the spatial domain $\Omega \subset \mathrm{R}^3$. Suppose that for an arbitrary point $\mathbf{x} \in \Omega$, a point $\mathbf{x}_{\mathrm{cp}}$ is the nearest to x point on the surface F,

$$\mathbf{x}_{\mathrm{cp}} = argmin_{\mathbf{y}\in\mathrm{F}}\ \mathrm{P}\,\mathbf{y} - \mathbf{x}\,\mathrm{P},$$

where $\mathrm{P}\cdot\mathrm{P}$ is the Euclidean norm in R^3. The point $\mathbf{x}_{\mathrm{cp}}$ is called the closest point projection of the point $\mathbf{x}$ onto the surface F, and the corresponding operator will be denoted by P,

$$\mathbf{x}_{\mathrm{cp}} = \mathrm{P}\mathbf{x}.$$

The operator P is vector-valued: it maps spatial domain Ω to the surface F, considered as a subset of $\Omega \subset \mathrm{R}^3$.

If the sign distance function $d_{\mathrm{F}}(\mathbf{x})$ can be specified for the surface F (for example, if F is oriented surface without boundary), then the following representaion for the operator P is valid:

$$\mathrm{P}(\mathbf{x}) = \mathbf{x} - d_{\mathrm{F}}(\mathbf{x})\nabla d_{\mathrm{F}}(\mathbf{x}), \quad d(\mathbf{x}) = \mathrm{P}\mathbf{x} - \mathrm{P}\mathbf{x}\,\mathrm{P}.$$

Just like the sign distance function (or a pair of such functions in the case of a surface with boundary), the projector P uniquely defines surface F,

$$\mathrm{F} = \{\mathbf{x} \in \Omega : \quad \mathbf{x} = \mathrm{P}\mathbf{x}\}.$$

However, the closest point projection approach is more general: it allows one to describe geometry of surface with boundary, non-orientable manifolds or manifolds of codimension greater than one (i.e. curves and points which are codimension 2 and 3 objects in 3d case – as well as union of objects of different codimension).

Using projector P, it is easy to construct an extension of an arbitrary function defined on the surface to the spatial domain Ω: for an arbitrary function u defined on a surface, its continuation $\mathrm{E}[u]$ in Ω is defined as

$$\mathrm{E}[u](\mathbf{x}) = u(\mathrm{P}\mathbf{x}), \quad \mathbf{x} \in \Omega.$$

In both cases, the operator E is a projector in the sense that $\mathrm{E}^2 = I$, where I is the identity operator. Note that:

- for an arbitrary function in Ω which is constant along the direction normal to F,

$$(\nabla u)|_{\mathrm{F}} = \nabla_{\mathrm{F}}\left(u|_{\mathrm{F}}\right);$$

- for an arbitrary vector field in Ω tangent to the surface F,

$$(\nabla\cdot\mathbf{q})|_{\mathrm{F}} = \nabla_{\mathrm{F}}\cdot\left(q|_{\mathrm{F}}\right).$$

Then, due to the properties of the projector P and the extension operator E:

$$\nabla\mathrm{E}[u](\mathbf{x}) = \nabla u(\mathrm{P}\mathbf{x}) = \nabla_{\mathrm{F}}\, u.$$

Since the $\mathrm{E}[u](\mathbf{x})$ is constant along directions normal to the surface, the vector field $\nabla\mathrm{E}[u](\mathbf{x})$ is tangent to F. Hence,

$$\nabla\cdot\left[\nabla\mathrm{E}[u](\mathbf{x})\right] = \nabla\cdot\left[\nabla u(\mathrm{P}\mathbf{x})\right] = \nabla_{\mathrm{F}}\cdot\nabla_{\mathrm{F}}\, u.$$

Similar extensions can be constructed for more complex elliptic operators defined in the surface, see [9].

Thus, the original equation (1) can be extended to the spatial domain Ω to define the so called embedding equation. Further, the embedding equation is approximated by a suitable difference method on a three-dimensional mesh introduced in Ω and, generally, inconsistent with the surface geometry. Solution of the original problem on the surface can restored as a trace of the solution of 3D problem on the surface. A rigorous justification for the constructions described above is presented in [9]. Details of the method are given in the papers cited above.

Following the described approach the embedding equation which corresponds to (1) reads:

$$\frac{\partial u}{\partial t} + \nabla \cdot (\nabla \mathrm{E}[u]) = \mathrm{E}[f], \quad \mathbf{x} \in \Omega_{\mathrm{F}} . \tag{2}$$

The trace of the solution of this equation on F is the solution of the equation (1).

Regarding domain Ω, it is assumed that (i) it is small in the sense that the closest point projection of an arbitrary point from Ω is well defined and (ii) the domain Ω includes the surface F, $\Omega \supset \mathrm{F}$, and the distances from the boundary points of Ω to surface are positive. In other words, all points on the surface are interior points of the domain.

If the surface F is a surface without boundary, the value of the solution $u(t,\mathbf{x})$ of the embedding PDE (2) at the points lying on the surface F will coincide with the solution of the original surface PDE (1) (in this case $\partial\mathrm{F} = \varnothing$ and the problem (1) is the Cauchy problem). Otherwise it is required to provide embedding PDE (2) with boundary conditions defined in a suitable way. This can be done in different ways. In [8] a convenient way to adress this issue in the discrete setting. In the context of the finite element method the corresponding questions were adressed in [10].

3 PROBLEM STATEMENT

This section presents the mathematical formulation of the problem. In subsequent section with consider a flow model itself and geometrical model for fracture mid-surface evolution.

3.1 Flow model in fracture

Consider a one-sided surface $\overline{\mathrm{F}}$ with boundary $\partial\mathrm{F}$ immersed into three-dimensional space R^3. Let $\overline{\mathrm{F}} = \mathrm{F} \cup \partial\mathrm{F}$, where an open domain F is an internal part of the surface, $\partial\mathrm{F}$ – its boundary. In some cases, we will identify the notation F and $\overline{\mathrm{F}}$. We assume that F and its boundary have the required smoothness.

The surface $\overline{\mathrm{F}}$ is assumed to be immersed into spatial domain Ω. Let for each point $\mathbf{x} \in \Omega$ a closest point projector P onto $\overline{\mathrm{F}}$ is uniquelly defined (see section 2). Depending on the location of the point, its projection onto $\overline{\mathrm{F}}$ belongs to either F or $\partial\mathrm{F}$. This allows us to represent Ω as a union $\Omega = \Omega_{\mathrm{F}} \cup \Omega_{\partial\mathrm{F}}$, where

$$\Omega_{\mathrm{F}} = \{\mathbf{x} \in \Omega : \mathrm{P}\mathbf{x} \in \mathrm{F}\}, \quad \Omega_{\partial\mathrm{F}} = \{\mathbf{x} \in \Omega : \mathrm{P}\mathbf{x} \in \partial\mathrm{F}\}.$$

The boundary $\partial\Omega_{\mathrm{F}}$ of the Ω_{F} can be represented as

$$\partial\Omega_{\mathrm{F}} = \Gamma_{\mathrm{F}} \cup \Gamma_{\partial\mathrm{F}}, \quad \Gamma_{\mathrm{F}} = \partial\Omega_{\mathrm{F}} \cap \partial\Omega, \quad \Gamma_{\partial\mathrm{F}} = \partial\Omega_{\mathrm{F}} \setminus \Gamma_{\mathrm{F}} .$$

In the case of evolving fracture geometry, the constructions above are generally the same, except that the computational domain now depends on time. That is, the problem is solved in the domain $\mathrm{F} = \mathrm{F}_t \equiv \mathrm{F}(t)$, where t is the time, $\Omega = \Omega_t \equiv \Omega(t)$, $\Omega_{\mathrm{F}} = \Omega_{\mathrm{F},t} \equiv \Omega_{\mathrm{F}}(t)$, *etc*. Additionally, we assume that for all moments of time, all previously considered time-dependent domains, as well as the mid-surface of the fracture are located inside some spatial and *time-independent* domain $\tilde{\Omega}$. Let us note that the dependency of the mid-surface F on time cannot to be an arbitrary. Corresponding clarifications will be given below.

The Reynold's lubrication equation of fluid flow in fracture has the form:

$$\frac{\partial \rho w}{\partial t} + div\left(-\frac{1}{12\nu}\rho w^3 \nabla p\right) = f_{\mathrm{m}}, \quad \mathbf{x} \in \mathrm{F}_t, \tag{3}$$

where ρ is the fluid density; $w = w(\mathbf{x})$ – fracture opening, $\mathbf{x} \in \mathrm{F}$ – point of the fracture mid-surface F; ν – fluid viscosity, $p = p(\mathbf{x})$ – its pressure, f_{m} – mass rate of external sources.

We assume that the fluid density linearly depends on pressure, i.e.

$$\rho = \rho_0[1 + c_{\mathrm{f}}(p - p_0)],$$

where ρ_0 and p_0 are reference values of density and pressure, c_{f} is compressibility of the fluid.

The equation (3) is supplemented with initial and boundary condition of the form

$$p(\mathbf{x}, t = 0) = p_{\mathrm{ini}}(\mathbf{x}), \quad p|_{\partial \mathrm{F}} = p_{\partial \mathrm{F}}.$$

Further, we assume that the opening w is always positive,

$$w = w(\mathbf{x}) .. w_0 > 0, \quad \mathbf{x} \in \mathrm{F} \cup \partial \mathrm{F}.$$

In the simplest case we can assume that fracture opening w does not depend on pressure. In a more realistic setting, opening is a function of pressure. This relationship reflects the fact that the fracture is located inside an elastic medium that deforms when the pressure in the fracture changes. In this paper it is assumed that this dependency is linear, i.e.

$$w(\mathbf{x}) = \mathrm{W}(\mathbf{x}; p(\mathbf{x})), \quad \mathrm{W}(\mathbf{x}; p) = w_{\mathrm{ref}}(\mathbf{x})\left(1 + c_{\mathrm{w}}[p(\mathbf{x}) - p_{\mathrm{ref}}]\right), \tag{4}$$

where $w_{\mathrm{ref}}(\mathbf{x})$ is fracture opening at reference pressure p_{ref}, c_{w} is coefficient describing "compressibility" of fracture (more precisely, the medium containing the fracture).

3.2. Model for surface evolution

Evolution of the surface F_t in the formulation (3) can not be an arbitrary. This section presents relevant assumptions, both of the fundamental and technical character. They reflect the specifics of the problem under consideration, namely, the fact that F_t is the mid-surface of an evolving fracture.

Let the problem (3) be solved on the time interval $t \in [0, T]$ and for $t_1 .. t_2$ condition $\mathrm{F}_{t_1} \subset \mathrm{F}_{t_2}$ holds. In other words, a family of surfaces corresponding to smaller times, is

contained inside a surface corresponding to any larger time. Or, which is the same, the surface may evolve only due to the movement of its boundary.

We assume that:

- at any time $t \in [0,T]$ the surface F_t is entirely located inside some spatial time-independent domain Ω;
- the evolution of the surface F is "smooth", i.e., at any time the surface can be smoothly and one-to-one mapped onto, for example, a unit disk in R^2. In particular, during its evolution F should not have self-intersections, and so on.

In this case, the family of surfaces $\{\mathrm{F}_t, t..0\}$ can be represented as a union of the "initial" surface $\mathrm{F} \equiv \mathrm{F}_0$ and the "trace" of the movement of the surface boundary, i.e.

$$\mathrm{F}_t = \{\mathrm{F}_0 \cup \gamma(t')|0,, t',, t\},$$

where $\mathrm{F} \equiv \mathrm{F}_0$ is the surface at $t = 0$, $\gamma(t) \equiv \partial \mathrm{F}_t$.

Thus, the evolution of the surface is determined by the motion of its boundary $\gamma(t)$. In the following we assume that at every moment of time, on the curve $\gamma(t)$ the velocity vector field $\mathbf{v} = \mathbf{v}(\mathbf{x},t)$, $\mathbf{x} \in \gamma(t)$ is defined which governs fracture evolution. Movement of the (Lagrangian) point of the boundary is described by the equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{v}(\mathbf{x},t), \quad \mathbf{x}|_{t=0} = \mathbf{x}_0 \in \gamma(0). \tag{5}$$

We assume that $\mathbf{v} = \mathbf{v}(\mathbf{x},t)$ (i) is smooth function of point $\mathbf{x} \in \gamma(t)$ for $t \in [0,T]$ and (ii) is a smooth function of time for each fixed (Lagrangian) point on the boundary.

Such a model of surface evolution corresponds to the problem of fracture dynamics. In this case the direction of surface evolution is known only at the points of its boundary (fracture front) and is determined by the appropriate fracture criteria (see e.g. [11, 12]).

Note that in a number of cases (in particular, for the purposes of theoretical analysis) it is convenient to assume that the velocity field $\mathbf{v}(\mathbf{x},t)$, $\mathbf{x} \in \partial \mathrm{F}_t$ is a trace on $\partial \mathrm{F}_t$ of some smooth vector field $\mathbf{V} = \mathbf{V}(\mathbf{x},t)$ without singular points which is defined in the domain Ω containing the family of surfaces F_t at all instants of time $t \in [0,T]$. The natural requirement for the field $\mathbf{V} = \mathbf{V}(\mathbf{x},t)$ is that that it is tangent to surfaces $\mathrm{F}_{\tilde{t}}$ for all $\tilde{t} < t$ (in other words, it does not change already formed fracture mid-surface). Under a suitable choice, the field $\mathbf{V}$ generates a smooth and one-to-one mapping of the surface $\mathrm{F} \equiv \mathrm{F}_0$ onto the surface F_t for any $t \in (0,T]$.

4 COMPUTATIONAL ALGORITHM

4.1. Formal time approximations

We write the problem (3) in the operator form as

$$\frac{\partial \varphi(u)}{\partial t} + \mathrm{A}(w(u);u) = f, \quad \mathbf{x} \in \mathrm{F}_t, \qquad u|_{\partial \mathrm{F}} = g, \quad \mathbf{x} \in \partial \mathrm{F}_t, \tag{6a}$$

with initial condition

$$u(\mathbf{x}, t = 0) = u_0(\mathbf{x}), \quad \mathbf{x} \in \mathrm{F}_0.$$

Let the problem (6a) be solved for $t \in [0,T]$. We divide interval $[0,T]$ it into intervals (time steps) Δt, so that

$$0 = t_0 < \Delta t < \ldots < n\Delta t = t_n < \ldots N\Delta t = t_N = T.$$

According to section 3.2 at time point t the solution $u_n = u(t_n)$ is defined in the domain $\mathrm{F}_{t_n} \equiv \mathrm{F}_n$, moreover, $\mathrm{F}_n \subset \mathrm{F}_{n+1}$.

Semi-discrete (in time) approximation of the problem (6a) on the interval $[t_n, t_{n+1}]$ can be defined as

$$\frac{\varphi(u_{n+1}) - \mathrm{E}[\varphi(u_n)]}{\Delta t} + \mathrm{A}(w(u_{n+1});u_{n+1}) = f_{n+1}, \quad \mathbf{x} \in \mathrm{F}_{n+1}. \tag{6b}$$

Here $\mathrm{E}[\cdot]$ is the continuation operator that maps functions, defined in F_n to functions defined in F_{n+1}. For the problem to be well-posed, the continuation operator must satisfy certain smoothness and boundedness properties. In terms of Sobolev spaces, such properties of the continuation operator are formulated in [13].

In this paper, a rigorous theoretical justification of the proposed method used is not given, but it can be expected that the conditions necessary for its correctness are satisfied since:

- for smooth domains with smooth boundary natural smoothness properties of the solution coincide with those ones for the case of planar domains (subsets of R^2). Moreover, the surface F_t for all t can be mapped smoothly to the two-dimensional domain $\tilde{\mathrm{F}}_t \subset \mathrm{R}^2$ (and even to a time-independent canonical domain, for example, a disk of a given radius)
- the vector field $\mathbf{V}$ that governs mid-surface evolution (see section 3.2) defines a smooth mapping $\mathrm{F} \equiv \mathrm{F}_0$ onto F_t.

We also note the following. In the works [13] and [14] the continuation operator is not used explicitly. Instead, an implicit way to prolongate solution from domain Ω_n to the domain Ω_{n+1} is used. Technically it is implemented by adding to the weak statement of the problem an additional penalty-type bilinear form. As a result, the required smoothness of the discrete extension is a consequence of the modified variational statement of the problem. This method is called “ghost penalty stabilization”. It can directly be used to solve the problem considered in this paper.

Below we use an algorithm, that assumes explicit definition of continuation $\mathrm{E}[u]$ of the solution u. Namely, the solution defined in the region F_n is extended to the domain F_{n+1} as a constant in the direction normal to the boundary of the domain F_n.

Alternatively, a continuation approach typical for the X-FEM method can be used (see, for example [15, 16]). Essentially it consist of two steps:

1. First, the vector field $\mathbf{v}$ defined at the boundary $\partial\mathrm{F}_t$ is extended into tubular neighborhood of $\partial\mathrm{F}_t$ (or the entire three-dimensional domain containing the surface);
2. Second, the extended velocity field is used to extend solution from the domain F_n into an "extended'' domain F_{n+1} by solving appropriate Hamilton-Jacoby type equation.

At each time step, the problem (6b) is solved on the surface F_{n+1}. A number of methods can be applied to do it, including variational version of the closest point projection method (see [10]).

In what follows, we assume that during its evolution, the surface F_t (and surfaces F_n, $n=\overline{0,N}$) are always located inside spatial domain $\tilde{\Omega}$. To proceed with spatial approximations, we assume that the finite element mesh T_h is introduced in Ω. This mesh is used to build spatial approximations of the problem (6b) for $n=\overline{1,N}$.

At each moment in time, a three-dimensional domain Ω_n^h is associated with the surface F_n in which solution to the problem (6b) will be approximated by the closest point projection method. Geometrically Ω_n^h can be constructed as a set of finite elements $\omega \in \mathrm{T}_h$, all nodes of which are distant from the surface F_n by a distance not exceeding the value of the given parameter δ, which is a multiple of the step of the computational mesh. Everywhere in the domain Ω_n^h we will assume that closest point projector P_n (the index "n'' indicates the number of time step) is well defined.

Afterwords to solve the problem in the domain Ω_{n+1}, algorithms from [17, 10, 20] can be used directly.

The sketch of the computational algorithm is as follows:

1. Initialize surface $\mathrm{F} \equiv \mathrm{F}_0$, velocity field $\mathbf{v}(\mathbf{x})$ defined at $\mathbf{x} \in \partial\mathrm{F}$; initial condition $u_0(\mathbf{x})$, $\mathbf{x} \in \mathrm{F}$.
2. Set domain Ω and the computational mesh T_h in it.
3. Set $n=0$, $t=0$.
4. Compute discrete approximations of the closest point projector P_0^h to the surface F_0..
5. For $n=\overline{1,N}$:
    a) Set $n := n+1$.
    b) Define mid-surface boundary evolution velocity $\mathbf{v}$.

c) Setup mesh domain Ω_n^h as a correct subset of the finite elements in T_h.
d) Based on the given field $\mathbf{v}$ and the mesh projector P_{n-1}^h determine discrete domain Ω_n^h and the corresponding closest point projector P_n^h. see [18].
e) Compute the extention $\mathrm{E}u_{n-1}$ of the solution from the domain Ω_{n-1}^h to the domain Ω_n^h.
f) Assemble and solve a finite-dimensional approximation of the problem (6b).
g) Go to step 5a.

6. Terminate algorithm.

Note that during the operation of the algorithm, the surface is specified directly only at its first initialization step. In the further steps, only the evolution over time of the projector P_n^h is computed. The image of this projector is an approximation of the surface; Its geometric characteristics, if necessary, can be calculated according to the algorithms described in [18].

To solve discrete non-linear problem at step (5f) of the algorithm, fixed point iterations are used.

As applied to time-implicit approximations of the problem (3) this algorithm, describing transition from the current state $\{p, \rho, w\}$ at the time t to the state $\{\hat{p}, \hat{\rho}, \hat{w}\}$ at the time $t + \Delta t$ has the following form:

1. For $i = 0$ set $\hat{p}_i = p$.
2. Compute $\hat{w}_{i+1}$ as

$$\hat{w}_{i+1}(\mathbf{x}) = \mathrm{W}(\mathbf{x}; \hat{p}_i), \quad \hat{\rho}_{i+1}(\mathbf{x}) = \rho(\hat{p}_i(\mathbf{x})).$$

3. Compute $\hat{p}_{i+1}$ as solution of the equation:

$$\frac{1}{\Delta t}(\rho(\hat{p}_{i+1})\hat{w}_{i+1} - \rho w) + div\left(-\frac{1}{12\nu}\hat{\rho}_{i+1}\hat{w}_{i+1}^3 \nabla \hat{p}_{i+1}\right) = f. \tag{7}$$

4. Check iterations stopping criteria. If the required accuracy is not achieved, set $i := i + 1$ and go to step 2. If the required accuracy is achieved, set

$$\{\hat{p}, \hat{\rho}, \hat{w}\} := \{\hat{p}_{i+1}, \rho(\hat{p}_{i+1}), \mathrm{W}(\hat{p}_{i+1})\}$$

It is assumed that the solution is obtained with the required accuracy, if the uniform norm of pressure increment is less then the given threshold value $\varepsilon_{\text{iter}}$:

$$\left\| \frac{p_i^{n+1} - p_{i-1}^{n+1}}{p_0} \right\|_\infty < \varepsilon_{\text{iter}}. \tag{8}$$

4.2 Spatial approximations with closest point projection method

To simplify notations let us write down equation (7) as

$$\frac{1}{\Delta t} M(u-\tilde{u}) + \nabla \cdot (A(\mathbf{x})\nabla u) = f, \quad \mathbf{x} \in \mathrm{F}. \tag{9}$$

Following [8] consider continuation of the equation (9) to the spatial domain $\mathbf{\Omega}_{\mathrm{F}}$:

$$\frac{1}{\Delta t} M(u-\tilde{u}) + \nabla \cdot (\mathrm{E}[A](\mathbf{x})\nabla \mathrm{E}[u]) = \mathrm{E}[f], \quad \mathbf{x} \in \mathrm{F}. \tag{10}$$

The trace of the solution of the embedded equation (10) on F is the solution of the equation (9).

Boundary conditions for the embedding equation (10) on $\Gamma_{\partial \mathrm{F}}$ (see section 2) can be constructed as an extension of the surface boundary conditions of the original problem (3) defined at $\partial \mathrm{F}$:

$$u|_{\Gamma_{\partial \mathrm{F}}} = \mathrm{E}[u|_{\partial \mathrm{F}}] = \mathrm{E}[g], \tag{11}$$

or, which is the same,

$$u(\mathbf{x}) = g(\mathrm{P}\mathbf{x}) = g(\mathbf{x}_{\mathrm{cp}}), \quad \mathbf{x} \in \Gamma_{\partial \mathrm{F}}, \ \mathbf{x}_{\mathrm{cp}} = \mathrm{P}\mathbf{x} \in \partial \mathrm{F}.$$

Note that the values of the solution of (10) on Γ_{F} are completely defined as an extension of u from interior points of the surface F . For this reason, the boundary conditions for (10) are not defined on this part of the boundary.

Let us briefly describe now the spatial approximations. Let $\mathrm{T}_h = \mathrm{T}_h(\Omega)$ be a partition of Ω into finite elements (tetrahedrons) ω,

$$\Omega_h = \bigcup_{\omega \in \mathrm{T}_h} \omega = \Omega,$$

that is, domain itself and its boundary are approximated exactly. Let N be a set of triangulation nodes, equipped with simplest continuous piecewise linear basis functions $\varphi_i = \varphi_i(\mathbf{x})$, $i \in \mathrm{N}$. The finite-dimensional space $V_h(\Omega) \subset V(\Omega)$ can be defined as

$$V_h(\Omega) = \underset{i \in \mathrm{N}}{span}\, \varphi_i(\mathbf{x}). \tag{12}$$

Recall that the boundary Γ of the domain Ω is approximated by the computational mesh exactly. Let $\mathrm{N} = \mathrm{N}_\Omega \cup \mathrm{N}_\Gamma$, N_Ω and N_Γ be sets of nodes inside domain Ω and on its boundary Γ, respectively. Then elements of the space $V_h = V_h(\Omega)$ have the form

$$v_h(\mathbf{x}) = \sum_{i \in \mathrm{N}_\Omega \cup \mathrm{N}_\Gamma} v_i \varphi_i(\mathbf{x}).$$

To proceed we construct discrete approximation P_h of the closest point projection operator P first. The details are covered in [18]) and are not considered here. Let $\mathbf{x}_i$ be a mesh node in Ω_h, $\mathbf{x}_i^{\mathrm{cp}} = \mathrm{P}_h \mathbf{x}_i$ be its projection, ω_i be a finite element such that $\mathbf{x}_i^{\mathrm{cp}} \in \omega_i$. Discrete extension operator defines the value of the (extended) function in the node $\mathbf{x}_i$ as $u_h(\mathrm{P}_h \mathbf{x}_i)$. This value

can be computes using linear interpolation over the values u_k, $k \in \omega_i$ of u_h over nodes ω_i. As a result, the value of u_h^{ext} can be defined as

$$u_h^{\text{ext}} = \mathbf{E}_h \mathbf{u}_h,$$

where $\mathbf{E}_h$ is a square matrix which nonzero rows are of the form

$$\mathbf{E}_h(i,[k_1,k_2,k_3,k_4]) = [\xi_1(\omega_i),\xi_2(\omega_i),\xi_3(\omega_i),\xi_4(\omega_i)].$$

Here $\xi_\alpha(\omega_i)$ are barycentric coordinates of the point $\mathrm{P}(\mathbf{x}_i)$ inside finite element ω_i.

Provided with discrete extension operator, the finite element approximations to (10) (in a case of homogeneous Neumann boundary conditions) reads:

$$\mathbf{M}_h \frac{\hat{\mathbf{u}}_h - \mathbf{u}_h}{\Delta t} = \mathbf{A}_h \mathbf{E}_h \mathbf{u}_h + \mathbf{E}_h \mathbf{f}_h,$$

where $\mathbf{M}_h$ is the mass matrix, $\mathbf{u}_h = \mathbf{u}_h(t)$ is a vector of degrees of freedom of the solution at time $t = t_n$; $\hat{\mathbf{u}}_h = \mathbf{u}_h(t+\Delta t)$ is vector of degrees of freedom of the solution at time $t = t_{n+1} = t + \Delta t$, $\mathbf{f}_h$ is the right side of the finite-dimensional problem, $\mathbf{A}$ – its stiffness matrix,

$$[\mathbf{M}_h]_{ij} = \int_\Omega M \phi_i \phi_j \, d\Omega, \quad [\mathbf{A}_h]_{ij} = \int_\Omega \mathrm{E}[A]\phi_i \cdot \nabla \phi_j \, d\Omega, \quad [\mathbf{f}_h]_i = \int_\Omega f \phi_i \, d\Omega, \quad i, j \in \mathrm{N}.$$

The resulting system of linear algebraic equations has the form:

$$(\mathbf{M}_h - \Delta t \mathbf{B}_h)\hat{\mathbf{u}}_h = \Delta t \mathbf{E}_h \mathbf{f}_h + \mathbf{M}_h \mathbf{u}_h.$$

As it is shown in [7], the constructed finite-dimensional problem can be unstable since matrix $\mathbf{B} = \mathbf{AE}$ may have eigenvalues with both positive or negative real part. The same work suggests modification that eliminates this effect. The solution is to modify $\mathbf{B}_h$ according to

$$\tilde{\mathbf{B}}_h = diag\mathbf{A}_h + (\mathbf{A}_h - diag\mathbf{A}_h)\mathbf{E}_h.$$

The same modification is used in this work.

A number of approaches can be used to account for boundary conditions in the discrete setting. In this paper, we use the simplest one: the values of the solution in the nodes of the computational mesh, which are projected onto the surface boundary, are assigned according to the given boundary values. Technically, this is done in a standard way, by modification of the appropriate rows of the system of equations above.

The described algorithm is applied at each step of the fixed point iterations to solve the equation (7).

5 SIMULATION RESULTS

5.1 Geometric constructions

This section provides an illustrations to the basic geometrical constructions needed for application of the closest point projection method (see section 2) using simple example.

Generally, an algorithmic steps used in closest point projection method are as follows. First, the spatial domain Ω is defined. The somain contains surface F with boundary $\partial\mathrm{F}$ such that $\bar{\mathrm{F}} = \mathrm{F} \cup \partial\mathrm{F} \subset \Omega$. We consider here the stationary case, i.e., the fracture mid-surface is assumed to be fixed in time and space. Therefore, in this section domains Ω and $\tilde{\Omega}$ are treated equally.

Hereafter in this work the domain Ω is asumed to be cube-shaped with an edge of the length l. In this domain a tetrahedral mesh is introduced which is constructed as follows: (i) a uniform mesh of $N_x \times N_y \times N_z$ smaller cubes in Ω is introduced; (ii) each small cube is further divided into five tetrahedra. Those tetrahedrons forms a computational mesh in Ω. In the following it is always assumed $N_x = N_y = N_z = N$. The mesh step size is defined as $h = l / N$. Note that it is not assumed *a-priory* that the surface F and the constructed mesh are consistent (i.e., the surface F is not represented as a union of the mesh faces).

The constructed meshed domain Ω^h is not used in the calculations directly, but is used to build a mesh Ω_{F}^h, in which solutions to the extended (prolongated) equation is to be found.

In accordance with the approach under consideration, the problem (surface PDE) is solved in the meshed domain Ω_{F}^h, which is the "submesh" of the Ω^h so that the distance (in the sense of closest point projection) from all nodes of the mesh in Ω_{F}^h to the surface does not exceed a given value δ. In all calculations below this value is selected as $\delta = N_\delta h$, where $N_\delta = 4$. Further in this section, the domains Ω and Ω^h (as well as Ω_{F}^h and Ω_{F}) are identified and the superscript "h" is dropped, if it does not lead to any misunderstanding.

For the example discussed in this section the surface F has the form of a flat disk of the diameter $L = 1$. The disk is located in the plane $\mathrm{O}xy$ and is centered at the origin. Domain Ω containing the disk is a cube with edge length $l = 1.5$.

On each edge of the cube Ω, a $N = 31$ nodes are defined. Thus, the characteristic mesh step size is $h = 0.05$.

According to the closest point projection method an equation defined on the surface F is extended and solved in the spatial domain $\Omega_{\mathrm{F}} \subset \Omega$. The region Ω_{F} consists of tetrahedrons from Ω with nodes located at the distance of $\delta = 0.2$ or smaller from F. So, the "thickness" of the domain Ω_{F} equals to $2\delta = 0.4$ ($\approx$ 8 grid nodes) and its largest diameter is 1.4 ($\approx$ 28 grid nodes).

The nodes of Ω_{F} are shown on figure 1 (on the left plot). It also shows a section of the domain Ω by the plane passing through the center of the disk F. The mesh with red edges shows the surface F, red spheres denote nodes located in the domain Ω_{F} ("cloud"). In the same figure (on the right plot) the vectors connecting the mesh nodes $\mathbf{x}$ with their projections $\mathrm{P}(\mathbf{x})$ onto F are shown. The color of each vector corresponds to its length.

Figure 2 shows the sections of the computational domain by the coordinate planes $\mathrm{O}xy$ and $\mathrm{O}yz$. Red spheres indicate mesh nodes located in Ω_{F} which projections belong to the boudary

∂F of the surface. On the same plots the white spheres are nodes in Ω_F, which projections are internal points of the surface F. Blue lines indicate the volumetric mesh defined in Ω.

5.2 Fluid flow in fracture with fixed mid-surface

Here we consider the case when the geometry of the fracture does not change in time and fracture opening is pressure dependent. The statement of the problem in this case has the form of (3).

Fracture opening, as a function of pressure, is defined by (4), where reference opening $w_{\text{ref}}(\mathbf{x})$ corresponds to the surface of bi-axial ellipsoid with axes l_0 and L_0, $l_0 = L_0$. Here l_0 is equal to maximum value of fracture opening, see below.

Let $\mathbf{x}_c$ be the center of the fracture, $\mathbf{x}_{cp} = P(\mathbf{x})$ – projection of a point $\mathbf{x}$ onto the fracture mid-surface, d is a distance from the center of the fracture to the projection of the point onto the fracture,

$$d = d(\mathbf{x}) = \| \mathbf{x}_c - \mathbf{x}_{cp} \|,$$

where $\| \cdot \|$ is the Euclidean norm in R^3. Then the reference opening as a function of the point at the fracture mid-surface is expressed as

$$w_{\text{ref}}(\mathbf{x}) = l_0 \sqrt{1 - \frac{\| \vec{x}_c - \vec{x}_{cp} \|^2}{L_0^2}}. \tag{13}$$

As it was mentioned above, Reynold's lubrication equation (3) is degenerate as opening vanishes at the fracture front (mid-surface boundary). To avoid this situation, we set $L_0 = (1+\varepsilon)L$ where L is actual radius of the disk-shaped fracture. Here $\varepsilon = l_0 = L$ is a small parameter.

At the initial state, the pressure is constant over the fracture mid-surface and is equal to p_0 (reference pressure value). Accordingly, the initial opening distribution equals to the reference one.

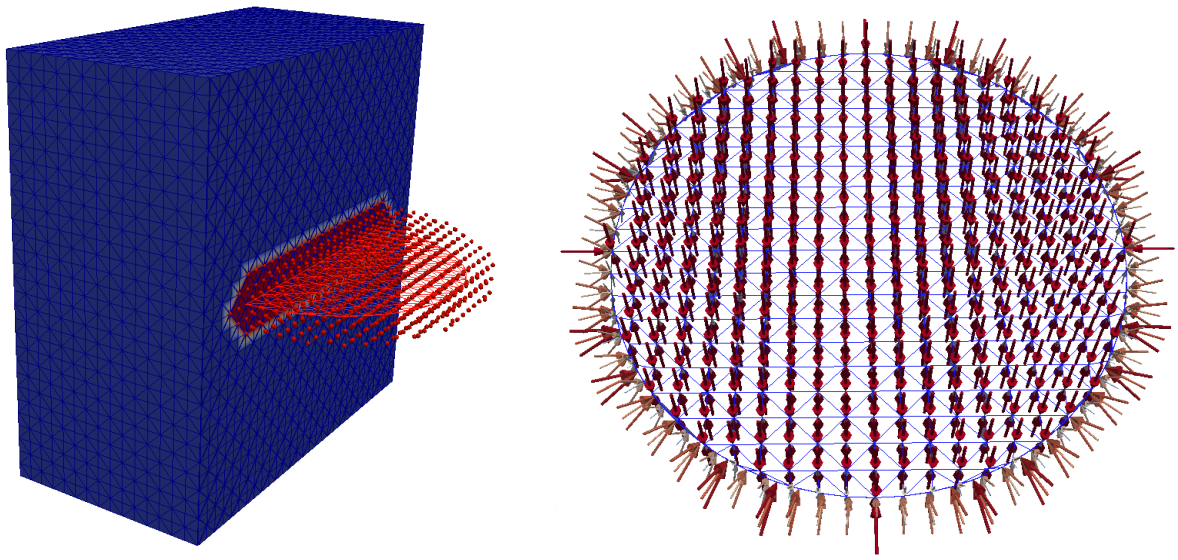
Boundary conditions at the fracture front read:

$$p_{\partial F} = p_{\partial F}(\mathbf{x}) = p_0 \left(1 + \gamma \frac{x - x_c}{L} \right), \quad \mathbf{x} \in F,$$

with $\gamma = 0.1$ being a parameter, x being projection of the point $\mathbf{x}$ onto the axis Ox.

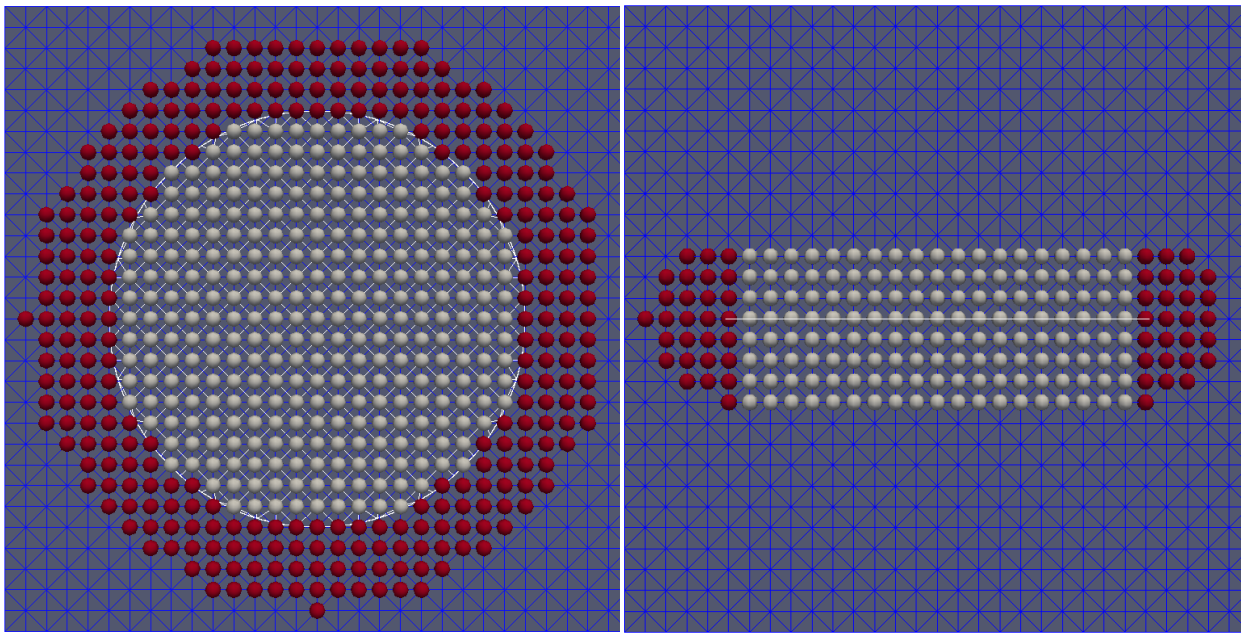
As before, the computational domain is a cube with edge $l = 30$ m. The mesh step size $h = 1$ m. The equation is solved in the domain Ω_F consisting of tetrahedrons, all nodes of which are distant from the fracture mid-surface no further then $\delta = 2$ m. Time step is $\Delta t = 0.5 \cdot 10^{-7}$ s.

The radius of the fracture is equal to $L = 10$ m. The fracture is a subset in the plane Oxy and is centered at the origin. The maximum fracture opening $l_0 = 1$ cm.



(a) (b)

Figure 1:(a) - section of the domain Ω, surface (red mesh) and mesh nodes in Ω_F (red markers). (b) - projection vectors of the nodes inside domain Ω_F on F.



(a) (b)

Figure 2: Projections of the mesh nodes onto the surface (white spheres) and onto its boundary (red speres). The blue lines are edges of computational mesh. (a) - top view, (b) - side view.

In the center of the fracture, inside disk Ω_{bc}^{h} of the radius $R = h$ a constant pressure is set being equal to $1.1p_0$.

Accuracy parameter in (8) is set to $\varepsilon_{\text{iter}} = 1.0 \cdot 10^{-6}$ and the maximum number of iterations is $N_{\text{iter}}^{\max} = 10$.

Physical parameters of the fluid in fracture are as follows: kinematic viscosity $\nu = 1 \cdot 10^{-3}$ Pa · s, reference pressure $p_0 = 300$ bar, reference opening $l_0 = 1$ cm, compressibility $c_{\text{f}} = 4.16 \cdot 10^{-10}$ Pa^{-1}, density $\rho_0 = 1000$ kg/m^3. Parameters of the equation (4) are $c_{\text{w}} = 1/p_0$, $p_{\text{ref}} = p_0$.

Pressure distribution and fracture opening for the computed solution for different moments of time are shown on figure 3.

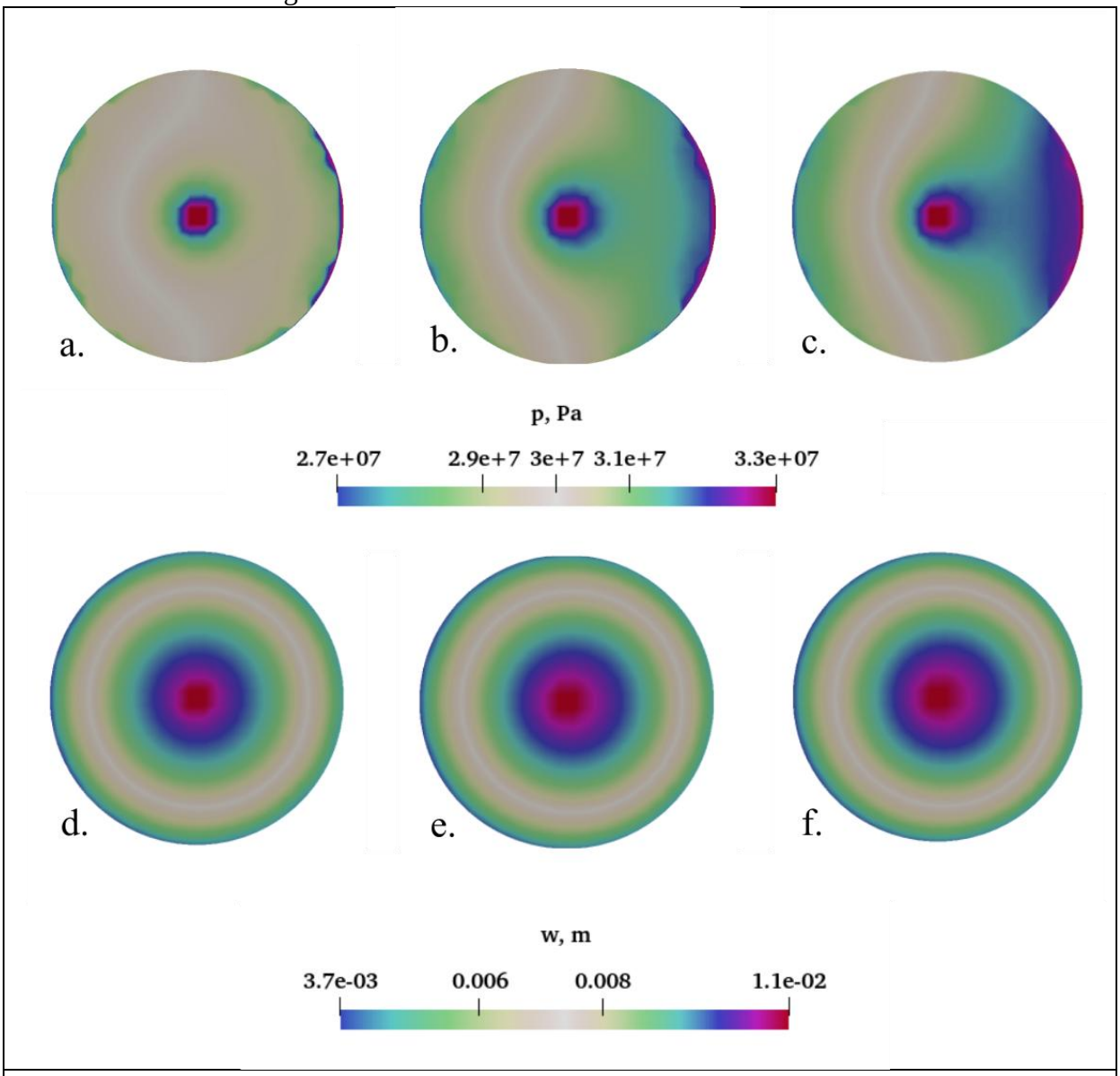


Figure 3: Pressure (on the top) and fracture opening (on the bottom) distribution at time $t = \Delta t$ (a, d), $t = 3\Delta t$ (b, e), $t = 5\Delta t$ (c, f).

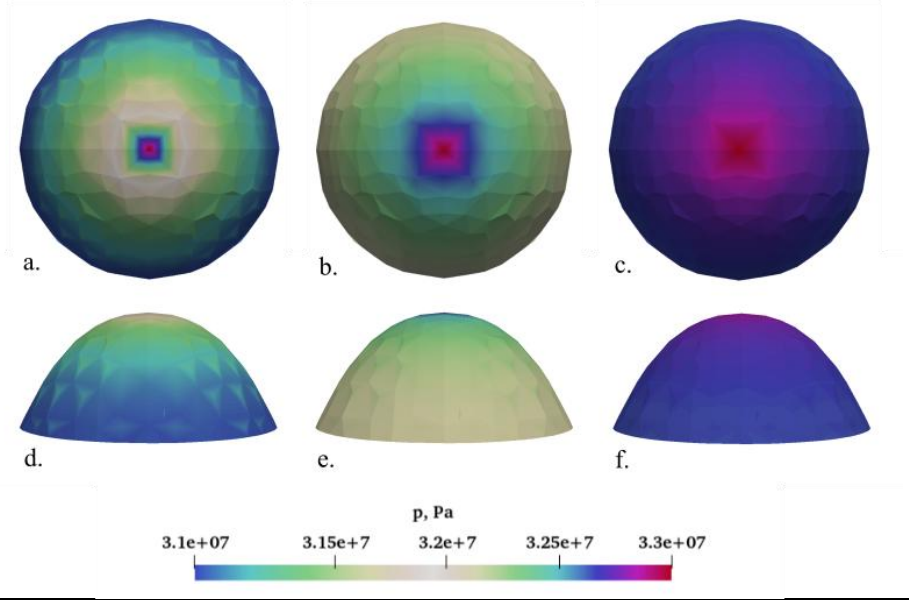


Figure 4: Pressure distribution at time $t = \Delta t$ (a, d), $t = 3\Delta t$ (b, e), $t = 5\Delta t$ (c, f). Top: top view, bottom: side view.

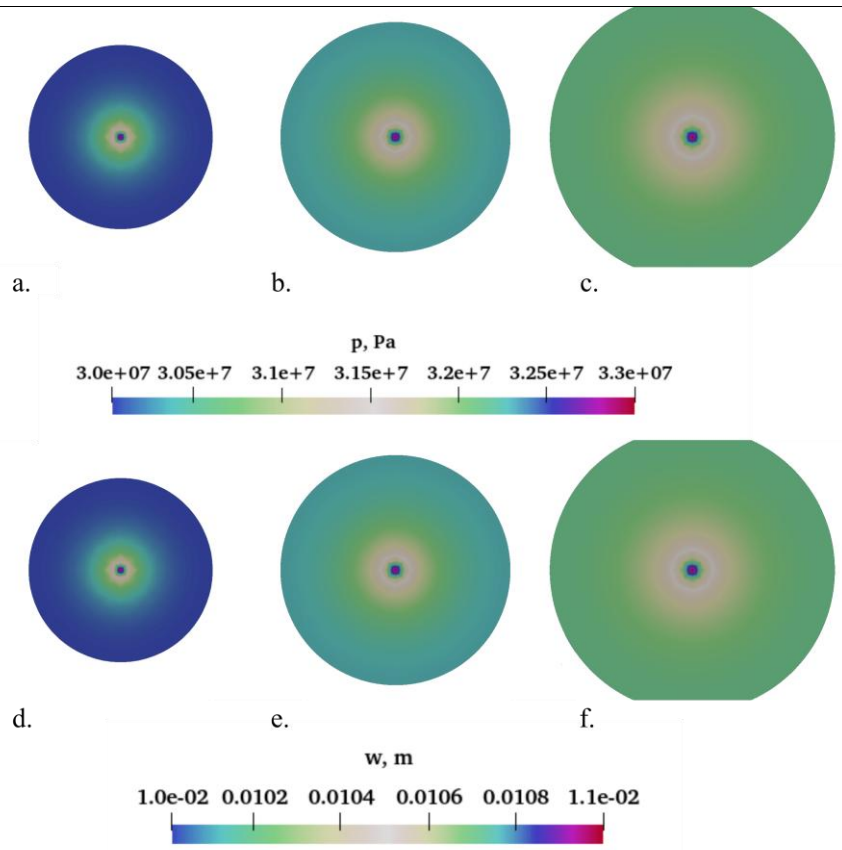


Figure 5: Pressure (top) and fracture opening (bottom) distribution at time $t = \Delta t$ (a, d), $t = 6\Delta t$ (b, e), $t = 12\Delta t$ (c, f).

The presented simulation results show that to achive the required accuracy using fixed-point iteration method requires no more than 10 iterations for each time layer. As expected, the number of iterations significantly decreased with time.

Next, we consider the case of fracture with fixed geometry but of the complex shape. The statement of the problem is the same in general.

At the initial moment of time, the pressure is constant over the middle surface of the fracture and is equal to p_0 (reference pressure value). The opening is constant as a function of the mid-surface point and is equal to l_0 (reference value of the opening). Dependency of the opening on the pressure is given by (4) with $w_{\text{ref}}(\mathbf{x}) = l_0 = \text{const}$.

The computational domain is a cube with edge length $l = 50$ m, mesh step size is $h = 1$ m. The equation is solved in the domain Ω_{F}, consisting of tetrahedrons, all nodes of which are distant from the fracture no further than $\delta = 2$ m. Time step size is $\Delta t = 0.00024$ s.

Accuracy parameter in fixed point iterations stopping criteria is equal to $\varepsilon_{\text{iter}} = 1.0 \cdot 10^{-6}$, the maximum number of iterations is set to $N_{\text{iter}}^{\max} = 3$.

Physical parameters of the fluid are as followos: kinematic viscosity $\nu = 1 \cdot 10^{-3}$ Pa·s, reference pressure value $p_0 = 300$ bar, reference opening $l_0 = 1$ cm, compressibility $c_{\text{f}} = 1 \cdot 10^{-6}$ Pa^{-1}, density $\rho_0 = 1000$ kg/m^3. Parameters in equation (4) are $c_{\text{w}} = 1/p_0$, $p_{\text{ref}} = p_0$.

The fracture mid-surface is part of the paraboloid of the height 3 m and radius of lower section 4 m. At the top of the paraboloid, in the disk Ω_{bc}^h of the radius $R = h$, where h is characteristic grid step, constant pressure equal to $1.1p_0$ is set over time. The simulation results are presented on figure 4.

5.3 Fluid flow in evolving fracture

In this section a case of fluid flow in evolving fracture is considered.

We assume that fracture opening depends on pressure in accordance with the equation (4).

Evolution of the fracture is driven by axially symmetric velocity field $\mathbf{v}$ given by

$$\mathbf{v}(\mathbf{x},t) = v_{\text{m}} \frac{\mathbf{x} - \mathbf{x}_{\text{m}}}{\text{P}\mathbf{x} - \mathbf{x}_{\text{m}}\text{P}}, \quad \mathbf{x} \in \partial \text{F}_t.$$

As it can be seen, fracture evolution occurs in the plane $\text{O}xy$. The velocity field is defined by the point of the "growth center" $\mathbf{x}_{\text{m}}$ and a constant (i.e., independent of the moment of time and point of space) velocity v_{m}. For time interval Δt the point $\mathbf{x}$ of the fracture front moves according to

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \Delta t v_{\text{m}} \frac{\mathbf{x}(t) - \mathbf{x}_{\text{m}}}{\left\| \mathbf{x}(t) - \mathbf{x}r_{\text{m}} \right\|}.$$

It is assumed that domain Ω is sufficiently large and contains fracture mid-surface at each moment of time during fracture evolution. Domain Ω_{F}, where the problem is actually solved, evolves according to fracture evolution.

The computational domain is a cube with edge $l = 50$ m. Accordingly, the mesh step size is $h = 1$ m. The equation is solved in the domain Ω_h^n, consisting of tetrahedrons, all nodes of which are distant from the fracture not further than $\delta = 2$ m. Time step is set to $\Delta t = 0.5 \cdot 10^{-7}$ s.

The radius of the fracture at the initial moment of time $L = 10$ m, the fracture belongs to plane $\mathrm{O}xy$ and is centered at the origin. The fracture growth velocity v_{m} is chosen so that $v_{\mathrm{m}} \Delta t = 0.5$ m. The reference opening is assumed to be constant over fracture surface, $w_{\mathrm{ref}}(\mathbf{x}) = \mathrm{const}$.

The physical parameters are as follows: kinematic viscosity of the fluid $\nu = 1004 \cdot 10^{-6}$ Pa·s, reference pressure value $p_0 = 300$ bar, fluid compressibility $c_{\mathrm{f}} = 4.16 \cdot 10^{-10}$ Pa^{-1}, density $\rho_0 = 1000$ kg / m^3. The reference value of the opening is $w_{\mathrm{ref}} = 10$ mm. Parameters of the equation (4) are $c_{\mathrm{w}} = 1/p_0$, $p_{\mathrm{ref}} = p_0$.

In the center of the fracture, in the disk Ω_{bc} of the radius $R = h$, with h being the characteristic mesh step size, a constant in time pressure equal to $1.1 p_0$ is defined. At the initial time, the pressure in the fracture is equal p_0.

Accuracy parameter in fixed point iterations stopping criteria is equal to $\varepsilon_{\mathrm{iter}} = 1.0 \cdot 10^{-6}$. The maximum number of iterations is $N_{\mathrm{iter}}^{\max} = 10$.

The simulation results are presented in the figure 5.

6 CONCLUSION

The present work is devoted to the numerical study of the finite element version of the closest point projection method applied to the numerical solution of the Reynol'd lubrication equations which describes fluid flow in fractures. Both stationary and evolving fractures are considered. The key idea of the approach is to apply closest point projection method to describe surface evolution and to construct embedding PDE. The fracture mid-surface evolution is described entirely in terms of discrete closest point projection operator without use of any other representation of the fracture. A number of test problems are considered, for stationary and evolving fractures. Numerical experiments demonstrated has demonstrated robustness and efficiency of the overall approach.

REFERENCES

[1] Economides Michael J., Oligney Ronald E., Valko P. *Unified Fracture Design: Bridging the Gap Between Theory and Practice*, Orsa Press, 2001.

[2] Savenkov E.B., Borisov V.E., "A mathematical model for hydraulic fracture propagation in three dimensional poroelastic medium", *PNRPU Mechanics Bulletin*, **1**, 5-17 (2018),

[3] Poveschenko Y.A., Podryga, V.O., Rahimly, P.I, "About one approach to free-volumetric approximation of a piezoconductive medium with gas hydrate inclusions", *Mathematica Montisnigri*, **40**, 68-89 (2017).

[4] Ruuth, S.J., Merriman, B., “A simple embedding method for solving partial differential equations on surfaces”, *J. Comput. Phys.*, **227**, 1943–1961 (2008).
[5] Merriman, B., Ruuth, S.J., “Diffusion generated motion of curves on surfaces”, *J. Comput. Phys.*, **225**, 2267–2282 (2007).
[6] Macdonald, C.B., Ruuth, S.J., “Level set equations on surfaces via the Closest Point Method”, *J. Sci. Comput.*, **35**, 219–240 (2008).
[7] Macdonald, C.B., Ruuth, S.J., “The implicit Closest Point Method for the numerical solution of partial differential equations on surfaces”, *SIAM J. Sci. Comput.*, **31**, 4330–4350 (2009).
[8] Macdonald, C.B., Brandman, J., Ruuth, S.J., “Solving eigenvalue problems on curved surfaces using the Closest Point Method”, *J. Comput. Phys.*, **230**, 7944–7956 (2011).
[9] März, T., Macdonald, C.B., “Calculus on Surfaces with General Closest Point Functions”, *SIAM J. Numer. Anal.*, **50**(6), 3303–3328 (2012).
[10] Savenkov, E. B., “Finite element framework for closest point projection method for surface PDEs”, *Keldysh Institute preprints*. **8** (2020).
[11] Ramazanov, M. M.,Kritskii, B. V., Savenkov, E. B., “Formulation of the j-integral for the biot elastic porous medium model”, *J Eng Phys Thermophys*, **91** (2018).
[12] Ramazanov, M., Borisov, V., Kritsky, B., Savenkov, E., “Fracture growth criterion for poroelastic media”, *AIP Conference Proceedings*, **2051 (**2018).
[13] Lehrenfeld, C., Olshanskii, M., “An Eulerian finite element method for pdes in time-dependent domains”, *ESAIM: M2AN*, **53**(2):585–614 (2019).
[14] Burman, E., Frei, S., Massing, A., “Eulerian time-stepping schemes for the non-stationary Stokes equations on time-dependent domains”, *arXiv preprint*, arXiv:1910.03054 (2019).
[15] Moës, N., Gravouil, A., Belytschko, T., “Non-planar 3D crack growth by the extended finite element and level sets – Part I: Mechanical Model”, *Int. J. Num. Meth. Eng.*, **53**(11), 2549-2568 (2002).
[16] Gravouil, A., Moës, N., Belytschko, T., “Non-planar 3D crack growth by the extended finite element and level sets – Part II: Level set update”, *Int. J. Num. Meth. Eng.*, **53**(11), 2569-2586 (2002).
[17] Savenkov, E. B., “Numerical techniques for surface PDEs. Review of approaches”, *Keldysh Institute preprints*, **5** (2020).
[18] Ivanov, A.V., Savenkov, E.B., “Simulation and visualization of the dynamics of a surface with a movable boundary on a stationary unstructured mesh”, *SciVis*, **9**(2), 64-81 (2017).
[19] Zipunova, E. V., Savenkov E. B., “Solving Reynolds lubrication equations using closest point projection method”, *Keldysh Institute preprints*, **10** (2020).
[20] Zipunova, E.V., Ivanov, A.V., Savenkov, E.B., “Solution of Reynolds lubrication equation on evolving surfaces”, *Keldysh Institute preprints*, **13** (2020).
[21] B.A. Dubrovin, A.T. Fomenko, S.P. Novikov, *Modern Geometry – Methods and Applications*, Springer, Vol.104, (1984-1990).

Received January 12, 2020

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

EQUATION OF STATE FOR NIOBIUM AT HIGH PRESSURES

K. V. KHISHCHENKO[1,2*]

[1] Joint Institute for High Temperatures of the Russian Academy of Sciences
Izhorskaya 13 Bldg 2, 125412 Moscow, Russia

[2] Moscow Institute of Physics and Technology
Institutskiy Pereulok 9, 141701 Dolgoprudny, Moscow Region, Russia

*Corresponding author. E-mail: konst@ihed.ras.ru

DOI: 10.20948/mathmontis-2020-47-10

Summary. An analytical function of pressure on specific volume and internal energy is developed for niobium. This function allows one to adequately describe the thermodynamic properties of this metal in a wide range of densities and pressures. A comparison of the calculated shock adiabat with experimental data at high dynamic pressures is made. The equation of state proposed for niobium can be used to model physical phenomena at high energy densities.

1 INTRODUCTION

The problem of a thermodynamic description of the properties of matter is of interest for both fundamental and applied investigations [1]. For the analysis and numerical simulation of physical processes at high energy densities, equations of state (EOSs) for materials are needed over the entire range of parameters that are realized in these processes [2]. For example, at high velocity impact [3–5], under the action of intense laser [6–8] and particle beams [9, 10], at an electrical explosion of conductors [11, 12], this range continues from normal conditions up to extremely high pressures and specific internal energies.

Niobium is a refractory material, has a low thermal neutron capture cross section. In particular, the EOS for this metal is required when modeling the operating modes of some nodes at nuclear power plants.

In this work, the EOS for niobium is proposed in the form of an analytic function of pressure on specific volume and internal energy. In this form, the EOS can be used in hydrodynamic simulations of adiabatic processes. To illustrate the quality of the EOS, the calculated shock adiabat of niobium is compared with experimental data at high pressures.

2 EOS MODEL

The model is formulated in the framework of the quasi-harmonic approximation. The general form of the EOS [13] is as follows:

$$P(V,E) = P_{\mathrm{c}}(V) + \frac{\Gamma(V,E)}{V}[E - E_{\mathrm{c}}(V)], \tag{1}$$

2010 Mathematics Subject Classification: 74A15, 74J40, 76L05, 80A10, 82D35.
Key words and phrases: equation of state, analytic function, niobium, high pressure, high energy density.

where P is the pressure; V is the specific volume, $V = 1/\rho$; ρ is the density; E is the specific internal energy; E_c is the specific internal energy at zero temperature, $T = 0$; P_c is the corresponding pressure at $T = 0$: $P_c = -dE_c/dV$.

The coefficient Γ is the ratio of thermal pressure to thermal energy density: $V[P - P_c]/[E - E_c]$. Its dependence on volume and internal energy is chosen as follows:

$$\Gamma(V,E) = \gamma_i + \frac{\gamma_c(V) - \gamma_i}{1 + \sigma^{-2/3}[E - E_c(V)]/E_a}, \tag{2}$$

where $\sigma = V_0/V$; V_0 is the specific volume under normal conditions, $E = E_0$ and $P = P_0$; γ_c is the Grüneisen coefficient $\gamma = V(\partial P/\partial E)_V$ at the case of $T = 0$; γ_i is the value of the Grüneisen coefficient at the case of high thermal energies, $E - E_c \gg E_a\sigma^{2/3}$; E_a is a parameter.

The coefficient γ_c is represented by the volume function

$$\gamma_c(V) = 2/3 + (\gamma_{0c} - 2/3)\frac{\sigma_n^2 + \ln^2\sigma_m}{\sigma_n^2 + \ln^2(\sigma/\sigma_m)}, \tag{3}$$

where the value of γ_{0c} corresponds to the normal volume V_0; σ_m and σ_n are parameters.

The cold energy is represented by a polynomial

$$E_c(V) = \frac{B_{0c}V_{0c}}{m - n}\left(\frac{\sigma_c^m}{m} - \frac{\sigma_c^n}{n}\right) + E_{sub}. \tag{4}$$

Here, $\sigma_c = V_{0c}/V$; V_{0c} and B_{0c} are the specific volume and the bulk modulus at $T = 0$ and $P = 0$; $E_{sub} = B_{0c}V_{0c}/(mn)$; m and n are parameters.

3 EOS FOR NIOBIUM

Under normal pressure, the solid phase of niobium has a body-centered cubic (bcc) structure; it melts at $T = 2740$ K [14]. At quasi-hydrostatic compression at room temperature, niobium was studied up to 134 GPa [15]; no transformations of the bcc phase were observed.

At shock compression, niobium was studied up to 180 GPa with traditional explosive systems [16–19]. Pressure up to 400 GPa in niobium was recorded in experiments with special explosive systems [17].

Figures 1–3 display the results of the calculation of the principal Hugoniot curve of niobium over entire range of measured shock and particle velocities, pressures and compression ratios [16–19]. The shock adiabat of the material is calculated using the energy conservation law at the shock front [20],

$$E = E_0 + \frac{1}{2}(P_0 + P)(V_0 - V), \tag{5}$$

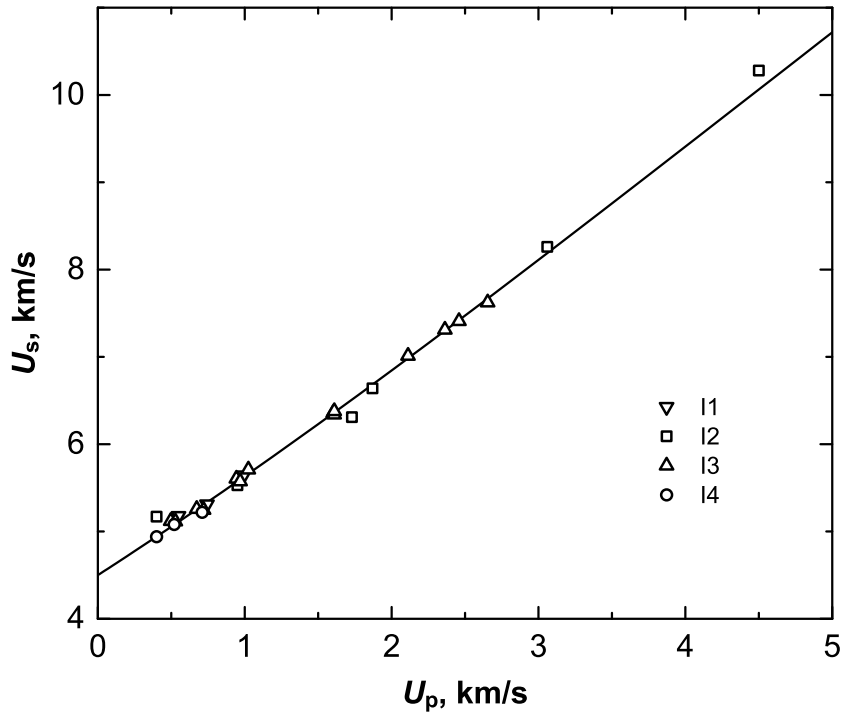


Figure 1: The principal Hugoniot adiabat of niobium: curve corresponds to the present calculations; markers—experimental data (I1—[16]; I2—[17]; I3—[18]; I4—[19]).

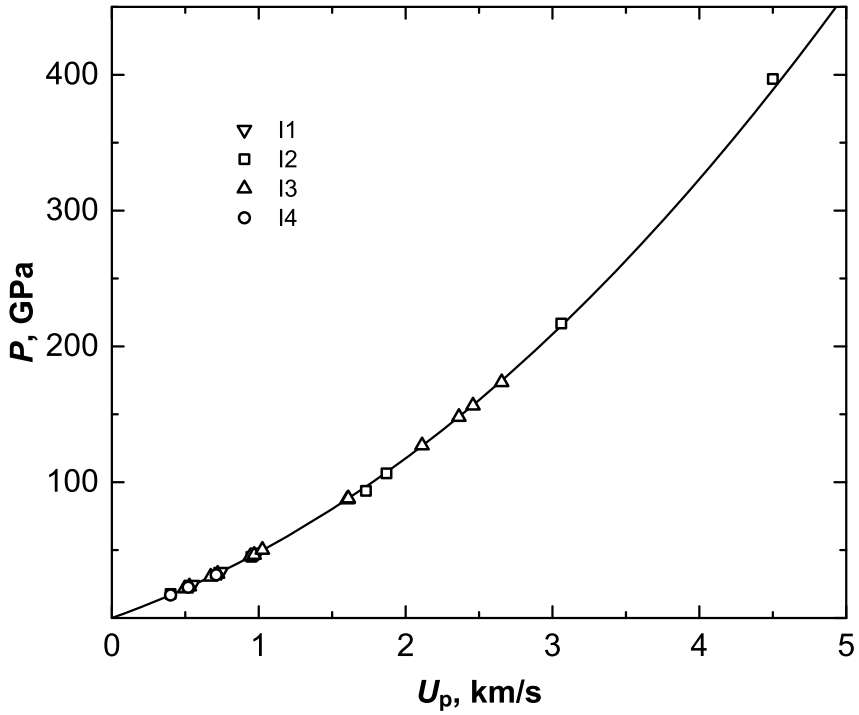


Figure 2: The principal Hugoniot adiabat of niobium: the notation is similar to figure 1.

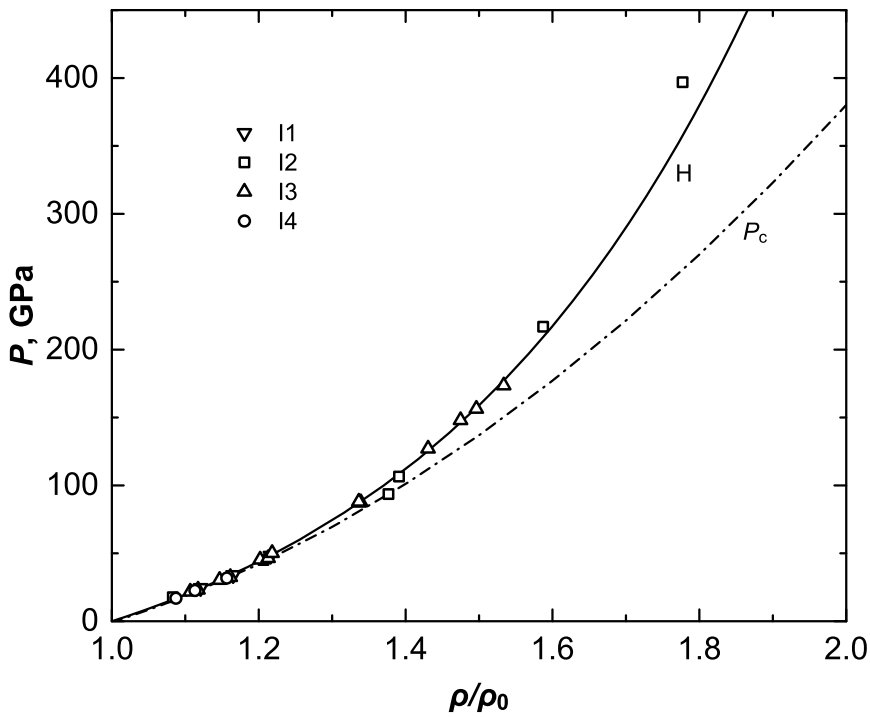


Figure 3: The cold curve (P_c) and the principal Hugoniot adiabat (H) of niobium: curves correspond to the present calculations; marker designations are similar to those used in figure 1.

along with the EOS (1)–(4). The velocities of the shock front (U_s) and particles behind it (U_p) are calculated with the use of the mass and momentum conservation laws [20]:

$$U_s = V_0\sqrt{(P-P_0)/(V_0-V)}, \tag{6}$$

$$U_p = \sqrt{(P-P_0)(V_0-V)}. \tag{7}$$

Comparison of the calculated adiabat with experimental data [16–19] is illustrated in figures 1–3. One can see that the EOS (1)–(4) adequately describes thermodynamic properties of niobium in the region studied in shock waves.

The coefficients of the EOS (1)–(4) for niobium are as follows: $V_0 = 0.11646$ cm^3/g, $V_{0c} = 0.116$ cm^3/g, $B_{0c} = 174.449$ GPa, $m = 0.66$, $n = 0.68$, $\sigma_m = 0.9$, $\sigma_n = 1.2$, $\gamma_{0c} = 1.6$, $\gamma_i = 0.45$ and $E_a = 60$ kJ/g.

4 CONCLUSIONS

Thus, EOS for niobium is developed, which is consistent with data from experiments with shock waves at high pressures. The EOS has a form suitable for use in the numerical simulation of adiabatic processes in a wide range of densities and internal energies.

Acknowledgments: The work is supported by the Russian Science Foundation (grant No. 19-19-00713).

The paper is based on the proceedings of the XXXV International Conference on Equations of State for Matter, Elbrus, the Kabardino-Balkar Republic of the Russian Federation, March 1 to 6, 2020.

REFERENCES

[1] V. Fortov, *Thermodynamics and Equations of State for Matter: From Ideal Gas to Quark–Gluon Plasma*, Singapore: World Scientific Publishing, (2016).
[2] I. V. Lomonosov and S. V. Fortova, "Wide-range semiempirical equations of state of matter for numerical simulation on high-energy processes", *High Temp.*, **55**(4), 585–610 (2017).
[3] E. I. Kraus and I. I. Shabalin, "The tool for high-velocity interaction and damage of solids", *Math. Montis.*, **39**, 18–29 (2017).
[4] K. K. Maevskii, "Thermodynamic parameters of mixtures with silicon nitride under shock-wave loading", *Math. Montis.*, **45**, 52–59 (2019).
[5] E. A. Strebkova, M. N. Krivosheina, and Ya. V. Mayer, "Features of the processes of elastic deformation in cubic crystals", *Math. Montis.*, **46** (2019).
[6] A. E. Zubko and A. A. Samokhin, "Modeling of thermoacoustic and evaporation pressure signals in absorbing liquids irradiated with nanosecond laser pulses", *Math. Montis.*, **36**, 78–85 (2016).
[7] A. A. Samokhin, V. I. Mazhukin, M. M. Demin, A. V. Shapranov, and A. E. Zubko, "On critical parameters manifestations during nanosecond laser ablation of metals", *Math. Montis.*, **43**, 38–48 (2018).
[8] S. Yu. Gus'kov, I. K. Krasyuk, A. Yu. Semenov, I. A. Stuchebryukhov, and K. V. Khishchenko, "Extraction of the shock adiabat of metals from the decay characteristics of a shock wave in a laser experiment", *JETP Lett.*, **109**(8), 516–520 (2019).
[9] M. E. Zhukovsky, M. B. Markov, S. V. Podolyako, I. A. Tarakanov, R. V. Uskov, A. M. Chlenov, and V. F. Zinchenko, "Researching the spectrum of bremsstrahlung generated by the RIUS-5 electron accelerator", *Math. Montis.*, **35**, 54–67 (2016).
[10] K. K. Inozemtseva, M. B. Markov, and F. N. Voronin, "The electromagnetic and thermomechanical effects of electron beam on the solid barrier", *Math. Montis.*, **39**, 79–100 (2017).
[11] S. A. Barengolts, E. V. Oreshkin, V. I. Oreshkin, and K. V. Khishchenko, "Simulation of the explosion of a surface microprotrusion during a radio frequency breakdown", *IEEE Trans. Plasma Sci.*, **47**(8), 3406–3411 (2019).
[12] V. N. Senchenko and R. S. Belikov, "Experimental investigation of thermophysical properties of rhenium near its melting point", *J. Phys.: Conf. Ser.*, **1370**, 012034 (2019).
[13] K. V. Khishchenko, M. V. Zhernokletov, I. V. Lomonosov, and Yu. N. Sutulov, "Dynamic compressibility, release adiabats, and the equation of state of stilbene at high energy densities", *Tech. Phys.*, **50**(2), 197–201 (2005).
[14] E. Yu. Tonkov, *Phase Diagrams of Elements at High Pressures*, Moscow: Nauka, (1979).
[15] T. Kenichi and A. K. Singh, "High-pressure equation of state for Nb with a helium-pressure medium: Powder x-ray diffraction experiments", *Phys. Rev. B*, **73**(22), 224119 (2006).
[16] J. M. Walsh, M. H. Rice, R. G. McQueen, and F. L. Yarger, "Shock-wave compressions of twenty-seven metals equations of state of metals", *Phys. Rev.*, **108**(2), 196–216 (1957).
[17] L. V. Al'tshuler, A. A. Bakanova, and I. P. Dudoladov, "Electron structure on the compressibility of metals at high pressure", *Sov. Phys. JETP*, **26**(6), 1115–1120 (1968).
[18] S. P. Marsh (ed.), *LASL Shock Hugoniot Data*, Berkeley, CA: University of California Press, (1980).
[19] L. V. Al'tshuler, A. A. Bakanova, I. P. Dudoladov, E. A. Dynin, R. F. Trunin, and B. S. Chekin, "Shock adiabatic curves of metals", *J. Appl. Mech. Tech. Phys.*, **22**(2), 145–169 (1981).
[20] Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, New York: Academic Press, (1967).

Received December 15, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

FUNDAMENTAL EQUATION OF STATE OF ARGON, SATISFYING THE SCALING HYPOTHESIS AND WORKING IN THE REGION OF HIGH TEMPERATURES AND PRESSURES

S. V. RYKOV[1*], V. A. RYKOV[1], I. V. KUDRYAVTSEVA[1], E. E. USTYUZHANIN[2], A. V. SVERDLOV[1]

[1] ITMO University
Kronvergskiy 49, 197101 Saint-Petersburg, Russia

[2] National Research University Moscow Power Engineering Institute
Krasnokazarmennaya 14, 111250 Moscow, Russia

* Corresponding author. E-mail: togg1@yandex.ru

DOI: 10.20948/mathmontis-2020-47-11

Summary. This paper discusses the problem of describing thermodynamic properties of a substance at high temperatures and pressures on the basis of the fundamental equation of state (FEoS). This FEoS has the following characteristics: it transforms into the virial equation of state in the region of low densities; it is converted into the Berestov equation in the vicinity of the critical point. FEoS testing has been carried out on known thermodynamic properties of argon and has allowed establishing its workspace: by the pressure up to 1000 MPa; by the temperature from the temperature of the triple point to 1200 K. It has been shown that our FEoS can qualitatively correctly describe the thermal surface of argon up to 17 000 K. A comparison of FEoS has been made with some well-known equations of state. When developing FEoS of argon, we have used elements of the similarity theory, which has allowed reducing the number of individual parameters of this FEoS.

1 INTRODUCTION

We investigate a problem of describing the thermophysical properties of substances in a wide range of temperatures and pressures including the critical region. The problem attracts the attention of many researchers [1–24]. In particular, this problem is actual when studying the behavior of substances:

- in the range of highly developed density fluctuations near the critical point;
- at high temperatures and high pressures.

To describe the properties of pure substances at high pressures and high temperatures, the authors of [25–32] have developed a number of fundamental equations of state (FEoS). When describing the liquid behavior in the vicinity of the critical point, we have used previously a number of approaches and developed some equations of state (EoS):

2010 Mathematics Subject Classification: 80A05, 80A10, 80M50.
Key words and phrases: mathematical modeling, thermodynamic surface, critical point, argon, equation of state.

- scaling EoSs and crossover EoSs in a parametric form and with (the density, the temperature variables) [1, 2, 10, 12, 16, 24];
- FEoS of the virial type [3, 6, 9, 17];
- FEoS [14] based on requirements of the scaling theory (ST) for the critical region [33];
- FEoS [4, 5, 7, 8, 11, 15, 18–20, 22, 23] converted into a Widom EoS and valid in the vicinity of the critical point.

We have analyzed approaches [1–12, 14–20, 22–24] and have got the following results. Scaling EoSs [1, 2, 16, 24] and crossover EoSs [10, 12] meet the requirements of ST [33], but they have a narrow work area limited by temperatures T ($0.9T_c < T < 2T_c$ [12], here T_c is the critical temperature), and, therefore, can not be used when modeling thermodynamic properties of a substance in the range of high temperatures and high pressures.

One of the disadvantages of the crossover EoS [12] is the need to use different critical temperatures: one T_c to calculate the pressure (p) and another T_c to calculate the isochoric heat capacity (C_V). FEoS [3, 6, 9, 17] do not meet the requirements of ST. Therefore, these EFoSs do not describe the sound velocity (w), C_V, the isobar heat capacity (C_p), and isothermal compressibility factor (K) in the critical region with acceptably small uncertainties. At the same time, these EFoSs describe the equilibrium properties of argon in the regular part of the thermodynamic surface with low uncertainties. For example, FEoS of argon is proposed [9] in this form. The workspace of FEoS FEoS [9] is (limited by pressures $0 \leqslant p \leqslant 1000$ MPa, by temperatures $83 \leqslant T \leqslant 700$ K) and can be successfully used when predicting thermal properties at high temperatures.

Bezverhiy *et al* [14] has developed FEoS, which takes into account the feature presence of $C_V(T,\rho)$ as a known function in the critical region, here ρ is the density. Our analysis shows that EFoS [14] reproduces power laws of ST qualitatively incorrectly. For example, the critical isotherm [14] follows $\Delta p \propto (\Delta\rho)^3$. It should be $\Delta p \propto \Delta\rho|\Delta\rho|^{\delta-1}$ [33], here $\Delta p = (p-p_c)/p_c$; $\Delta\rho = (\rho-\rho_c)/\rho_c$; p_c is the critical pressure; ρ_c is the critical density; δ is the critical index of the critical isotherm.

Empirical FeoSs are proposed in [5, 7, 8, 11]. They qualitatively correctly reproduce all of ST power laws when describing properties in the vicinity of the critical point.

On the basis of the phenomenological theory of the critical point [34], the authors of [4, 15, 18–20, 22, 23] developed FEoS which is not inferior to scaling EoS and crossover EoS when describing the asymptotic vicinity of the critical point. We mark that FEoSs [4, 5, 7, 8, 11, 15, 18–20, 22, 23] do not satisfy the theory of extended scaling [2]. Indeed, the function, $C_V(T,\rho_c)$ [4, 5, 7, 8, 11, 14, 15, 18–20, 22, 23], follows $C_V \simeq A\tau^{-\alpha} + C\tau$. It is shown in [2] that

$C_V(T,\rho_c)$ should be $C_V \simeq A\tau^{-\alpha} + B\tau^{-\alpha+\Delta} + C\tau$, here $\tau = (T - T_c)/T_c$; α and Δ are the critical indexes. In addition, our analysis shows that FEoSs [4, 5, 7, 8, 11, 14, 15, 18–20, 22, 23] are inferior to FEoS [9] when describing properties at high temperature region. For example, there are a discrepancy between p values calculated by EFoS [9] and p values calculated by [20] at 17 000 K these deviations exceed 50%.

In this paper on the basis of the approach [35], we plan to develop a FEoS that meets the following requirements:

- satisfies ST requirements [33] and does not inferior to EoSs [10, 12] when describing properties in the critical region;
- simulates the thermal surface of argon at temperatures up to 17 000 K and by pressures up to 12 GPa;
- can be converted into the Berestov equation [2] in the critical region.

2 STRUCTURE OF FEOS

By analogy with [4, 15, 18–20, 22, 23], this FEoS has the following structure:

$$F(\rho,T) = F_{\text{reg}}(\rho,T) + F_{\text{nreg}}(\rho,T), \tag{1}$$

where $F(\rho,T)$ is the Helmholtz free energy; $F_{\text{reg}}(\rho,T)$ is a regular function; $F_{\text{nreg}}(\rho,T)$ is an irregular component of the Helmholtz free energy:

$$F_{\text{nreg}}(\rho,T) = RT_c\phi(\omega,t)\Big(|\Delta\rho|^{\delta+1}a_0(x) + |\Delta\rho|^{\delta+1+\frac{\Delta}{\beta}}a_1(x)\Big), \tag{2}$$

where $\phi(\omega,t)$ is the regular function; R is the gas constant; $\omega = \rho/\rho_c$; $x = \tau/|\Delta\rho|^{1/\beta}$ is the scaling variable; $t = T/T_c$; β is the critical index.

We notice: there is a principal difference of our FEoS from FEoS [4, 15, 18–20, 22, 23]. We have included an additional component in $F_{\text{nreg}}(\rho,T)$. There is a special scale function, $a_1(x)$, in this additional component recommended in [35]. This principal modernization has allowed us to improve a FEoS structure and to meet the requirements [2].

The scaling functions $a_0(x)$ and $a_1(x)$ are calculated based on the following:

- a new representation of the scaling hypothesis [34, 35];
- the Benedek hypothesis [36];
- the Berestov equation [2].

i	$j=0$	$j=1$	$j=2$
0	0	0	2.727 031 612 1447
1	0	0	−2.180 917 085 2935
2	0	0	2.018 179 285 6405
3	0	−1.651 807 350 2083	1.951 547 138 476
4	0	3.235 0 97 0 27 9 45 2	5.861 967 866 4433
5	0	0.203 261 164 281 07	−2.078 078 708 984
6	−0.283 648 592 739 017	−1.850 670 154 3516	−2.801 735 606 0172
	−0.031 673 399 139 638	2.497 843 489 6566	9.620 211 455 1673
	−0.117 319 511 789 66	−1.164 995 874 2581	−3.476 213 158 3227
9	0.413 193 730 791 89	2.810 166 015 2324	−0.811 288 614 251 57
10	− 0.765 606 737 657 49	−2.801 124 973 5011	0.766 664 260 646 57
11	− 0.650 491 354 2378	−2.592 742 798 4863	0.450 684 903 397 98
12	1.808 588 644 5017	5.578 734 268 4796	−0.239 259 405 790 52
13	−1.042 305 956 028	−2.485 069 647 1961	−1.024 947 033 0846
14	−0.813 944 971 192 75	−2.392 997 971 8019	0.759 164 862 584 33
15	1.328 057 607 1621	3.119 794 125 8801	0
16	−0.486 803 106 500 06	−0.933 985 969 4002	−0.221 489 078 823 57
17	−0.243 474 625 433 64	−0.585 625 628 796 48	0.091 451 137 589 177
18	0.327 530 667 992 16	0.660 860 378 937 56	0
19	−0.154 068 046 320 52	−0.286 546 955 475 54	−0.008 667 473 663 7731
20	0.039 124 504 337 479	0.068 175 553 922 501	0.001 865 695 143 8862
21	−0.005 346 887 409 843	−0.008 772 962 958 1014	0
22	0.000 310 067 180 058 02	0.000 478 732 897 948 04	$-2.5448089017224 \times 10^{-5}$

Table 1: Coefficients $C_{i,j}$ of FEoS (1).

These functions are written the following form:

$$a_0(x) = -\frac{u_0 k \gamma_1 x_0^{2-\alpha}}{2\alpha b^2 \alpha_1 (1-\varepsilon)} \left[(\varphi+\varphi_1)^{2-\alpha} - \varepsilon(\varphi+\varphi_2)^{2-\alpha} \right] + \frac{u_0 x_0^{\gamma}}{2k} (\varphi+\varphi_3)^{\gamma} + u_0 C_0, \quad (3)$$

$$a_1(x) = -\frac{u_1 k \gamma_2 x_0^{2-\alpha+\Delta}}{2\alpha b^2 \alpha_2 (1-\varepsilon)} \left[(\varphi+\varphi_1)^{2-\alpha+\Delta} - \varepsilon(\varphi+\varphi_2)^{2-\alpha+\Delta} \right] + \frac{u_1 x_0^{\gamma+\Delta}}{2k} (\varphi+\varphi_3)^{\gamma+\Delta} + u_1 C_1, \quad (4)$$

where $\varepsilon_0 = x_1/x_2$; $\alpha_1 = (2-\alpha)(1-\alpha)$; $\gamma_1 = \gamma(\gamma-1)$; $\gamma_2 = (\gamma+\Delta)(\gamma+\Delta-1)$; $\alpha_2 = (2-\alpha+\Delta)(1-\alpha+\Delta)$; $\varphi = x/x_0$; $\varphi_i = x_i/x_0$, $i \in \{1,2,3\}$; $b^2 = (\gamma-2\beta)/[\gamma(1-2\beta)]$; $k = [(b^2-1)/x_0]^{\beta}$; γ is the critical index; x_0, u_0 and u_1 are the individual parameters; $Z_c = p_c/(R\rho_c T_c) \times 10^3$; C_0

i	$j=3$	$j=4$	$j=5$
0	4.482 248 574 7539	2.332 643 055 2399	1.808 465 772 8776
1	−3.225 639 106 0006	−1.048 810 609 669	−0.932 531 831 731 91
2	−3.143 085 800 7921	−7.416 650 230 6154	−4.156 240 517 2991
3	7.651 653 302 7528	6.805 576 926 7176	3.940 459 100 9914
4	6.118 667 623 2535	4.737 381 837 8476	0.741 441 138 784 28
5	−8.029 797 760 4914 6	−8.232 510 077 0624	−2.172 672 507 2028
	−0.081 651 952 400 293	1.488 772 709 3593	0.930 235 432 967 88
7	8.388 156 700 3335	2.510 816 288 7711	−0.130 873 463 355 37
8	−4.059 017 137 1799	−1.089 247 218 7001	0
9	−0.103 759 991 449 93	−0.277 754 425 60302	0
10	1.129 529 996 898	0.093 168 194 589 203	0
11	−0.543 765 088 527 54	0.151 336 027 7963	0
12	0.008 645 103 248 2461	−0.082 484 913 633 882	0
13	0.001 821 779 428 3432	0.012 151 299 548 948	0

Table 2: Coefficients $C_{i,j}$ of FEoS (1).

and C_1 are the constant coefficients which value is found from the equations

$$(\delta+1)a_0|_{\varphi=-1}+\frac{x_0}{\beta}a'_0|_{\varphi=-1}=0, \tag{5}$$

$$\left(\delta+1+\frac{\Delta}{\beta}\right)a_1|_{\varphi=-1}+\frac{x_0}{\beta}a'_1|_{\varphi=-1}=0. \tag{6}$$

We have selected the regular component (1) in the form [19]:

$$\begin{aligned} F_{\text{reg}}(\rho,T) &= F^0(\rho,T)+RT\omega y_2+RT\omega(Z_c-0.2)y_6+RT\omega D_3(y_4-y_6) \\ &+RT\omega\tau_1\Big[D_1(\omega-3)+D_2\big(\omega^2-2\omega\big)\Big]+RT\omega\sum_{i=0}^{22}\sum_{j=0}^{20}\big(C_{i,j}\tau_1^j\Delta\rho^i\big), \end{aligned} \tag{7}$$

where $F^0(\rho,T)$ is the ideal gas component of $F(\rho,T)$; $\tau_1=T_c/T-1$; functions y_2, y_4, y_6 have the following form: $y_2=-7.7/6+2.9/6\Delta\rho-1.1/6\Delta\rho^2+0.05\Delta\rho^3$, $y_4=5-4\Delta\rho+3\Delta\rho^2-2\Delta\rho^3+\Delta\rho^4$, $y_6=4-3\Delta\rho+2\Delta\rho^2-\Delta\rho^3+\Delta\rho^5$.

We have calculated φ_1, φ_2, φ_3 values according to the method detailed in [15]. It let us got $\varphi_1=2.807\,223\,47$, $\varphi_2=14.471\,7304$, $\varphi_3=5.732\,468\,25$.

i	$j=6$	$j=7$	$j=8$
0	1.942 056 320 0621	3.245 246 493 1065	−8.239 406 700 9885
1	−1.034 640 564 3285	−0.970 159 560 317 12	−0.011 163 693 637 208
2	−0.944 139 567 2871	0.298 859 602 686 75	0.039 252 086 979 538
3	1.204 115 946 6534	0	0
4	−0.327 499 512 264 19	0	0
i	$j=9$	$j=10$	$j=11$
0	−18.746 448 404 883	51.077 633 966 366	68.645 329 452 91
1	0.541 072 550 799 12	0	0
2	−0.178 044 619 880 26	0	0
i	$j=12$	$j=13$	$j=14$
0	−182.047 371 3271	−144.870 071 874 34	383.406 155 478 06
i	$j=15$	$j=16$	$j=17$
0	174.017 641 515 55	−472.418 838 330 36	−110.347 173 018 13
i	$j=18$	$j=19$	$j=20$
0	314.552 869 844 35	28.506 239 206 301	−87.384 487 306 415

Table 3: Coefficients $C_{i,j}$ of FEoS (1).

We have chosen the crossover function in accordance with the recommendations [7]:

$$\phi(\omega,t)=\phi_0(\omega)\phi_1(t), \qquad \phi_0(\omega)=\Big[\big(1-\omega\big)^m-1\Big]^2, \qquad \phi_1(t)=1/t^2, \tag{8}$$

where $m \in N$.

We have tested FEoS (1) with components (2)–(4) and (7) on the example describing the equilibrium properties of argon [37–54].

3 FEOS OF ARGON

Select the ideally-gas component of argon $F^0(T,\rho)$ according to the recommendations of [9]:

$$F^0(\rho,T)=RT\big(\ln\omega+a_1^0+a_2^0t^{-1}-1.5\ln t\big), \tag{9}$$

where $a_1^0=58.316\,662\,43$ and $a_2^0=524.946\,511\,64$.

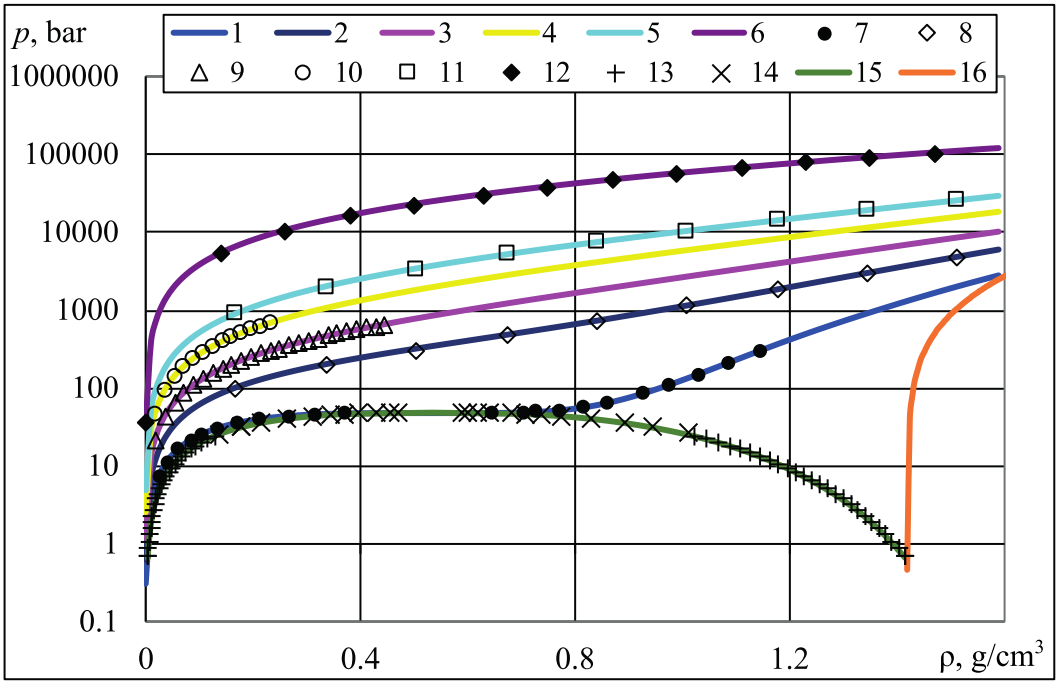


Figure 1: Isotherms, the liquid–gas coexistence curve and the melting line of argon. Isotherms calculated using FEoS (1): 1—critical isotherm; 2—300 K; 3—573.15 K; 4—1223.15 K; 5—2300 K; 6—17 000 K. Experimental data: 7—150.65 K [42]; 8—300 K [3]; 11—2300 K [3]. Tabulated data: 9—573.15 K [43]; 10—1223.15 K [43]; 12—17 000 K [9]; 13—tabulated data on the density at the saturation line [9]; 14—experimental data on the density of a saturated liquid and saturated vapor [49]; 15—data on the density at the saturation line calculated by FEoS (1); 16—data on the density at the melting line [9].

We have calculated expressions for compressibility Z on the basis of FEoS (1) with components (2)–(4) and (7):

$$\begin{aligned} Z(\rho,T) = {} & 1 + y_1\omega^2 + y_2\omega + D_3\left(y_3\omega^2 + y_4\omega - y_5\omega^2 - y_6\omega\right) + \left(y_5\omega^2 + y_6\omega\right)(Z_c - 0.2) \\ & + \omega\sum_{i=0}^{22}\sum_{j=0}^{20} C_{i,j}{\tau_1}^j\Delta\rho^{i-1}(i\omega + \Delta\rho) + D_1\omega\tau_1(2\omega - 3) + D_2\omega^2\tau_1(3\omega - 4) \\ & + Z_c\omega|\Delta\rho|^{\delta}\phi_1(t)t\left(\phi_0(\omega)\mathrm{sign}(\Delta\rho)h_0(x) + {\phi'}_0(\omega)|\Delta\rho|a_0(x)\right) \\ & + Z_c\omega|\Delta\rho|^{\delta+\frac{\Delta}{\beta}}\phi_1(t)t\left(\phi_0(\omega)\mathrm{sign}(\Delta\rho)h_1(x) + {\phi'}_0(\omega)|\Delta\rho|a_1(x)\right), \end{aligned} \tag{10}$$

where $y_{2i-1} = {y'}_{2i}(\omega)$ $(i \in \{1,2,3\})$; $h_n(x)$ are scale functions of chemical potential [27]:

$$h_0(x) = (\delta+1)a_0(x) - \frac{x}{\beta}{a'}_0(x), \qquad h_1(x) = \left(\delta + 1 + \frac{\Delta}{\beta}\right)a_1(x) - \frac{x}{\beta}{a'}_1(x). \tag{11}$$

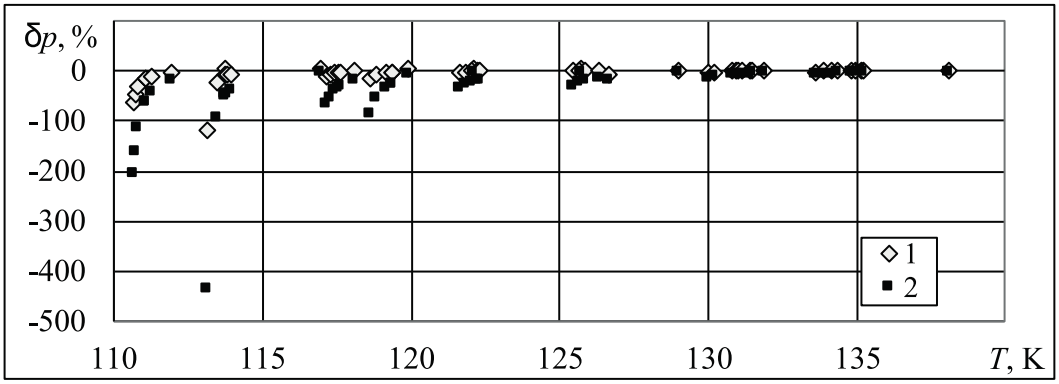


Figure 2: Relative deviations $\delta p = (p_{\text{exp}} - p_{\text{cal}})/p_{\text{exp}}100\%$ corresponded to p_{cal} values calculated with the help of FEoS (1) and FEoS [9] in the metastable states of argon: (p_{exp}, ρ_{exp}, T_{exp}) data are taken from [45] over isochoric lines 1231.9, 1210.9, 1180.2, 1165.6, 1140.9, 1099.8, 1050.8 and 1010.7 kg/m^3; 1—p_{cal} values calculated with the help of FEoS (1); 2—p_{cal} values calculated with the help of FEoS [9]. On each of the isochors, two experimental points corresponding to large values of the pressure are located in the single-phase range, the rest of experimental points are located in the metastable range.

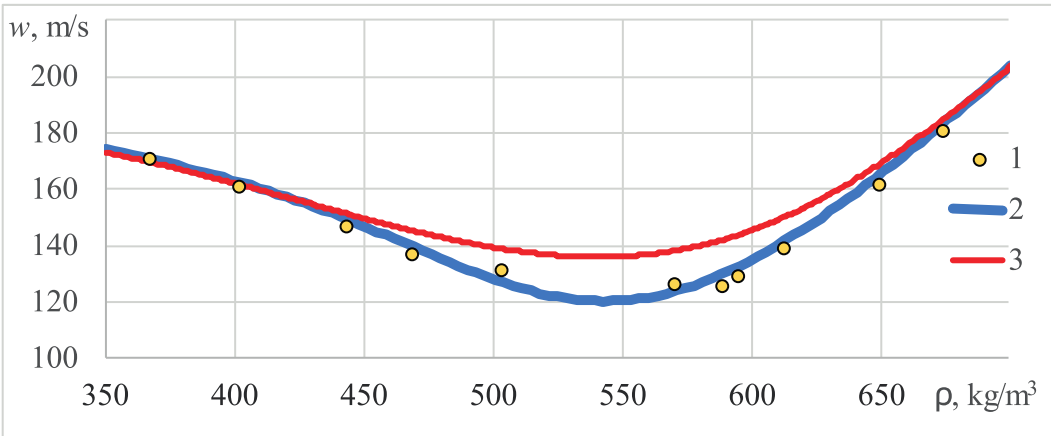


Figure 3: Argon sound speed at 150.8 K isotherm: 1—experimental data [46]; 2—calculation by FEoS (1); 3—calculation by FEoS [9].

Coefficients and parameters of FEoS (1) with components (2)–(4) and (7), (8) were determined on the basis of an array of experimental data [37–54] among them: $T_c = 150.66$ K, $p_c = 4.8634$ MPa, $\rho_c = 535.1$ kg/m^3, $R = 0.208\,133\,32$ J/(g K), $u_0 Z_c = 4.549\,364\,19$, $u_1 Z_c = 0.052\,429\,623\,1552$, $\alpha = 0.11$, $\beta = 0.3255$, $\gamma = 1.239$, $\delta = 4.806$, $\Delta = 0.51$, $m = 3$, $D_1 = 0.528\,541\,695\,546\,02$, $D_2 = 0.874\,668\,218\,972\,52$, $D_3 = -7.913\,173\,555\,7194 \times 10^{-3}$ and $x_0 = 0.311\,220\,376\,399\,66$. The values of coefficients $C_{i,j}$ are presented in tables 1, 2 and 3.

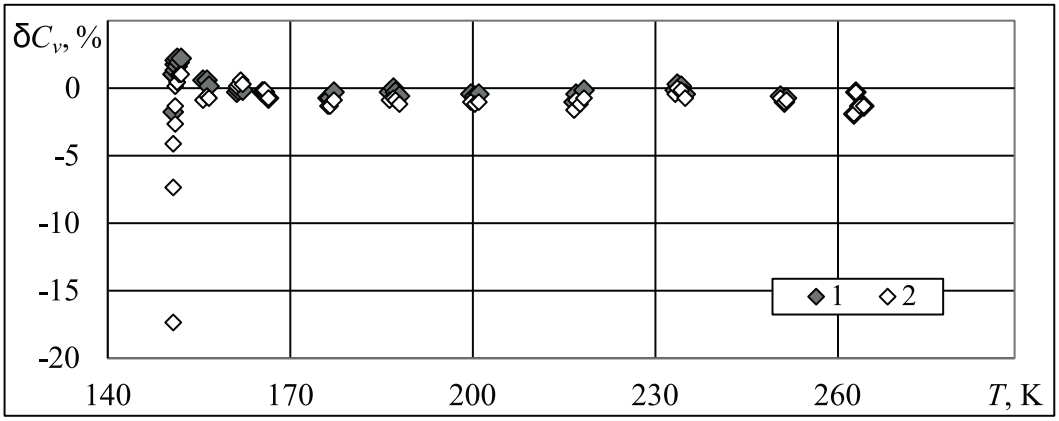


Figure 4: Relative deviations $\delta C_V = (C_{V,\text{exp}} - C_{V,cal})/C_{V,\text{exp}}100\%$; p_{cal}: $C_{V,cal}$ corresponded to values calculated with the help of (1) in the single phase range; $C_{V,\text{exp}}$ corresponded to data [48] over isochoric line 473.6 kg/m^3; 1—C_V calculated with the help of FEoS (1); C_V calculated with the help of FEoS [9].

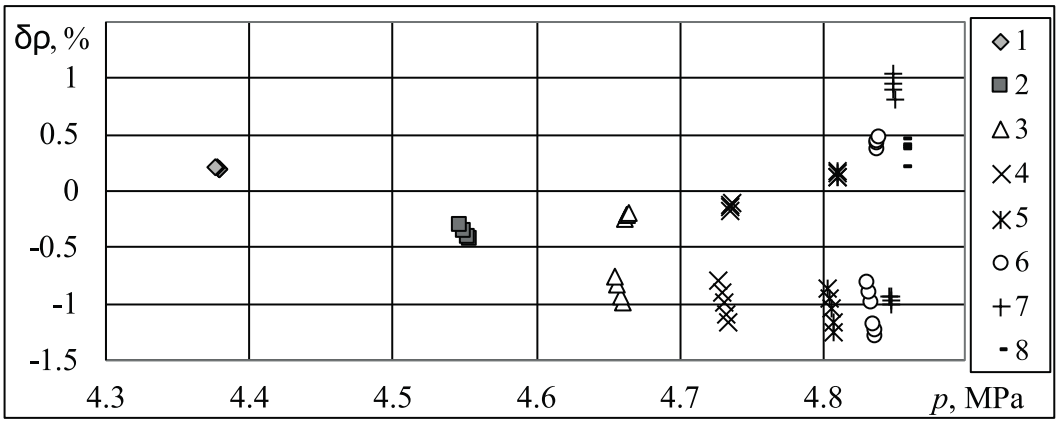


Figure 5: Relative divergence values of density $\delta\rho = (\rho_{\text{exp}} - \rho_{\text{calc}})/\rho_{\text{exp}}100\%$ calculated as per the equations pre-sented in this study as compared with the experimental data [44]. Isothermal lines: 1—148.007 K; 2—149.006 K; 3—149.598 K; 4—149.983 K; 5—150.372 K; 6—150.52 K; 7—150.579 K; 8—150.621 K.

Based on compressibility Z (10), we have calculated the thermal surface of argon (figure 1). As one can see, FEoS (1) transmits the thermal surface of argon in the temperature range from the saturation line and the melting line to 2300 K and it can be extrapolated by temperature up to 17 000 K and by pressure up to 12 GPa. FEoS (1) describes the experimental (p, ρ, T)-data in the metastable range [45], experimental data about C_V [48] and about the speed of sound w [46] in the vicinity of the critical point with less uncertainty than FEoS NIST [9] (figures 2, 3 and 4). Note that when searching for the coefficients of FEoS (1), experimental data [45, 46] were not used. The FEoS (1) represents experimental (p, ρ, T)-data [44] within the range of the experimental error (figure 5). Experimental data on C_V [48, 50] are transmitted within the experimental error (figure 6) in a wide range of state parameters including the vicinity of the critical point.

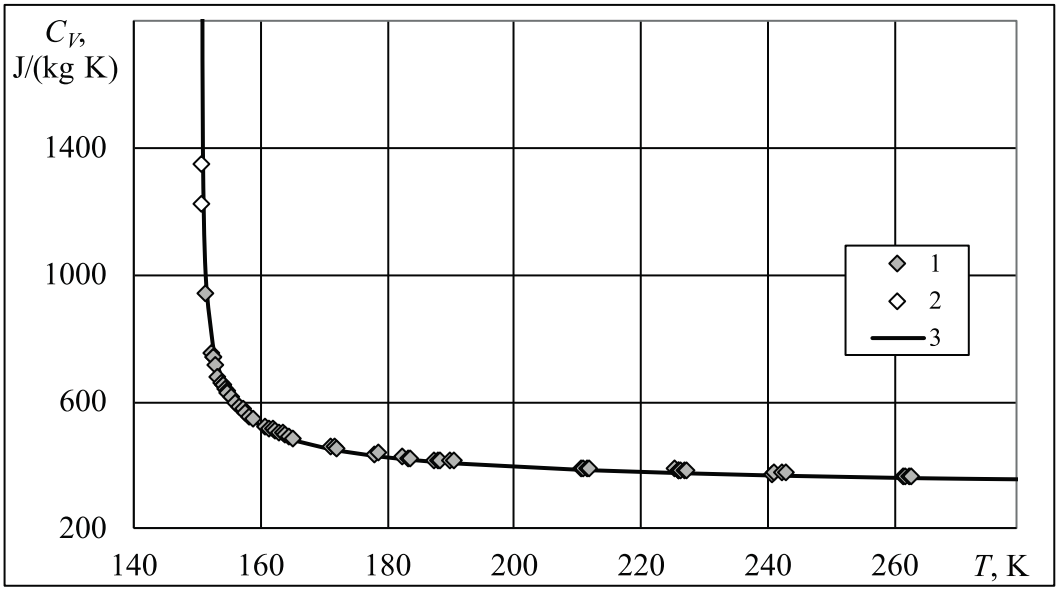


Figure 6: Behavior of the isochoric heat capacity of argon in a single-phase region: 1—experimental points [48] over the isochor of 531 kg/m^3; 2—experimental points [50] over the isochor 530 kg/m^3; 3—C_V calculated with a help of the FEoS (1) over isochor 530.5 kg/m^3.

4 CONCLUSIONS

On the basis of a new representation of the scaling hypothesis [34, 35] and the Berestov equation [2], FEoS (1) was developed. This FEoS primarily works satisfactorily in a wide range of pressures and temperatures including the critical range and the range of high temperatures and pressures. FEoS (1) with components (2)–(4) and (7) has the properties of the virial series in the regular part of the thermodynamic surface as well as the properties of the Berestov equation in the critical range.

Argon FEoS (1) can be used to calculate the equilibrium properties in various technology processes. We have analyzed properties calculated with the help of FEoS (1) in the vicinity of the critical point. Our values significantly exceeds the accuracy of the data generated with the help of known FeoSs and known crossover EoSs [6, 9, 12, 17].

The proposed method of constructing FEoS can be recommended for developing EoSs of substances, which have reliable experimental data, for example, it carbon dioxide and sulfur hexafluoride.

Calculated values of properties from the FEoS (1) to verify computer code are $T = 400$ K, $\rho = 1000$ kg/m^3, $p(T,\rho) = 168\,974.25$ kPa, $C_V(T,\rho) = 0.392\,0699$ J/(g K).

Acknowledgments: The paper is based on the proceedings of the XXXIV International Conference on Interaction of Intense Energy Fluxes with Matter, Elbrus, the Kabardino-Balkar Republic of the Russian Federation, March 1 to 6, 2019.

REFERENCES

[1] P. Schofield, J. D. Litster, and J. T. Ho, "Correlation between critical coefficients and critical exponents", *Phys. Rev. Lett.*, **23**(19), 1098–1102 (1969).

[2] A. T. Berestov, "Equation of state in the critical region with inclusion of non-asymptotic terms", *Zh. Eksp. Teor. Fiz.*, **72**(1), 348–353 (1977).

[3] C. Ronchi, "Extrapolated equation of state for rare gases at high temperatures and densities", *J. Nucl. Mater.*, **96**(3), 314–328 (1981).

[4] V. A. Rykov and G. B. Varfolomeeva, "Method of determining a structural form of the free energy satisfying the requirements of the scaling hypothesis", *J. Eng. Phys. Thermophys.*, **48**(3), 341–345 (1985).

[5] V. A. Rykov, "Method of constructing a single equation of state satisfying the requirements of the scaling hypothesis", *J. Eng. Phys. Thermophys.*, **48**(4), 476–481 (1985).

[6] R. B. Stewart and R. T. Jacobsen, "Thermodynamic properties of argon from the triple point to 1200 K with pressures to 1000 MPa", *J. Phys. Chem. Ref. Data*, **18**(2), 639–798 (1989).

[7] A. D. Kozlov, V. F. Lysenkov, P. V. Popov, and V. A. Rykov, "Unified nonanalytical equation of state for freon-218", *J. Eng. Phys. Thermophys.*, **62**(6), 611–617 (1992).

[8] V. F. Lysenkov, A. D. Kozlov, P. V. Popov, and M. V. Yakovleva, "Nonanalytical unified equation of state of freezant R23", *J. Eng. Phys. Thermophys.*, **66**(3), 286–294 (1994).

[9] Ch. Tegeler, R. Span, and W. Wagner, "A new equation of state for argon covering the fluid region for temperatures from the melting line to 700 K at pressures up to 1000 MPa", *J. Phys. Chem. Ref. Data*, **28**(3), 779–850 (1999).

[10] L. Sun, S. B. Kiselev, and J. F. Ely, "Multiparameter crossover equation of state: Generalized algorithm and application to carbon dioxide", *Fluid Phase Equilib.*, **233**(2), 204–219 (2005).

[11] I. V. Kudryavtseva, V. A. Rykov, and S. Rykov, "Asymmetric unified equation of state for R134a", *Journal of IAR*, **2**, 36–39 (2008).

[12] A. Rizi and A. Abbaci, "A thermodynamic equation of state for the critical region of argon", *J. Mol. Liq.*, **171**, 64–70 (2012).

[13] V. I. Mazhukin, A. V. Shapranov, O. N. Koroleva, and A. V. Rudenko, "Molecular dynamics simulation of critical point parameters for silicon", *Math. Montis.*, **31**, 64–77 (2014).

[14] P. P. Bezverkhy and V. G. Martynets, "Calculation of thermodynamic properties of CO_2 using the combined thermal equation of state with a small number of adjustable parameters", *High Temp.–High Pressures*, **45**(2), 145–154 (2016).

[15] S. V. Rykov, I. V. Kudryavtseva, V. A. Rykov, M. I. Poltoratskiy, and A. V. Sverdlov, "Equation of state for refrigerant R32", *Kholod. Tekh.*, **11**, 34–37 (2016).

[16] I. V. Kudryavtseva and S. V. Rykov, "A nonparametric scaling equation of state, developed on the basis of the Migdal's phenomenological theory and Benedek's hypothesis", *Russ. J. Phys. Chem.* A, **90**(7), 1493–1495 (2016).

[17] A. B. Kaplun and A. B. Meshalkin, "Simple fundamental equation of state for liquid, gas, and fluid of argon, nitrogen, and carbon dioxide", *Thermophys. Aeromech.*, **55**, 513–522 (2017).

[18] M. I. Poltoratskiy, *Method of constructing a fundamental equation of state and thermodynamic tables for hexafluoropropane (R236ea)*, PhD thesis, Saint-Petersburg: ITMO University, (2018).

[19] I. V. Kudryavtseva, V. A. Rykov, S. V. Rykov, and E. E. Ustyuzhanin, "A new variant of a scaling hypothesis and a fundamental equation of state based on it", *J. Phys.: Conf. Ser.*, **946**, 012118 (2018).

[20] S. V. Rykov, I. V. Kudryavtseva, V. A. Rykov, and A. V. Zaitsev, "Methods for calculating equilibrium properties of pure substances, considering the critical point features", *J. Phys.: Conf. Ser.*, **1128**, 012106 (2018).

[21] O. N. Koroleva, M. M. Demin, V. I. Mazhukin, and A. V. Mazhukin, "Modeling of thermal conductivity of Si in the range from the normal to near-critical conditions", *Math. Montis.*, **45**, 85–94 (2019).

[22] I. V. Kudryavtseva, V. A. Rykov, and S. V. Rykov, "The method for constructing the fundamental equation of state for SF_6", *J. Phys.: Conf. Ser.*, **1385**, 012009 (2019).

[23] V. A. Rykov, S. V. Rykov, and A. V. Sverdlov, "Fundamental equation of state for R1234yf", *J. Phys.: Conf. Ser.*, **1385**, 012013 (2019).
[24] S. V. Rykov, I. V. Kudryavtseva, V. A. Rykov, and E. E. Ustyuzhanin, "Scaling Migdal model and a nonparametric equation of state for argon", *J. Phys.: Conf. Ser.*, **1147**, 012018 (2019).
[25] V. E. Fortov, K. V. Khishchenko, P. R. Levashov, and I. V. Lomonosov, "Wide-range multi-phase equations of state for metals", *Nucl. Instr. Meth. Phys. Res.* A, **415**, 604–608 (1998).
[26] K. V. Khishchenko, "Equations of state for two alkali metals at high temperatures", *J. Phys.: Conf. Ser.*, **98**, 032023 (2008).
[27] I. V. Lomonosov and S. V. Fortova, "Wide-range semiempirical equations of state of matter for numerical simulation on high-energy processes", *High Temp.*, **55**, 585–610 (2017).
[28] K. V. Khishchenko, "Equation of state of sodium for modeling of shock-wave processes at high pressures", *Math. Montis.*, **40**, 140–147 (2017).
[29] K. V. Khishchenko, "Equation of state for lithium in shock waves", *Math. Montis.*, **41**, 91–98 (2018).
[30] K. K. Maevskii, "Thermodynamic parameters of lithium deuteride in pressure range 5–1000 gigapascals", *Math. Montis.*, **41**, 123–130 (2018).
[31] K. V. Khishchenko, "Equation of state for magnesium hydride under conditions of shock loading", *Math. Montis.*, **43**, 70–77 (2018).
[32] K. V. Khishchenko, "Equations of state for rubidium and cesium at high pressures in shock waves", *J. Phys.: Conf. Ser.*, **1147**, 012001 (2019).
[33] Sh. Ma, *Modern Theory of Critical Phenomena*, New York, NY: Roudedge, (1976).
[34] S. V. Rykov, I. V. Kudryavtsev, and V. A. Rykov, "The physical basis of the method of pseudocritical points", *Scientific and Technical Bulletin of Povolzhie*, **2**, 44–47 (2014).
[35] S. V. Rykov, "Fundamental equation of state taking into account liquid asymmetry", *Scientific and Technical Bulletin of Povolzhie*, **1**, 33–36 (2014).
[36] G. B. Benedek, in: *Polarization Matiere et Rayonnement, Livre de Jubile en l'Honneur du Professeur A. Kastler* Paris: Presses Universitaires de Paris, 71 (1968).
[37] A. Michels, J. M. Levelt, and G. J. Wolkers, "Thermodynamic properties of argon at temperatures between 0 °C and −140 °C and at densities up to 640 amagat (pressures up to 1050 atm.)", *Physica*, **24**(6), 769–794 (1958).
[38] S. L. Robertson, S. E. Babb, and G. J. Scott, "Isotherms of argon to 10 000 bars and 400 °C", *J. Chem. Phys.*, **50**(5), 2160–2166 (1969).
[39] W. B. Streett and L. A. K. Staveley, "Experimental study of the equation of state of liquid argon", *J. Chem. Phys.*, **50**(6), 2302–2307 (1969).
[40] A. Van Itterbeek and O. Verbeke, "Density of liquid nitrogen and argon as a function of pressure and temperature", *Physica*, **26**(11), 931–938 (1960).
[41] E. Whalley, Y. Lupien, and W. G. Schneider, "The compressibility of gases: VII. Argon in the temperature range 0–600 °C and the pressure range 10–80 atmospheres", *Can. J. Chem.*, **31**(8), 722–733 (1953).
[42] A. Michels, J. M. Levelt, and W. D. Graaff, "Compressibility isotherms of argon at temperatures between −25 °C and −155 °C, and at densities up to 640 amagat (pressures up to 1050 atmospheres)", *Physica*, **24**(6), 659–671 (1958).
[43] A. Lecocq, "Determination experimentale des equations d'etat de l'argon jusqua 1000 °C et 1000 kg/m^3", *J. Recherches Centre National Recherche Scientific*, **55**, 55–82 (1960).
[44] R. Gilgen, R. Kleinrahm, and W. Wagner, "Measurement and correlation of the (pressure, density, temperature) relation of argon I. The homogeneous gas and liquid regions in the temperature range from 90 K to 340 K at pressures up to 12 MPa", *J. Chem. Thermodyn.*, **26**(4), 383–398 (1994).
[45] V. P. Skripov, E. N. Sinitsyn, P. A. Pavlov, G. V. Ermakov, G. N. Muratov, N. V. Bulanov, and V. Baidakov, *Teplofizicheskiye Svoystva Zhidkostey v Metastabil'nom Sostoyanii (Thermophysical Properties of Liquids in a Metastable State)*, Moscow: Atomizdat, (1980).
[46] J. Thoen, E. Vangeel, and W. V. Dael, "Experimental investigation of the sound velocity in the critical region of argon", *Physica*, **52**(2), 205–224 (1971).
[47] C. Gladun, "The specific heat of liquid argon", *Cryogenics*, **11**(3), 205–209 (1971).

[48] M. A. Anisimov, B. A. Koval'chuk, V. A. Rabinovich, and V. A. Smirnov, *Teplofizicheskiye Svoystva Veshchestv i Materialov (Thermophysical properties of substances and materials)*, **12**, 85–106 (1978).
[49] R. Gilgen, R. Kleinrahm, and W. Wagner, "Measurement and correlation of the (pressure, density, temperature) relation of argon II. Saturated-liquid and saturated-vapour densities and vapour pressures along the entire coexistence curve", *J. Chem. Thermodyn.*, **26**(4), 399–413 (1994).
[50] A. V. Voronel' and Yu. R. Chashkin, "Specific heat C_v of argon as a function of density near the critical point", *Sov. Phys. JETP*, **24**(2), 263–267 (1967).
[51] J. Klimeck, R. Kleinrahm, and W. Wagner, "An accurate single-sinker densimeter and measurements of the (p, ρ, T) relation of argon and nitrogen in the temperature range from (235 to 520) K at pressures up to 30 MPa", *J. Chem. Thermodyn.*, **30**(12), 1571–1588 (1998).
[52] A. V. Voronel', V. G. Snigirev, and Yu. R. Chashkin, "Behavior of the specific heat C_v of pure substances near the critical point", *Sov. Phys. JETP*, **21**(3), 653–655 (1965).
[53] A. Michels, H. Wijker, and H. Wijker, "Isotherms of argon between 0 °C and 150 °C and pressures up to 2900 atmospheres", *Physica*, **15**(7), 627–633 (1949).
[54] R. K. Crawford and W. B. Daniels, "Equation-of-state measurements in compressed argon", *J. Chem. Phys.*, **50**(8), 3171–3183 (1969).

Received November 15, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

MOLECULAR-DYNAMIC MODELING OF THERMOPHYSICAL PROPERTIES OF PHONON SUBSYSTEM OF COPPER IN WIDE TEMPERATURE RANGE

M.M. DEMIN, O.N. KOROLEVA*, A.A. ALEKSASHKINA, V.I. MAZHUKIN

Keldysh Institute of Applied Mathematics, Russian Academy of Science.
*Corresponding author. E-mail: koroleva.on@mail.ru

DOI: 10.20948/mathmontis-2020-47-12

Summary. Copper is a noble metal and has unique properties, due to which it is widely used in scientific research, industrial production and, more recently, in the problems of biomedicine. Using the molecular dynamics method, a series of calculations was performed to determine the lattice thermophysical properties of copper in a wide temperature range of 300K ≤ T ≤ 5800K. In the calculations, special attention is paid to the melting-crystallization and near-critical regions, in which cardinal changes in the thermophysical properties of the substance occur. The temperature dependences of the specific heat *Cp(T)*, thermal conductivity *κ(T)*, and density *ρ(T)* were among the studied characteristics of the phonon subsystem of Cu. Molecular - dynamic modeling was carried out using the potential of the embedded atom method (EAM). A comparison of the results with the results of experiments and alternative calculations showed a good agreement. The obtained calculation results were approximated by polynomials of low degrees..

1 INTRODUCTION

Copper, a noble metal, occupies an important place among metals in importance and prevalence in many branches of scientific research and innovative technological applications [1-5]. Such a relatively new, rapidly developing direction is the production of metal-based nanomaterials, including copper, which attracts general attention by its wide applicability [5-8]. Due to its unique properties, copper nanoparticles have gained the possibility of being used primarily in problems of theranostics and nanomedicine [1-3, 5]. One of the rapidly developing methods for the production of nanomaterials is pulsed laser ablation of metals (PLA) [5-9]. The increasing possibilities of using PLA in the production of nanoparticles make this direction attractive for basic research, the main tool of which is mathematical modeling. Continuous, atomistic, and combined models are used for the mathematical description of fast and highly nonequilibrium processes induced by ultrashort laser pulses in the metal targets [9-15]. For continuum and combined models, the properties of the substance are input parameters. Therefore, one of the most important problems of mathematical modeling is the need to determine for each of the subsystems the thermophysical, optical, and thermodynamic characteristics in a wide temperature range - from room temperature T_0 = 300 K to critical T_{cr}.

The most important thermophysical properties characterizing heat transfer in metals are density *ρ(T)*, specific heat *Cp(T)* and thermal conductivity *κ(T)*.

2010 Mathematics Subject Classification: 82-08, 82D35, 82B27.
Key words and Phrases: Molecular Dynamics Simulation, Near-Critical Region, Copper, EAM potential.

The specific heat of metals at the temperature T> 300 K is mainly determined by the lattice vibrations. The contribution of the electronic subsystem to the specific heat is noticeable only at a low temperature T <10 K [16].

Unlike specific heat, the contribution of the electronic component to the thermal conductivity of copper is significant. According to [17], the contribution of the electronic component to the total thermal conductivity is ≈ 95%. Since in metals the main part of the heat flux is carried by conduction electrons, it was believed that lattice thermal conductivity does not play a significant role. Therefore, it was not necessary to separate the total thermal conductivity into electronic and phonon components.

The interest in quantifying the thermophysical characteristics of the phonon subsystem of metals was stimulated primarily by the need for a deeper understanding of the mechanisms of thermal transfer during nonequilibrium energy transfer in a number of applications, for example, [18, 19].

Due to the limited possibilities of instrumental measurement of the thermophysical characteristics of the material under study at high temperatures ($T > T_m$, where T_m is the melting temperature), computational approaches become relevant.

Significant progress, first of all, in the development of (numerical) atomic modeling methods (and computational algorithms) allows one to determine the thermophysical characteristics (phonon specific heat, phonon thermal conductivity, etc.) in a wide temperature range for most metals [20-30] and semiconductors [24, 25 , 31-34] with a sufficient degree of accuracy.

The aim of this work is to obtain lattice thermophysical characteristics ($\rho(T)$, $Cp(T)$, $\kappa(T)$) of copper in a wide temperature range $T_0 \leq T \leq T_{cr}$ using the molecular dynamics method with the EAM potential [21].

2 METHODS AND APPROACHES

The determination of the thermophysical properties of the phonon subsystem of copper in this work is based on the atomistic approach. Atomistic models rely on the molecular dynamics (MD) method. The MD method is based on a model representation of a polyatomic molecular system in which all atoms are represented by material points, and the motion is described in the classical case by Newton's equations. Because of this, atomistic models are a system of differential equations, the integration of which requires knowledge of the coordinates and velocities of all particles at the initial time t = 0. The resulting ODE system is solved using the Verlet finite-difference scheme [35].

When using atomistic models to study various properties of substances, the most important role is played by the choice of interaction potentials between particles, since the reliability of the results obtained directly depends on it. In molecular dynamics modeling of the properties of metals, the empirical and semi-empirical potentials of the "embedded atom method" (EAM) are mainly used as the interparticle interaction potentials [22, 23]. Since pair and collective interactions are taken into account in the EAM potentials, the potential energy of the metal is the sum of the embedding potential of the ith atom, which depends on the effective electron density in the region where the center of the atom and the pair potential are located. However, the EAM potentials do not take into account the phonon-electron interaction, which is their drawback. The disadvantages of these potentials include a large number of fitting parameters included in them (up to two dozen). When choosing the potential

for atomistic modeling, careful testing of the potential used is necessary, since not all the EAM potentials used allow a good description of both the crystalline and liquid phases of the metal. In this work, we used the EAM potential for copper developed and tested in [21].

Of all the thermophysical properties, the determination of phonon thermal conductivity in the framework of classical molecular dynamics occupies a special place, being a complex problem.

The direct method (DM) was chosen as an approach for determining the temperature dependence of the phonon thermal conductivity of copper using molecular dynamics (MD). This method is the most simple and economical from a computational point of view.

The direct method (DM) [24-26] is one of the most common methods for calculating thermal conductivity. DM is a nonequilibrium method of molecular dynamics (NEMD); it is based on applying a temperature gradient to the modeling cell, for which it received its name "heat source - sink". Due to this, the direct method is similar to the experimental situation. One of the advantages of DM is the saving of computing resources, which is very important, sometimes determining the choosing of the modeling method. For example, as noted in [24], for the direct method, the simulation time of 1 ns is sufficient to obtain a smooth temperature profile, and the value of κ converges with an accuracy of ± 10%. The method demonstrates the finite-size effects, which are its drawback. These effects arise if the mean free path of phonons is comparable to the size of the simulation cell. In this regard, the necessary size of the computational domain to achieve a completely convergent value of κ may be beyond the reach of atomistic modeling and it becomes necessary to impose a restriction on the smallest length of the computational domain. In this connection, the thermal conductivity of copper can be obtained by the direct method from modeling systems of different sizes and extrapolating the results to a system of an infinite size.

Along with classical molecular dynamics, the ab-initio approach is also used to determine the phonon thermal conductivity [17, 36]. The ab initio methods have appeared recently and are considered the most promising. They do not require specifying the interparticle potential and can be applied to any material. However, ab-initio methods have limitations associated with an increase in computational costs with an increase in model size. The use of these methods in the calculation of phonon thermal conductivity allows one to take into account the influence of both phonon-phonon (p-p) and phonon-electron (p-e) interactions, which can significantly increase the reliability of the results. However, the number of calculations of the phonon thermal conductivity of metals and, in particular, copper [17, 36], is currently relatively small. As a rule, all calculations are limited to the solid-state phase in the temperature range T ~ (300 - 1000) K. There are no systematic results of experimental-theoretical studies of the properties of liquid metals in a wide temperature range (from the beginning of melting to the critical region).

The wide temperature range 300K $\leq T \leq$ 5800K, in which the thermophysical properties of copper are determined in this work, covers the first-order phase transition (melting-crystallization) and the near-critical region in which drastic changes in the thermophysical properties of the substance occur. Therefore, calculations of the properties of copper in this range cannot be carried out without knowledge of such important characteristics as the melting temperature T_m and critical parameters: temperature T_{cr}, density ρ_{cr}, pressure P_{cr}.

The equilibrium melting temperature used in the calculations was obtained from molecular dynamics calculations in [27] with the EAM potential [21], which is also used in this work, by the two-phase method [28]. We used a system with an ensemble of particles of 8000 atoms.

The obtained value is T_m = 1330 K, slightly lower than the reference (T_m = 1356 K) [37] and experimental value (T_m = 1357.7 K) [38] and deviates from these values by 1.9% and 2.06%, respectively. The error is quite acceptable for modeling. In this work, the following values of the critical parameters of copper were taken: $T_{cr} \approx 6550$ K, $\rho_{cr} \approx 1.895$ g / cm3, $P_{cr} \approx 0.16$ GPa, obtained in [30] using the liquid – vapor coexistence curve.

The specific heat and thermal conductivity of the phonon subsystem, as well as the copper density characterizing heat transfer, were simulated in the temperature range 300 ≤ T ≤ 5800 K using the well-known LAMMPS application package (large-scale atomic-molecular massively parallel simulator) [39]. It implements many paired and multiparticle potentials, the ability to save atomic configurations in a text file, as well as built-in thermostats and barostats. The velocity and pressure for the ensemble of particles were controlled using the thermostat and barostat of Berendsen [40].

3 MODELING RESULTS

For convenience of further use, the calculation results are approximated by polynomials of the degree of m:

$$P_{\mathrm{m}}(\mathrm{x}) = \sum_{k=0}^{m} a_k x^k , \tag{1}$$

where a_k are the polynomial coefficients.

The approximation error was calculated by the least squares criterion [41]:

$$\Delta\left(P_m\left(t_j\right), y_j\right) = \sqrt{\left(1/n+1\right)\sum_{j=0}^{n}\left(P_m\left(t_j\right) - y_j\right)^2} \to \min, \tag{2}$$

where y_j are the values of the variable from the results of calculations for t_j $(j=0,...,n)$.

3.1 Calculation of specific heat and density of copper

The traditional way to determine the thermophysical properties of metals is experiment. For copper, the experimental values of density [42–44] and heat capacity in the solid [45,46] and liquid [46, 47] phases are known. The experimental approach has limitations, primarily for the temperature range. The copper density was obtained experimentally [42] in a wide temperature range of 300 K ≤ T ≤ 5000 K, and the heat capacity of the liquid phase was obtained up to 2000K [46]. A scatter of values is also observed in various experiments, which was noted in [44].

However, when simulating the laser ablation processes, the modeling enters the higher temperature region, including the region of the critical point, so the known data becomes insufficient. In addition, for working with mathematical models, it is relevant to obtain temperature dependences in a wide range of parameters.

The temperature dependences of the density $\rho(T)$ and specific heat $C_p(T)$ of copper were determined from a series of molecular dynamics calculations within the framework of one computational experiment.

We used a cubic computational region of 30×30×30 unit cells containing an fcc crystal of 108,000 particles (fluctuations that are too large arise with smaller sizes). Periodic boundary conditions were set. The relaxation procedure preceding the simulation was carried out at a

temperature of 300 K and zero pressure. After that, the slow heating of the sample with a constant rate of approximately 0.5 K/ps continued to a temperature of 6000 K. During MD calculations, the temperature dependences were recorded: density $\rho(T)$ and enthalpy $H(T)$. The experiment was carried out at a constant zero pressure $P = 0$.

3.1.1 Density of copper

As a result of MD calculations in the range 300K <T<5620K, the temperature dependence of the density of copper $\rho(T)$ was obtained, which, after additional statistical processing, is shown in Figure 1. The markers in this figure show the experimental data [42]. The vertical dashed lines indicate the melting temperature T_m [27] and the critical temperature T_{cr} [30] of copper. Figure 1 shows the changes in the density of copper at the equilibrium melting temperature (T_m = 1330K). The density of the copper melt is lower than the density of the crystal at the same temperature T_m, i.e. copper melts with decreasing density, similar to what was observed in experiments [42, 43]. At the equilibrium melting temperature T_m, the density changes abruptly. The density difference between the solid and liquid phases in the calculations is 5.2%, and in the experiment [42] - 4.4%, which shows a fairly good agreement between the results.

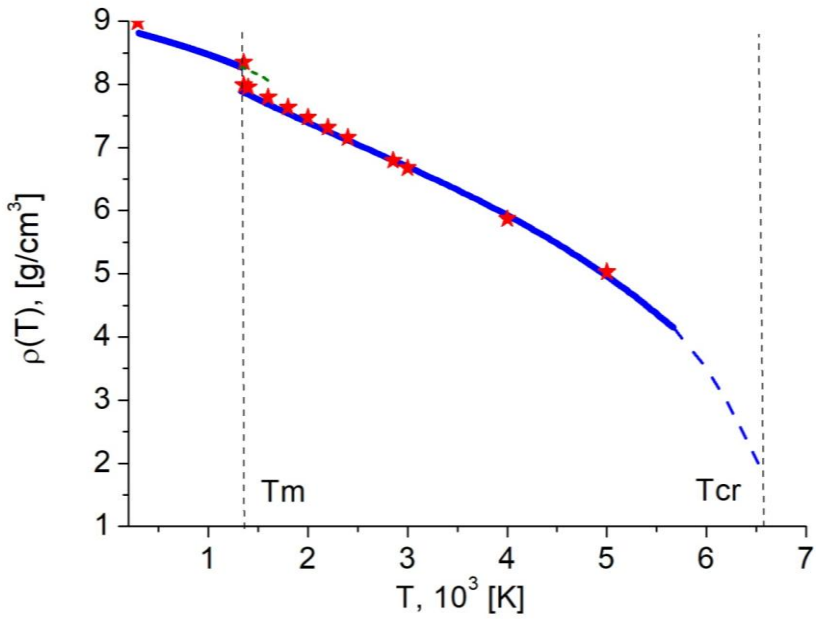


Fig. 1. The temperature dependence of the density of copper. The markers show the experimental data [42].

It is also noticeable that the density decreases to $\rho(T) \approx 8$ g/cm^3 upon overheating of the solid phase in the temperature range $T_m < T < 1.2T_m$ (green dashed line in Figure 1). In the liquid phase, with increasing temperature $T > T_m$, the copper density decreases. The results obtained show good agreement with the experimental results [42]. At $T \approx 3000$ K, the density value in this work is $\rho(T)$ = 6.678 g/cm^3, in the experiment - $\rho(T)$ = 6.675 g/cm^3 the values differ by 0.04%. At $T \approx 5000$ K in this work, $\rho(T)$ = 4.956 g/cm^3, in the experiment - $\rho(T)$ = 5.03 g/cm^3, the difference is 1.49%. At $T \approx 5620$ K, a temperature close to the critical

temperature T_{cr} = 6550 K [30], the density $\rho(T)$ = 4.207 g/cm^3 was obtained. In Fig. 1, the dashed line shows the continuation of the function $\rho(T)$, which extrapolates the density of copper to the critical point. At T_{cr} = 6550 K, the density $\rho_{cr} \approx$ 1.895 g/cm^3 was obtained in [30].

The results obtained in the calculations are more convenient to use as an analytical dependence of the form (1). For solid and liquid phases, dependences of the third degree are obtained

$$\rho(T) = a_0 + a_1(T - T_0) + a_2(T - T_0)^2 + a_3(T - T_0)^3$$

For the solid phase, ($300\ K \le T \le T_m$), $T_0 = 300$ K, for the liquid phase, ($T_m \le T \le 5620K$), $T_0 = T_m$. The values of the coefficients a_k and the approximation errors by the least squares criterion (2) are presented in table 1.

k	Solid	Liquid
a_0	8,81	7.89
a_1	-4.28×10^{-4}	-7.96×10^{-4}
a_2	-6.12×10^{-8}	8.69×10^{-8}
a_3	-2.77×10^{-11}	$-2{,}383\times10^{-11}$
$\Delta(P(x_j), y_j)$	0.001	0.014

Table 1. The coefficients a_k of the function, which approximates the calculation results of the copper density $\rho(T)$ in g/cm^3.

3.1.2 Specific heat of copper

The temperature dependence of the specific heat of the lattice $C_p(T)$ in the temperature range 300 K < T <5800 K at constant pressure P in this work was determined from the enthalpy $H(T)$ obtained during the computational MD experiment considered above. The values of $H(T)$ were approximated for the liquid and solid phases by polynomials $\tilde{H}(T)$ (1). The temperature dependence of the specific heat $C_p(T)$ for each phase was determined by differentiating the corresponding dependence $\tilde{H}(T)$:

$$C_p(T) = \left(\frac{\partial \tilde{H}(T)}{\partial T}\right)_P \tag{3}$$

Fig. 2. shows the temperature dependence of the specific heat of copper $C_p(T)$ according to the results of calculations of the present work, the markers show the reference and experimental results [37, 38]. The vertical dashed lines indicate the melting temperature T_m and the critical temperature T_{cr} of copper. The region of the solid – liquid phase transition, with zooming, is shown in the inset of Fig. 2. It is seen that at the equilibrium melting temperature T_m, an insignificant jump by about ~ 3.128% occurs, an abrupt decrease in the heat capacity of copper. According to the experiment [38], this value is ~ 1.529%. On the inset of fig. 2 one can clearly see an increase in the specific heat to $C_p(T) \approx 39$ J·mol^{-1}·K^{-1} upon overheating of the solid phase in the temperature range $T_m < T < 1.2T_m$ (green dashed

line in Figure 2). With increasing temperature $T_m < T < 2.63T_m$ in the liquid phase, the specific heat is almost constant and amounts to $C_p(T) \approx 31.0$ J·mol^{-1}·K^{-1}, which is 8% less than $C_p(T) \approx 33.84$ [38]. At the temperature $T > 4000$ K, the specific heat of copper increases. At $T = 5800$ K, its value is $C_p(T) = 47.698$ J·mol^{-1}·K^{-1}. In the near-critical region in Fig. 2, the dashed line shows the extrapolation of the temperature dependence of the specific heat to the critical point.

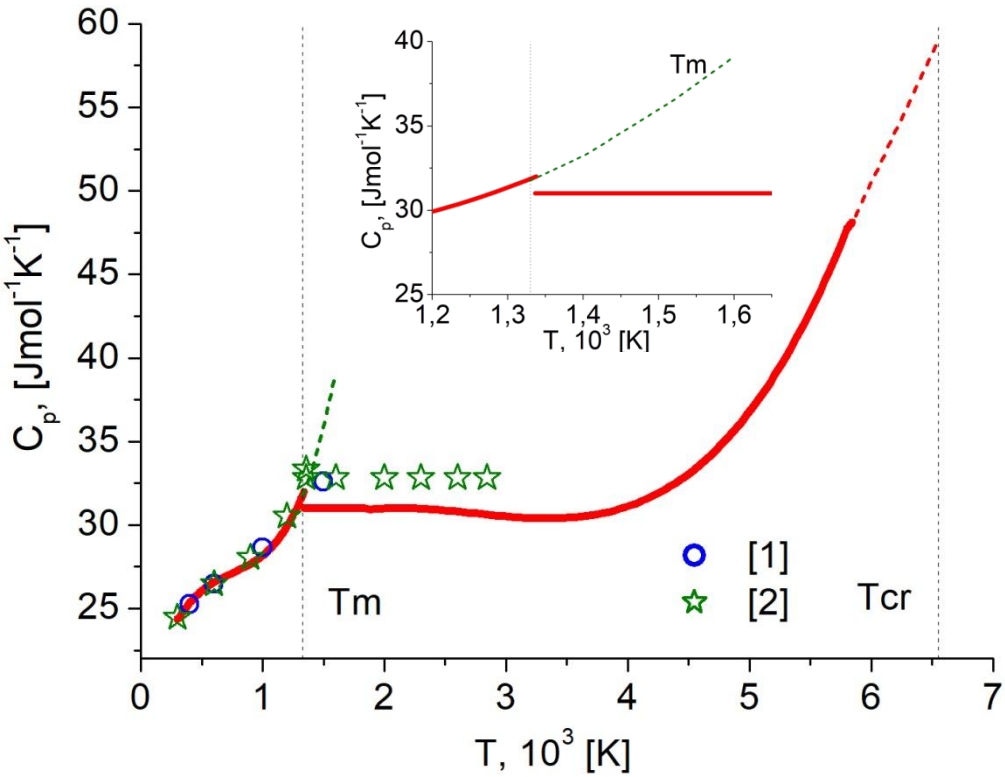


Fig.2. The temperature dependence of the specific heat Cp (T) of copper according to the calculation results (solid line). Markers: (1), (2) - experimental results [37, 38].

For the use in further calculations, the results obtained are more convenient to use in the form of an analytical dependence of the form (1). For the solid and liquid phases, the results were approximated by power dependences of the 4th degree:

$$C_p(T) = a_0 + a_1(T - T_0) + a_2(T - T_0)^2 + a_3(T - T_0)^3 + a_4(T - T_0)^4,$$

where $T_0 = 300$ K for the solid phase ($300\ K \le T \le T_m$), $T_0 = T_m$ for the liquid phase ($T_m \le T \le 5800K$). The values of coefficients a_k and approximation error according to the least squares criterion (2) are shown in the Table 2.

k	Solid	Liquid
a_0	24.27	31.0018
a_1	1.23×10^{-2}	3.28×10^{-3}
a_2	-2.05×10^{-5}	-2.91×10^{-6}
a_3	$1{,}53\times10^{-8}$	7.06×10^{-10}
a_4	-2.88×10^{-12}	-4.68×10^{-15}
$\Delta(P(x_j), y_j)$	0.021	0.038

Table 2. The values of coefficients a_k for the function, which approximates the calculation results of specific heat of copper $C_p(T)$ $J\cdot mol^{-1}\cdot K^{-1}$

3.2 Calculation of thermal conductivity of copper

To determine the thermal conductivity of the phonon subsystem of copper, a series of calculations was carried out based on molecular dynamics modeling. The phonon thermal conductivity was determined using the direct method (DM) [24.25].

When using DM, the heat source and sink areas are created in the modeling domain to apply a constant heat flux along the direction of interest.

At each time step, a fixed amount of heat dQ_N was deposited in the heating region, and the same amount was taken from the sink region. The heat flux W was calculated as

$$W = dQ/(SNdt)/2, \tag{4}$$

where $dQ = N\times dt\times\delta Q_N$ is the total deposited energy, δQ_N is the energy deposited during one timestep, N is the number of timesteps, dt is the size of the timestep, S is the domain cross-section. The timestep size was chosen depending on the temperature, from 3fs at 300K to 1fs at 2000K and above. The division by 2 is used due to periodic boundary conditions, i.e. heat flux goes in two directions. Then, the resulting temperature gradient is calculated, and the thermal conductivity coefficient κ_{lat} was determined from the known heat flux by the Fourier law [16]

$$W = -\kappa_{lat}\frac{\partial T}{\partial x} \tag{5}$$

where W is the heat flux, x is the coordinate in the direction of the flux.

The difficulty in applying the direct method to solids lies in the fact that the size of the modeling region should be much larger than the mean free path of phonons in a substance. For a crystal, this is difficult to do, because requires a very large size of the computational domain and, accordingly, a very large number of atoms. Therefore, when calculating with a small number of atoms, the thermal conductivity coefficient is dependent on the length of the region due to phonon scattering at the boundary. To limit the size of the simulation region, a scaling procedure is used, in which the thermal conductivity is determined for several lengths of the simulation region L_n (n is the number of unit cells in the computational domain) along the x direction. Then, the inverse dependence of the thermal conductivity $1/\kappa_{lat}$ is constructed with respect to the reciprocal of the length of the simulation region, $1/L_n$, and the thermal conductivity is determined by extrapolating the data $1/L_n \to 0$ [24-26].

To determine the thermal conductivity of copper, the simulation domain in the form of a parallelepiped was considered. The initial sizes of the region were 10×10×20 unit cells (lattice constant 0.361 nm), corresponding to 8000 particles. Periodic boundary conditions were set along the three axes. As the interaction potential, the EAM potential is used [21]. The particle velocities were set as random variables with a Maxwell distribution corresponding to a double temperature of 600 K. Then the sample was equilibrated at 300 K using a thermostat and barostat.

The sample was divided along the x axis into the number of cells corresponding to the number of particles. At each step, a certain amount of heat is deposited in the first interval, and the same amount of heat is taken from the middle of the sample, where the drain is located. After some time (for small samples 5 ns, for large 10 ns), a stationary equilibrium is established. The temperature difference is calculated at 0.8 of the entire length between the heat source and the sink over the last 0.5 ns and averaged.

3.2.1 Modeling results of thermal conductivity

To calculate the thermal conductivity from the Fourier law (5), it is necessary to determine the value of the heat flux *W* (4) from the spatial temperature distribution obtained from the MD modeling. Figure 3 shows the time-averaged spatial temperature profile for an average temperature of 300 K. In a small region (~ 6 nm) in the immediate vicinity of the source, a very strong nonlinear temperature profile is observed. The same strong nonlinear temperature profile is also observed near the sink in the middle of the computational domain. In the intermediate region, the behavior of the temperature profile is close to a linear dependence. The intermediate region in Fig. 3 is indicated by dashed lines. In this interval between the heat source and sink, the temperature gradient was measured. The presence of the heat source and heat sink, the use of periodic boundary conditions creates a current in two opposite directions.

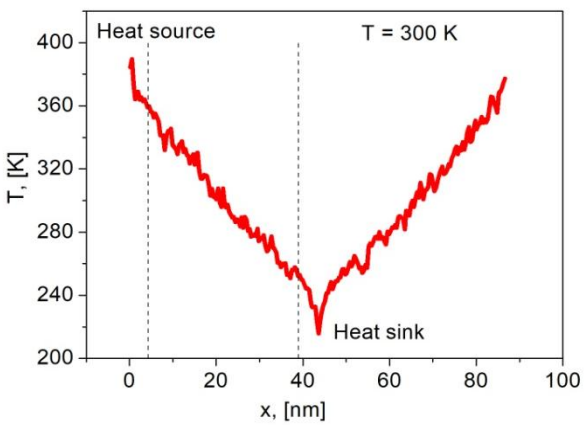


Fig. 3. Spatial temperature profile at one moment in time. The dashed lines indicate the interval in which the temperature gradient was measured.

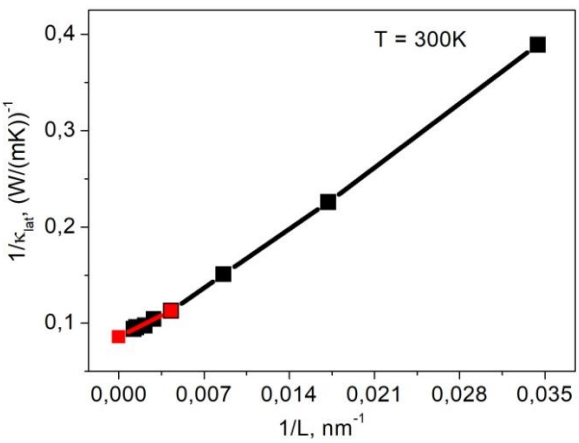


Fig. 4. The dependence of the reciprocal of the thermal conductivity on the reciprocal of the size of the region for the temperature T = 300K.

To overcome the effects of the finite size, in accordance with the scaling procedure, the heat flux was determined by a series of calculations for various sizes of the computational

domain. The number of computational regions and their sizes depended on the temperature for which the thermal conductivity was calculated. The lower the temperature, the more areas were selected. For the range 300K ≤ T ≤ 900K, the calculations were carried out for 8 regions of different sizes L_n: 20, 40, 80, 160, 240 320, 480, 560 unit cells corresponding to the number of particles. For the range 1200K ≤ T ≤ 2000K, 3 calculations were performed for the sizes of the L_n region: 80, 160, 240 cells, and for T ≥ 4000K, only one calculation was performed for the size of the region L_n = 80 cells (for larger size, the results were the same). The cross section of the region was constant: S = 10 × 10 cells. The heat flux was determined from the temperature difference between the heating and heat sink areas, for which the instantaneous temperature difference was averaged over the entire calculation time after establishing a stationary distribution. To increase the accuracy of the calculations, the temperature difference was calculated not over the entire interval between the source and the sink, but in its central part, 0.8 of the total length.After a series of calculations, for each temperature from the range 300K <T <5700K, the scaling procedure was performed and thermal conductivity was calculated. Let us consider the scaling procedure using the example of calculating the thermal conductivity of copper for 300 K. Figure 4 shows the dependence of the reciprocal of the thermal conductivity on the reciprocal of the size of the region for a temperature of 300 K. The results of the calculations on the graph are shown by black lines with markers.

Five values of the inverse thermal conductivity corresponding to the longest lengths of the computational domain were approximated by a linear dependence (1)

$$\frac{1}{\kappa_{lat}}(x) = 0.086 + 6.13x \ , \qquad (6)$$

where $x = 1/L_n$. The approximation error, according the least squares criterion (2) was $\Delta = 0.972$. At $x = 1/L_n = 0$ from the dependence (6) one can obtain the inverse value of thermal conductivity $1/\kappa_{lat} = 0.086$, from which the thermal conductivity of copper was obtained $\kappa_{lat}(T=300K)$ = 11.627 W/mK. It corresponds to the infinite value of L_n. Fig.4 shows the linear dependence (6) as a red line with markers.

The results of calculation of the thermal conductivity of copper are presented in Fig. 5. The vertical dashed lines indicate the melting temperature T_m and the critical temperature T_{cr} of copper. At the temperature of 300 K, according to the above calculations, the phonon thermal conductivity is κ_{lat} = 11.627 W/mK. With increasing temperature, the thermal conductivity of copper decreases. The region of the solid – liquid phase transition is shown in enlarged form in the inset of Fig. 5. It is seen that at the equilibrium melting temperature T_m = 1330 K, the value of thermal conductivity decreases stepwise. In the solid phase, the thermal conductivity is κ_{lat} = 1.55 W/mK, and in the liquid phase at the same temperature the thermal conductivity is κ_{lat} = 1.29 W/mK. The difference in thermal conductivity between solid and liquid phases is 16.77%. The calculation was carried out up to the temperature T = 5700 K, at which the thermal conductivity is κ_{lat} = 0.647 W/mK. Such a change in the phonon thermal conductivity with increasing temperature does not contradict the ideas about the behavior of the phonon thermal conductivity of metals. In the near-critical region in Fig. 5. the dashed line shows the extrapolation of the temperature dependence of thermal conductivity to the critical point.

A comparison with alternative ab-inito calculations in the range 300K≤T <1000K made by the Generalized Gradient Approximation (GGA) method in [36] showed good agreement. At

low temperatures (300K ≤ T < 600K), the largest difference with [36] is $\Delta\kappa$ ~ 14% at 300K. With increasing temperature, the difference in results becomes smaller. At T = 600 K, the difference is $\Delta\kappa$ ~ 4%, and at T = 1000 K, $\Delta\kappa$ ~ 2% - the results almost completely match. In the temperature range T > 1000K, there is no data for comparison. In general, such comparison results suggest that the selected method and potential describe the model with good accuracy and are applicable for further studies.

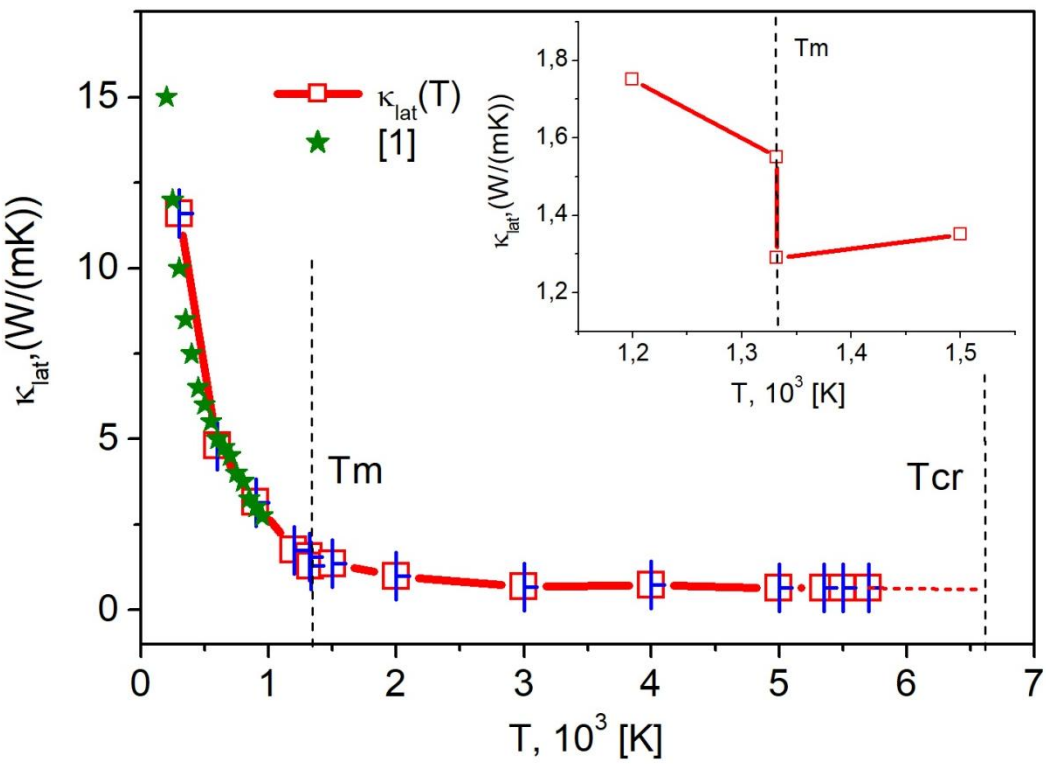


Fig. 5. Temperature dependence of the phonon thermal conductivity of copper. Markers [1] show the results of calculations from [36]. The fragment shows an abrupt decrease in thermal conductivity at the solid – liquid phase transition.

For the use in further calculations, the results obtained are more convenient to write as an analytical dependence of the form (1). For the solid phase, the results were approximated by a polynomial dependence of the 4th degree, for the liquid phase by a polynomial of the 3rd degree

$$\kappa_{lat,sol}(T) = a_0 + a_1(T - T_0) + a_2(T - T_0)^2 + a_3(T - T_0)^3 + a_4(T - T_0)^4,$$
$$\kappa_{lat,liq}(T) = a_0 + a_1(T - T_0) + a_2(T - T_0)^2 + a_3(T - T_0)^3$$

where $T_0 = 300$ K for the solid phase ($300\ K \le T \le T_m$), $T_0 = T_m$ for the liquid phase ($T_m \le T \le 5700K$). The values of the coefficients a_k and approximation errors according to the least squares criterion (2) are shown in the Table 3.

k	solid	liquid
a_0	11.627	1.29
a_1	-4.2781×10^{-2}	-5.01×10^{-4}
a_2	9.2403×10^{-5}	1.31×10^{-7}
a_3	-9.6577×10^{-8}	-1.2×10^{-11}
a_4	3.686×10^{-11}	
$\Delta(P(x_j), y_j)$	1.201×10^{-11}	0.021

Table 3. The values of the coefficients a_k of the function, which approximates the calculation results of the thermal conductivity of copper $\kappa_{lat}(T)$, $W\cdot m^{-1}\cdot K^{-1}$

4 CONCLUSION

The temperature dependences were obtained for the thermophysical characteristics of copper: phonon thermal conductivity, heat capacity and density using molecular dynamics simulation with the EAM potential [21]. The wide temperature range 300K ≤ T ≤ 5800K, in which the thermophysical properties of copper were determined in this work, covers the first-order phase transition (melting-crystallization) and the near-critical region in which drastic changes in the thermophysical properties of the substance occur. The calculation results were approximated by polynomials of low degrees.

The temperature dependences of the density $\rho(T)$ and specific heat $C_p(T)$ of copper were determined from a series of molecular dynamics calculations within the framework of one computational experiment.

The temperature dependence of the density of copper $\rho(T)$ was obtained in the range 300K<T <5620K. At the melting point, the density is calculated for two states of matter: solid and liquid. Based on the simulation results, the temperature range of the overheating of the solid phase T_m <T <1.2T_m and the density values in this range were obtained. The results obtained show good agreement with the experimental results [42].

The temperature dependence of the specific heat of copper $C_p(T)$ was obtained in the range 300K <T <5800K. In the region of the solid – liquid phase transition at an equilibrium melting temperature T_m, a small stepwise decrease in the heat capacity of copper occurs, amounting to ~ 3.128% (according to the experiment [38], ~ 1.529%). In the temperature range of the overheating of the solid phase T_m<T<1.2T_m, the specific heat was obtained. In the temperature range T_m <T <2.63T_m in the liquid phase, the specific heat is constant and amounts to $C_p(T)$ ≈ 31.0 $J\cdot mol^{-1}\cdot K^{-1}$, which is 8% less than in [38]. At temperatures above 4000 K, the heat capacity of copper increases and at T = 5800 K its value is $C_p(T)$ = 47.698 $J\cdot mol^{-1}\cdot K^{-1}$. The results obtained show good agreement with the experimental results [37, 38].

Based on a series of calculations by the direct method using molecular dynamics simulation, the temperature dependence of the phonon thermal conductivity of copper was obtained in the temperature range 300 <T <5700 K. The EAM potential [21] was used in the simulation. In the region of the solid – liquid phase transition, at the equilibrium melting temperature T_m = 1330 K, the value of thermal conductivity decreases stepwise. The difference in thermal conductivity between solid and liquid phases is 16.77%. Comparison with alternative ab-inito calculations in the range 300K ≤ T <1000K made by the Generalized Gradient Approximation (GGA) method in [36] showed good agreement.

In general, such comparison results suggest that the selected interparticle interaction potential and calculation methods for copper of the phonon specific heat, phonon thermal conductivity, and density show good accuracy and can be used for further studies.

Acknowledgements: This study was supported by the Russian Science Foundation (project no. 18-11-00318).

REFERENCES

[1] Mónica Fernández-Arias, Mohamed Boutinguiza, Jesús del Val, Antonio Riveiro, Daniel Rodríguez , Felipe Arias-González, Javier Gil and Juan Pou, "Fabrication and Deposition of Copper and Copper Oxide Nanoparticles by Laser Ablation in Open Air", *Nanomaterials*, **10**, 300 (2020).

[2] S. Fernando, T. Gunasekara, J. Holto, "Antimicrobial Nanoparticles: applications and mechanisms of action", *Sri Lankan Journal of Infectious Diseases*, **8**(1), 2-11 (2018)

[3] Ali K. Yetisen, Yunuen Montelongo, Fernando da Cruz Vasconcellos, J.L. Martinez-Hurtado, Sankalpa Neupane, Haaider Butt, Malik M. Qasim, Jeffrey Blyth, Keith Burling, J. Bryan Carmody, Mark Evans, Timothy D. Wilkinson, Lauro T. Kubota, Michael J. Monteiro, and Christopher R. Lowe, "Reusable, Robust, and Accurate Laser-Generated Photonic Nanosensor", *Nano Lett.*, **14**, 3587-3593 (2014).

[4] Muhammad Imran Din, Rida Rehan, "Synthesis, characterization, and applications of copper nanoparticles", *Analytical Letters*, **50**(1), 50-62, (2017).

[5] Amir Reza Sadrolhosseini, Mohd Adzir Mahdi, Farideh Alizadeh and Suraya Abdul Rashid, "Laser Ablation Technique for Synthesis of Metal Nanoparticle in Liquid", Chapter 4 in *Laser Technology and its Applications*, Yufei Ma (Ed.), InTech (2020).

[6] S. Barcikowski, A. Hahn, A. V. Kabashin, and B. N. Chichkov, "Properties of nanoparticles generated during femtosecond laser machining in air and water," *Appl. Phys. A*, **87**, 47-55 (2007).

[7] P.V. Kazakevich, V.V. Voronov, A.V. Simakin, G.A. Shafeev. "Production of copper and brass nanoparticles upon laser ablation in liquids", *Quantum Electronics*, **34**(10), 951-956 (2004)

[8] N. Tsakiris, K. K. Anoop, G. Ausanio, M. Gill-Comeau, R. Bruzzese, and S. Amoruso, "Ultrashort laser ablation of bulk copper targets: Dynamics and size distribution of the generated nanoparticles," *J. Appl. Phys.*, **115**, 243301 (2014)

[9] C. M. Rouleau, C.-Y. Shih, C. Wu, L. V. Zhigilei, A. A. Puretzky, and D. B. Geohegan, "Nanoparticle generation and transport resulting from femtosecond laser ablation of ultrathin metal films: Time-resolved measurements and molecular dynamics simulations," *Appl. Phys. Lett.*, **104**, 193106 (2014).

[10] A. Amouye Foumani, A. R. Niknam, "Atomistic simulation of femtosecond laser pulse interactions with a copper film: Effect of dependency of penetration depth and reflectivity on electron temperature", *J. of Appl. Phys.*, **123**, 043106 (2018)

[11] V.I. Mazhukin, A.V. Shapranov, A.V. Mazhukin, O.N. Koroleva, "Mathematical formulation of a kinetic version of Stefan problem for heterogeneous melting/crystallization of metals", *Mathematica Montisnigri*, **36**, 58-77 (2016)

[12] A.A. Samokhin, V.I. Mazhukin, A.V. Shapranov, M.M. Demin, A.E. Zubko, "Molecular dynamics modeling of nanosecond laser ablation: transcritical regime", *Mathematica Montisnigri*, **38**, 78-89 (2017)

[13] V. I. Mazhukin, A. V. Mazhukin, M. M. Demin, A. V. Shapranov. “Nanosecond laser ablation of target Al in a gaseous medium: explosive boiling”. *Applied Physics A: Material Science and Processing*, **124** (3), 237(10), (2018)

[14] V.I. Mazhukin, M.M. Demin, A.V. Shapranov. “High-speed laser ablation of metal with pico- and subpicosecond pulses”. *Applied Surface Science*, **302**, 6–10 (2014).

[15] C. Wu and L. V. Zhigilei, “Microscopic mechanisms of laser spallation and ablation of metal targets from large-scale molecular dynamics simulations”, *Appl. Phys. A*, **114**, 11-32 (2014).

[16] Neil W. Ashcroft, N. David Mermin, *Solid State Physics*, Holt, Rinehart and Winston, (1976).

[17] Z. Tong, S. Li, X. Ruan, and H. Bao, “Comprehensive first-principles analysis of phonon thermal conductivity and electron-phonon coupling in different metals”, *Physical review*, **B 100**, 144306 (2019)

[18] A.V. Mazhukin, V.I. Mazhukin, M.M. Demin, “Modeling of femtosecond laser ablation of Al film by laser pulses”, *Applied Surface Science*, **257**, 5443–5446 (2011).

[19] R. Venkatasubramanian, E. Siivola, T. Colpitts and B. O. Quinn, “Thin Film Thermoelectric Devices with High Room Temperature Figures of Merit”, *Nature*, **413,** 597-602 (2001). http://dx.doi.org/10.1038/35098012

[20] M.M. Demin, V.I. Mazhukin, A.A. Aleksashkina, “Molecular dynamic calculation of lattice thermal conductivity of gold in the melting-crystallization region”, *Mathematical Montisnigri*, **46**, 105-113 (2019)

[21] Y. Mishin, M. J. Mehl and D. A. Papaconstantopoulos, A. F. Voter, J. D. Kress, “Structural stability and lattice defects in copper: Ab initio, tight-binding, and embedded-atom calculations”, *Phys. Rev. B*, **63**, 224106 (2001)

[22] S.M. Foiles, M.I. Baskes, and M.S. Daw, "Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys", *Physical Review B*, **33(12)**, 7983-7991 (1986). DOI: 10.1103/physrevb.33.7983.

[23] S.V. Starikov, A.Y. Faenov, T.A. Pikuz, I.Y. Skobelev, V.E. Fortov, S. Tamotsu, M. Ishino, M. Tanaka, N. Hasegawa, M. Nishikino, T. Kaihori, T. Imazono, M. Kando, and T. Kawachi, "Soft picosecond X-ray laser nanomodification of gold and aluminum surfaces", *Applied Physics B*, **116(4)**, 1005-1016 (2014). DOI: 10.1007/s00340-014-5789-y.

[24] P. K. Schelling, S. R. Phillpot, P. Keblinski, “Comparison of atomic-level simulation methods for computing thermal conductivity”, *Phys. Rev. B*, **65**, 144306 (2002)

[25] L. Hu, W. J. Evans, P. Keblinski, “One-dimensional phonon effects in direct molecular dynamics method for thermal conductivity determination”, *J. Appl. Phys.*, **110**, 113511 (2011)

[26] Florian Müller-Plathe, “A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity”, *J. Chem. Phys.*, **106**, 6082 (1997).

[27] V.I. Mazhukin, M.M. Demin, A.A. Aleksashkina, “Atomistic modeling of thermophysical properties of cooper in the region of the melting point”, *Mathematical Montisnigri*, **41**, 99-111 (2018)

[28] V.S. Dozhdikov, A.Y. Basharin, P.R. Levashov, “Two-phase simulation of the crystalline silicon melting line at pressures from -1 to 3 GPa”, *J. Chem. Phys.*, **137**(5), 054502 (2012)

[29] Yang Chun Zou, Shi Kai Xiang, Cheng Da Dai. “Investigation on the efficiency and accuracy of methods for calculating melting temperature by molecular dynamics simulation”, *Computational Materials Science*, **171**, 109156(1-11) (2020).

[30] M.M. Demin, O.N. Koroleva, A.V. Shapranov, A.A. Aleksashkina, “Atomistic modeling of the critical region of copper using a liquid-vapor coexistence curve”, *Mathematical Montisnigri*, **46**, 61-71 (2019)

[31] O.N. Koroleva, M.M. Demin , V.I. Mazhukin, A.V. Mazhukin, "Modeling of thermal conductivity of Si in the range from the normal to near-critical conditions", *Mathematical Montisnigri*, **45**, 85-94 (2019)

[32] G. Sebastian, Gang Chen Volz, "Molecular-dynamics simulation of thermal conductivity of silicon crystals", *PHYS. REV. B*, **61**(4), 2651-2656 (2010)

[33] D.P. Sellan, E.S. Landry, J.E. Turney, A.J.H. McGaughey, and C.H. Amon, "Size effects in molecular dynamics thermal conductivity predictions", *PHYS. REV. B*, **81**, 214305 (2010)

[34] P.C. Howell, "Comparison of molecular dynamics methods and interatomic potentials for calculating the thermal conductivity of silicon", *J. Chem. Phys.*, **137**, 224111 (2012).

[35] L. Verlet, "Computer "Experiments" on Classical Fluids. I. Thermodynamically Properties of Lennard-Jones Molecules", *Phys. Rev.*, **159**, 98-103 (1967).

[36] Y. Wang, Z. Lu, and X. Ruan, "First principles calculation of lattice thermal conductivity of metals considering phonon-phonon and phonon-electron scattering", *Journal of applied physics*, **119**, 225109 (2016)

[37] *Fizicheskie velichiny*, Spravochnik pod red. I.S. Grigoreva, E.Z. Meilihova, M: Energoatomizdat (1991)

[38] M. W. Chase, Jr., C. A. Davies, J. R. Downey, Jr., D. J. Frurip, R. A. McDonald, and A. N. Syverud, "JANAF Thermochemical Tables, Third Edition", *J. Phys. Chem. Ref. Data*, **14**(1), (1985)

[39] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics", *J. Comput. Phys.*, **117**(1), 1-19 (1995)

[40] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, A. DiNola, J.R. Haak, "Molecular dynamics with coupling to an external bath", *J. Chem. Phys.*, **81**, 3684-3690 (1984).

[41] A.A. Samarskii, F.I. Gulin, *Chislennye metody*, M.: Fizmatlit, (1989)

[42] J.A. Cahill, A. D. Kirshenbaum, "The density of liquid copper from its melting point (1356K) to 2500K and an estimate of its critical constants", *J. Phys. Chem.*, **66**, 1080–1082 (1962)

[43] P.M. Nasch and S.G. Steinemann, "Density and thermal expansion of molten manganese, iron, nickel, copper, aluminum and tin by means of the gamma-ray attenuation technique", *Phys. Chem. Liq.*, **29**, 43 (1995).

[44] M.J. Assael, A.E. Kalyva, K.D. Antoniadis, R.M. Banish, I. Egry, J. Wu, E. Kaschnitz, W.A. Wakeham, "Reference Data for the Density and Viscosity of Liquid Copper and Liquid Tin", *J. of Phys. and Chem.*, **39**, 033105 (2010)

[45] G.T. Furukawa, W.G. Saba, M.L. Reilly, *Critical analysis of the heat-capacity data of the literature and evaluation of thermodynamic properties of copper, silver, and gold from 0 to 300 K*, National Standard Reference Data SeriesNational Bureau of Standards **18**, Washington, p. 64 (1968).

[46] R. Hultgren, P.D. Desai, D.T. Hawkins, M. Gleiser, K. K. Kelley, and D. D. Wagman, *Selected Values of the Thermodynamic Properties of the Elements*, Am. Soc. Metals, Metals Park, OH, (1973).

[47] A.K. Chaudhuri, D.W. Bonnell, L.A. Ford, J.L. Margrave, *High Temp. Sci.*, **2(**3), 203-212 (1970)

Received November 15, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

COMPARATIVE STUDY OF HIGH PERFORMANCE SOFTWARE RASTERIZATION TECHNIQUES

V. F. FROLOV[1,2], V. A. GALAKTIONOV[1,*] AND B. H. BARLADYAN[1]

[1]Keldysh Institute of Applied Mathematics of RAS,
Miusskaya Sq. 4, Moscow, Russia, 125047

[2]Lomonosov Moscow State University,
GSP-1, Leninskie Gory, Moscow, 119991, Russian Federation

*Corresponding author. E-mail: vlgal@gin.keldysh.ru, web page: http://keldysh.ru/

DOI: 10.20948/mathmontis-2020-47-13

Abstract. This paper provides a comparative study and performance analysis of different rasterization algorithms and approaches. Unlike many other papers, we don't focus on rasterization itself, but investigate complete graphics pipeline with 3D transformations, Z-buffer, perspective correction and texturing that, on the one hand, allow us to implement a useful subset of OpenGL functionality and, on the other hand, consider various bottlenecks in the graphics pipeline and how different approaches manage them. Our ultimate goal is to find a scalable rasterizer technique that on the one hand effectively uses current CPUs and on the other hand is accelerating with the extensive development of hardware. We explore the capabilities of scan-line and half-space algorithms rasterization, investigate different memory layout for frame buffer data, study the possibility of instruction-level and thread-level parallelism to be applied. We also study relative efficiency of different CPU architectures (in-order CPUs vs out-of-order CPUs) for the graphics pipeline implementation and tested our solution with x64, ARMv7 and ARMv8 instruction sets. We were able to propose an approach that could outperform highly optimized OpenSWR rasterizer for small triangles. Finally, we conclude that despite a huge background high-performance software rasterization still has a lot of interesting topics for future research.

2010 Mathematics Subject Classification: 78-04, 65C05, 65C20.
Key words and Phrases: software rasterization, graphics pipeline, real-time rendering.

1 INTRODUCTION

Modern hardware-accelerated graphics pipeline consists of dozen stages and has great flexibility [1]. However, it is not always possible to rely on existing graphics hardware for various reasons. At first, various embedded applications do not have dedicated graphics processors and thus forced to use software implementation. Second, a huge amount of popular Linux distributives uses open-source graphics drivers with partial or complete software implementation of rasterization (Mesa OpenGL). Particular applications use specific hardware anyway [2, 3]. In such cases, software implementation should be able to provide real-time rendering sacrificing graphics quality for the sake of correctness or clarity of the displayed information. In such situations programmable functionality of OpenGL shaders, for example, can be excluded or restricted.

At the same time, processors are greately evolved over the past decades and therefore software rasterization methods that were relevant a couple of decades ago may not be the best ones for today. This gives rise to a fundamental contradiction in the design of the rasterizer: it is necessary to pay attention to efficient loading of hardware units of a modern CPU when we come to its peak performance, but on the other hand we don't want to depend too much on any particular hardware. At last, software rasterization is still remaining a widespread challenge in graphics community and thus, have a scientific interest to study within itself.

1.1 Need for software rasterization

Today, almost all rendering techniques have become GPU based. Software solutions, however, do not lose their relevance. For example, Linux uses widely open-source software graphics drivers (Mesa OpenGL [4]). GPU driver installation is not always easy and even not possible on some Linux systems (running, for example, on a custom CPU development board which is quite common for embedded systems). Microsoft also has its own software rasterizer in DirectX10 and DirectX11 called "WARP". WARP rasterizer scales well into multiple threads, and in some cases is known to be faster than low-end GPUs [5]. Besides, software graphics pipilene is more flexible and can directly use system memory. Thus it is useful in scientific visualization of large data sets [6, 7]. At last, the recent development of CPUs sets a new round in software rendering research since many applications for which it was previously impossible to achieve high speed pure in software are enabled now.

1.2 Graphics pipeline

Before moving on, we would like to shortly describe a subset of graphics pipeline that we took for our research and point out why this subset is important and challenging to accelerate on CPU. Useful graphics pipeline requires at least 5 stages:

1. vertex processing;
2. primitive assembly;
3. triangle rasterization;

4. pixel processing;
5. alpha blending.

Both first and second stages are quite simple, especially if we don't have to consider triangle clipping. Vertex processing consists of multiplying points by a matrix and is implemented trivially. Primitive assembly consists of the formation of triangles by indices of its vertices and thus, is mostly trivial. Also, these stages are rarely a bottleneck due to vertices and triangles amount is considerably less than pixel amount.

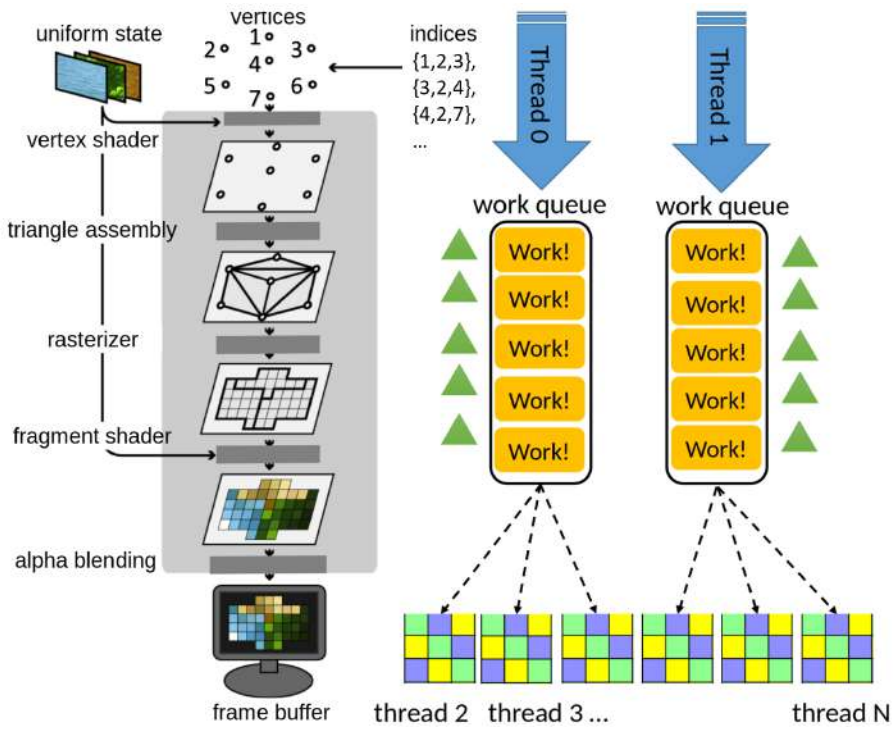


Fig. 1. Graphics pipeline forming producer-consumer scheme where some threads (0 and 1) push triangles (or some other portion of work) to queue and other threads process pixels and behave like consumers taking work from the queue.

However, the following 3 stages are not so simple. It becomes especially noticeable for multi-core implementation where triangle rasterization became a sort of work distribution for pixel processing forming a producer-consumer scheme (fig 1). Alpha blending should be mentioned separately due to it assumes fixed order for processing of pixels for different triangles. The situation is complicated by the fact that not all rasterization algorithms and not all methods of efficient pixel processing (using instruction level parallelizm for example) can be easily used together. This happens due to algorithms have different optimal data structure layout and different access patterns to frame-buffer data. When performance is a goal, these problems became essential. Programmable functionality of OpenGL shaders, on the other side, can be excluded from consideration without loss of generality due to it influences mostly on the pixel processing computation complexity. Thus, we can model its influence if consider heavy pixel processing cases (heavy in comparison to vertex processing and triangle assembly, for example).

1.3 Scientific problems

With the extensively developed graphics hardware last decades many research topics in the area of real-time software rendering became abandoned. At the same time CPUs have evolved significantly:

1. deep out-of-order pipelines, speculative execution, SIMD and various CPU architectures;
2. multi-level caches and tremendous gap between memory and processor speed;
3. true multi-core systems, the number of cores increases significantly;
4. The "relaxed memory model" have appeared and efficient sharing of the cache by many threads has become non-trivial task, especially when increasing number of cores.

Thus, many algorithms and optimizations that were popululer 20 years ago (the dawn of graphics hardware development) mostly useless and even performance-harmful for modern CPUs. The goal of our work is to explore different techniques together (considering the influence of all factors upon each other) and find the most practical and scalable approach for software implementation of OpenGL graphics pipeline on modern multicore CPUs which is, in our opinion, is not solved.

2 HIGH PERFORMANCE SOFTWARE RASTERIZATION TECHNIQUES

2.1 Triangle rasterization basics

Before considering triangle rasterization algorithms, we should note that in the existing graphics pipelines (for example OpenGL, DirectX or Vulkan) there is a certain agreement about drawing triangles. A pixel is considered as overlapped by a two-dimensional triangle if its **center lies inside the triangle**. Thus, the pixel-triangle overlap test is called a "coverage test" (fig. 2).

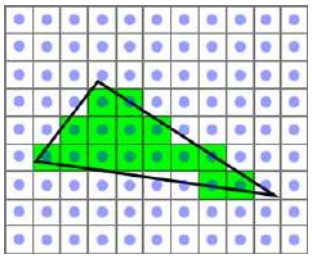


Fig. 2. Standard agreement about covered pixels. A pixel is considered as overlapped by two-dimensional triangle if its center lies inside the triangle.

Probably the most well known **scanline** algorithm [8] sudbidives a triangle into 2 adjacent triangles with horizontal edges. Then it is proposed in some way to move along the edges of the triangle and paint the area between the edges line by line. A straitforward way is to move along edges using finite differences (equations 1 and 2) [9].

$$\Delta_{xy1} = \frac{v2.x - v1.x}{v2.y - v1.y}; \quad \Delta_{xy2} = \frac{v3.x - v1.x}{v3.y - v1.y}. \tag{1}$$

$$\begin{aligned} y &:= y + 1; \\ x1 &:= x1 + \Delta_{xy1}; \\ x2 &:= x2 + \Delta_{xy2}. \end{aligned} \tag{2}$$

We will refer to this algorithm **"scanline"**. Despite the simple idea, we should pay attention to the fact that the algorithm has certain problems:

- The known algorithms for moving along edges (Brezenham [10], Fujimoto [11], or algorithm with finite differences discussed above) do not allow us to say whether the edge pixel is covered by a triangle or not. This means that such a rasterization algorithm itself does not comply with the agreement adopted in OpenGL. For its correct implementation it is necessary to add a pixel-triangle overlap test (so-called "coverage test", fig. 2).
- The algorithm should be additionally limited to a rectangle (built around a triangle), because scanline uses division by the difference between the coordinates of 2 vertices, which under certain conditions became a small number (though zero, as a rule, is excluded by a separate condition that the triangle does not degenerate into a line). This leads to the fact that the offset in y by 1 pixel gives a huge offset in x, which can even go beyond the limits of the screen. The reason for this problem is that according to the OpenGL standard, the coordinates of the triangle's vertices when moving to screen space should be floating point numbers (or at least, have 4-bits subpixel precision [12]). They can not be just integer pixel coordinates. Therefore, strictly speaking, the Bresenham algorithm cannot be used to move along edges.

2.2 Related Work

An improved scanline implementation can be found in [13]. It moves along the longest edge, drawing lines between edges. In comparison to the prevoius naive scanline approach, this algorithm is simpler for CPU due to it has less branches and special cases and it doesn't have a *near zero* division problem because it doesn't use finite differences. However, it does not eliminates the need for the coverage test and the original version does not implement it. We will refer to this algorithm as **"scanline(fast)"** and will test its original implementation without coverage test. Such algorithm would be equivalent to the classic version using Bresenham for movement along the edges.

In [14] **half-space** rasterization was proposed. This paper introduces the concept of *edge-function* (equations 3–6) which was later adopted as a kind of standard agreement for "coverage test" that we discussed before. This method is based on the fact that a line in 2D subdivides the space (plane) into two half-spaces (half-planes). If we substitute the coordinates of the center of the pixel P into the equation of a line, we can obtain the sign distance to this line (equation 3). The *edge-function* is a special case of well known cross

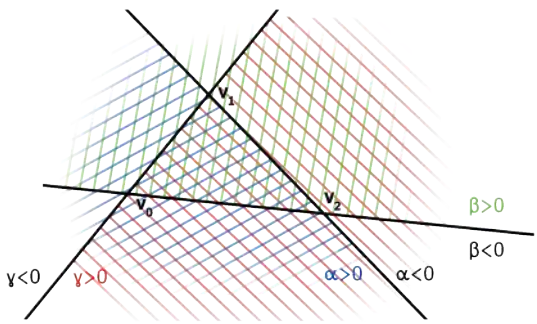


Fig. 3. Half-space algorithm idea

product and it allows calculating the signed distance from pixel center (x, y) to some edge — $(\alpha, \beta, \gamma$; equations 4–6). If all signed distances are greater than zero, the point lies inside the triangle (fig. 3).

$$E(A, B, P) = (P.x - A.x)(B.y - A.y) - (P.y - A.y)(B.x - A.x) \tag{3}$$

$$\begin{aligned} E_\alpha(x, y) = E(A, B, P) &= (x - A.x)(B.y - A.y) - (y - A.y)(B.x - A.x); && (4)\\ E_\beta(x, y) = E(B, C, P) &= (x - B.x)(C.y - B.y) - (y - B.y)(C.x - B.x); && (5)\\ E_\gamma(x, y) = E(C, A, P) &= (x - C.x)(A.y - C.y) - (y - C.y)(A.x - C.x). && (6) \end{aligned}$$

The most useful property of the *edge-function* is that it can be **evaluated incrementally** when rasterizer moves along pixels (figure 4) [14]. Besides, baricentric coordinates (u, v, w) also can be evaluated directly from *edge-function* by multiplying its value with inverse triangle double area which is also evaluated with the *edge-function* (equations 7–9).

$$\begin{aligned} u(P) &= \frac{E(A, B, P)}{E(A, B, C)}; && (7)\\ v(P) &= \frac{E(B, C, P)}{E(A, B, C)}; && (8)\\ w(P) = 1 - u(P) - v(P) &= \frac{E(C, A, P)}{E(A, B, C)}. && (9) \end{aligned}$$

The most significant advantage of half-space rasterizer is extremely simple kernel of the algorithm, especially in comparison with scanline approach. No more difficult to fill the rectangle (fig. 4). This property allows branch prediction mechanisms working efficiently and this is also the reason for the popularity of hardware solutions. The disadvantage of half-space approach (in comparison to scanline for example) is the presence of idle iterations since inside the bounding rectangle; there can be a rather large area which is not covered by a triangle. However, this disadvantage is easily fixed by a serpentine traversal algorithm [14] or Blocked based version of half-space rasterization [15].

```
for y in range minY .. maxY do
    Cx1 := Cy1;
    Cx2 := Cy2;
    Cx3 := Cy3;
    for x in range minX .. maxX do
        if Cx1 > 0 and Cx2 > 0 and Cx3 > 0 then
            u = Cx1*TriAreaInv;
            v = Cx2*TriAreaInv;
            framebuffer[x,y] := DrawPixel(u, v, 1-u-v);
        end if;
        Cx1 := Cx1 - Dy12;
        Cx2 := Cx2 - Dy23;
        Cx3 := Cx3 - Dy31;
    end for;
    Cy1 := Cy1 + Dx12;
    Cy2 := Cy2 + Dx23;
    Cy3 := Cy3 + Dx31;
end for;
```

Fig. 4. Half-space rasterization kernel. $Cx*$ and $Cy*$ variables store edge-functions for line and colum respectively. $TriAreaInv = 1/E(A, B, C)$ is a constant inverse triangle double area. A triplet of $(u, v, 1-u-v)$ represents baricentric coordinates of a pixel center.

Blocked based half-space method was also suggested in [14] but well-developed much later in [15]. The main idea of blocked version is that if we perform coverage test check (via evaluating *edge-function*) for 4 corner points of a pixel block (4x4 or 8x8 for example) and all tests have passed then the block is covered by triangle and we can fill/process all internal pixels in parallel (for example using SIMD instructions). Several blocked versions of half-space rasterizer were proposed and tested in [15]. The most complex version (called "Block-based Bisector Half-Space Rasterization") processes triangle in such a way that it minimizes checks for empty blocks due to a quick cut of empty space from inside triangle bounding box. The advantages of "bisector" algorithm appear only on extremely large triangles and simple fill modes (without texture for example) because incremental *edge-function* evaluation is quite cheap in comparison to pixel processing for a fully-covered or even partially-covered block. At the same time average amount of blocks for most of triangles is usually just a little: 4-8 blocks. As a result complication of the algorithm leads to poor performance due to branch misprediction simultaneously with winning of empty blocks tend to zero. We will refer to the blocked version of half-space rasterizer as **"blocked half-space"**. The main advantage of blocked version (over previous half-space approach) is the possibility of parallel processing of pixels via SIMD instructions. Besides, blocked half-space algorithm processes empty space faster. The disadvantage of blocked version appears with small triangles — not all calculations that were performed for 4x4 tile (for example) are useful.

2.2.1 Floating point vs fixed point

When choosing between a floating point and a fixed point, two cases should be distinguished: (1) rasterization algorithm itself and (2) pixel operations. When speaking about rasterization — current graphics hardware uses "28.4" or "15.8" (or other) fixed point

format with 4 or 8 bit subpixel precision [12, 1, 16] and there is simply no any reason for using floating point to process the triangle in the rasterization algorithm. This is so because a fixed point has deterministic behavior and is not subject to rounding errors; therefore, it's not even about speed, but rather about correctness. Both half-space and scanline approaches are known to be implemented in fixed point well [17, 18].

Fig. 5. PlayStation1 (right) didn't have correct texture mapping due to absence of floating point for pixel operations [19].

While speaking about pixel processing — it depends on the hardware. Early versions of gaming consoles didn't have floating point support [20] so they had visible problems with texture mapping and Z-buffer (fig. 5). There are still processors without a floating point and SIMD support (or its performance may be not enough), therefore, fixed-point can be an option [18]. Also, if we do not need rendering in three dimensions, we can get by with a fixed point. Otherwise, we believe that for pixel operations it's better to use floating point in conjuction with SIMD. Here are our reasons:

1. Rendering in 3D is difficult to be correct without a floating point (fig. 5).

2. SIMD and floating point can be used together. If SIMD instructions are enabled, there should be no need in complex and chip-expensive Out Of Order execution mechanisms to speed-up floating point operations. Blocked based half-space always has a lot of independent work (at leats 16 operations for 4x4 pixel block), so coarse-grained instruction parallelism [21] can be used. GPUs actively use this idea sending commands to the pipeline from different micro-threads [22]. This is why they are so good at floating point operations and have high memory bandwidth. Thus, even straitforward implementation of SIMD floating point should work well.

3. Almost all CPUs have different register sets for integer and floating point numbers. Using both (we must use integer registers for fixed point rasterization anyway) will increase the effective number of processor registers and in this case reduce register pressure.

4. A CPU may not have SIMD for integers (for example, SSE1 doesn't have them).

5. Half precision reduces architectural state by a half and thus more pixels can be processed in parallel or we may use less transistors for the CPU or, at least, reduce necessary memory bandwidth. Processors with half precision support for neural networks are currently becoming popular (for example late ARM and Intel CPUs).

Thus in our experiments we used fixed point for triangle rasterization algorithms and SIMD floating point operations for pixel processing.

2.2.2 Multi-threaded implementations

Multithreaded implementation of a graphics pipeline is a challenging task. Figure 1 shows it in general. An unknown number of triangles of arbitrary size is fed to the input of the graphics pipeline in general (so it is hard to say in advance exactly at which stage of the pipeline there will be the bottleneck). Non uniform work distrubution is easy arising here. Triangles could significantly overlap each other. Moreover, if alpha blending is enabled, a certain order of pixel processing for triangles must be preserved: if the triangle A was filed into the graphics pipeline before the triangle B, then A must be drawn before B and its pixels must be processed before the pixels of the triangle B. Otherwise, we will get an incorrect image.

One of the first papers about software rasterization on modern CPUs is [23]. In this paper, SSE instructions and multithreading capabilities were exploited. Binned implementations of rasterizer was used (which is known as a **"sort-middle"** approach [1]). In this paper, screen is subdivided into large bins/tiles (in size of 64x64, 128x128 or 256x256 pixels). Once all primitives are binned, threads switch over to tiles for rasterization and fragment processing work. Thus, in this paper, for each bin there is its own queue of triangles, which is first completely filled with all the threads, and then all the queues are emptied in parallel. One tile is processed at a time by only one stream. The blocked version of half-space rasterization was used with 8x8 block size for SIMD processing of pixels. An advantage of *sort-middle* approach from [23] is the correct alpha blend support by default due to each bin is processed in a single thread. The disadvantage is a limited parallelization capability due to different bins could have significantly different numbers of triangles and thus some bins will hang for a while in a single thread when all the others bins/threads have already finished. A performance growth demonstrated in [23] was measured on a quite heavy pixel operations (which reduces the described disadvantage) with shadow mapping, and even in this case was not perfect. Authors of [24] simply split screen in 4 parts and [25] also didn't introduce any new technique.

In [18] disadvantages of *sort-middle* approach was also noted and a solution was proposed that is parallelized almost perfectly — render different frames completely in different threads. This idea is similar to Nvidia SLI and AMD Cross-Fire GPU solutions [26]. The reason for such successful results is that this work bypasses the Amdahl law, making sequential calculations parallel via pipelining. Unfortunately, it has at least 2 drawbacks. First, this method of parallelization does not reduce the latency of rendered information. It makes the animation smoother, but the user sees the information on the screen with such a delay as if the whole rendering has occurred in a single thread. In automotive and avionics applications, for example, such disadvantage became serious, because a person in critical situation may wrong react to displayed information due to a time lag. Second,

a processor memory bus has limited bandwidth and thus SLI-method has a physical limitation on parallelizability on a single device (so, Nvidia and AMD use it for multi-GPU setup) due to each thread accesses its' own frame buffer and the total amount of memory moved along the bus increases with the number of threads.

Unlike previously discussed papers, an older work GRAMPS [27] uses the approach which is known as **"sort-last"**[1]. This approach parallelizes individual operations on pixels or groups of pixels, and unlike *sort-middle* does not require screen to be split into bins. Thus, different triangles can be processed by different threads. The main focus of [27] was done on prototyping and simulating graphics hardware. So, there was no information about efficiency of this approach for software implementation on practice.

2.2.3 Hardware solutions: sort-middle vs sort-last

Modern graphics hardware has a tremendous amount of parallellism inside. However, before fragments/pixels finally got to the frame-buffer they have to be sorted in some way to form a correct image. This becomes especially important if alpha blending is used. Current graphics hardware can be divided into 2 large classes based on what stage of the graphics pipeline this sorting takes place: *sort-middle* and *sort-last* [1].

Desktop GPUs have a high memory bandwidth and uses *sort-last* approach implementing the ordering of fragments inside Render Output Unit ("ROP") hardware units. Same units are known to be used for atomic operations in GPGPU, so, ROPs are useful units anyway. Mobile GPUs are aimed more at energy efficiency than at performance and use a *sort-middle* method (except Nvidia Tegra). This approach is more energy efficient because it allows performing fewer operations to DRAM keeping a small piece of framebuffer (for example 64x64) in the on-chip memory (cache). The disadvantage of *sort-middle* approach for GPUs is lower performance with a large number of triangles due to vertex shader and triangle set up executes several times (thus multiplying the cost of geometry stages with the number of tiles).

2.2.4 Software rasterization on GPUs

First succesfull software GPU implementation "in compute" (i.e. without using dedicated rasteriszation units) was proposed in [12]. This implementation was a three-level (bin-raster, coarse-raster, fine-raster) and used *sort-middle* on desktop GPUs. More advanced approach was suggested in [28] which reduces memory transactions in comparison to [12]. Due to efficient usage of shared memory and the extremely high computing power of the GPU, good results were obtained in both papers described above. Combined with a heavy pixel shader software rasterizations may have almost the same speed than hardware implementation but it may have higher flexibility.

Larabee [29] uses 4x4 blocked half-space with 16-wide vector instructions and the algorithm was recursive: each triangle evaluates 16 blocks of pixels at a time to figure out which blocks are even touched by the triangle, then descended into each block that's at least partially covered, evaluating 16 smaller blocks within it, continuing to descend recursively until we had identified all the pixels inside the triangle [16]. Thread paralellizm used *sort-middle* approach.

2.3 PERFORMANCE EXPERIMENTS AND ANALYSIS

We tested various methods on the fixed set of scenes. However, the purpose of our experiments was to select successful methods for a wide range of scenes. Therefore one of the most important criteria for an objective study is the correct choice of test scenes.

2.3.1 Test scenes

Our test scenes are presented at fig. 6 and 7. We chose these scenes so that the bottlenecks are presented in different parts of the graphics pipeline. Here is the description of these scenes and their rasterization modes/states:

1. T1: 18 triangles, color interpolation with perspective correction and Z-buffer;
2. T2: 8K triangles, color interpolation without perspective correction (2D mode);
3. T3: 92 triangles, texture with bilinear fetch, perspective correction and Z-buffer;
4. T4: 4K triangles, same rasterizer state than a previous one;
5. T5: 37K triangles, same rasterizer state than a previous one;
6. T6: 131K triangles. same rasterizer state, lighting was baked in the texture.

T1 scene is simple in all stages: geometry, rasterization and pixel processing. T2 scene is simple in pixel and geometry processing, but more complex for rasterizer itself due to it draws 8K small triangles. T3 scene is complex in pixel processing but simple at geometry and rasterization stages. T4 (4K triangles) and T5 (36K triangles) scenes are more or less balanced. T6 scene contains 131K triangles and is positioned as a complex scene for all stages. T6 scene has baked lighting. Therefore, having a small number of test scenes, we are able to study different bottlenecks in graphics pipeline ignoring irrelevant details of a complete OpenGL implementation in the same time.

2.3.2 Investigated and proposed techniques

Thus, we have implemented minimal but useful graphics pipeline subset. Such things as attribute interpolation, perspective correction and depth buffer during triangle rasterization are implied. Pixel processing includes texture mapping with bilinear filtering. However, we don't evaluate differentials (dFdx/dFdy [30]) for texture coordinates and avoid using MIP levels. For each OpenGL state we have implemented code generator using C++ templates for pixel processing excluding unnecessary code explicitly.

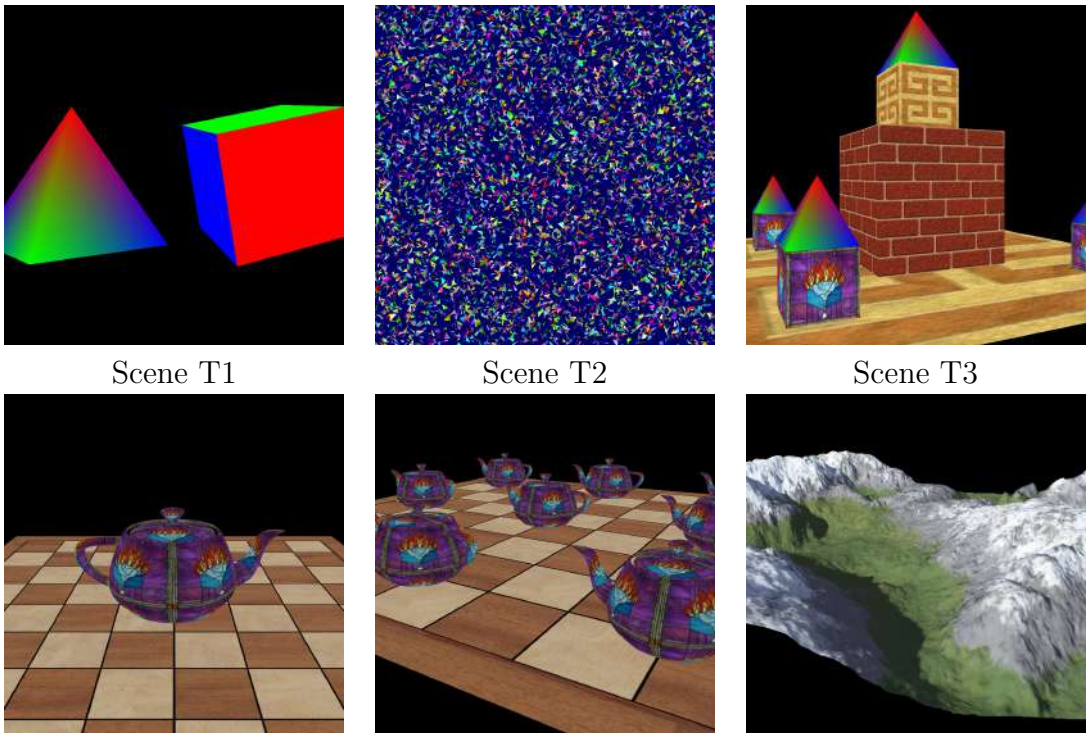


Fig. 6. Our test scenes rendered in solid mode to demonstrate their actual appearence.

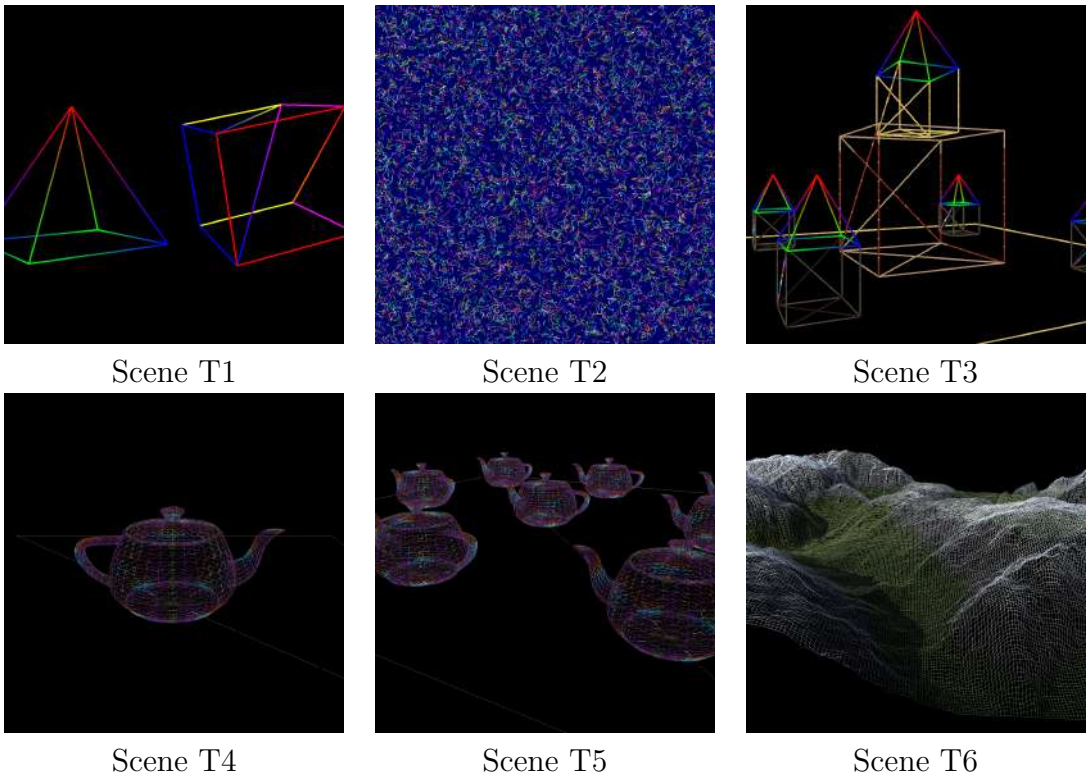


Fig. 7. Our test scenes rendered in wire frame to demonstrate triangles.

Thus, our pixel processing code doesn't have any branches except a depth test. For experiments we used three states/filing modes: (1) 2D color interpolation without texture (T2 scene); (2) 3D color interpolation with perspective correction and depth test (scene T1), and (3) 3D mode with texture mapping (other scenes), perspective correction, depth test and bilinear texture fetch.

Using compiler explorer [31], we have estimated that the first mode consists of approximately 68 instructions per pixel and the second takes 290 instructions (though each instruction processes line of 4 or 8 pixels for blocked half-space algorithm) for x64 CPU architecture (table 6). It may seem that having 68 instructions for just interpolating colors is too much. This is partly true; here we can see the disadvantage of the blocked half-space algorithm: it must evaluate half-space distances and baricentrics for all pixels in block while the iterative half-space evaluates them incrementally. On the other hand, texture mapping introduces significant amount of computation making this disadvantage irrelevant.

Rasterization algorithm: scan-line vs half-space. Our first experiment was about comparison of existing rasterizations algorithms on a single core (table 1). We used SSE processor instructions to accelerate computations where possible. For scan-line and half-space columns we vectorized the calculations by coordinates and image channels (we call such approach "sse1" in table 2). For blocked half-space we used pixel vectorization (i.e. single command processes a bunch of pixels; we call this approach "simd(sse4)" and "simd(avx8)" depending on instruction length). Rasterization algorithms themselves were implemented in a fixed point. We further studied optimal tile size (which is related to vector length) in our experiments (table 2).

scene	half-space	blocked half-space	scan-line	scan-line (fast)	fill_color
T1	286 FPS	294 FPS	158 FPS	**400 FPS**	625 FPS
T2	**650 FPS**	417 FPS	83 FPS	117 FPS	667 FPS
T3	68 FPS	**91 FPS**	61 FPS	73 FPS	500 FPS
T4	76 FPS	**87 FPS**	48 FPS	53 FPS	400 FPS
T5	**57 FPS**	51 FPS	35 FPS	46 FPS	250 FPS
T6	**50 FPS**	40 FPS	19 FPS	22 FPS	116 FPS

Table 1: Time for different rasterization algorithms. Each implementation was accelerated with SSE instructions. All numbers (FPS, Frames Per Second) are measured for single thread and 1024x1024 resolution. The higher is better. The last column fill_color is a tiled half-space algorithm filling all pixels with white color (like memset). We consider the performance of this case as the best possible one and compare the rest with respect to it. For this experiment we have used Intel Core i7 (3770, 3.4Ghz) CPU.

Experimental results show that the scan-line approach does have an advantage over half-space on large triangles and simple filling modes if a coverage test is removed (table 1, first row, scene T1). However, this advantage is easily eliminated by increasing block size in blocked half-space algorithm (table 2, fist row, avx8 column): blocked half-space gives

448 FPS versus 400 FPS (*this comparison, though is not quite correct since the numbers in tables 1 and 2 were measured on different processors, but we can rely on it because Xeon with a lower frequency in a single thread is usually slower than the Core-i7*) for scan-line (fast). In all other cases, half-space and blocked half-space show absolute advantage over scan-line approach.

Comparing half-space and blocked half-space approaches we can say that blocked half-space algorithm is usually better (table 1). The exceptions are scenes T2 and T6 where common half-space algorithms substantially defeated the vectorized version. This result is explained quite simply: T2 and T6 scenes contain a lot of small triangles which result in a large amount of partially-covered blocks for a block based rasterizer.

scene	pure_cpp	simd (sse1)	simd (sse4)	simd (avx8)	fill_color (sse4)
T1	147 FPS	297 FPS	427 FPS	**448 FPS**	588 FPS
T2	108 FPS	**204 FPS**	102 FPS	96 FPS	137 FPS
T3	35 FPS	61 FPS	83 FPS	**92 FPS**	500 FPS
T4	35 FPS	65 FPS	**74 FPS**	62 FPS	323 FPS
T5	26 FPS	**44 FPS**	42 FPS	33 FPS	119 FPS
T6	17 FPS	**36 FPS**	16 FPS	13 FPS	30 FPS

Table 2: Frames per second for different acceleration techniques for half-space (pure_cpp and sse1) and tiled half-space (simd(sse4), simd(avx8)) rasterizers. All numbers are measured for single thread and 1024x1024 resolution. The higher is better. The last column fill_color (sse4) is a tiled half-space algorithm filling all pixels with white color (like memset). We consider the performance of this case as the best possible one. For this experiment we have used Intel Xeon (5-2690 v4 2,6Ghz) CPU.

Combined approach. Such a result encourages us to combine sse1 and sse4 implementations: if a block is fully-covered, we used vectorized pixel processing; if a block is partially-covered we render its pixels subsequently using vectorization by coordinates or color channels (table 3, column "sse1+sse4"). It can be seen from table 3 that combined approach is good in average, but was not the best in all cases. We explain this by saying that blocked half-space implementation (and combined algorithm as follows) is much more complicated for branch prediction and speculative execution mechanisms. So, combined approach can be further improved: for triangles with small area use simple half space (sse1) and for other — cobmined (sse1+sse4) algorithms. This approach allowed us to fix performance for scenes with a large number of small triangles (T2 and T6).

Threads: sort-middle vs sort-last. As can be obvious from the previous work, most existing implementations use straitforward *sort-middle* approach subdividing image into bins. This approach supposes that pixel work dominates over geometry and rasterization itself. We also began with *sort-middle* approach but we have found that adding bins is in itself introducing essential overhead (table 4, second column). This happens due to essential duplicating of triangles that overlapped several bins and it becomes noticeable on geometrically-heavy scenes (T2, T5 and T6). Then we decided to try a different approach.

Having 4x4 blocked half-space algorithm, we decided to use spin-locks for 4x4 tile and thus implemented *sort-last*. We used *std::atomic_flag* [32] for spin-lock implementation.

The *sort-last*, in general (if we do not take into account the locks), should scale better due to it processes separate triangles in parallel. An additional advantage of this algorithm is locality and cache efficiency for triangles data: rasterized triangles are formed on the top the stack (or triangle queue) memory and they are in the cache.

If go further, *sort-last* could be optimized in such a way that it reads data directly from user input pointers, rasterizes triangles and immediately discards them (thus turning into a memory-compact and cache-effitient way). However, we did not do this because OpenGL has tremendous amount of ways for input user data layout.

scene	pure_cpp	simd (sse1)	simd (sse4)	simd (sse1+sse4)	fill_color (sse4)
T1	147 FPS	297 FPS	427 FPS	**430 FPS**	588 FPS
T2	108 FPS	**204 FPS**	102 FPS	197 FPS	137 FPS
T3	35 FPS	61 FPS	83 FPS	**84 FPS**	500 FPS
T4	35 FPS	65 FPS	74 FPS	**79 FPS**	323 FPS
T5	26 FPS	44 FPS	42 FPS	**46 FPS**	119 FPS
T6	17 FPS	**36 FPS**	16 FPS	31 FPS	30 FPS

Table 3: Comparison of suggested combined implementation (sse4+sse1). All numbers are measured for single thread and 1024x1024 resolution. The higher is better. The last column fill_color (sse4) is a tiled half-space algorithm filling all pixels with white color (like memset). We consider the performance of this case as the best possible one. For this experiment we have used Intel Xeon (5-2690 v4 2,6Ghz) CPU.

Although, the *sort-last* can be implemented in different ways, we used the simplest approach: a thread performs lock of 4x4 tile, processes pixels and then immediately unlocks the tile. For parallel processing of triangles we used a lock-free concurrent queue [33]. Some threads act as *producers* and push triangles into queue (1 or 2), while the others act as *consumers*, taking out triangles from the queue and performing rasterization. We did not limit the size of the queue, although we believe that for better cache efficiency it is worth doing, switching producer threads to consuming triangles when a limit has been exceeded.

Fig. 8 shows our experiment results. The *sort-middle* approach, as expected, was better for pixel-heavy scenes. However, for cases where pixel work was not enough, *sort-last* approach has won. The exception is T6 scene. This result seemed strange for us, especially in combination with the fact that *sort-last* has shown almost linear scaling on T2 scene. Nevertheless, this result may be explained. Scene T2 consists of 8K small random triangles (which bounding boxes overlap only slightly) where each next triangle is located at random position on the screen. Scene T6 consists of successive *triangle strips* and also triangle bounding boxes overlap much more. We were able to achieve a slight performance increase (15-20%) by increasing the pulling portion size for the consumer up to 4 triangles (this reduces conflicts of threads if they process a single trip). However, threads that handle different strips still conflict much. Moreover, T6 scene is heavier for pixel processing than T2, so *sort-middle* has won here.

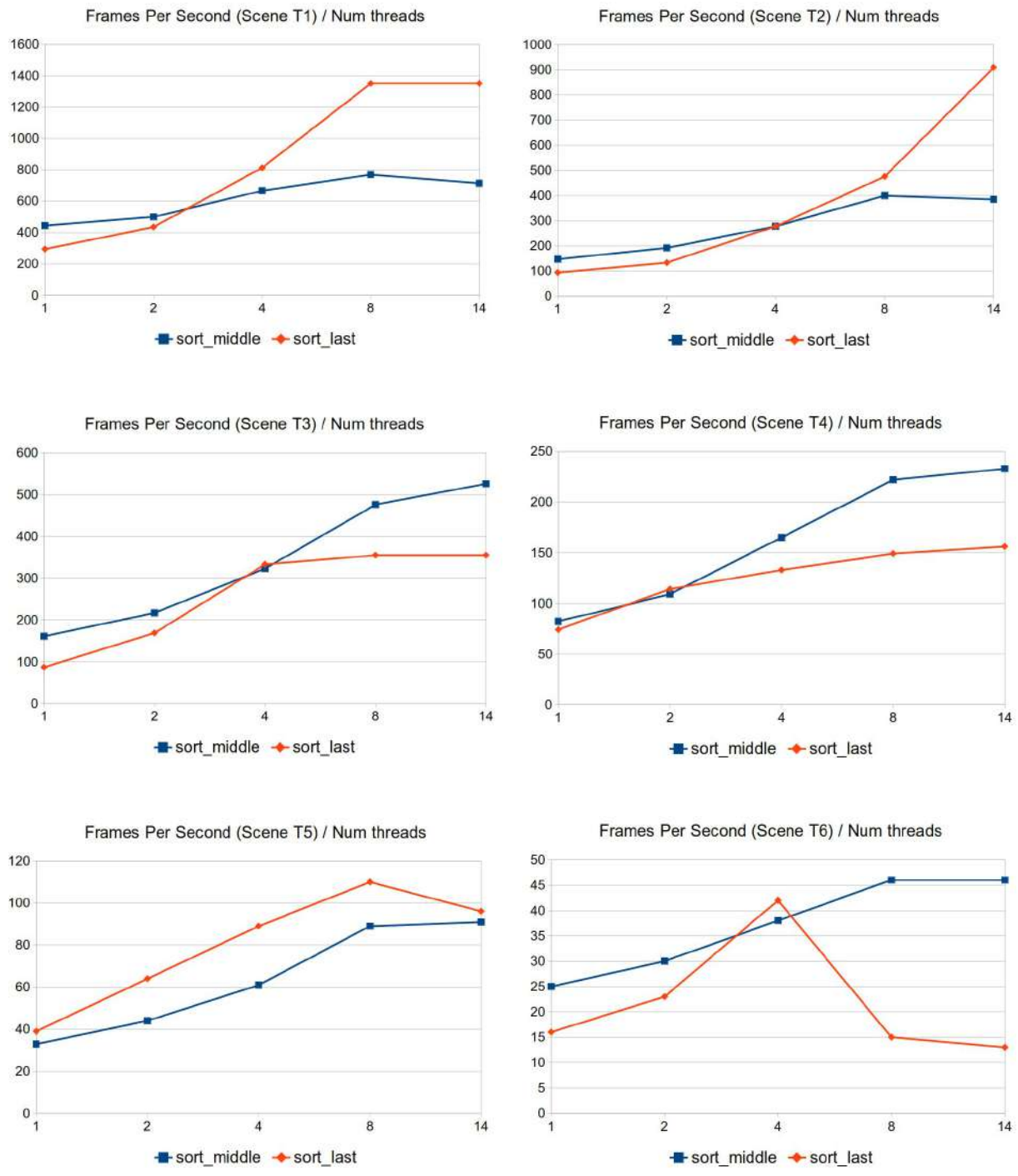


Fig. 8. Multithreading experiment. X axis — number of threads. Intel Xeon (5-2690 v4 2,6Ghz, 14 cores) CPU.

At last, we should make a note that in our comparison, a *sort-last* approach used pitch-linear buffer, and the *sort-middle* used a binned/tiled one. After comparing these two methods (pitch-linear vs tiled) in next subsection we can state that a *sort-last* approach can be even more efficient if it uses a tiled frame-buffer.

Framebuffer layout: pitch-linear vs tiled. Our next experiment was targeted to investigate memory subsystem efficiency when access frame buffer data. We assumed that frame buffer (and also depth buffer) access can be a bottelneck due to these buffers a priori can not be fit into the cache. Thus, some tiled frame buffer layout might be helpful because of less cache misses when accessing different rows (fig. 9).

We have investigated 4 different implementations (table 4):

1. pitch-linear frame and depth buffers layout. Default layout of 2d image by rows.
2. pitch-linear + binning overhead. This implementation has the same memory layout as a prevous one. However, it has bins for different 64x64 tiles and thus triangles that overlap several tiles should be duplicated. This implementation will show use binning overhead.
3. big tiles (64x64), i.e. bins. For this layout we split screen into 64x64 bins. For each bin inside we used pitch-linear layout.
4. Two-level tiling. At the first level, we split screen into 64x64 bins. At the second level we split each bin into 16x16 tiles thus making address linear inside the whole tile. Such layout will also allow wide vectors (for example, AVX512) being used for the whole 4x4 tile.

scene	pitch-linear	pitch-linear + bins	bins (64x64)	bins (64x64) + tiles (16x16)
T1	340 FPS	345 FPS	444 FPS	**476 FPS**
T2	113 FPS	103 FPS	147 FPS	**145 FPS**
T3	98 FPS	99 FPS	161 FPS	**169 FPS**
T4	99 FPS	82 FPS	132 FPS	**141 FPS**
T5	85 FPS	46 FPS	82 FPS	**91 FPS**
T6	40 FPS	20 FPS	33 FPS	**34 FPS**

Table 4: Comparison of pitch-linear and tiled frame buffer layouts. The higher is better. Blocked half-space algorithm was used (4x4). For this experiment we have used single thread and Intel Xeon (5-2690 v4 2,6Ghz) CPU. First column shows a default pitch-linear framebuffer layout. Second column demonstrates overhead we got from binned approach by itself: some triangles are duplicated due to they overlap several bins. Third column shows performance for binned approach and the last one — for two-level bins (64x64) + small tiles (4x4) approach.

Thus, memory layout has an extremely large impact on performance and tiled layout sould be definitely used.

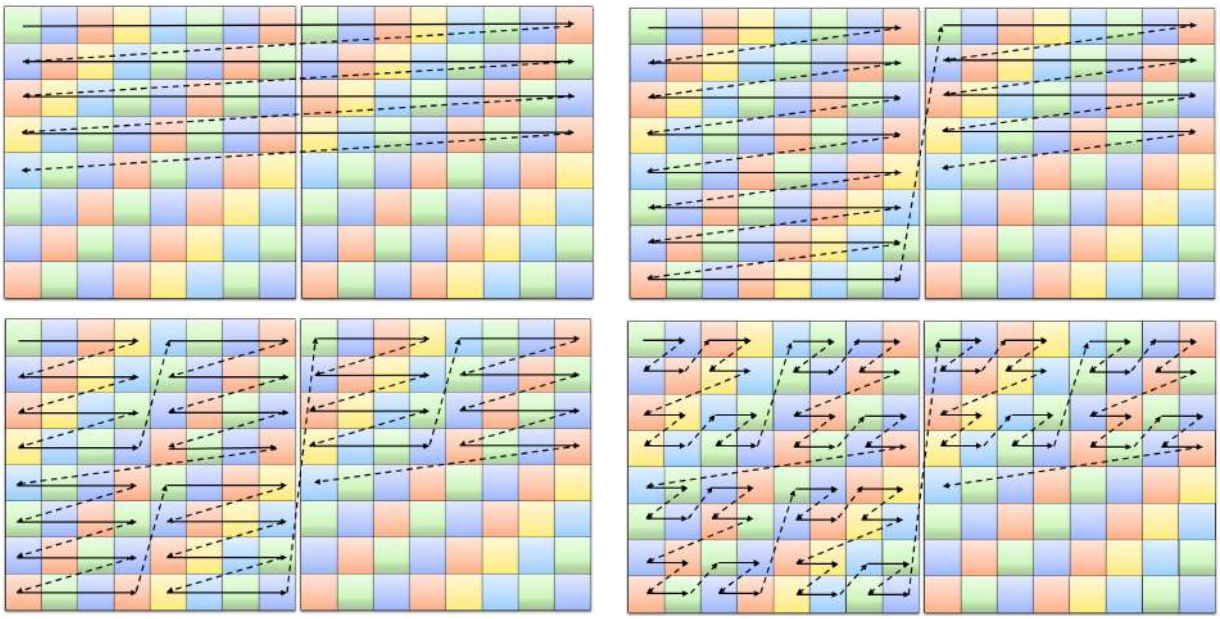


Fig. 9. Different framebuffer layout illustration. On this image, a bin size is shown to be 8x8, however on practice it was 64x64. Default pitch-linear is shown at top left image. Binned/tiled — top right. Two level (64x64 bins + 4x4 tiles) layout is shown at bottom left and Morton code layout [34] is shown at bottom right. With algorithmic point of view, the last one has better 2D locality [34]. However, Morton code evaluation is expensive and will also complicate half-space distances evaluation. At the same time, we would like to guarantee that all pixels in line have a subsequent addresses. This allow us reading/writing line of pixels with the single instruction and easily change length of instruction to test both SSE (for 4x4 tiles) and AVX (for 8x8 tiles).

CPU architecture: In Order vs Out Of Order. Our last experiment was aimed to study efficiency of different processor architectures for software graphics pipeline and rasterization. A trade off between performance and other CPU characteristics (such as energy efficiency, heat dissipation and cost) is essential for embedded systems. It is well known that the most significant performance gained on modern CPUs gives super scalar Out Of Order execution pipeline. This mechanism, at the same time, is the most expensive one. Our assumption is that with a large number of vector operations and independent instruction flow, software graphics pipeline should work well even on an in-order processor. Another reason we make this comparison is that in-order processors are more easily implementing precise exceptions which are important for safety-critical applications.

Since our blocked half-space algorithm is implemented via platform-independent light-weight vector library, we could easily port it to ARM. Unfortunalely our SSE1 implementation is heavily platform dependent (though, various options are exists [35]), so in this experiment we tested only pixel vectorization (blocked half-space algorithm). Using compiler explorer [31], we have counted instructions for different arhitectures and pixel processing modes (table 6). This information would allow us to more accurately evaluate how well the pipeline was loaded by the arithmetic instructions.

For this test we have selected several CPUs (table 7). First two processors (A83T and Cortex A53) are 2-way super scalar in-order machines. The i.MX6 (Cortex A9) has

scene	pitch-linear	bins (64x64)	bins (64x64) + tiles (16x16)
T1	340 FPS	444 FPS	**476 FPS**
T2	113 FPS	**167 FPS**	164 FPS
T3	98 FPS	161 FPS	**169 FPS**
T4	99 FPS	150 FPS	**155 FPS**
T5	85 FPS	102 FPS	**114 FPS**
T6	40 FPS	60 FPS	**63 FPS**

Table 5: Comparison of pitch-linear and tiled frame buffer layouts **without binning overhead**. The higher is better. Half-space block (4x4) algorithm was used. For this experiment we have used single thread and Intel Xeon (5-2690 v4 2,6Ghz) CPU. First column shows a default pitch-linear framebuffer layout. Second column shows performance for binned approach and the last one — for two-level bins (64x64) + small tiles (4x4) approach.

2-way super scalar out of order pipeline. The Core-i5 2410M (Sandy Bridge) has 4-way super scalar out of order pipeline. In addition to a wider pipeline, Sandy Bridge has many floating point ALUs, so it can execute 16 single precision floating point operations per clock (4 SIMD instructions per clock, each of 4 floats).

CPU arch/mode	Colored2D	Colored3D	Textured3D
x86/x64	68	100	290
ARMv7	79	110	500
ARMv8	60	86	250

Table 6: Comparison of instruction count per pixel for different rasterization states and CPU architectures. GCC compiler. 4x4 tiles were used. Colored2D includes color interpolation only. Colored3D — color interpolation with the perspective correction and a depth test. Textured3D adds bilinear texture fetch and perspective correction of texture coordinates to the previous mode. We have observed a significant increase in the number of instructions for ARMv7 and *Textured3D* mode due to spilling registers to memory. We used GCC 5.4.0 for both ARM cases.

We further introduce a special metric (equation 10, fig. 10) to compare *in-order* vs *out of order* from measured frames per second (table 7). We do this because in our experiments we used different CPUs with different architectures, manufacturing technology (for example 14 and 28 nm) and frequency. Our reason is straitforward: we don't want to compare the absolute performance values for different processors like table 7 does. Instead of that, we would like to approximately match instructions per clock for different CPUs to know whether *out of order* gives a benefit for our problem or not. Thus, if for a some CPU we have more instructions than for the other, we do not consider this a disadvantage for our comparison and we also do not want to take into account any inefficiencies introduced by the compiler. For this reason, the instruction count is in the numerator. At the same time frequency should be in denominator to bring all measures to a single frequency.

scene	A83T(ARMv7)		Cort.A53(ARMv8)		i.MX6 (ARMv7)		Core-i5 (x86/x64)	
	pure_cpp	SIMD	pure_cpp	SIMD	pure_cpp	SIMD	pure_cpp	SIMD
T1	16,7	17,5	26.3	35,9	13.9	19.2	96	191
T2	17,9	21,2	33,2	19,0	14,1	6,6	79	77
T3	5,8	6,5	7,4	14,5	4,5	6,3	27	67
T4	6,0	6,2	8,0	12,2	4,8	5,1	28	55
T5	4,5	4,6	6,1	7,0	3,7	3,0	22	33
T6	2,8	2,2	3,6	2,7	4,1	1,3	14	13

Table 7: Performance of a single-pixel (pure_cpp) and vectorized (SIMD) versions. Frames Per Second (FPS). Single thread, 1024x1024 and offscreen rendering. Binned frame-buffer (64x64 pixels) is used. A83T and Cortex A53 are in order machines; i.MX6 and Core-i5 are out of order ones. For this test we have used a laptop version of Core-i5 CPU (2410M, 2.3 GHz).

$$Efficiency = \frac{FPS * Instructions}{Frequency}. \tag{10}$$

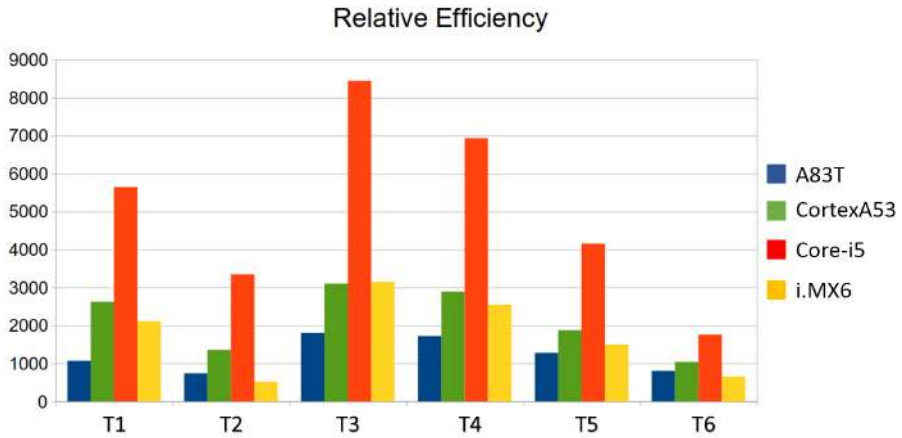


Fig. 10. Relative CPUs efficiency (equation 10). This efficiency could be thought as a relative instruction per clock (IPC). All histogram columns were obtained from SIMD columns of table 7.

Fig. 10 shows that the out of order (OOO) execution mechanism in itself gives only a very little benefit in average (compare A83T over i.MX6 — they both have 2-way execution pipeline, but i.MX6 have OOO and the A83T don't have it). The loss of the i.MX6 on T2 and T6 scenes can be easily explained — this is a result of expensive pipeline flush for the out of order CPU due to large amount of branch misprediction and complex code path in the blocked half-space algorithm; we rendered partially covered blocks with common (not vectorized) C++ code and therefore branch misprediction forces the CPU to flush pipeline and start executing another piece of code. Core-i5 has speculative execution mechanism and thus amortizes this problem. At the same time, in-order machine with greater amount of registers (ARMv8) shown better IPC (fig 10). Therefore, more registers combined with better code density for ARMv8 in Cortex A53 shown much better absolute performance than OOO execution added to ARMv7 in i.MX6 CPU (table 7).

2.3.3 Comparison with other implementations

We have compared our implementation to Mesa on A83T (ARMv7) and Mesa-OpenSWR on Intel Core-i7 CPUs (x86/x64). For Core-i7 we used high performance OpenSWR [6] implementation on Windows 7 and for A83T we used default Mesa 10.5.4 software rasterizer on Ubuntu Linux 16.04.6 LTS (xenial), BPI-M3 dev-board [36]. All comparisons were done in 1024x1024 resolution for windows and in 1024x640 for Linux on BPI-M3 due to maximum resolution limitation; please also note that this time (table 8) we have to include frame buffer display time into the comparison and therefore our numbers for A83T CPU in tables 8 and 7 are slightly differ.

Scn/CPU	A83T(ARMv7), 1 and 4 threads			Core-i7, 4 threads	
Scn/OGL	Mesa (1 thread)	Ours (1 thread)	Ours, 4 threads	OpenSWR	Ours
T1	5.6 FPS	7.0 FPS	**16.5 FPS**	**400 FPS**	240 FPS
T2	4.2 FPS	5.7 FPS	**15.0 FPS**	136 FPS	**150 FPS**
T3	1.0 FPS	6.7 FPS	**14.8 FPS**	**270 FPS**	210 FPS
T4	0.77 FPS	7.5 FPS	**8.2 FPS**	**220 FPS**	101 FPS
T5	0.45 FPS	5.0 FPS	**6.1 FPS**	**110 FPS**	63 FPS
T6	0.26 FPS	3.3 FPS	**4.8 FPS**	33 FPS	**40** FPS

Table 8: Comparison of our implementation to Mesa and OpenSWR OpenGL implementations. In this comparison, we used several optimizations altogether (such as tiled frame buffer and multithreading).

On x86/x64 our implementation [37] could not beat OpenSWR on pixel-heavy scenes (table 8). However, we were faster on T2 and T6 scenes where our combined approach (sse4+sse1, section 2.3.2) has shown its advantage. Our code was designed to quickly test the maximum number of different rendering techniques. So, considering that OpenSWR is made by Intel for the x86/x64 architecture only (and it simply can not run on the others), it would be naive to expect excellence from our experimental implementation for all cases. We believe that OpenSWR generates better vectorized code (processing a half of 4x4 tile with a single AVX instruction, for example). Also OpenSWR could proceed better with multithreading due to our experiments revealed problems for both studied methods (*sort-middle* and *sort-last*).

On the other hand, with the same software implementation, we can significantly outperforms default Mesa rasterizer on ARM which was the only avaliable software solution for BPI-M3 board during our work with it; according to our information there is no working graphics driver for Ubuntu Linux on BPI boards and therefore the whole rendering is performed actually in the software. Many other developent boards suffer the same problem on practice (along with patent issues [38]).

3 CONCLUSIONS

In this article we investigated various high performance graphics rasterization algorithms and techniques to be accelerated on different modern processor architectures. Prac-

tical and scalable solutions for software implementation of OpenGL graphics pipeline on modern multicore CPUs were elaborated.

The experimental results demonstrated unexpected results — even on fairly simple test scenes popular methods (*sort-middle* and blocked half space) substantially lost to rarely used ones (*sort-last* and simple half space). Relying on these results, a combined approach (blocked half-space + half-space) accelerated with SIMD instructions was introduced that have beaten extremely optimized OpenSWR implementation for 2 scenes. At the same time our implementation outperform default Mesa rasterizer on ARM CPUs an order of magnitude, which demonstrates the relevance of this area of research. We also offered a special metric for benchmarking relative Instructions Per Clock (IPC) for different CPUs without special tools, and this metric shows relatively low efficiency of the out of order mechanism itself for pixel processing. The more particular conclusions are shown further:

1. Half-space rasterization methods are absolutely better than scanline ones;
2. SIMD pixel processing for blocked half-space rasterizer gives essential benefit, but has limitations:
   (a) small triangles degrade performance, so the combined approach should be used;
   (b) wide vectors on architectures with low amount of vector registers may not have benefit due to high register pressure, increased number of instructions and spilling intermediate results to memory (table 7).
3. Even in such a computationally intensive task as pixel processing during rasterization (where the ratio of computational instructions to memory operations is greater than 100:1), memory access is still a seriously performance limit. Tiled frame buffer and depth buffers layouts increase performance up to 60%;
4. Despite our multithreading implementation is far from perfect (we don't have linear acceleration for most cases), we believe that the *sort-last* approach is more perspective, although it is non trivial.
5. For the considered problem out of order (OOO) machines have essential benefit if the OOO machine has significantly larger maximum instructions per clock than an in order one. It is more essential to have larger maximum thoroughput of floating point instructions (i.e. have for floating point ALUs).

When we first started our work, we were sure that it would be more technical and that all the research that could be done in this area had already been done due to the popularity of GPUs today. However, on practice, everything turned out to be differently. We could not find a single optimal approach for the implementation of software rasterization and graphics pipeline. Moreover, we found that with the advent of GPUs, researchers mostly ignore real-time software rendering. At the same time, processors were actively developing, so we believe that this field is the fertile ground for the future performance research.

REFERENCES

[1] T. Akenine-Moller, E. Haines, N. Hoffman, P. Angelo, I. Michal, and S. Hillaire. *Real-Time Rendering*, 4 ed. Natick, MA, USA: A. K. Peters, Ltd. (2018).

[2] G. K. Borovin and V. V. Lapshin, "Motion control of a space robot," *Mathematica Montisnigri*, **41**, 166–173 (2018). http://www.montis.pmf.ac.me/vol41/13.pdf

[3] K. V. Ryabinin and S. I. Chuprina, "Using scientific visualization systems to automate monitoring of data generated by lightweight programmable electronic devices," *Programming and Computer Software*, **44** (4), 278–285 (2018).

[4] OpenGL, "The mesa 3d graphics library." (2019) https://www.mesa3d.org/ (accessed August 8, 2019)

[5] Microsoft, "Windows advanced rasterization platform (warp) guide." (2018) https://docs.microsoft.com/en-us/windows/win32/direct3darticles/directx-warp (accessed August 8, 2019)

[6] OpenSWR, "A high performance, highly scalable software rasterizer for opengl." (2019) https://www.openswr.org/index.html (accessed August 8, 2019)

[7] M. M. Zieojevic, D. S. Jokanavic, and D. Baralic, "Sofware "cindrella" and its application in visualization of physics and mathematics," *Mathematica Montisnigri*, **34**, 86–93 (2015). http://www.montis.pmf.ac.me/vol34/6.pdf

[8] W. J. Bouknight, "A procedure for generation of three-dimensional half-toned computer graphics presentations," *ACM*, **13**(9), 527–536 (2019). http://doi.acm.org/10.1145/362736.362739

[9] B. Molkenthin, "Software rasterization algorithms for filling triangles." (2010) http://www.sunshine2k.de/coding/java/TriangleRasterization/TriangleRasterization.html (accessed August 8, 2019)

[10] J. E. Bresenham, "Algorithm for computer control of a digital plotter," *IBM Syst. J.*, **4** (1), 25–30 (1965). http://dx.doi.org/10.1147/sj.41.0025

[11] A. Fujimoto, T. Tanaka, and K. Iwata, "Arts: Accelerated ray-tracing of system tutorial," *Computer Graphics, Image Synthesis, Computer Science Press, Inc, New York, NY, USA,.*, 148–159 (1988). http://dl.acm.org/citation.cfm?id=95075.95111

[12] S. Laine and T. Karras, "High-performance software rasterization on gpus." 79–88 *New York, NY, USA*: *ACM*. (2011) http://doi.acm.org/10.1145/2018323.2018337

[13] J. Beam, "Tutorial – introduction to software-based rendering: Triangle rasterization." (2009) https://www.joshbeam.com/articles/triangle_rasterization/ (accessed August 8, 2019)

[14] J. Pineda, "A parallel algorithm for polygon rasterization," *SIGGRAPH Comput. Graph.*, **22** (4), 17–20 (1988). http://doi.acm.org/10.1145/378456.378457

[15] P. Mileff, K. Nehéz, and J. Dudra, "Accelerated half-space triangle rasterization," *Acta Polytechnica Hungarica*, **12** (7), 217-236 (2015). https://www.uni-obuda.hu/journal/Mileff_Nehez_Dudra_63.pdf

[16] M. Abrash, "Rasterization on larrabee," *Dr. Dobbs*, **37** (2009). http://drdobbs.com/high-performance-computing/217200602

[17] Scratchapixel, "Rasterization: a practical implementation," (2016). https://www.scratchapixel.com/lessons/3d-basic-rendering/rasterization-practical-implementation/rasterization-practical-implementation (accessed August 8, 2019)

[18] B. Barladian, A. Voloboy *et al.*, "Efficient implementation of opengl sc for avionics embedded systems," *Programming and Computer Software*, **44** (4), 207–212 (2018). DOI: 10.1134/S0361768818040059

[19] VCDECIDE, "Fighting force (nintendo 64 vs playstation) side by side comparison," (2019). https://www.youtube.com/watch?v=izOPJA2MyC0 (accessed August 8, 2019)

[20] Wikipedia, "Playstation technical specifications." (2019) https://en.wikipedia.org/wiki/PlayStation_technical_specifications (accessed August 8, 2019)

[21] E. Koskinen, M. Parkinson, and M. Herlihy, "Coarse-grained transactions," *SIGPLAN Not.*, **45** (1), 19–30 (2010). http://doi.acm.org/10.1145/1707801.1706304

[22] M. Rhu, M. Sullivan, J. Leng, and M. Erez, "A locality-aware memory hierarchy for energy-efficient gpu architectures," *Proceedings of the 46th Annual IEEE/ACM International Symposium on Microarchitecture*, 86–98 (2013). http://doi.acm.org/10.1145/2540708.2540717

[23] Z. Bethel, "A modern approach to software rasterization." University Workshop, Taylor University (2011).

[24] P. Mileff and J. Dudra, *Advanced 2D Rasterization on Modern CPUs.* Springer International Publishing, 63–79 (2014). https://doi.org/10.1007/978-3-319-01919-2_5

[25] P. Mileff and J. Dudra, "Modern software rendering," *Production Systems and Information Engineering*, **6**, 55-66 (2012).

[26] Y. Kim, J.-E. Jo, H. Jang, M. Rhu, H. Kim, and J. Kim, "Gpupd: A fast and scalable multi-gpu architecture using cooperative projection and distribution," *Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture*, 574–586 (2017).

[27] J. Sugerman, K. Fatahalian, S. Boulos, K. Akeley, and P. Hanrahan, "Gramps: A programming model for graphics pipelines," *ACM Trans. Graph.*, **28**(1), 4:1–4:11 (2009). http://doi.acm.org/10.1145/1477926.1477930

[28] M. Kenzel, B. Kerbl, D. Schmalstieg, and M. Steinberger, "A high-performance software graphics pipeline architecture for the gpu," *ACM Trans. Graph.*, **37** (4), 140:1–140:15 (2018).

[29] L. Seiler, D. Carmean, E. Sprangle, T. Forsyth, M. Abrash, P. Dubey, S. Junkins, A. Lake, J. Sugerman, R. Cavin, R. Espasa, E. Grochowski, T. Juan, and P. Hanrahan, " *ACM SIGGRAPH 2008 Papers* "Larrabee: A many-core x86 architecture for visual computing *SIGGRAPH '08.* 18:1–18:15 *New York, NY, USA*: *ACM.* (2008) http://doi.acm.org/10.1145/1399504.1360617

[30] K. Group, "dfdx, dfdy – return the partial derivative of an argument with respect to x or y." (2019) https://www.khronos.org/registry/OpenGL-Refpages/gl4/html/dFdx.xhtml (accessed August 8, 2019)

[31] M. Godbolt, "Compiler explorer," (2019). https://godbolt.org/ (accessed August 8, 2019)

[32] cppreference.com, "std::atomic_flag." (2019) https://en.cppreference.com/w/cpp/atomic/atomic_flag (accessed August 8, 2019)

[33] C. Moody Camel, "A fast multi-producer, multi-consumer lock-free concurrent queue for c++11." (2019) https://github.com/cameron314/concurrentqueue (accessed August 8, 2019)

[34] K. P. Lorton and D. S. Wise, "Analyzing block locality in morton-order and morton-hybrid matrices," *SIGARCH Comput. Archit. News*, **35** (4), 6–12 (2007). http://doi.acm.org/10.1145/1327312.1327315

[35] K. A. Batuzov, "The use of vector instructions of a processor architecture for emulating the vector instructions of another processor architecture," *Programming and Computer Software*, **43** (6), 366–372 (2017).

[36] Sinovoip, "Banana pi m3 octa core development board." (2019) http://www.banana-pi.org/m3.html (accessed August 8, 2019)

[37] V. Frolov, B. Barladian, and V. Galaktionov, "Swgl: Experimental opengl1 software implementation; github repository," (2019). https://github.com/FROL256/SWGL (accessed August 8, 2019)

[38] T. A. Polilova, "Ethical norms and legal framework of scientific publication," *Mathematica Montisnigri*, **45**, 129–136 (2019).

Received October 5, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

ВЫВОД В РАМКАХ НЕЭКСТЕНСИВНОЙ КИНЕТИКИ КРИТЕРИЯ ГРАВИТАЦИОННОЙ НЕУСТОЙЧИВОСТИ ДЖИНСА ДЛЯ ДОПЛАНЕТНОГО ВРАЩАЮЩЕГОСЯ ОБЛАКА С УЧЕТОМ РАДИАЦИИ И МАГНИТНОГО ПОЛЯ

А.В. КОЛЕСНИЧЕНКО

Институт прикладной математики им. М.В. Келдыша РАН. Москва, Россия
*Ответственный автор. E-mail: kolesn@keldysh.ru,
web page: http: keldysh.ru/kolesnichenko/person.htm

DOI: 10.20948/mathmontis-2020-47-14

Ключевые слова: Неэкстенсивная кинетика Тсаллиса, гравитационный критерий Джинса, допланетное газовое облако, чернотельное излучение.

Аннотация. В настоящее время неэкстенсивная статистическая механика Тсаллиса успешно применяется к космическим системам с дальним силовым взаимодействием, которое и является причиной их аномальности (стататистической и термодинамической неаддитивности). Известно, что гравитационная неустойчивость является фундаментальным процессом фрагментации гравитирующего допланетного облака. В конечном счете, именно она вызывает формирование устойчивых астрофизических объектов, таких как звезды, туманности, допланетные пылевые сгущения, аккреционные диски и т. д. При этом в случае нормальных звезд большую роль играет давление излучения как фактор их гидростатического равновесия.

В данной работе на основе кинетики Тсаллиса рассмотрена проблема гравитационной неустойчивости Джинса для протяженного самогравитирующего плазменного облака, заполнявшего все пространство прото-солнечной системы, с учетом влияния неэкстенсивности среды, вращения и магнитного поля на критическую длину волны возмущения, ведущей к неустойчивости. Обобщённые критерии гравитационной неустойчивости Джинса найдены из соответствующих дисперсионных соотношений, полученных как для нейтрального вещества, состоящего из смеси совершенного q-газа и чернотельного излучения, так и для плазмы. Определены функциональные зависимости критического значения длины возмущающей волны от энтропийного индекса деформации q, размерности пространства скоростей D и коэффициента β, характеризующего долю излучения в полном давлении системы. Эти свободные параметры должны задаваться в каждом конкретном случае из статистических или экспериментальных данных. Показано, что и радиационное давление стабилизирует вещество неэкстенсивных допланетных облаков. Для вращающейся намагниченной плазмы критерии неустойчивости Джинса модифицируются силой Кориолиса и магнитным полем только в поперечном режиме распространения волн возмущения. Полученные здесь результаты помогут, по мнению автора, лучше понять некоторые астрофизические проблемы, связанные, в частности, с моделированием процессов образования звезд и экзопланет из звездных туманностей.

2010 Mathematics Subject Classification: 85A35, 91B50, 82C40.
Key words and Phrases: Non-extensive Kinetics of Tsallis, Gravitational Criterion of Jeans, Pre-planet Gas Cloud, Blackbody Radiation.

CONCLUSION IN THE FRAMEWORK OF THE NON-EXTENSIVE KINETICS OF JEANS` GRAVITATIONAL INSTABILITY CRITERION FOR A PREPLANETARY ROTATING CLOUD WITH ACCOUNT OF RADIATIONS AND MAGNETIC FIELD

A.V. KOLESNICHENKO

Keldysh Institute of Applied Mathematics, Russian Academy of Science
*Corresponding author. E-mail: kolesn@keldysh.ru,
web page: http: keldysh.ru/kolesnichenko/person.htm

DOI: 10.20948/mathmontis-2020-47-14

Summary. At the present time, non-extensive statistical mechanics of Tsallis is successfully applied to space systems with long-range force interaction, which is the reason for their anomaly (statistical and thermodynamic non-extensity). It is known that gravitational instability is a fundamental process of fragmentation of gravitating cosmic matter. It causes the formation of stable astrophysical objects, such as stars, nebulae, pre-planetary dust condensations, accretion disks, etc. Wherein, in the case of normal stars, the radiation pressure as a factor in their hydrostatic equilibrium plays an important role.

It is on the basis of statistical mechanics of Tsallis that the paper considers the problem of Jeans gravitational instability for an extended self-gravitating plasma cloud that fills the entire space of the proto-solar system, taking into account the influence of medium nonextension, rotation, and magnetic field on the critical wavelength of the perturbation leading to instability. The generalized criteria for Jeans' gravitational instability are found from the corresponding dispersion relations obtained both for a neutral substance consisting of a mixture of perfect q-gas and blackbody radiation, and for plasma. The functional dependences of the critical value of length of the perturbing wave on the entropy strain index q, the dimension of the velocity space D, and the coefficient β, characterizing the fraction of radiation in the total pressure of the system are determined. These free parameters should be specified in each case from statistical or experimental data. It was shown that radiation pressure stabilizes the matter of non-extensive preplanet clouds. For a rotating magnetized plasma, the Jeans instability criteria are modified by the Coriolis force and magnetic field only in the transverse mode of propagation of perturbation waves. The results obtained here will help, according to the author, a better understanding of some astrophysical problems related, in particular, to modeling the processes of formation of stars and exoplanets from stellar nebulae.

ВВЕДЕНИЕ

Как теперь стало понятно, статистическая механика Больцмана−Гиббса и классическая статистическая термодинамика не являются вполне универсальными теориями, поскольку они имеют ограниченные области применимости. Это связано, в частности, с тем, что в основе этой статистики лежит гипотеза молекулярного хаоса. А это, в свою очередь, означает, что любой выделенный объем приобретает по истечении времени настолько хорошо развитую хаотическую структуру, что при $t \to \infty$ его точки могут располагаться в произвольной части фазового пространства. Таким образом, фазовое пространство в классической статистике не содержит запрещенных состояний и обладает обычными свойствами непрерывности, гладкости, евклидовости. При этом стохастический процесс имеет марковский характер, а гипотеза перемешивания, дополненная предположением о бесконечном числе степеней свободы, приводит, в конечном счете, к каноническому (экспоненциальному) распределению вероятности состояний Больцмана−Гиббса, из которого следует свойство аддитивности экстенсивных термодинамических переменных, таких как внутренняя энергия, энтропия и т.п., а в случае кинетической теории - к максвелловскому распределению скоростей.

Вместе с тем, в физике и в других естественных науках, использующих методы статистической механики, известны многочисленные примеры аномальных систем с дальним силовым взаимодействием, фрактальным характером фазового пространства и значительными корреляциями между отдельными их частями. Сложная пространственно-временная структура подобных систем приводит к нарушению принципа аддитивности для таких важнейших термодинамических величин, как энтропия или внутренняя энергия. Моделирование эволюции подобных систем, обладающих произвольным фазовым пространством, возможно, в частности, в рамках так называемой неэкстенсивной статистической механики Тсаллиса, важным преимуществом которой является асимптотический степенной закон распределения вероятностей.

В настоящее время теории разнообразных неэкстенсивных систем развиваются в ускоренном темпе, при котором появляются новые идеи, позволяющие глубже понять их природу, возможности и ограничения. Каждая такая теория имеет широкий спектр важных приложений, связанных с физикой статистических систем, вероятностные свойства которых описываются не гиббсовыми (и не гауссовыми), а степенными распределениями. В частности, неэкстенсивная статистическая механика успешно применяется к космическим системам с дальним силовым взаимодействием, которое и является причиной их аномальности (стататистической и термодинамической неэкстенсивности).

Как известно, при неустойчивости неравновесных систем (в частности, различных астрофизических газопылевых объектов) возникает динамический хаос, что делает возможным образование более сложных упорядоченных (в общем случае фрактальных) структур. Возникновение фрактальных структур подтверждается для многих астрофизических систем, в частности, у звезд, межзвездных молекулярных облаков, аккреционных допланетных дисков и т.д. При учете сильного гравитационного поля в моделях эволюции подобных аномальных структур возникают принципиальные трудности, поскольку для них традиционные газодинамические и термодинамические методы описания часто неприемлемы. Преодоление этих трудностей требует нового подхода к решению эволюционных задач в космогонии. Один из возможных подходов к моделированию эволюции космогонических систем может быть основан на методах неэкстенсивной статистической механики

Тсаллиса[i)], как раз и предназначенной для описания эволюции газопылевых сред с дальним (сильным) гравитационным воздействием, которое и является причиной их аномальности (см., например, [1-16]). Важным отличием неэкстенсивной статистики Тсаллиса от классической статистики Больцмана–Гиббса является наличие асимптотического степенного закона распределения вероятностей (появляющегося при максимизации параметрической энтропии Тсаллиса), который не зависит от экспоненциального поведения, обусловленного распределением Гиббса. Тем не менее, основанная на параметрической энтропии неэкстенсивная статистика Тсаллиса представляет собой всё же обобщение, а не альтернативу статистике Больцмана–Гиббса, поскольку она распространяет область применимости классической статистической теории на неэкстенсивные системы только путём расширения математической формы их энтропийного функционала.

Самогравитирующая среда становится гравитационно-неустойчивой, если возникшие в ней сколь угодно малые возмущения плотности неограниченно растут со временем вследствие тяготения и равновесие нарушается, если соответствующие длины волн превышают определенное значение. В частности, с джинсовской гравитационной неустойчивостью связан процесс фрагментации самогравитирующего околозвёздного облака. Именно она вызывает, в конечном счете, образование и эволюцию астрофизических объектов, таких как аккреционные диски, допланетные пылевые сгущения, планетезимали и т. д. (см., [18-24]). Проблеме гравитационной неустойчивости космических объектов в последнее время посвящено большое число публикаций, среди которых можно выделить следующие публикации [25-50]. Во всех этих работах рассмотрены различные аспекты джинсовской неустойчивости самогравитирующих газовых сред как в рамках классических уравнений Навье–Стокса и МГД-уравнений, так и на основе бесстолкновительного уравнения Больцмана при наличии гравитационных полей и уравнения Пуассона.

Вместе с тем в работах [6,7,9,10,12,14-16] были развиты термодинамический и газодинамический (на основе модифицированного кинетического уравнения с интегралом столкновений в форме Бхатнагара–Гросса–Крука) подходы, позволяющие моделировать эволюцию космогонических систем в рамках формализма деформированной статистической механики Тсаллиса. С учетом полученных в них результатов в представленной работе выполнено в рамках неэкстенсивной кинетики Тсаллиса рассмотрение влияния радиации на гравитационную неустойчивость Джинса для допланетного вращающегося плазменного облака (точнее его экваториальной части, в которой практически все излучение является длинноволновым, поскольку оно уже успело пройти через многократное поглощение и переизлучение частицами среды). Именно в этой области возможно существование локального термодинамического равновесия, при котором температура частиц практически совпадает с температурой черного тела.

[i)] Обзорам исследований в рамках неэкстенсивной статистики Тсаллиса посвящены многочисленные журнальные статьи, сборники и монографии. Кроме этого, имеется постоянно обновляющаяся полная библиография (Nonextensive statistical mechanics and thermodynamics: Bibliography/ http://tsallis.cat.cbpf.br/biblio.htm), которая на сегодняшний день состоит из более 5600 ссылок [17].

1. ИСХОДНЫЕ УРАВНЕНИЯ q- ГИДРОДИНАМИКИ

Рассмотрим далее газообразную динамическую неэкстенсивную систему с нормированным распределением частиц $f(\mathbf{r},\mathbf{c},t)$ в геометрическом пространстве $\mathbf{r}$ и в пространстве скоростей $\mathbf{c}$ с размерностью D. Предлагаемое Тсаллисом обобщение статистической механики (в случае статистики Курадо−Тсаллиса) лучше всего описывается следующими двумя аксиомами [4,6,]:

Аксиома 1. Функционал энтропии, связанный с нормированным распределением функции вероятностей $f(\mathbf{z},t)$ равен

$$\mathcal{S}_q[f]=\frac{k_{\mathcal{B}}}{q-1}\int d\mathbf{z}\left\{f(\mathbf{z})-[f(\mathbf{z})]^q\right\}, \tag{1}$$

где q − параметр деформации – число, связанное с фрактальной размерностью, а для неэкстенсивных систем, являющееся мерой их неаддитивности [3]; $\mathbf{z}=(\mathbf{r},\mathbf{c})$ − элемент объёма фазового пространства; $d\mathbf{z}\equiv d\mathbf{r}d^D\mathbf{c}$, где D − размерность пространства скоростей; $k_{\mathcal{B}}$ − постоянная Больцмана.

Аксиома 2. Экспериментально измеряемое значение любой макроскопической величины $\langle\mathcal{A}\rangle_q$ (термодинамической характеристики q-системы) задаётся соотношением

$$\langle\mathcal{A}\rangle_q\equiv\int d\mathbf{z}\mathcal{A}(\mathbf{r},t)[f(\mathbf{z})]^q, \tag{2}$$

где $\mathcal{A}(\mathbf{r},t)$ − соответствующая микроскопическая величина.

Важно подчеркнуть, что энтропия $\mathcal{S}_q(\mathcal{A}\cup\mathcal{B})$ двух независимых систем не является аддитивной термодинамической переменной при $q\neq 1$, поскольку [3]

$$\mathcal{S}_q(\mathcal{A}\cup\mathcal{B})=\mathcal{S}_q(\mathcal{A})+\mathcal{S}_q(\mathcal{B})+k^{-1}(1-q)\mathcal{S}_q(\mathcal{A})\mathcal{S}_q(\mathcal{A}).$$

Несмотря на это обстоятельство, в литературе было показано, что существует значительное количество обычных статистических и термодинамических свойств, которые q-инвариантны, т. е. справедливы для любого q. К ним, в частности, относятся свойство выпуклости энтропии, структура равновесных канонических ансамблей, неаддитивная термодинамика, структура преобразования Лежандра и многое другое (см. [17].

Основные определения. Энтропия Тсаллиса влечёт за собой не только обобщение статистической физики и термодинамики, но и обобщение физической кинетики и гидродинамики [12,51,52]. Простейшей макроскопической величиной является q-плотность числа частиц, которая определяется соотношением

$$n_q(\mathbf{r},t)\equiv\int[f(\mathbf{z})]^q d^D\mathbf{c}. \tag{3}$$

Тогда массовая q-плотность равна $\rho_q(\mathbf{r},t)\equiv m\,n_q(\mathbf{r},t)$. Поскольку частица, движущаяся со скоростью $\mathbf{c}$, обладает импульсом $m\mathbf{c}$, то выражение

$$\mathbf{u}_q(\mathbf{r},t) \equiv \int m\mathbf{c}\left[f(\mathbf{z})\right]^q d^D\mathbf{c} / \rho_q(\mathbf{r},t) \tag{4}$$

определяет гидродинамическую скорость элемента объёма. Величина

$$\varepsilon_q(\mathbf{r},t) = \rho_q^{-1}\int \tfrac{m}{2}\left|\mathbf{c}-\mathbf{u}_q\right|^2 \left[f(\mathbf{z})\right]^q d^D\mathbf{c} \tag{5}$$

является удельной внутренней q-энергией (на единицу массы) неэкстенсивной системы. Потоки

$$\mathcal{P}_q(\mathbf{r},t) = m\int(\mathbf{c}-\mathbf{u}_q)(\mathbf{c}-\mathbf{u}_q)[f(\mathbf{z})]^q d^D\mathbf{c}, \tag{6}$$

$$\mathcal{J}_q(\mathbf{r},t) \equiv \tfrac{1}{2}m\int\left|\mathbf{c}-\mathbf{u}_q\right|^2(\mathbf{c}-\mathbf{u}_q)[f(\mathbf{z})]^q d^D\mathbf{c} \tag{7}$$

представляют собой соответственно тензор давлений и поток тепла. Гидростатическое q-давление определяется как

$$p_q(\mathbf{r},t) = \tfrac{1}{3}\mathcal{P}:\mathcal{I} = \tfrac{1}{3}m\int\left|\mathbf{c}-\mathbf{u}_q\right|^2 \left[f(\mathbf{z})\right]^q d^D\mathbf{c}, \tag{8}$$

где $\mathcal{I}$ − единичный тензор второго ранга. В частности, если сдвиговые напряжения равны нулю, а нормальные напряжения равны между собой, то $\mathcal{P}_q = p_q\mathcal{I}$.

Система уравнений q гидромеханики. В рамках неэкстенсивной статистической механики Тсаллиса в работах [12,51] было проведено методом моментов конструирование гидродинамических и квазигидродинамических уравнений на основе модифицированного кинетического уравнения Больцмана[ii)] с интегралом столкновений в форме Бхатнагара−Гросса− Крука):

$$\left(\frac{\partial}{\partial t} + \mathbf{c}\cdot\nabla + \mathcal{F}_q\cdot\nabla_{\mathbf{c}}\right)\left[f(\mathbf{r},\mathbf{c},t)\right]^q = -\frac{\left[f(\mathbf{r},\mathbf{c},t)\right]^q - \left[f^{(0)}(\mathbf{r},\mathbf{c},t)\right]^q}{\tau}. \tag{9}$$

Здесь $\nabla_{\mathbf{c}} \equiv \mathbf{i}_x\partial/\partial c_x + \mathbf{i}_y\partial/\partial c_y + \mathbf{i}_z\partial/\partial c_z$; $\mathcal{F}_q(\mathbf{r},t) = \mathbf{f}/m - grad\ \Psi_q(\mathbf{r},t)$ − не зависящая от скорости внешняя сила (сила тяжести) отнесённая к единице массы; $\mathbf{f}$ − сила негравитационного происхождения (например, электромагнитная сила Лоренца); $\Psi_q(\mathbf{r},t) \equiv -G\int\frac{m}{\left|\mathbf{r}-\mathbf{r}'\right|}[f(\mathbf{z}',t)]^q \mathrm{d}\mathbf{z}'$ − гравитационный потенциал, удовлетворяющий уравнению Пуассона $\Delta\Psi_q(\mathbf{r}) = 4\pi G\int mf^q d^D\mathbf{c}$; G − гравитационная постоянная; τ − положительный параметр, который интерпретируется как характерное время релаксации произвольной функции распределения f к обобщённому локально- максвелловскому распределению (величина τ совпадает по порядку величины со средним временем свободного

ii) В цитируемой работе кинетическая теория была основана на операторе столкновений Бхатнагера−Гросса−Крука (BGK), который был обобщён для произвольного значения параметра q.

пробега частиц в системе). Равновесное распределение $f^{(0)}(\mathbf{r},\mathbf{c})$, в случае когда $q>1$, определяется следующей формулой (см., например, [10])

$$f^{(0)}(\mathbf{r},\mathbf{c})=\left\{c_{q,D}\frac{\rho_q}{m}\left(\frac{m}{2\pi k_{\mathcal{B}}T_q}\right)^{D/2}\right\}^{1/q}\left\{1-(1-q)\frac{m(\mathbf{c}-\mathbf{u}_q)^2}{2k_{\mathcal{B}}T_q}\right\}^{1/(1-q)}, \tag{10}$$

где $c_{q,D}=\dfrac{(1-q)^{D/2}\Gamma(\frac{q}{1-q})}{\Gamma(\frac{q}{1-q}-\frac{D}{2})}$; $\Gamma(x)=\int_0^\infty t^{x-1}\mathrm{e}^{-t}dt$ – Гамма-функция.

В результате были получены следующие моментные уравнения q- гидродинамики, которые являются обобщением на произвольное значение параметра q обычных гидродинамических уравнений Навье–Стокса:

$$\frac{\partial\rho_q}{\partial t}+\nabla\cdot(\rho_q\mathbf{u}_q)=0, \tag{11}$$

$$\frac{\partial(\rho_q\mathbf{u}_q)}{\partial t}+\nabla\cdot\left(\mathcal{P}_q+\rho_q\mathbf{u}_q\mathbf{u}_q\right)=n_q\mathbf{f}-\rho_q\nabla\psi_q\ , \tag{12}$$

$$\frac{\partial(\rho_q\varepsilon_q)}{\partial t}+\nabla\cdot\left\{\mathcal{J}_q+\rho_q\varepsilon_q\mathbf{u}_q\right\}+\mathcal{P}_q:\nabla\mathbf{u}_q=0. \tag{13}$$

Уравнения (11)-(13) не является в общем случае замкнутыми, поскольку отсутствует необходимая связь (определяющие соотношения) потоковых величин ($\mathcal{P}_q$ и $\mathcal{J}_q$) и скалярных характеристик течения (ρ_q, $\mathbf{u}_q$ и T_q). Эта связь может быть найдена с помощью решения модельного кинетического уравнения (9) методом Чепмена–Энскога при использовании общего асимптотического разложения функции распределения по числу Кнудсена. Этот метод был использован, в частности, в работе [51]; в результате были найдены определяющие соотношения, замыкающие систему (11)-(13). В случае приближения нулевого порядка, когда распределение $f\equiv f^{(0)}$ (т.е. является обобщённым локально-максвелловским распределением (10)), было показано, что тензор напряжения $\mathcal{P}_q$ сводится к шаровому тензору $\mathcal{P}_q^{(0)}\equiv p_q\mathcal{I}$, а поток тепла $\mathcal{J}_q=0$. При этом внутренняя энергия ε_q и гидростатическое давление p_q определяются соотношениями

$$\varepsilon_q=\frac{Dk_{\mathcal{B}}T_q}{2m}\left[1+(1-q\tfrac{D}{2})\right]^{-1}, \tag{14}$$

$$p_q=\frac{\rho_q k_{\mathcal{B}}T_q}{m[1+(1-q)\frac{D}{2}]}=\frac{2}{D}\rho_q\varepsilon_q. \tag{15}$$

Заметим, что поскольку определение температуры в q-кинетике Тсаллиса достаточно произвольно (оно зависит от довольно произвольного определения температуры с точки зрения множителей Лагранжа (см., например, [10]), то далее величина T_q интерпретируется как обобщённая температура сложной неаддитивной системы. Естественно, что эта температура, в корне отличается от абсолютной термодинамической температуры T, характеризующей интенсивность хаотизации (т.е. беспорядочного движения) частиц системы. Заметим, что если определить формулой $T_{eff} \equiv T / \left[1+(q-1)\frac{D}{2}\right]$ эффективную температуру q-системы, то для величины ε_q получим соотношение $\varepsilon_q = DkT_{eff} / 2m > 0$ (совпадающее при $q \to 1$ и $D = 3$ с определением внутренней энергии в статистике Больцмана−Гиббса), которое соответствует равному распределению энергии идеального газа по степеням свободы для всех q. Если сохранить обычные представления температуры и для обобщённой температуры T_q, то тогда неравенство $\varepsilon_q > 0$ накладывает жёсткое ограничение на величину параметра деформации q: в этом случае энтропийный индекс удовлетворяет неравенству $1 < q < 1 + 2/D$.

В приближении первого порядка определяющие уравнения для потока тепла $\mathcal{J}_q$ и тензора вязких напряжений $\mathcal{T}_q \equiv \mathcal{P}_q - p_q \mathcal{I}$ имеют вид:

$$\mathcal{J}_q(\mathbf{r},t) = -\lambda_q \nabla T\,, \tag{16}$$

$$\mathcal{T}(\mathbf{r},t) = \mu_q \left(\nabla \mathbf{u} + (\nabla \mathbf{u})^T - \frac{2}{3}\mathcal{I}\nabla \cdot \mathbf{u} \right), \tag{17}$$

где $\lambda_q = \tau \frac{k_{\mathcal{B}} p_q}{m} \frac{1+D/2}{1+(1-q)(1+D/2)}$ и $\mu_q = \tau\, p_q = \tau \frac{\rho_q k_{\mathcal{B}} T}{m[1+(1-q)\frac{D}{2}]}$ − соответственно коэффициенты теплопроводности и сдвиговой вязкости.

2. ЗАМКНУТАЯ СИСТЕМА УРАВНЕНИЙ q-ГИДРОДИНАМИКИ ДЛЯ ДОПЛАНЕТНОГО ОБЛАКА С РАВНОВЕСНЫМ ИЗЛУЧЕНИЕМ

В эволюции многих астрофизических объектов большую роль играет давление излучения, как фактор их гидростатического равновесия. Впервые анализ неустойчивости в аккреционных дисках относительно осесимметричных возмущений с учетом давления излучения был проведен в работе Шакуры и Сюняева [30]. В последующих работах рассматривались общие политропные модели [31], учитывались неосесимметричные возмущения [53], звуковые и эпициклические колебания [24,54] и т.д.

Ниже мы используем приведенную выше систему уравнений q- гидродинамики для моделирования неустойчивости околосолнечного допланетного облака (толстого диска), вещество которого состоит из смеси с q-газа и чёрнотельного изотропного излучения при температуре T, распространяющегося по всем направлениям. Будем предполагать, что допланетное облако оптически толстое и распределение поля излучения близко к равновесному. Подчеркнём также, что облако в значительной мере обладает осевой симметри-

ей, что является следствием его вращения вокруг центральной звезды. Далее будем также предполагать, что облако- самогравитирующее, для которого вертикальная структура (вдоль оси вращения) определяется балансом сил давления и гравитации самого диска.

В случае пренебрежения гидродинамическими диссипативными процессами и нагревом космического вещества, обусловленным диссипацией и процессами ионизации и возбуждения, исходная система q-уравнений, состоящая из аналога уравнений Эйлера и уравнения Пуассона, имеет вид[iii)] (см., например, [10]):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{18}$$

$$\frac{d\mathbf{u}}{dt} = -\frac{1}{\rho}\nabla P - \nabla \psi, \tag{19}$$

$$\Delta \psi = 4\pi G \rho, \tag{20}$$

$$\frac{d\mathcal{E}}{dt} = -\frac{P}{\rho}\nabla \cdot \mathbf{u} + \frac{dQ}{dt}. \tag{21}$$

где соотношением $d\mathcal{A}/dt \equiv \partial \mathcal{A}/dt + (\mathbf{u}\cdot\nabla)\mathcal{A}$ определяется полная производная структурной величины $\mathcal{A}(\mathbf{r},t)$ по времени. Здесь

$$P(\mathbf{r},t) = p_q + p_{rad} \equiv p_q + aT^4/3, \tag{22}$$

$$\mathcal{E}(\mathbf{r},t) = \varepsilon_q + \varepsilon_{rad} \equiv \varepsilon_q + aT^4/\rho \tag{23}$$

− соответственно полное давление и полная внутренняя энергия (на единицу массы) смеси идеального q-газа и чёрнотельного излучения; $\rho dQ/dt = -\nabla \cdot \mathcal{J}_Q$; $\mathcal{J}_Q$ − суммарный вектор теплового потока, учитывающий в принципе все термодинамически обратимые процессы, которые могут уносить тепло из элемента среды при его движении; $\varepsilon_q(\mathbf{r},t) = c_{vq}T(\mathbf{r},t) = \frac{D}{2+(1-q)D}\frac{k_{\mathcal{B}}T(\mathbf{r},t)}{m}$ − внутренняя энергия (на единицу массы газовой составляющей допланетного диска); $\varepsilon_{rad} = aT^4/\rho$ − энергия излучения чёрного тела, находящаяся в единице массы; $p_q(\mathbf{r},t) = \frac{2}{2+(1-q)D}\frac{k_{\mathcal{B}}}{m}T(\mathbf{r},t)\rho(\mathbf{r},t) = \frac{2}{D}\rho\varepsilon_q$ − газовое давление в неэкстенсивной дисковой системе (аналог закона состояния в кинетической теории идеальных газов); T − абсолютная температура; $p_{rad} \equiv aT^4/3$ − лучевое давление;

[iii]) Здесь и далее индекс “q” у ряда гидродинамических и термодинамических переменных мы будем опускать.

a − постоянная излучения Стефана−Больцмана; $\psi(\mathbf{r},t) = -G\int_V \frac{\rho(\mathbf{r}',t)}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}'$ − гравитационный потенциал, являющийся решением уравнения Пуассона (8) (интеграл здесь берётся по всему объёму V, занимаемому допланетным облаком); G − гравитационная постоянная; $c_{vq} = \frac{D}{2+(1-q)D}\frac{k_B}{m}$ − удельная изохорная теплоёмкость газовой составляющей смеси. Определим также показатель адиабаты газового вещества диска, как отношение $\gamma_q \equiv \gamma_{gas} = c_{pq}/c_{vq}$. Тогда $\gamma_q \equiv \gamma_{gas} = 2 - q + 2/D$, $\gamma_1 = (2+D)/D$.

Уравнение для полной внутренней энергии смеси (21) удобно переписать, используя уравнение неразрывности (18), в форме первого начала термодинамики $dQ/dt = d\mathcal{E}/dt + Pdv/dt$, которое остаётся справедливым и для неэкстенсивных систем [10, 11], или в виде тождества Гиббса

$$Td\mathcal{S}/dt \equiv dQ/dt = d\mathcal{E}/dt + Pdv/dt, \tag{24}$$

выражающего скорость $d\mathcal{S}/dt$ изменения энтропии $\mathcal{S}$ (на единицу массы) дискового вещества и излучения при движении элемента среды вдоль его траектории (здесь $v(\mathbf{r},t) = 1/\rho$ − удельный объём).

Изоэнтропические изменения в среде, содержащей q-газ и радиацию. Далее мы будем рассматривать такие движения космического вещества (находящегося в состоянии идеального q-газа) и чёрнотельного излучения, для которых энтропия каждой частицы среды остается в первом приближении постоянной на протяжении всего пути частицы, т.е. $d\mathcal{S}/dt \equiv \partial\mathcal{S}/\partial t + \mathbf{u}\cdot\nabla\mathcal{S} = 0$. Подобные обратимые и адиабатические движения являются изоэнтропическими. Для них энергетическое уравнение (21) сводится к уравнению

$$\rho d\mathcal{E}/dt + P\nabla\cdot\mathbf{u} = 0, \tag{25}$$

выражающему тот факт, что скорость изменения полной внутренней энергии движущегося элемента среды равна работе по сжатию этого элемента, совершаемой окружающей средой.

Вместе с тем, для астрофизических целей часто удобно использовать другие формы уравнения (25) (которые впервые были выведены Эддингтоном [55] и Чандрасекхаром [20]. Эти формы справедливы, когда давление P и внутреннюю энергию $\mathcal{E}$ можно вычислить из соответствующих уравнений состояния как функций от удельного объёма v и температуры T (или энтропии $\mathcal{S}$) в зависимости от исследуемого процесса. Для «медленного» процесса, характеризуемого временем, много большим времени теплопередачи, любые возмущения профиля температуры будут успевать релаксировать. Следовательно, этот процесс можно рассматривать как изотермический, при котором $P = P(v,T_0) = P(v)$. «Быстрый» процесс (по сравнению с процессом теплопереноса) можно считать адиабатическим в силу нехватки времени для обмена теплом двух соседних областей:

$\mathcal{S} = \mathcal{S}_0 = const$ и $P = P(v, \mathcal{S}_0) = P(v)$.

Из энергетического уравнения (25) для квазистатического процесса следует

$$\left(\frac{\partial \mathcal{E}}{\partial T}\right)_v dT + \left(\frac{\partial \mathcal{E}}{\partial v}\right)_T dv + Pdv = \frac{v}{T}\left(12 p_{rad} + \frac{c_{vq}}{c_{pq} - c_{vq}} p_q\right) dT + (4 p_{rad} + p_q) dv. \tag{26}$$

Следовательно, для изоэнтропических изменений имеем

$$\left(12 p_{rad} + \frac{1}{\gamma_q - 1} p_q\right) d \ln T + (4 p_{rad} + p_q) d \ln v = 0. \tag{27}$$

Введём теперь адиабатические показатели смеси вещества и излучения Γ_1, Γ_2 и Γ_3 соотношениями

$$\frac{d}{dt} \ln P = \Gamma_1 \frac{d}{dt} \ln \rho, \tag{28}$$

$$\frac{d}{dt} \ln T = (\Gamma_3 - 1) \frac{d}{dt} \ln \rho = \frac{\Gamma_2 - 1}{\Gamma_2} \frac{d}{dt} \ln P, \tag{29}$$

которые могут быть использованы вместо энергетического уравнения (25). С учётом уравнения состояния «идеального q-газа» (15) можно записать

$$dP = d(p_{rad} + p_q) = (4 p_{rad} + p_q) d \ln T - p_q d \ln v. \tag{30}$$

Следовательно, (28) есть не что иное, как

$$\frac{4 p_{rad} + p_q}{T} d T + \left[\Gamma_1 \frac{(p_{rad} + p_q) - p_q}{v}\right] dv = 0. \tag{31}$$

Из (27) и (31) следует, что

$$\frac{12 p_{rad} + (\gamma_q - 1)^{-1} p_q}{4 p_{rad} + p_q} = \frac{4 p_{rad} + p_q}{\Gamma_1 (p_{rad} + p_q) - p_q}. \tag{32}$$

Введём теперь в рассмотрение величину $\beta \equiv p_{gas} / P$ – коэффициент, характеризующий долю вещества в полном давлении системы[iv)]. При использовании этого параметра, соотношение (32) можно переписать в виде:

[iv)] На особую важность отношения $(1 - \beta)$ для теории звездной структуры впервые указал Эддингтон. В известном отрывке из его книги «Внутреннее строение звезд» Эддингтон связывал это отношение с «явлением звезды» («happening of the stars»).

$$\Gamma_1 = \beta + \frac{(4-3\beta)^2(\gamma_q - 1)}{\beta + 12(\gamma_q - 1)(1-\beta)}, \quad (\gamma_q - 1 = 1 - q + 2/D). \tag{33}$$

Можно легко показать, что имеют место следующие соотношения

$$\Gamma_2 = \frac{(4-3\beta)\Gamma_1}{\beta + 3(1-\beta)\Gamma_1} = 1 + \frac{(4-3\beta)(\gamma_q - 1)}{3(\gamma_q - 1)(1-\beta)(4+\beta)},$$

$$\Gamma_3 = 1 + \frac{\Gamma_1 - \beta}{4-3\beta} = 1 + \frac{\Gamma_1(\Gamma_2 - 1)}{\Gamma_2} = 1 + \frac{(4-3\beta)(\gamma_q - 1)}{\beta + 12(\gamma_q - 1)(1-\beta)}.$$

Если $p_{rad} \ll p_q$, то все обобщённые показатели адиабаты Γ_j для «q-газа + излучение» совпадают с показателем адиабаты чистого q-газа $\left(\gamma_q = 2/D + 2 - q\right)$, а когда присутствует одно лишь излучение абсолютно чёрного тела ($p_q \ll p_{rad}$), то они равны $4/3$. Таким образом, для смеси «идеального q-газа» и радиации обобщенные показатели адиабаты принимают промежуточные значения от $4/3$ до γ_q.

3. ДЖИНСКОВСКАЯ ГРАВИТАЦИОННАЯ НЕУСТОЙЧИВОСТЬ В НЕЭКСТЕНСИВНОЙ КИНЕТИЧЕСКОЙ ТЕОРИИ

Рассмотрим сначала простейшую задачу возникновения неустойчивости в бесконечной покоящейся сферически однородной газовой среде. Напомним, что при рассмотрении гравитационной неустойчивости Дж. Джинс рассматривал однородное состояние самогравитирующей среды в состоянии покоя, что не совсем корректно, так как такое состояние не является состоянием равновесия. Тем не менее, его вывод критерия неустойчивости можно рассматривать как первое приближение, которое в наиболее простых случаях дает правильный порядок нижней критической длины волны возмущения, ведущего к неустойчивости (см., например, [22,24].

Линеаризованные основные дифференциальные уравнения (18)-(21) для случая чисто радиального сферически симметричного движения с учетом допущений, что невозмущенное состояние является равновесным $(u = u_0 + u',\ u_0 = 0)$ и что уравнение Пуассона (20) можно применить лишь к возмущениям плотности (условие $\psi_0 \cong 0$ называют иногда «мошенничеством» Джинса [18,19], имеют вид:

$$\frac{\partial \rho'}{\partial t} + \frac{\partial \rho_0 u}{\partial r} = 0, \tag{34}$$

$$\frac{\partial u}{\partial t} = \frac{1}{\rho_0}\frac{\partial P'}{\partial r} - \frac{\rho'}{\rho_0^2}\frac{\partial P_0}{\partial r} - \frac{\partial \psi'}{\partial r}, \tag{35}$$

$$\frac{d}{dt}\left(\frac{P'}{P_0}\right)=\Gamma_{1,0}\frac{d}{dt}\left(\frac{\rho'}{\rho_0}\right), \tag{36}$$

$$\frac{\partial^2\psi'}{\partial r^2}=4\pi G\rho'. \tag{37}$$

Здесь и далее индекс «0» относится к невозмущенным величинам.

Уравнение (36) тривиально интегрируется. Выбирая постоянную интегрирования так, чтобы $P'=0$ при $\rho'=0$, получим

$$P'/P_0)=\Gamma_{1,0}\rho'/\rho_0. \tag{38}$$

Допустим теперь, что характерная длина, связанная с пространственными изменениями величин P_0 и ρ_0, велика по сравнению с другими характерными длинами задачи (это так называемое приближение коротковолновой акустики), т.е. можно пренебречь производными $\partial P_0/\partial r$ и $\partial\rho_0/\partial r$. При этих дополнительных упрощающих предположениях уравнение неразрывности, импульса и энергии легко объединить в одно уравнение для адиабатической звуковой волны[v)] [56]

$$\frac{\partial^2\rho'}{\partial t^2}+v_{S,q}^2\frac{\partial^2\rho'}{\partial r^2}-4\pi G\rho_0\rho'=0. \tag{39}$$

Здесь возмущенная производная давления $\partial P'/\partial r$ выражается, согласно (38), через возмущенную производную плотности $\partial\rho'/\partial r$ в виде $\partial P'/\partial r=(\Gamma_{1,0}P_0/\rho_0)\partial\rho'/\partial r=v_{S,q}^2\partial\rho'/\partial r$, где

$$v_{S,q}\equiv\sqrt{\Gamma_{1,0}\frac{P_0}{\rho_0}}=\left\{\frac{p_{q0}}{\rho_0}\left[1+(\Gamma_{3,0}-1)\frac{4-3\beta_0}{\beta_0}\right]\right\}^{\frac{1}{2}}=$$

$$=\left\{\frac{1}{(\gamma_q-1)D/2}\frac{k_B T_0}{m}\left[1+\frac{(4-3\beta_0)^2(\gamma_q-1)}{\beta_0^2+12\beta_0(\gamma_q-1)(1-\beta_0)}\right]\right\}^{\frac{1}{2}} \tag{40}$$

− адиабатическая (или лапласова) скорость звука в неэкстенсивной радиационной гидродинамике. При написании (40) учтено, что

$$\frac{P_0}{\rho_0}=\frac{p_{q,0}+p_{rad,0}}{\rho_0}=\frac{1}{\beta_0}\frac{p_{q,0}}{\rho_0}=\frac{1}{\beta_0}\frac{1}{(\gamma_q-1)D/2}\frac{k_B T_0}{m}=\frac{1}{\beta_0}\frac{1}{1+(1-q)D/2}\frac{k_B T_0}{m}. \tag{41}$$

[v)] Отметим, что при изучении возмущённых состояний самогравитирующего космического вещества часто приходится иметь дело с разновидностью звуковых волн.

В частном случае, когда $q=1$ и $D=3$, имеем $\gamma_1 = 5/3$ (классический идеальный одноатомный газ). Тогда из (40) следует, что

$$v_{S,1} \equiv \left\{ \frac{k_B}{m} T_0 \left[1 + \frac{2(4-3\beta_0)^2}{3\beta_0(8-7\beta_0)} \right] \right\}^{\frac{1}{2}}. \qquad (40^*)$$

Если излучение также отсутствует, то $(v_{S,1})_{\beta_0=1} \equiv v_{gas,1} = \sqrt{\gamma_1 k_B T_0 / m}$ – адиабатическая скорость звука в идеальном газе.

В случае когда $q \neq 1$ (идеальный q-газ), а излучение отсутствует ($\beta_0 = 1$), будем иметь

$$(v_{S,q})_{\beta_0=1} = \left[\frac{k_B T_0}{m} \frac{2\gamma_q}{(\gamma_q - 1)D} \right]^{\frac{1}{2}} = \left[\frac{k_B T_0}{m} \frac{2-q+2/D}{(1-q)D/2+1} \right]^{\frac{1}{2}}. \qquad (40^{**})$$

Уравнение (39) является линейным и однородным уравнением в частных производных, следовательно, к нему применим метод нормальных колебаний (метод мод). Решая уравнения (39) для возмущенной плотности в виде $\rho' \sim \exp(-i\omega t + i\mathrm{k} r)$, описывающем волны с угловой частотой ω, волновым вектором k в направлении r [vi)] и длиной волны $\lambda_r = 2\pi/\mathrm{k}$, получим следующее дисперсионное уравнение для бегущей волны

$$\omega^2 - \mathrm{k}^2 \frac{p_{q,0}}{\rho_0} \left\{ 1 + \frac{\Gamma_{1,0} - \beta_0}{4 - 3\beta_0} \left(1 + 4\frac{1-\beta_0}{\beta_0} \right) \right\} + 4\pi G \rho_0 = 0, \qquad (42)$$

которое с учетом соотношений (40) и (41) принимает «стандартный» вид

$$\omega^2 = \mathrm{k}^2 v_{S,q}^2 - 4\pi G \rho_0. \qquad (42^*)$$

Здесь адиабатическая скорость звука $v_{S,0}$ определяется формулой (40).

Для устойчивых волн с частотами ω имеем $\omega^2 > 0$, тогда как неустойчивость соответствует условию $\omega^2 < 0$. Эти два класса разделяет случай нейтральной устойчивости $\omega^2 = 0$, что соответствует модам с критической длиной волны возмущения

$$\lambda_{cr} = 2\pi/\mathrm{k}_{cr}, \quad \mathrm{k}_{cr}^2 = \omega_{cr}^2 / v_{Sq}^2, \quad \omega_{cr}^2 = 4\pi G \rho_0. \qquad (43)$$

Из уравнения (42^*) следует, что граничное значение $\mathrm{k} = \mathrm{k}_{cr}$ разделяет устойчивые ($\mathrm{k} > \mathrm{k}_{cr}$) и неустойчивые ($\mathrm{k} < \mathrm{k}_{cr}$) пульсации плотности. При малых k (длинные волны) пульсации будут расти со временем и появляется нестабильность Джинса, а коротковол-

[vi)] Следует заметить, что линеаризованное уравнение импульса требует, чтобы скорость $\mathbf{u}$ была параллельна волновому вектору $\pm\mathbf{k}$ [56]. Следовательно, скорости частиц жидкости, связанные с адиабатическими звуковыми волнами, параллельны направлению распространения волн.

новые пульсации плотности (большие k, малые длины волн) колеблются, т.е. распространяются в виде звуковых волн.

Таким образом, критическая длина волны возмущения

$$\lambda_{cr} = \frac{2\pi v_{Sq}}{\omega_{cr}} = \sqrt{\frac{\pi v_{Sq}^2}{G\rho_0}} \equiv \left\{ \frac{2\pi k_Б T_0}{mG\rho_0 D} \left[\frac{1}{\gamma_q - 1} + \frac{(4-3\beta_0)^2}{\beta_0^2 + 12\beta_0(\gamma_q - 1)(1-\beta_0)} \right] \right\}^{\frac{1}{2}} \tag{44}$$

является размером мельчайших «капель» рассматриваемой «фрактальной» газовой среды с излучением, которые могут удерживаться вместе собственным гравитационным притяжением. Следовательно, модифицированный в рамках неэкстенсивной кинетической теории критерий неустойчивости Джинса для смеси q-газа и чернотельной радиации будет выглядеть следующим образом: длина неустойчивой волны возмущения λ_r должна удовлетворять неравенству

$$\lambda_r \geq \lambda_{cr} = v_{Sq}\sqrt{\frac{\pi}{G\rho_0}} \equiv \left\{ \frac{\pi k_Б T_0}{mG\rho_0} \frac{2}{(\gamma_q - 1)D} \left[1 + \frac{(4-3\beta_0)^2(\gamma_q - 1)}{\beta_0^2 + 12\beta_0(\gamma_q - 1)(1-\beta_0)} \right] \right\}^{\frac{1}{2}}. \tag{45}$$

В традиционной литературе длину

$$\lambda_ج = \sqrt{\frac{\pi v_{gas}^2}{G\rho_0}} = \left(\gamma_1 \frac{\pi k_Б T_0}{mG\rho_0} \right)^{\frac{1}{2}}, \tag{46}$$

соответствующую размеру области сжатия самогравитирующего идеального газа, называют длиной Джинса. С учетом (45) критерий неустойчивости Джинса в неэкстенсивной кинетике может быть переписан в виде:

$$\frac{\lambda_r}{\lambda_ج} \geq \frac{v_{Sq}}{v_{gas}} = \left\{ \frac{1}{\gamma_1} \frac{2}{(\gamma_q - 1)D} \left[1 + \frac{(4-3\beta_0)^2(\gamma_q - 1)}{\beta_0^2 + 12\beta_0(\gamma_q - 1)(1-\beta_0)} \right] \right\}^{\frac{1}{2}} =$$

$$= \left\{ \frac{1}{\gamma_1} \frac{2/D}{(1-q+2/D)} \left[1 + \frac{(4-3\beta_0)^2(1-q+2/D)}{\beta_0^2 + 12\beta_0(1-q+2/D)(1-\beta_0)} \right] \right\}^{\frac{1}{2}} \equiv \Xi_q. \tag{45*}$$

Отсюда следует:

1. Если $q = 1$ (при этом $\gamma_1 = 1 + 2/D$), то фактор

$$\Xi_1 \equiv \left[\frac{1}{\gamma_1} \left(1 + \frac{(4-3\beta_0)^2 2/D}{\beta_0^2 + 24\beta_0(1-\beta_0)/D} \right) \right]^{\frac{1}{2}} > 1. \tag{47}$$

Следовательно, критическая длина волны возмущения λ_r в рассматриваемом случае больше джинсовской длины волны $\lambda_{\mathcal{J}}$, т.е. благодаря давлению излучения облачная среда стабилизируется, причем равенство соответствует предельной устойчивости.

2. Если $q \neq 1$, но излучение отсутствует $\beta_0 = 1$, то фактор

$$\Xi_q = \left[\frac{1}{\gamma_1} \left(2/D + \frac{2/D}{1 - q + 2/D} \right) \right]^{\frac{1}{2}}, \quad 0 < q < 1 + 2/D. \tag{48}$$

В этом случае критерий гравитационной неустойчивости зависит от численных значений индекса энтропийной деформации q и размерности пространства скоростей D. При этом возможна ситуация, при которой гравитационно-устойчивое (на основе классической статистики Больцмана−Гиббса) облако газа, будет неустойчивым согласно неэкстенсивной статистики Тсаллиса [14,15].

Связанная с λ_{cr} критическая масса (масса, содержащаяся внутри сферы диаметром λ_{cr}) определяется соотношением

$$\mathcal{M}_{cr} = (\pi/6)\rho_0 \lambda_{cr}^3 = \mathcal{M}_{\mathcal{J}} \Xi^3, \tag{49}$$

где

$$\mathcal{M}_{\mathcal{J}} \equiv (\pi/6)\rho_0 \lambda_{\mathcal{J}}^3 = (\pi/6)\rho_0 \left(\frac{\gamma_1 \pi k_B T_0}{mG\rho_0} \right)^{3/2} \tag{50}$$

− критическая масса Джинса. Возмущения с массой $\mathcal{M}_r$, превышающей критическую массу Джинса $\mathcal{M}_{\mathcal{J}}$ $(\Xi > 1)$ могут расти, формируя гравитационно-ограниченные структуры, в то время как возмущения с массой $\mathcal{M}_r$ меньше $\mathcal{M}_{\mathcal{J}}$ не растут и ведут себя как акустические волны. При этом для самогравитирующих неэкстенсивных сред с излучением критические значения длины волны и массы явно зависят от энтропийного индекса q, размерности пространства скоростей D и коэффициента β, которые, являясь свободными параметрами, должны определяться в каждом конкретном случае эмпирическим путем из экспериментальных данных. Это позволяет при исследовании неустойчивости самогравитирующих космических объектов в рамках неэкстенсивной статистики более обосновано моделировать реально складывающуюся ситуацию.

Заметим, что дальнейшее развитие предложенного здесь подхода может быть связано с учетом влияния на джинсовскую неустойчивость вращения среды, магнитного поля, вязкости и других диссипативных эффектов.

4. ГРАВИТАЦИОННАЯ НЕУСТОЙЧИВОСТЬ ВРАЩАЮЩЕГОСЯ ПЛАЗМЕННОГО ОБЛАКА С ЧЁРНОТЕЛЬНЫМ ИЗЛУЧЕНИЕМ

Поскольку вращение космогонических плазменных объектов является весьма распространенным феноменом во Вселенной, возникает вопрос: как эти факторы действует на джинсовскую неустойчивость? В связи с этим рассмотрим в упрощенной постановке проблему влияния силы Кориолиса на гравитационную неустойчивость неэкстенсивной среды допланетного плазменного облака с излучением. Исходные бездиссипативные уравнения в этом случае состоят из следующих уравнений: уравнений Эйлера в q-гидродинамике, уравнения Пуассона и уравнения магнитной индукции:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{51}$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{\nabla P}{\rho} + 2\mathbf{u} \times \mathbf{\Omega} + \frac{1}{\rho c}\mathbf{j} \times \mathcal{B} - \nabla \psi, \tag{52}$$

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = (\Gamma_3 - 1)\frac{T}{\rho}\left\{\frac{\partial \rho}{\partial t} + \mathbf{u} \cdot \nabla \rho\right\}, \tag{53}$$

$$\nabla^2 \psi = 4\pi G \rho, \tag{54}$$

$$\frac{\partial \mathcal{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathcal{B}), \quad \nabla \cdot \mathcal{B} = 0. \tag{55}$$

Здесь $\mathbf{\Omega} = \Omega_x \mathbf{i}_x + \Omega_z \mathbf{i}_z$; $\mathcal{B} = \mathcal{B}_0 \equiv \mathcal{B}_0 \mathbf{i}_z$ – магнитное поле; c – скорость света; $\mathbf{j} = \frac{c}{4\pi}\nabla \times \mathcal{B}$ – сила тока; $\Gamma_3 = 1 + \frac{(4-3\beta)(\gamma_q - 1)}{\beta + 12(\gamma_q - 1)(1-\beta)}$ – адиабатический показатель смеси вещества и чёрнотельного излучения; $\gamma_q \equiv \gamma_{gas} = c_{pq} / c_{vq} = \left(2\frac{D+1}{D} - q\right)$ – показатель адиабаты газового вещества диска.

Для изучения малых возмущений линеаризуем систему (51)-(55). Для этого представим входящие в эту систему переменные в виде сумм равновесных и возмущенных величин. В предположении, что для невозмущенного облака состояние его среды является однородным и равновесным $(\mathbf{u} = \mathbf{u}_0 + \mathbf{u}',\ \mathbf{u}_0 = 0)$ и что уравнение Пуассона (65) можно применять только к возмущениям плотности, линеаризованные уравнения (51)-(55) принимают вид:

$$\frac{\partial \rho'}{\partial t} + \rho_0 \nabla \cdot \mathbf{u}' = 0, \tag{56}$$

$$\frac{\partial \mathbf{u}'}{\partial t} - 2\mathbf{u}' \times \mathbf{\Omega} + \frac{\beta_0 P_0}{\rho_0}\left\{\nabla\left(\frac{\rho'}{\rho_0}\right) + \frac{4-3\beta_0}{\beta_0}\nabla\left(\frac{T'}{T_0}\right)\right\} + \frac{1}{4\pi\rho_0}\mathcal{B}_0 \times (\nabla \times \mathcal{B}') + \nabla \psi' = 0, \tag{57}$$

$$\frac{\partial T'}{\partial t}-(\Gamma_{3,0}-1)\frac{T_0}{\rho_0}\frac{\partial \rho'}{\partial t}=0, \tag{58}$$

$$\nabla^2\psi'-4\pi G\rho'=0, \tag{59}$$

$$\frac{\partial \mathcal{B}'}{\partial t}-\nabla\times(\mathbf{u}'\times\mathcal{B}_0)=0, \quad \nabla\cdot\mathcal{B}'=0. \tag{60}$$

Здесь $v_{S,q}=\left\{\frac{p_{q0}}{\rho_0}\left[\frac{4-3\beta_0}{\beta_0}(\Gamma_{3,0}-1)+1\right]\right\}^{\frac{1}{2}}$ – адиабатическая скорость звука в неэкстенсивной газовой среде с излучением; $\frac{\beta_0 P_0}{\rho_0}=\frac{k_{\mathcal{B}}}{m_0}\frac{T_0}{1+(1-q)D/2}$; $2\mathbf{u}'\times\mathbf{\Omega}=\{2u_y\Omega,-2u_x'\Omega,0\}$. Величины $\rho_0, T_0, P_0, \mathbf{u}_0$, $\mathcal{B}_0$ и β_0 описывают некоторое стационарное решение системы (56)-(60), а величины ρ', T', $\mathbf{u}'$, $\mathcal{B}'$ и ψ' – суть малые возмущения магнито-гидродинамических параметров, слабо нарушающих невозмущенное состояние.

В результате объединения уравнений (57) и (58) будем иметь

$$\frac{\partial \mathbf{u}'}{\partial t}-2\mathbf{u}'\times\mathbf{\Omega}+v_{S,q}^2\nabla\rho'+\nabla\psi'+\frac{1}{4\pi\rho_0}\mathcal{B}_0\times(\nabla\times\mathcal{B}')=0. \tag{61}$$

Система уравнений (56), (59)-(61) описывает развитие малых адиабатических возмущений во фрактальной плазменной среде с излучением на фоне основного решения в пространстве и во времени. Она является системой линейных и однородных уравнений в частных производных, следовательно, к ней применим метод нормальных колебаний (метод мод). Предполагая далее цилиндрическую симметрию движения $\mathbf{u}'(=\mathbf{i}_x u_x'+\mathbf{i}_z u_z')$[vii], а также что возмущенные параметры $\rho', \mathbf{u}', \psi'$ и $\mathcal{B}'$, эволюционируют по закону $\sim\exp(-i\omega t+i\,\mathrm{k}_x x+i\,\mathrm{k}_z z)$, где ω — частота гармонических колебаний (в общем случае комплексная величина), а $|\mathbf{k}|=\sqrt{\mathrm{k}_x^2+\mathrm{k}_z^2}$ — волновое число, в результате получим:

$$-\omega\rho'+\rho_0\mathbf{k}\cdot\mathbf{u}'=0, \tag{73}$$

[vii] Известно, что проблему устойчивости самогравитирующего газового облака в принципе нельзя описывать в рамках двумерного приближения, поскольку оно заведомо является сильно неустойчивым (см., например, [24]). Однако при наличие сильного внешнего гравитационного поля с цилиндрической геометрией и с образующей вдоль оси вращения облака, возможно обеспечить его устойчивость в случае, когда угловая скорость вращения достаточно велика. В этом случае структура допланетного облака вдоль оси вращения будет определяться исключительно его самогравитацией. Разумеется, этот случай искусственный, поскольку в реальных астрофизических системах такие цилиндрические поля если и встречаются, то без вложенных дисков. Вместе с тем, рассмотрение такого вложенного в цилиндр самогравитирующего газового диска представляет определённый математический интерес, поскольку только в этом случае можно выделить эффекты, к которым приводит самогравитация в чистом виде. Именно такие модели рассматривались в большинстве классических работ по астрофизическим дискам (см., например, [27,28,57].

$$-\omega\rho_0\mathbf{u}' + i2\mathbf{u}'\times\mathbf{\Omega} + v_{S,0}^2\rho'\mathbf{k} - \rho_0\mathbf{k}\psi' + \frac{\mathcal{B}_0}{4\pi}\left[\mathbf{i}_x(\mathrm{k}_x\,\mathcal{B}_z' - \mathrm{k}_z\,\mathcal{B}_x') - \mathbf{i}_y\,\mathrm{k}_z\,\mathcal{B}_y'\right] = 0 \qquad (74)$$

$$|\mathbf{k}|^2\,\psi' - 4\pi G\rho' = 0, \qquad (75)$$

$$-\omega\mathcal{B}_x' - \mathcal{B}_0 k_z u_x' = 0\,,\quad -\omega\mathcal{B}_y' - \mathcal{B}_0 k_z u_y' = 0\,,\quad -\omega\mathcal{B}_z' + \mathcal{B}_0 k_x u_x' = 0\,,\quad k_x\mathcal{B}_x' + k_z\mathcal{B}_z' = 0\,. \qquad (76)$$

Система уравнений (73)-(76) является исходной для дальнейшего анализа динамики малых возмущений в модели вращающегося плазменного облака с радиацией.

Дисперсионные уравнения в изэнтропическом плазменном облаке. Условие существования нетривиальных решений системы (73)-(76) приводит к следующему дисперсионному уравнению 6-го порядка относительно комплексной величины $\omega(\mathbf{k})$:

$$\omega^6 - \omega^4\left[4|\mathbf{\Omega}|^2 + v_{\mathcal{A}}^2\left(|\mathbf{k}|^2 + \mathrm{k}_z^2\right) + \left(v_{S,0}^2|\mathbf{k}|^2 - 4\pi G\rho_0\right)\right] + \omega^2\left\{\frac{4}{|\mathbf{k}|^2}\left(v_{S,0}^2|\mathbf{k}|^2 - 4\pi G\rho_0\right)\times\right.$$

$$\left.\times\left(\Omega_x\,\mathrm{k}_x^2 + \Omega_z\,\mathrm{k}_z^2\right)^2 + v_{\mathcal{A}}^2\,\mathrm{k}_z^2\left[v_{\mathcal{A}}^2|\mathbf{k}|^2 + 2\left(v_{S,0}^2|\mathbf{k}|^2 - 4\pi G\rho_0\right)\right]\right\} -$$

$$-v_{\mathcal{A}}^4\,\mathrm{k}_z^4\left(v_{S,0}^2|\mathbf{k}|^2 - 4\pi G\rho_0\right) = 0\,, \qquad (77)$$

где $v_{\mathcal{A}} = \mathcal{B}_0 / \sqrt{4\pi\rho_0}$ – альфвеновская (магнитогидродинамическая) скорость волн, обусловленных квазиупругим натяжением магнитных силовых линий.

Методом Кардана возможно получение точного решения этого алгебраического уравнения (кубического относительно величины $n = \omega^2$). Однако это решение, к сожалению, не приводит к наглядным формулам для различных показателей роста. Вместе с тем качественный анализ системы уравнений (73)-(76) возможен на основе рациональной аппроксимации отдельных их членов.

Исключая с этой целью из системы уравнений (73)-(76) возмущенные параметры ρ', ψ' и $\mathcal{B}'$, в результате получим следующее алгебраическое соотношение

$$\omega^2\mathbf{u}' - i2\omega\mathbf{u}'\times\mathbf{\Omega} - v_{S,0}^2\mathbf{k}(\mathbf{k}\cdot\mathbf{u}') + 4\pi G\rho_0\mathbf{k}\frac{\mathbf{k}\cdot\mathbf{u}'}{|\mathbf{k}|^2} - V_{\mathcal{A}}^2|\mathbf{k}|^2(\mathbf{u}' - \mathbf{i}_z u_z') = 0\,. \qquad (78)$$

При использовании векторного тождества $\mathbf{u}' = \dfrac{\mathbf{k}(\mathbf{k}\cdot\mathbf{u}')}{|\mathbf{k}|^2} + \dfrac{1}{|\mathbf{k}|^2}\mathbf{k}\times(\mathbf{u}'\times\mathbf{k})$ [58], принимающего для продольных звуковых волн в жидкости следующий вид $\mathbf{u}' = \mathbf{k}(\mathbf{k}\cdot\mathbf{u}')/|\mathbf{k}|^{-2}$ (см. сноску «VI»), соотношение (77) можно переписать следующим образом:

$$\omega^2\mathbf{u}' - i\,\omega 2\rho_0^{-1}(\mathbf{u}'\times\mathbf{\Omega}) - |\mathbf{k}|^2 v_{S,q}^2\mathbf{u}' + 4\pi G\rho_0\mathbf{u}' - v_{\mathcal{A}}^2|\mathbf{k}|^2(\mathbf{u}' - \mathbf{i}_z u_z') = 0\,. \qquad (78^*)$$

Проанализируем теперь это уравнение.

1. Рассмотрим сначала случай самогравитирующего незаряженного газового облака. Тогда

$$\omega^2\mathbf{u}' - |\mathbf{k}|^2 v_{S,q}^2\mathbf{u}' + 4\pi G\rho_0\mathbf{u}' = i\,\omega 2(\mathbf{u}'\times\mathbf{\Omega}). \tag{79}$$

При скалярном умножении этого соотношения на возмущенную скорость $\mathbf{u}'$ получим дисперсионное соотношение для звуковых волн во вращающемся облаке

$$\omega^2 - |\mathbf{k}|^2 v_{S,q}^2 + 4\pi G\rho_0 = 0\ , \tag{80}$$

из которого следует, что кориолисова сила не преодолевает стабилизирующего эффекта излучения для вращающегося облака, поскольку в этом случае справедлив рассмотренный выше критерий неустойчивости Джинса (42*) для самогравитирующего газового облака с излучением.

Если волна возмущения распространяется в плоскости xz перпендикулярно направлению оси вращения облака $\mathbf{\Omega} = \mathbf{i}_z\Omega$, то из (79) следует алгебраическое соотношение:

$$|\mathbf{u}'|^2(\omega^2 - |\mathbf{k}|^2 v_{S,q}^2 + 4\pi G\rho_0)^2 = -4\omega^2(\mathbf{u}'\times\mathbf{\Omega})\cdot(\mathbf{u}'\times\mathbf{\Omega}) = 4\omega^2|\mathbf{u}'|^2|\mathbf{\Omega}|^2, \tag{81}$$

записанное здесь с использованием условия $\mathbf{u}'\cdot\mathbf{\Omega} = 0$ и формулы векторной алгебры $(\boldsymbol{a}\times\boldsymbol{b})\cdot(\boldsymbol{a}\times\boldsymbol{b}) = (\boldsymbol{a}\cdot\boldsymbol{b})^2 - \boldsymbol{a}^2\boldsymbol{b}^2$ [58]. Из (81) вытекает следующее дисперсионное уравнение

$$\omega^4 + 2\omega^2(4\pi G\rho_0 - |\mathbf{k}|^2 v_{S,q}^2 - 2\Omega^2) + (4\pi G\rho_0 - |\mathbf{k}|^2 v_{S,q}^2)^2 = 0, \tag{82}$$

Пусть ω_1^2 и ω_2^2 − корни уравнения (82); тогда

$$\omega_1^2 + \omega_2^2 = -2(-|\mathbf{k}|^2 v_{S,q}^2 + 4\pi G\rho_0 - 2\Omega^2), \qquad \omega_1^2\omega_2^2 = (|\mathbf{k}|^2 v_{S,q}^2 - 4\pi G\rho_0)^2. \tag{83}$$

Отсюда следует, что условие неустойчивости облака $\omega_{1,2}^2 < 0$ для совокупности волн возмущения имеет вид

$$v_{S,q}^2|\mathbf{k}|^2 < 4\pi G\rho_0 - 2\Omega^2. \tag{84}$$

В этом случае критическая длина волны возмущения $\lambda_{cr} = 2\pi/\mathrm{k}_{cr}$ и критическое волновое число $\mathrm{k}_{cr} = |\mathbf{k}|_{cr}$, разделяющее устойчивые $(\mathrm{k}_r > \mathrm{k}_{cr})$ и неустойчивые $(\mathrm{k}_r < \mathrm{k}_{cr})$ возмущенные волны, определяются соотношениями

$$|\mathbf{k}|_{cr} = \frac{1}{v_{S,q}}\left(4\pi G\rho_0 - 2\Omega^2\right)^{1/2} = 2\left(\frac{\pi G\rho_0}{v_{S,q}^2}\right)^{1/2}\left(1 - \frac{\Omega^2}{2\pi G\rho_0}\right)^{1/2}, \tag{85}$$

$$\lambda_{cr} = \frac{2\pi}{|\mathbf{k}|_{cr}} = \sqrt{\frac{\pi v_{S,q}^2}{G\rho_0}}\left(1 - \frac{\Omega^2}{2\pi G\rho_0}\right)^{-1/2}. \tag{86}$$

Следует иметь в виду, что критерий (84) имеет смысл только в случае, если выполняется условие $\Omega^2 / 2\pi G\rho_0 < 1$ (условие устойчивости вращающегося облака по Тумре [57]).

Таким образом, для критерия джинсовской неустойчивости вращающегося газового облака с учетом излучения для волн возмущения распространяющихся в направлении перпендикулярном направлению оси вращения облака, получим следующее представление:

$$\lambda_r > \lambda_{cr} = v_{S,q}\sqrt{\frac{\pi}{G\rho_0}}\left(1-\frac{\Omega^2}{2\pi G\rho_0}\right)^{-1/2} = v_{S,q}\sqrt{\frac{\pi}{G\rho_0}}\left(1-\frac{\Omega^2}{2\pi G\rho_0}\right)^{-1/2}, \tag{87}$$

которое, с учетом формулы (46) для длины Джинса, может быть записано в виде

$$\frac{\lambda_r}{\lambda_J} > \frac{v_{S,q}}{v_{gas}}\left(1-\frac{\Omega^2}{2\pi G\rho_0}\right)^{-1/2} =$$

$$= \left\{\frac{1}{\gamma_1}\frac{1}{(\gamma_q-1)D/2}\left[1+\frac{(4-3\beta_0)^2(\gamma_q-1)}{\beta_0^2+12\beta_0(\gamma_q-1)(1-\beta_0)}\right]\right\}^{1/2}\left(1-\frac{\Omega^2}{2\pi G\rho_0}\right)^{-1/2}. \tag{88}$$

2. Пусть теперь вращение плазменного облака отсутствует. Тогда из (77) следует

$$\omega^2\mathbf{u}' - v_{S,0}^2|\mathbf{k}|^2\frac{\mathbf{k}(\mathbf{k}\cdot\mathbf{u}')}{|\mathbf{k}|^2} + 4\pi G\rho_0\frac{\mathbf{k}(\mathbf{k}\cdot\mathbf{u}')}{|\mathbf{k}|^2} - v_{\mathcal{A}}^2|\mathbf{k}|^2(\mathbf{u}'-\mathbf{i}_z u_z') = 0, \tag{89}$$

Рассмотрим два простых случая:

a). Для поперечного распространения волн возмущения (когда $\mathrm{k}_x = \mathrm{k}$, $u_z' = 0$) уравнение (89) сводится к простому дисперсионному соотношению (сравни с (42*))

$$\omega^2 - v_{\mathcal{A}}^2\mathrm{k}_x^2 - v_{S,0}^2\mathrm{k}_x^2 + 4\pi G\rho_0 = 0, \tag{90}$$

для которого, с учётом (42*), критерий гравитационной неустойчивости самогравитирующей плазмы с магнитным полем и радиационным давлением принимает вид:

$$\mathrm{k}_x^2\left\{v_{\mathcal{A}}^2 + \frac{1}{(\gamma_q-1)D/2}\frac{k_{\mathcal{B}}T_0}{m}\left[1+\frac{(4-3\beta_0)^2(\gamma_q-1)}{\beta_0^2+12\beta_0(\gamma_q-1)(1-\beta_0)}\right]\right\} > 4\pi G\rho_0. \tag{91}$$

b). В случае продольного (к направлению магнитного поля) распространения пульсационных волн (для которых $\mathrm{k}_z = \mathrm{k}$, $\mathrm{k}_x = 0$) уравнение (89) записывается следующим образом:

$$\omega^2\mathbf{u}' - v_{S,0}^2(\mathrm{k}_z^2 u_z')\mathbf{i}_z + 4\pi G\rho_0 u_z'\mathbf{i}_z + v_{\mathcal{A}}^2\mathrm{k}_z^2\left(\mathbf{u}'-\mathbf{i}_z u_z'\right) = 0. \tag{92}$$

Отсюда для волны возмущения, направленной вдоль направления вектора магнитного поля $(\mathbf{u}' = \mathbf{i}_z u_z')$, получим дисперсионное соотношение

$$\omega^2 - v_{S,0}^2\mathrm{k}_z^2 + 4\pi G\rho_0 = 0. \tag{93}$$

Если $u_z' = 0$, то из (92) следует

$$\omega^2 - v_A^2 \, k_z^2 = 0 . \tag{93}$$

Таким образом, в поперечном режиме распространения волны возмущения критерий неустойчивости Джинса для плазмы модифицируется магнитным полем и радиационным давлением. В случае продольного режима магнитное поле не влияет на джинсовский критерий, поскольку этот режим обеспечивает Альфвен-режим движения отдельно от гравитационного режима.

ЗАКЛЮЧЕНИЕ

Имея в виду большое космогоническое значение проблемы гравитационной неустойчивости, в представленной работе в рамках неэкстенсивной кинетики исследовано влияние неэкстенсивности среды на критерий гравитационной неустойчивости Джинса для самогравитирующего допланетного облака, вещество которого состоит из смеси идеального q -газа и чёрнотельного излучения. Выведены дисперсионные уравнения, на основе которых выполнен анализ осесимметричных колебаний космических самогравитирующих объектов с излучением и размерностью пространства скоростей. Для неэкстенсивных сред получены модифицированные критерии гравитационной неустойчивости Джинса как для бесконечной покоящейся сферически однородной среды, состоящей из идеального q-газа и излучения, так и для бездиссипативной намагниченной плазмы с учётом вращения и радиационного давления.

Рассмотренный здесь подход к описанию в рамках неэкстенсивной кинетики эволюции относительно простых (модельных) астрофизических объектов может быть распространён на более реалистичные физические ситуации, связанные, в частности, с учетом динамики возмущений в неоднородных и неизотропных дисковых фрактальных средах, с исследованием гравитационных возмущений диссипативных дисков, с исследованием собственных частот колебаний вертикально неоднородных магнитных дисков и т.п. (см.[24]). Это позволяет более обоснованно моделировать реальные астрофизических газо-пылевые структуры и находить соответствующие критерии их гравитационной неустойчивости.

Поскольку физический смысл и численные значения индекса энтропийной деформации q играют существенную роль в понимании эволюции многих аномальных астрофизических объектов, то проблема их определения представляется чрезвычайно важной. К сожалению, эта проблема всё ещё остаётся открытой. Вместе с тем, в настоящее время имеются серьёзные успехи в современной гелиосейсмологии, которая надёжно исследует внутреннюю структуру и динамику Солнца [59]. В солнечной атмосфере установлены и изучены миллионы резонансных мод колебаний. Их частоты измерены с достаточно большой точностью, что позволяет исследовать внутреннюю структуру Солнца на больших глубинах [60]. Эти результаты позволяют решить не только некоторые известные проблемы космологии, но и поднимают ряд теоретических вопросов, ответы на которые необходимы для понимания того, как на самом деле эволюционирует обычная звезда. В частности, гелиосейсмология позволяет, вообще говоря, найти экспериментальные доказательства присутствия неэкстенсивных эффектов в недрах звезды по определяемым скоростям звука. Следовательно, есть уверенность, что в самое ближайшее время можно будет получить астрономические данные по численным значениям параметра q, отличным от единицы.

REFERENCES

[1] C. Tsallis, ”Possible Generalization of Boltzmann-Gibbs-Statistics”, *J. Stat. Phys*., **52**(1-2), 479-487 (1988).

[2] C. Tsallis, “Nonextensive Statistic: Theoretical, Experimental and Computational Evi-dences and Connections”, *Brazilian J. Phys.*, **29**(1), 1-35 (1999).

[3] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics. Approaching a Complex World.* New York: Springer. (2009).

[4] E.M.F. Curado, C. Tsallis. “Generalized statistical mechanics: connection with thermodynamics”, *J. Physica. A.*, **24**, L69-72 (1991).

[5] C. Tsallis, R.S. Mendes, A.R. Plastino,”The role of constraints within generalized Nonextensive statistics”, *J. Physica A.*, **261**, 534-554 (1998).

[6] A.V. Kolesnichenko, “Modifikatsiya v ramkakh statistiki Tsallisa kriteriev gravitatsionnoy neustoychivosti astrofizicheskikh diskov s frartal’noy strukturoy fazovogo prosranstva”, *Mathematica Montisnigri*, **32**, 93-118 (2015).

[7] A.V. Kolesnichenko, “Kriteriy termicheskoy ustoychivosti i zakon raspredeleniy chastits dlya samogravitiruyushchikh astrofizicheskikh system v ramkakh statistiki Tsallisa”, *Mathematica Montisnigri*, **37**, 45-75 (2016).

[8] A.V. Kolesnichenko, “Power distributions for self-gravitating astrophysical systems based on nonextensive Tsallis kinetics”, *Solar System Research*, **51**(2), 127-144 (2017).

[9] A.V. Kolesnichenko, “Dvukhparametricheskiy entropiynyy funktsional Sharma-Mittala kak osnova semeystva obobshchyenykh termodinamik neekstensivnykh system”, *Mathematica Montisnigri*, **42**, 74-101 (2018).

[10] A.V. Kolesnichenko, *Statisticheskaya mekhanika i termodinamika Tsallisa neadditivnykh system: Vvedenie v teoriyu i prilozheniya*. Moskow: LENAND. (Sinergetika ot proshlogo k budushchemu. № 87), (2019).

[11] A.V. Kolesnichenko, “K postroeniyu neadditivnoy termodinamiki slozhnykh sistem na osnove statistiki Kurado−Tsallisa”, *Keldysh Institute Preprints*, **25**, 1-40 (2018).

[12] A.V. Kolesnichenko, B.N. Chetverushkin, “Kinetic derivation of a quasi-hydrodinamic system of equations on the base of nonextensive statistics”, *RJNAMM (Russian Journal of Numerical Analysis and Mathematical Modelling*), **28**(6), 547-576 (2013).

[13] A.V. Kolesnichenko, M.Ya. Marov, “Modeling of aggregation of fractal dust clusters in a laminar protoplanetary disk”, *Solar System Research*, **47**(2), 80-98 (2013).

[14] A.V. Kolesnichenko, M.Ya. Marov, **“**Modification of the jeans instability criterion for fractal-structure astrophysical objects in the framework of nonextensive statistics”, *Solar System Research*, **48** (5), 354-365 (2014).

[15] A.V. Kolesnichenko, M.Ya. Marov, “Modification of the Jeans and Toomre instability criteria for astrophysical fractal objects within nonextensive statistics”, *Solar System Research*, **50**(4), 251-261 (2016).

[16] A.V. Kolesnichenko, M.Ya. Marov, “Renyi Thermodynamics as a Mandatory Basis to Model the Evolution of a Protoplanetary Gas−Dust Disk with a Fractal Structure”, *Solar System Research*, **53**(6), 443-461 (2019).

[17] Nonextensive statistical mechanics and thermodynamics: *Full bibliography/ http://tsallis.cat.cbpf.br/biblio.htm.* (accessed 12 February 2020).

[18] J.H. Jeans, “The stability of a spherical nebula”, *Philosophical Transactions of the Royal Society of London. Series A*. Containing Papers of a Mathematical or Physical Character. **199**, 1-53 (1902).

[19] J.H. Jeans, *Astronomy and Cosmogony*, Cambridge Univ. Press. (2009).

[20] S. Chandrasekkhar, *Vvedeniye v ucheniye o stroyenii zvezd*. M.: Izdatelystvo Inost. Liter. (1950).

[21] S. Chandrasekkhar, "O zvezdakh, ikh evolyutsii i ustoychivosti ", *UFN*. **145**(3), 489-506 (1985).
[22] V.S. Safronov, *Evolyutsiya doplanetnogo oblaka i obrazovaniye Zemli i planet,* M.: Nauka. (1969).
[23] N.N. Gor'kavyy, A.M. Fridman, *Fizika planetnykh kolets*. M.: Nauka. (1994).
[24] A.M. Fridman, A.V. Khoperskov, *Fizika galakticheskikh diskov*. M.: Fizmatlit. (2011).
[25] S. Chandrasekhar, E. Fermi, "Problems of gravitational stability in the Presence of a magnetic field", *Astrophysical Journal*, **118**, 116-141 (1953).
[26] W. B. Bonnor, "Jeans' Formula for Gravitational Instability", *Monthly Notices of the Royal Astronomical Society*, **117**(1), 104-117 (1957).
[27] C. Hunter, "Self-gravitating gaseous disks", *Ann. Rev. Fluid Mech.*, **4**. 219-242 (1972).
[28] P. Goldreich, D. I. Lynden-Bell, "Gravatational stability of uniformly rotating disks", *MNRAS*, **130**, 97-124 (1965).
[29] C. Low, D. Lynden-Bell, "The minimum Jeans mass or when fragmentation must stop", *Monthly Notices of the Royal Astronomical Society,* **176**(2), 367-390 (1976).
[30] N.I. Shakura, R.A. Sunyaev, "A theory of the instability of disk accretion onto black holes and the variability of binary X-ray sources, galactic nuclei and quasars", *Mon.Not.RAS, astr.Soc.*, **175**, 613-632 (1976).
[31] M. Camenzind, F. Demole, N. Straumann, "The stability of radiation–pressure–dominated accretion discs", *Astron. Astrophys.*, **158**, 212-216 (1986).
[32] V.M. Cadez, "Applicability problem of Jeans criterion to a stationary self-gravitating cloud", Astron. Astrophys., **235**, 242-244 (1990).
[33] B.P. Pandey, K. Avinash, "Jeans instability of a dusty plasma", *Physical Review E* (*Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*), **49**(6), 5599-5606 (1994).
[34] J. M. Owen, J. Villumsen, V. Baryons, "Dark Matter, and the Jeans Mass in Simulations of Cosmological Structure Formation", *J. Astroph.*, **481**(1), 1-21 (1997).
[35] D. Tsiklauri, "Jeans Instability of Interstellar Gas Clouds in the Background of Weakly Interacting Massive Particles", *J. Astroph.*, **507**(1), 226-228 (1998).
[36] R. L. Mace, V. Frank, M. A. Hellberg, "Jeans stability of dusty space plasmas", *Physics Letters A.*, **237**, 146-151 (1998).
[37] J.A. S. Lima, R. Silva, J. Santos, "Jeans' gravitational instability and nonextensive kinetic theory", *Astronomy and Astrophysics*. **396**, 309-313 (2002).
[38] S. A. Trigger, A. I. Ershkovich, G. J. F. van Heijst, P. P. J. M. Schram, "Kinetic theory of Jeans instability", *Phys. Rev. E.*, **69**, 066403 -066405 (2004).
[39] M. Sakagami, A.Taruya, "Self-gravitating stellar systems and non-extensive thermostatistics", *Continuum Mechanics and Thermodynamics,* **16** (3), 279-292 (2004).
[40] P. K. Shukla, L. Stenflo, "Jeans instability in a self-gravitating dusty plasma", *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **462**, 403-407 (2006).
[41] N. L. Tsintsadze , R. Chaudhary, H. A. Shah, G.Murtaza, "Jeans instability in a magneto-radiative dusty plasma", *Journal of Plasma Physics*., **74**(6), 847-853 (2008).
[42] A.E. Radwan, "Variable streams self-gravitating instability of radiating rotating gas cloud", *Applied mathematics and computation*,. **148**, 331-339 (2004).
[43] V.M. Cadez, "Instabilities in stratified magnetized Stellar atmospheres", *Publ. Astron. Obs. Belgrade*., **90**, 121-124 (2010).
[44] J.S. Dhiman, R. Dadwal, "On the Jeans Criterion of a Stratified Heat Conducting Gaseous Medium in the Presence of Non-uniform Rotation and Magnetic Field", *Journal of Astrophysics and Astronomy*, **33**(4), 363-373 (2012).
[45] A.M. Fridman, V.L. Polyachenko, *Physics of gravitating system. V.1-2*. - N.Y.: Springer-Verlag. 1984.

[46] A.M. Fridman, V.L. Polyachenko, *Physics of Gravitating Systems I: Equilibrium and Stability*. Springer Science & Business Media. (2012).

[47] S. Kaothekar, R.K. Chhajlani, "Jeans Instability Of Self Gravitating Partially Ionized Hall Plasma With Radiative Heat Loss Functions And Porosity", *AIP Conference Proceedings,* **1536**(1), 1288-1289 (2013)**.**

[48] H. Joshi, R. K. Pensia, "Effect of rotation on Jeans instability of magnetized radiative quantum plasma", *Physics of plasmas,*. **24,** 032113-1−032113-8 (2017).

[49] R. K. Pensia, D. L. Sutar, S. Sharma, "Analysis of Jeans Instability of Optically Thick Quantum Plasma under the Effect of Modified Ohms law", *2nd International Conference on Condensed Matter and Applied Physics* (*ICC 2017*).*AIP Conf. Proc.,* 1**953**(1), 060044-1–060044-4 (2018).

[50] V. Kumar, D. L. Sutar, Pensia, S. Sharma, "Effect of fine dust particles and finite electron inertia of rotating magnetized plasma", *2nd International Conference on Condensed Matter and Applied Physics (ICC 2017).AIP Conf. Proc.,* **1953**(1), 060036-1–060036-4 (2018).

[51] B. M. Boghosian, "Navier-Storts Equations for Generalized Thermostatistics", *Bras. J. Phys*., **29**(1), 91-107 (1999).

[52] D.S. Oliveira, R. M. O. Galvao, "Transport equations in magnetized plasmas for non-Maxwellian distribution functions", *Physics of plasmas,* **25,** 102308-1─102308-13 (2018).

[53] M.R. McKee, "The radial-azimuthal stability of accretion disks around black holes", *Astron. Astrophys*, **235**, 521-525 (1990).

[54] A.V. Khopepskov, S.S. KHpapov, "Heustoychivost' zvukovykh voln v tonkom gazovom diske", *Pis'ma v AZH*., **21**, 388-393 (1995).

[55] A. S. Eddington, *The Internal Constitution of the Stars*. Cambridge. England: Cambridge University Press. (1988).

[56] L.D. Landau, Ye.M. Lifshits, *Statisticheskaya fizika*. CH. I. M.: Nauka. (1976).

[57] A. Toomre, "On the gravitational stability of a disk of stars", *J. Astroph*., **139**, 1217-1238 (1964).

[58] N.Ye. Kochin, *Vektornoye ischisleniye i nachala tenzornogo ischisleniya.* M.: Izd-vo Akad. Nauk SSSR. (1961).

[59] D.O. Gough, "Heliophysics Gleaned from Seismology", *Progress in solar/stellar Physics with Helio- and Asteroseismology, Proc. 65th Fujihara Seminar, Astron. Soc. Pacific Conf. Ser.,* **462,** 429-454 (2011).

[60] D. O. Gough, B. Hindman, "Helioseismic Detection of Deep Meridional Flow", *J. Astroph*., **714**(1), 960-970 (2010).

Received December 1, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

V. V. BELETSKY is an OUTSTANDING MECHANICAL SCIENTIST, ONE of the CLASSICS of SPACE FLIGHT MECHANICS (For the 90th birthday)

G.K. BOROVIN*, Yu.F. GOLUBEV, A.V. GRUSHEVSKII, A.G. TUCHIN

Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, Russia

*Corresponding author: E-mail: borovin@keldysh.ru

DOI: 10.20948/mathmontis-2020-47-15

Summary. The article briefly presents the life and scientific activities of an outstanding mechanical scientist, a major scientist, corresponding member of the RAS, honored professor of Lomonosov Moscow State University Vladimir Vasilyevich Beletsky. He is rightly considered one of the founding fathers of the Soviet and Russian school of space flight dynamics in the theory of rotational movements of artificial and natural celestial bodies.

1 INTRODUCTION

In 2017, Vladimir Beletsky (Fig.1), an outstanding scientist and unsurpassed teacher, the corresponding member of the RAS, honored Professor of Moscow state University, passed away. He is rightly considered one of the founding fathers of the Soviet and Russian school of space flight dynamics in the field of the theory of rotational movements of artificial and natural celestial bodies.

Vladimir Vasilyevich was born on may 2, 1930 in Irkutsk. He spent his childhood in this city and in the villages on the banks of the Angara and lake Baikal. At the age of 12, he lost his hearing after a severe form of meningitis. In the postwar years, the family moved to Smolensk, where Bielecki graduated from the 7th high school with a gold medal. In 1949, he entered the mechanics and mathematics faculty of Moscow state University. In 1954 he graduated with honors from the faculty of mechanics and mathematics of Lomonosov Moscow State University. In the same year, V.V. Beletsky was assigned to work in the Department of Applied Mathematics of the Steklov Mathematical Institute of the USSR AS USSR, created by M.V. Keldysh (now the Keldysh Institute of Applied Mathematics of RAS). This determined his scientific destiny as one of the galaxy of brilliant scientists - representatives of the world-recognized school of space flight dynamics, founded by M.V. Keldysh and D.E. Okhotsimsky. In his book "Theoretical mechanics and modern technology" A.A. Kosmodemyansky wrote: "I Think that for some well-known nowadays scientists interest in certain problems of modern mechanics arose as a result of work in scientific circles and seminars of the mechanics and mathematics faculty of Moscow State University. I can name, for example, the following comrades: corresponding members of the USSR Academy of Sciences D.E. Okhotsimsky and T.M. Eneev, doctors of physical-mathematics sciences V.A. Egorov, V.V. Beletsky, V.A. Sarychev...".

Here is how Vladimir Vasilyevich himself wrote about that time: "I felt that I was in the midst of the brewing events in space exploration and that these events are not least maturing

2010 Mathematics Subject Classification: 37Q05, 70Q05, 70M20, 70F15.
Key words and Phrases: spacecraft, space vehicle, orbiter, artificial earth satellite, artificial moon satellite, artificial martian satellite, gravity assist maneuver, quasi-synchronous orbit.

thanks to the authoritative, businesslike and purposeful activity of M.V. Keldysh. He defined the style of research direction of the dynamics of space flight and in General the scientific program of space research. This was done by M.V. Keldysh at all levels, up to the state. After all, he was Chairman of the Interdepartmental Commission on space research, and then President of the USSR Academy of Sciences. Later E.L. Akim, A.K. Platonov came to our team, and in 1957 M.L. Lidov was appeared". Characteristic of all his work was the practical direction of research. Most of the results of spacecraft orbit design were implemented in specific missions. Perhaps for this reason, the direction in which the employees of the Keldysh - Okhotsmskiy school acted was called as the "applied celestial mechanics".

Fig. 1. V.V. Beletsky (02.05.1930-20.07.2017)

Academician D.E. Okhotsimsky wrote about that time: "When in 1953 the Department of applied mathematics of Steklov Moscow Institute Academy Sciences USSR was organized, Mstislav Vsevolodovich offered me to go to the Institute together with the team as the head of the Department. Research on our subject has always been carried out here; first, it was aimed at the development of missile technology, and then, when the air blew the possibility of space launches, we joined in these cases from the very beginning. In 1954, when it became clear that the time of the space age was approaching, she was already knocking at the door, Mstislav Vsevolodovich convened a meeting of scientists and leaders of missile technology. Apparently, as a result of the discussion at this meeting with academician P.L. Kapitsa, Dmitry Evgenievich had the idea of passive gravitational stabilization of artificial earth satellites, i.e. orientation of satellites at the expense of natural forces without any fuel costs

for orientation control. The fifties years – the years of preparation and implementation of the breakthrough into space – were years of unprecedented takeoff initiated and led by M.V. Keldysh research in the Department of D.E. Okhotsimsky.

2 SCIENTIFIC ACHIEVEMENTS

The first researches of V. V. Beletsky became known and recognized by specialists. Speaking on September 14, 1956 at a meeting of the Presidium of the USSR Academy of Sciences, M. V. Keldysh in his report, talking about the stability of the relative equilibrium of the satellite in orbit, said: "... This interesting problem of solid mechanics was solved by a very young employee V. V. Beletsky in the Department of applied mathematics". The results of these studies are summarized in the monograph of 1965, which, being translated into English, and is now a Handbook of specialists.

In the introduction to his monograph V.V. Beletsky indicates that the theory of motion of celestial bodies near the mass center in classical mechanics has developed with respect to specific bodies (Earth, Moon) and therefore used a number of simplifications. In this case, the influence of gravitational moments was mainly considered. The problem of the rotational motion of artificial space objects is much more complex, because "due to the arbitrariness of the shape and mass distribution of the object, the arbitrariness of the initial data, the many factors affecting the movement. In addition to the gravitational moments should take into account more aerodynamic and electromagnetic moments.

V.V. Beletsky practically opened a new branch of the celestial mechanics, simplifying the classical formulation of problems and received as a result of such simplification the main "scrolls" for the mechanics of rotational motion of satellites. On successful examples, he showed how they should be used, came up with the number of interesting problems outside these schemes, solved for the first time standard problems in the new situation. In this sense, Vladimir Beletsky has being brilliantly solved a number of such problems, and become the father-founder for a whole area of such problems solving, giving the scientific community the tools for such solving.

The main results of this cycle of work are as follows. A theorem on the stability conditions of the relative equilibrium of a satellite in a gravitational field is proved. The theory of oscillations of a satellite in an elliptical orbit under the action of the gravitational moment was developed. The problems has been formulated and the theory of the evolution of satellite rotation under the influence of disturbing moments caused by the gravity gradient, the influence of the Earth’s magnetic field, atmosphere and light pressure forces has been developed. This theory has been applied to describe the motion of a number of the particular artificial satellites.

V.V. Beletsky was the first who posed and considered the problem of the dynamics of the orbital tethered bodies as a system with release from coupling.

In the same years, V.V. Beletsky for the first time in the world has been setting the general task of determining the actual orientation of the satellite and clarifying the parameters of the perturbing acting moments based on the results of processing the measurements of the orientation sensors installed on Board. He developed and applied an effective method of solving this problem (the third Soviet satellite, the “Proton” satellite, the “Electron” satellites). This approach has been successfully used today.

Since the early 1970s, V.V. Beletsky has conducted a series of studies of nonlinear problems of the dynamics of rotational motion of artificial satellites and planets, taking into account the existing resonances in their orbital and rotational motion, as well as the effect of energy dissipation (tidal effect) on the formation of a modern picture of the rotation of planets, taking into account the probability of capture in existing resonances [11]. The resonance theory of "generalized Cassini's laws" of planetary rotation was developed by him, which gives a rigorous justification of the empirical Cassini's laws for the moon's rotation (1693).

The main results of Beletsky's scientific activity also include deservedly the theory of tidal evolution of the rotational motion of celestial bodies; the solution of the optimal problems of spacecraft' flights with low-thrust engines; the formulation and analysis of problems of the orbital tether systems dynamics; the construction of models and the study of the of bipedal devices dynamics.

3 RESULTS

Here is how briefly V.V. Beletsky presented his main scientific achievements, with links to his own anthology.

1. A theorem on the stability conditions of the relative equilibrium of a satellite in a gravitational field is proved [1], [2]. This result is used in the theory and practice of the passive gravitational stabilization of artificial satellites.
2. The theory of oscillations of a satellite at the elliptical orbit in the gravitational field is developed [1], [2], [3], [4].
3. The problem is posed and the theory of evolution of satellite rotation under the influence of disturbing moments of forces (gravitational, magnetic, aerodynamic, light pressure) [5],[2], the moment of tidal forces [11] is developed.
4. The problem of determining the actual orientation of satellites and the acting moments by on-Board measurements is posed and solved [6], [2], [7].
5. The theory of orbital tether system motion and motion randomization is formulated and developed [8], [9], [10].
6. A resonant theory of generalized "Cassini's laws of rotation" of the natural and artificial celestial bodies has been created [12], [13]. This theory, in particular, justifies the empirical laws of J.D. Cassini, established more than 300 years ago (1693).

Vladimir V. Beletsky, as a chief scientific officer of the KIAM of RAS, was the Member of his Scientific Council, also has been a Member of the specialized dissertation councils of KIAM RAS and mechanics and mathematics faculty of Lomonosov Moscow State University, a Member of the Russian National Committee on Theoretical and Applied Mechanics (1976). In 1997 he was awarded the title corresponding member of RAS. He was a Full member of the International Academy of Astronautics (1992) and a Full member of the Russian Academy of cosmonautics (1994), was a Member of the editorial Board of the journal "Regular and chaotic dynamics". V.V. Beletsky prepared 26 candidates and 5 Doctors of Sciences.

Beletsky's scientific achievements are highly appreciated in Russia and abroad. He was awarded Honored Professor of Lomonosov MSU (2002), and laureate of A. von Humboldt

prize (Germany). He honored the F.A. Tsander prize of RAS. The minor planet No. 14790 (discovered July 30, 1970) has been named after V.V. Beletsky “Beletskij” (Fig.2).

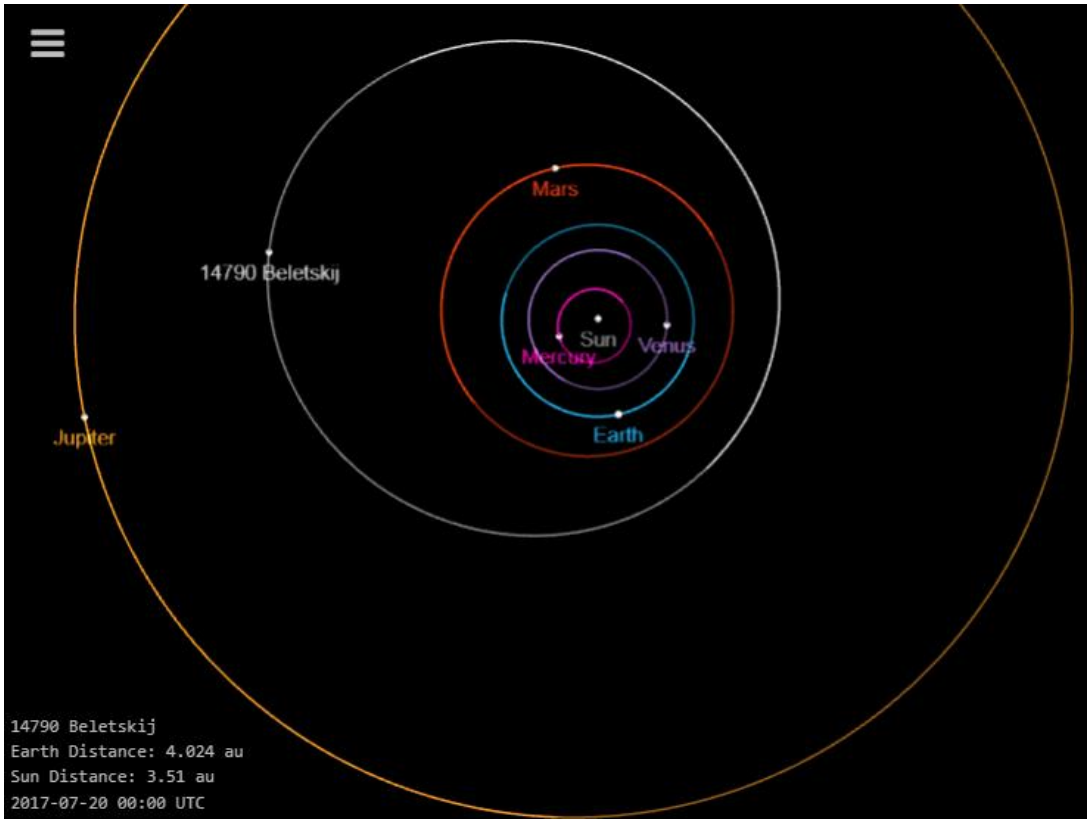


Fig. 2. Small planet № 14790 (opened July 30, 1970), named after V.V. Beletsky

Summarizing, it can be argued that V.V. Beletsky practically opened a new the scientific school and the branch of the celestial mechanics, simplifying the classical formulation of problems and developed in such simplification main "science scrolls" to describe the mechanics of the rotational motion of a satellites and planets.

REFERENCES

[1] V.V. Beletsky, “O libratcii sputneyka”, *Iskusstvennye sputneyki Zemli. M.: Izd-vo AN SSSR,* **3**, 13-31, (1959).

[2] V.V. Beletsky, *Dvizhenie iskusstvennogo sputneyka otnositelno centra mass,* M.: Nauka, (1965).

[3] V.A. Zlatoustov, D.E. Okhotsimsky, V.F. Sarychev, A.P. Torzhevsky, “Periodicheskie resheniya v zadachah o ploskih kolebaniyah sputneyka na ellipticheskoy orbite”, *Kosmicheskie issledovaniya,***2**(5), 658-666 (1964).

[4] A.D. Bryuno, “Semeistvo periodicheskih reshenii uravneniya Beletskogo”, *Kosmicheskie issledovaniya,***40**(3), 295-316 (2002).

[5] V.V. Beletsky, “Dvizhenie iskusstvennogo sputneyka otnositelno centra mass”, *Kosmicheskie issledovaniya,***1**, 25-43 (1957).

[6] V.V. Beletsky, Yu.V. Zonov, “Vrashenie i orientaciya tretego sovetskogo sputneyka”, *Iskusstvennye sputneyki Zemli. M.: Izd-vo A*N SSSR, **7**, 32-55, (1961).

[7] V.V. Beletsky, V.V. Golubkov, E.A. Stepanova, I.G. Hatckevich, “Rezultaty opredeleniya orientacii sputnika “Proton-2” i opisanie ego dvizheniya otnositelno centra mass”, *Kosmicheskie issledovaniya*,**7**(4**)**, 522-533 (1969).
[8] V.V. Beletsky, E.T. Novikova, “Ob otnositelnom dvizhenii cvyazki dvuh tel na orbite”, *Kosmicheskie issledovaniya*,**7**(3), 377-384 (1969).
[9] V.V. Beletsky, E.M. Levin, *Dinamika kosmicheskih trosovyh system*, M.: Nauka, (1990).
[10] V.V. Beletsky, D.V. Pankov, “Connected bodies in the orbit as a dynamic billiard”, *Regular and chaotic dynamics*, **1**, 87-103 (1996).
[11] V.V. Beletsky, A.V. Grushevskii, “Model formirovaniya vrashatelnyh dvizhenii nebesnyh tel s ogranicheniyami na poryadok rezonansov”, *Astronomicheskii vestnik*, **24**(2), 140-147 (1990).
[12] V.V. Beletsky, “Resonance rotation of celestial bodies and Cassini's laws”, *Celest. Mech.*, **6**(3), 356-378 (1972).
[13] V.V. Beletsky, *Dvizhenie sputneyka otnositelno centra mass v gravitatcionnom pole*, M.: Izd. MGU, (1975)

Received November 15, 2019

MATHEMATICA MONTISNIGRI
Vol XLVII (2020)

БЕЛЕЦКИЙ В.В. – ВЫДАЮЩИЙСЯ УЧЁНЫЙ–МЕХАНИК, ОДИН ИЗ КЛАССИКОВ МЕХАНИКИ КОСМИЧЕСКОГО ПОЛЁТА (К 90-летию со дня рождения)

Г.К. БОРОВИН*, Ю.Ф. ГОЛУБЕВ, А.В. ГРУШЕВСКИЙ, А.Г. ТУЧИН

Институт прикладной математики им. М.В. Келдыша РАН. Москва, Россия

*Ответственный автор: E-mail: borovin@keldysh.ru

DOI: 10.20948/mathmontis-2020-47-15

Ключевые слова: космический аппарат, искусственный спутник Земли, искусственный спутник Луны, искусственный спутник Марса, планета Венера, гравитационный манёвр, гравитационное поле

Аннотация. В статье кратко представлена жизнь и научная деятельность выдающегося ученого-механика, крупного ученого, члена-корреспондента РАН, заслуженного профессора МГУ имени М.В.Ломоносова Владимира Васильевича Белецкого. Он по праву считается одним из основателей советской и российской школы динамики космических полетов в теории вращательных движений искусственных и естественных небесных тел.

1 ВВЕДЕНИЕ

Владимир Васильевич Белецкий (Рис.1.) родился 2 мая 1930 года в городе Иркутске. Детство его прошло в этом городе и в деревнях на берегах Ангары и Байкала. В 12-летнем возрасте потерял слух после перенесённой им тяжёлой формы менингита. В послевоенные годы семья переехала в Смоленск, где Белецкий закончил 7-ю среднюю школу с золотой медалью.

В 1949 году он поступил на механико-математический факультет МГУ. В 1954 г. он окончил с отличием механико-математический факультет МГУ им. М.В. Ломоносова. В том же году В.В. Белецкий был распределен на работу в Отделение прикладной математики МИАН СССР, только что созданное М.В. Келдышем (теперь Институт прикладной математики им. М.В. Келдыша РАН). Это определило его научную судьбу как одного из плеяды блестящих ученых - представителей всемирно признанной школы динамики космического полета, основанной М.В. Келдышем и Д.Е. Охоцимским. В своей книге «Теоретическая механика и современная техника» А.А. Космодемьянский писал: «Думаю, что для некоторых известных в наши дни ученых интерес к определенным проблемам современной механики зародился в результате работы в научных кружках и семинарах механико-математического факультета МГУ. Я могу назвать, например, следующих товарищей: члены-корреспонденты АН СССР Д.Е. Охоцимский и Т.М. Энеев, доктора физ.-мат. наук В.А. Егоров, В.В. Белецкий, В.А. Сарычев…».

2010 Mathematics Subject Classification: 37Q05, 70Q05, 70M20, 70F15.
Key words and Phrases: spacecraft, space vehicle, orbiter, artificial earth satellite, artificial moon satellite, artificial martian satellite, gravity assist maneuver, quasi-synchronous orbit

Рис. 1. В.В.Белецкий (02.05.1930-20.07.2017)

Вот как сам Владимир Васильевич писал о том времени: «Я почувствовал, что попал в самую гущу назревающих событий в исследовании космоса и что эти события не в последнюю очередь зреют благодаря авторитетной, деловитой и целеустремленной деятельности М.В. Келдыша. Он определял стиль направление исследований динамики космического полета и вообще научной программы исследования космического пространства. Это делалось М.В. Келдышем на всех уровнях, вплоть до государственного. Ведь он был председателем Межведомственной комиссии по исследованию космоса, а потом и президентом Академии наук СССР. Несколько позже в наш коллектив пришли Э.Л. Аким, А.К. Платонов, а в 1957 г. появился М.Л. Лидов». Характерным для всего его творчества была практическая направленность исследований. Большая часть результатов по проектированию орбит космических аппаратов была реализована в конкретных полетах. Может быть, по этой причине направление, на котором действовали сотрудники школы Келдыша - Охоцимского, получило название "прикладная небесная механика".

Академик Д.Е. Охоцимский писал о том времени: «Когда в 1953 г. организовывалось Отделение прикладной математики, Мстислав Всеволодович предложил мне перейти в ОПМ вместе с коллективом в качестве руководителя отдела. Исследования по нашей тематике здесь проводились всегда; сперва они были направлены на развитие ракетной техники, а затем, когда в воздухе повеяло возможностью космических запусков, мы с самого начала подключились к этим делам. В 1954 г., когда уже стало ясным, что

приближается время космической эры, она уже стучится в дверь, Мстислав Всеволодович созвал совещание ученых и руководителей ракетной техники. Видимо, в результате дискуссии на этом совещании с академиком П.Л. Капицей у Дмитрия Евгеньевича родилась идея пассивной гравитационной стабилизации искусственных спутников Земли, т.е. ориентации спутников за счет природных сил без всяких затрат топлива на управление ориентацией. Пятидесятые годы – годы подготовки и реализации прорыва в космос – были годами невиданного взлета инициированных и руководимых М.В. Келдышем научных исследований в отделе Д.Е. Охоцимского.

2 НАУЧНЫЕ ДОСТИЖЕНИЯ

Уже первые исследования В.В. Белецкого приобрели известность и признание специалистов. Выступая 14 сентября 1956 г. на заседании президиума Академии наук СССР, М.В. Келдыш в своем докладе, рассказывая об устойчивости относительного равновесия спутника на орбите, заметил: "... Эта интереснейшая задача механики твердого тела была решена совсем еще молодым сотрудником В.В. Белецким в Отделении прикладной математики". Результаты этих исследований подытожены в монографии 1965 г., которая, будучи переведенной на английский язык, и сейчас является настольной книгой специалистов.

Во введении к монографии В.В. Белецкий указывает, что теория движения небесных тел около центра масс в классической механике развивалась применительно к конкретным телам (Земля, Луна) и посему использует ряд упрощений. При этом рассматривалось в основном влияние гравитационных моментов. Задача о вращательном движении искусственных космических объектов гораздо более сложна, поскольку «обусловливается произвольностью формы и распределения масс объекта, произвольностью начальных данных, многочисленностью факторов, влияющих на движение. Кроме гравитационных моментов следует учитывать ещё аэродинамические и электромагнитные моменты.

В.В. Белецкий практически открыл новую отрасль механики, упростив классические постановки задач и получив в результате такого упрощения основные «скрижали» для механики вращательного движения спутников. На удачных примерах он показал, как их надо использовать, придумал ряд интересных задач вне этих схем, впервые решил стандартные задачи в новой ситуации. В этом смысле Владимир Васильевич Белецкий, с блеском решив ряд указанных задач, явился отцом-основателем целого направления по решению подобных задач, дав научному сообществу инструментарий для их решения.

Основные результаты этого цикла работ состоят в следующем. Доказана теорема об условиях устойчивости относительного равновесия спутника в гравитационном поле. Развита теория колебаний спутника на эллиптической орбите под действием момента градиента силы тяжести. Выполнена постановка проблема и разработана теория эволюции вращения спутников под влиянием возмущающих моментов, вызванных градиентом силы тяжести, влиянием магнитного поля Земли, атмосферы и сил светового давления. Эта теория нашла свое применение для описания движения целого ряда конкретных спутников.

В.В. Белецкий впервые поставил и рассмотрел проблему динамики орбитальной "связки тел" как системы с освобождающей связью.

В эти же годы В.В. Белецкий впервые в мире поставил общую задачу об определении фактической ориентации спутника и уточнения параметров действующих на него возмущающих моментов по результатам обработки измерений датчиков ориентации, установленных на борту. Он разработал и применил эффективную методику решения этой задачи (третий советский спутник, спутник "Протон", спутники "Электрон"). Этот подход успешно используется и в наши дни.

С начала 1970-х годов В.В. Белецкий провел цикл исследований нелинейных проблем динамики вращательного движения искусственных спутников и планет с учетом существующих резонансов в их орбитальном и вращательном движении, а также влияния диссипации энергии (приливной эффект) на формирование современной картины вращения планет с учетом вероятностей захвата в существующие резонансы [11]. Им создана резонансная теория "обобщенных законов Кассини" вращения планет, которая дает строгое обоснование эмпирических законов Кассини вращения Луны (1693).

В.В. Белецкий опубликовал свыше 200 научных работ, в том числе 11 монографий, переиздававшихся в стране и за рубежом. Его работы легко читать, потому что они написаны с любовью к читателю и сочетают в себе строгость анализа и прекрасный стиль изложения.

К числу замечательных научных достижений В.В. Белецкого относится его монография "Очерки о движении космических тел" (второе издание в 1977 г.), переведенная на многие языки. В этой книге, написанной живым и красочным языком, ясно и доступно излагались как классические, так и современные результаты исследований многих учёных (и самого автора) в области небесной механики.

К основным результатам научной деятельности В.В. Белецкого заслуженно также относятся: теория приливных эффектов во вращении и ориентации небесных тел; решение оптимальных задач космических перелетов с двигателями малой тяги; постановка и анализ проблем динамики орбитальных тросовых систем; построение моделей и исследование динамики двуногоходящих устройств.

3 ИТОГИ

Вот как вкратце сам В.В. Белецкий представлял свои основные научные достижения со ссылками на собственную антологию.

1. Доказана теорема об условиях устойчивости относительного равновесия спутника в гравитационном поле [1], [2]. Этот результат используется в теории и практике систем пассивной гравитационной стабилизации спутников.

2. Развита теория колебаний спутника на эллиптической орбите в гравитационном поле [1], [2], [3],[4].

3. Поставлена проблема и разработана теория эволюции вращения спутников под влиянием возмущающих моментов сил (гравитационных, магнитных, аэродинамических, светового давления) [5],[2], момента приливных сил [11].

4. Поставлена и решена проблема определения по бортовым измерениям фактической ориентации спутников и действующих на него моментов [6], [2], [7].

5. Поставлена и развита теория движения орбитальной тросовой системы и хаотизации движения [8], [9], [10].

6. Создана резонансная теория обобщенных законов Кассини «вращения» естественных и искусственных небесных тел [12], [13]. Эта теория, в частности, обосновывает эмпирические законы Дж. Д. Кассини, установленные более 300 лет тому назад (в 1693г.).

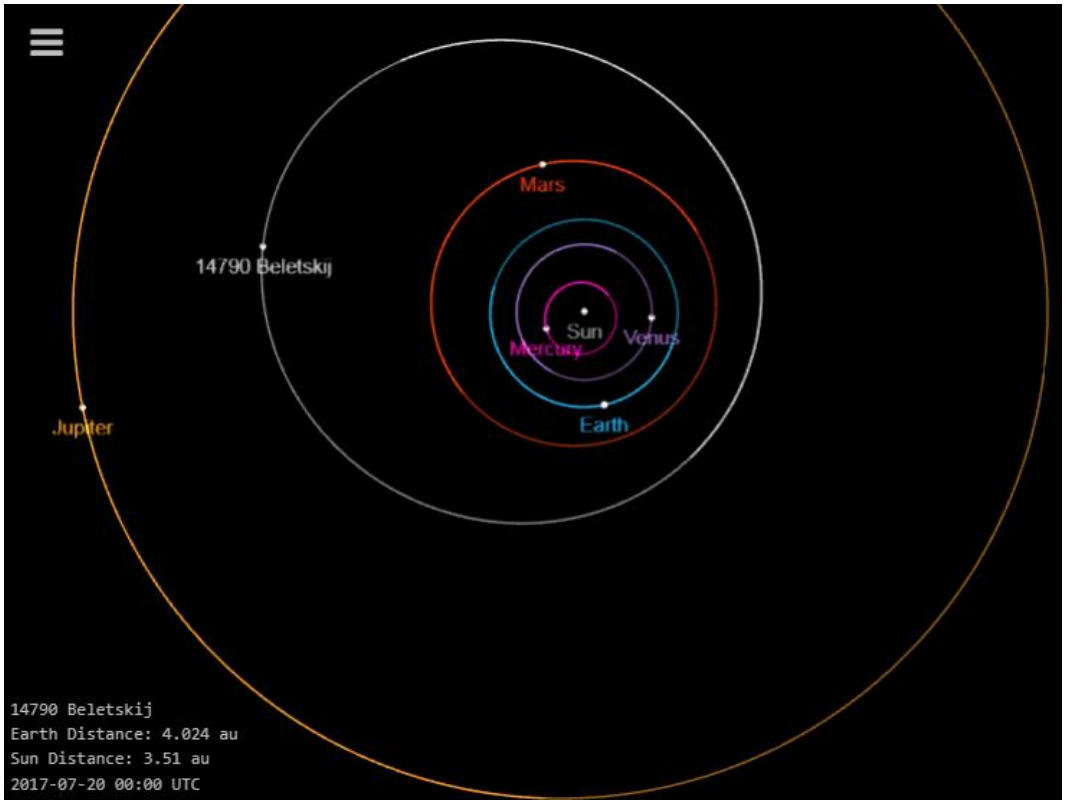


Рис. 2. Малая планета № 14790 (открыта 30 июля 1970 г.), названная именем В.В. Белецкого

Владимир Васильевич Белецкий, будучи главным научным сотрудником ИПМ им. М.В. Келдыша РАН, состоял Членом его Ученого совета, являлся также Членом специализированных диссертационных советов ИПМ им. М.В. Келдыша РАН и механико-математического факультета МГУ им. М.В. Ломоносова, Членом Российского национального комитета по теоретической и прикладной механике (1976). Он был избран Членом-корреспондентом РАН (1997), являлся Действительным членом Международной академии астронавтики (1992) и Действительным членом Российской академии космонавтики (1994), состоял Членом редколлегии журнала “Регулярная и хаотическая динамика”.

Научные достижения В.В. Белецкого высоко оценены в России и за рубежом. Он был удостоен звания Заслуженный профессор МГУ (2002), стал лауреатом премии А. фон Гумбольдта (Германия), лауреат премии РАН им. Ф.А. Цандера. Малая планета № 14790 (открыта 30 июля 1970 г.), названа именем В.В. Белецкого “Beletskij” (рис. 1).

Резюмируя, можно утверждать, что В.В. Белецкий практически открыл новую отрасль механики, упростив классические постановки задач и получив в результате такого упрощения основные «скрижали» для описания механики вращательного движения спутников и планет.

Это был блистательный учёный и непревзойдённый учитель. Его по праву считают одним из отцов-основателей советской и российской школы динамики космического полёта в области теории вращательных движений искусственных и естественных небесных тел. Владимир Васильевич Белецкий скончался 20 июля 2017 г. Владимир Васильевич похоронен на Троекуровском кладбище в Москве.

REFERENCES

[1] V.V. Beletsky, “O libratcii sputneyka”, *Iskusstvennye sputneyki Zemli. M.: Izd-vo AN SSSR,* **3**, 13-31, (1959).

[2] V.V. Beletsky, *Dvizhenie iskusstvennogo sputneyka otnositelno centra mass,* M.: Nauka, (1965).

[3] V.A. Zlatoustov, D.E. Okhotsimsky, V.F. Sarychev, A.P. Torzhevsky, “Periodicheskie resheniya v zadachah o ploskih kolebaniyah sputneyka na ellipticheskoy orbite”, *Kosmicheskie issledovaniya,***2**(5), 658-666 (1964).

[4] A.D. Bryuno, “Semeistvo periodicheskih reshenii uravneniya Beletskogo”, *Kosmicheskie issledovaniya,***40**(3), 295-316 (2002).

[5] V.V. Beletsky, “Dvizhenie iskusstvennogo sputneyka otnositelno centra mass”, *Kosmicheskie issledovaniya,***1**, 25-43 (1957).

[6] V.V. Beletsky, Yu.V. Zonov, “Vrashenie i orientaciya tretego sovetskogo sputneyka”, *Iskusstvennye sputneyki Zemli. M.: Izd-vo A*N SSSR, **7**, 32-55, (1961).

[7] V.V. Beletsky, V.V. Golubkov, E.A. Stepanova, I.G. Hatckevich, “Rezultaty opredeleniya orientacii sputnika “Proton-2” i opisanie ego dvizheniya otnositelno centra mass”, *Kosmicheskie issledovaniya,***7**(4**)**, 522-533 (1969).

[8] V.V. Beletsky, E.T. Novikova, “Ob otnositelnom dvizhenii cvyazki dvuh tel na orbite”, *Kosmicheskie issledovaniya,***7**(3), 377-384 (1969).

[9] V.V. Beletsky, E.M. Levin, *Dinamika kosmicheskih trosovyh system,* M.: Nauka, (1990).

[10]V.V. Beletsky, D.V. Pankov, “Connected bodies in the orbit as a dynamic billiard”, *Regular and chaotic dynamics,* **1**, 87-103 (1996).

[11]V.V. Beletsky, A.V. Grushevskii, “Model formirovaniya vrashatelnyh dvizhenii nebesnyh tel s ogranicheniyami na poryadok rezonansov”, *Astronomicheskii vestnik,* **24**(2), 140-147 (1990).

[12]V.V. Beletsky, “Resonance rotation of celestial bodies and Cassini's laws”, *Celest. Mech.,* **6**(3), 356-378 (1972).

[13]V.V. Beletsky, *Dvizhenie sputneyka otnositelno centra mass v gravitatcionnom pole,* M.: Izd. MGU, (1975)

Received November 15, 2019