MATHEMATICA MONTISNIGRI
Vol XLVIII (2020) 5-170

CONTENTS

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

SOME PROPERTIES OF CENTRAL RIGID RINGS

DUŠAN S. JOKANOVIĆ*

University of East Sarajevo
Faculty of Production and Management Trebinje
Stepe Stepanovića bb, 89 101, Bosnia and Herzegovina
*Corresponding author. E-mail: dusan.jokanovic@fpm.ues.rs.ba

DOI : 10.20948/mathmontis-2020-48-1

Summary. In this paper we deal with central rigid ring which is generalization of rigid ring. We consider how the notion of central rigidity can be transferred from the ring to the corresponding matrix and polynomial extension. Quasi-rigid rings will be also considered.

1 INTRODUCTION

In this article R denotes an associative ring with identity. Mehrabadi i Shakebi in [1] introduced a notion of central rigid ring. For endomorphism σ, a ring R is σ central rigid if a $\sigma(a) = 0$ implies $a \in C(R)$, where C(R) is center of ring R. We know that a class of central rigid rings is generalization of class of rigid rings. An ideal I of a ring R is called σ-ideal if $\sigma(I) \subseteq I$. We say that σ-ideal I is quasi σ-rigid if $aR\sigma(a) \subseteq I$ implies $a \in I$, for all $a \in R$ [2]. A ring R with an endomorphism σ is quasi σ-rigid if ideal I=0 is quasi σ-rigid which is equivalent to condition $aR\sigma(a) = 0$ implies $a=0$. It is well known from [1] that the class of central rigid rings is closed for direct products, localizations, direct limits with injective maps, and isn't closed for homomorphic images. If σ is endomorphism of a ring R then the map σ can be naturally extended to an endomorphism σ' of the ring $R[x]$ by $\sigma'(\sum_{i=0}^{n} a_i x^i) = \sum_{i=0}^{n} \sigma(a_i)x^i$. In [1] is also shown that the notion of central rigidity can be transferred from ring R to corresponding ring of polynomials in linear variant. Let σ be an endomorphism of R. We use $R[x;\sigma]$ to denote skew polynomial ring with the ordinary addition and the multiplication subject to the relation $xr = \sigma(r)x$ (see [3]).

2 CENTRAL RIGIDITY VERSUS QUASI RIGIDITY IN POLYNOMIAL RINGS

Recall that [4] the notion of quasi rigidity transferees from the ring R to the ring R[x]. If σ is endomorphism of a ring R then the map σ can be n extended to an endomorphism σ' of the ring R[x] by equation

$$\sigma'\left(\sum_{i=0}^{n} a_i x^i\right) = \sum_{i=0}^{n} \sigma(a_i)x^i.$$

THEOREM 2.1. [3] If R is quasi σ-rigid then R[x] is quasi σ'-rigid ring.

PROOF. Let $f(x) = a_0 + a_1 x + \ldots + a_n x^n$ and $f(x)R[x]\sigma'(f(x)) = 0$.
Let

$$f(x) = b_0 + b_1 x + \ldots + b_m x^m,$$

be an element from the ring R[x]. At first glance from the

2010 Mathematics Subject Classification: 00A00, 00B00, 00C00.
Key words and Phrases: Adaptive Modeling, Meshing and Remeshing, Goal oriented Adaptivity.

$$\sum_{i=0}^{n} a_i x^i \sum_{j=0}^{m} b_j x^j \sum_{i=0}^{n} \sigma\,(a_i) x^i$$

we obtain $a_0 b_0 \sigma\,(a_0) = 0$ so that we have $a_0 = 0$.

$$a_0 b_0 \sigma\,(a_2) + a_0 b_1 \sigma\,(a_1) + a_0 b_2 \sigma(a_0) + a_1 b_0 \sigma(a_1) + a_2 b_0 \sigma(a_0) + a_2 b_2 \sigma(a_0) = 0$$

so that we obtain $a_1 b_0 \sigma\,(a_1) = 0$. Since R is quasi σ-rigid, we have $a_1 = 0$. Continuing in this way, since the coefficient of x^{2n+m-2} has to be zero and $a_{n-2} = 0$ is obtained in the previous step, we have $a_{n-1} b_m \sigma(a_{n-1}) = 0$. Using the quasi rigidity argument we have $a_{n-1} = 0$. Finally at the end the coefficient of x^{2n+m} has to be zero, we obtain $a_n b_m \sigma\,(a_n) = 0$, which means that $a_n = 0$ and so $f(x) = 0$. □

From the fact that a class of quasi-rigid rings is generalization of class of rigid rings we obtain next result. We know from [4] that a class of quasi-rigid rings is closed for direct product constructions.

COROLLARY 2.1. If R is σ-rigid then R[x] is σ'-rigid ring.

Now we note a result from [5] about generating rigid rings with inner automorphism.

PROPOSITION 2.1. Let σ be an inner automorphism of a reduced ring R, then R is σ-rigid.

PROOF. Suppose that $r\sigma(r) = 0, r \in R$. Since σ is an inner automorphism there exists an invertible element $u \in R$, such that $r\sigma(r) = ru^{-1}ru = 0$, so $ru^{-1}r = 0$. This implies that $(ru^{-1})^2 = 0$ and since R is reduced we get $ru^{-1} = 0$, which means $r = u$. Then $\sigma(r) = r^{-1}r^2$, so $r\sigma(r) = r^2 = 0$, hence $r = 0$. Therefore R is σ-rigid. □

In the class of central rigid ring we have an extension of central rigidity notion only in a case when polynomial is linear. The ring R[x] is called linear central σ-rigid if for any $f(x) = a_0 + a_1 x \in R[x], f(x)\sigma(f(x)) = 0$ implies that $f(x) \in C(R[x])$.

THEOREM 2.2 Let σ be an endomorphism of a ring R. Then R is central σ-rigid if and only if R[x] is linear central σ-rigid.

PROOF. Assume that R[x] is linear central σ-rigid. Then R is central σ-rigid as a subring of R[x]. Conversely, assume that R is central σ-rigid and $f(x) = a_0 + a_1 x \in R[x]$ such that $f(x)\sigma(f(x)) = 0$. Then $a_0\sigma(a_0) = 0$ and $a_1\sigma(a_1) = 0$ and so $a_0;\ a_1 \in C(R)$, since R is central σ-rigid. Therefore, $f(x) \in C(R[x])$ and hence R[x] is linear central \sigma-rigid. □

Noting that the class of central rigid rings is closed for localization, and $R[x;\ x^{-1}] = RS^{-1}$, for $S = \{1, x, x^2, x^3, x^4, \dots\}$, we recall next corollary from [1].

COROLLARY 2.2. Let R be a ring and σ an automorphism of R. Then the following are equivalent:

(1) R is central σ-rigid.
(2) R is linear central σ-rigid.

(3) $R[x;\ x^{-1}]$ is linear central σ-rigid.

3 MATRIX CENTRAL RIGID RINGS

In this section we give an example of matrix central rigid ring. At first glance for a ring R we consider a following set of triangular matrices

$$T_n(R) = \left\{ \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix} \mid a_{ij} \in R \right\},$$

and

$$T(R,n) = RI_n + \sum_{i=1}^{n} \sum_{k=i+1}^{n} RE_{ij},$$

where E_{ij} is the matrix unit for all i,j and I_n is the identity matrix.

$$T(R,n) = \left\{ \begin{bmatrix} a_0 & a_1 & a_2 & \dots & a_{n-1} \\ 0 & a_0 & a_1 & \dots & a_{n-2} \\ 0 & 0 & a_0 & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_0 \end{bmatrix} \mid a_i \in R \right\},$$

$T_n(R)$ and $T(R,n)$, as we know, are subrings of the triangular matrix rings with matrix addition and multiplication. Let α be endomorphism of ring R. It is well known that endomorphism α can be naturally extended to an endomorphism

$$\overline{\alpha}: T_n(R) \to T_n(R),$$

and

$$\overline{\alpha}: T(R,n) \to T(R,n).$$

In the next part we use the notation from [3].

Let $E_{ij} = (e_{st} : 1 \leq s,t \leq n)$ denotes $n \times n$ unit matrices over ring R, in which $e_{ij} = 1$ and $e_{st} = 0$ when $s \neq i$ or $t \neq j$, $0 \leq i,j \leq n$, for all $n \geq 2$. If $V = \sum_{i=1}^{n} E_{i,i+1}$, then

$$V_n(R) = RI_n + RV + \dots + RV^{n-1}$$

is the subring of upper triangular skew matrices.In the next section we will show that under the assumption that $R[x]/(x^n)$ is central rigid we obtain that a ring T(R,n) has the same property.

PROPOSITION 3.1. [1] Let α be an endomorphism of a ring R. Let S be a ring and $\phi : R \to S$ an isomorphism. Then R is central α-rigid if and only if S is central $\phi\alpha\phi^{-1}$-rigid ring.

THEOREM 3.1.Suppose that α is an endomorphism of ring R. If the factor ring $R[x]/(x^n)$ is weak $\tilde{\alpha}$-central rigid, then T(R,n) is $\tilde{\alpha}$-central rigid.

PROOF. Suppose that $R[x]/(x^n)$ is $\tilde{\alpha}$- central rigid. In the next step we define the ring isomorphism $\theta : V_n(R) \to R[x]/(x^n)$ by

$$\theta(r_0 I_n + r_1 V + .. + r_{n-1} V^{n-1}) = r_0 + r_1 x + .. + r_{n-1} x^{n-1} + (x^n).$$

We obtain that $V_n(R)$ is $\theta^{-1}\tilde{\alpha}\theta$-central rigid and calculate

$$\begin{aligned}\theta^{-1}\tilde{\alpha}\theta\theta(r_0 I_n + r_1 V + .. + r_{n-1} V^{n-1}) &= \theta^{-1}\tilde{\alpha}\big(r_0 + r_1 x + .. + r_{n-1} x^{n-1} + (x^n)\big) = \\ &= \theta^{-1}(\alpha(r_0) + \alpha(r_1)x + .. + \alpha(r_{n-1})x^{n-1} + (x^n)) = \\ &= \alpha(r_0) I_n + \alpha(r_1) V + \cdots + \alpha(r_{n-1}) V^{n-1} = \tilde{\alpha}(r_0 I_n + r_1 V + \cdots + r_{n-1} V^{n-1}).\end{aligned}$$

which means that $V_n(R)$ is weak $\tilde{\alpha}$-central rigid [1]. In the next step we use isomorphism f defined by isomorphism $f: R[x]/(x^n) \to T(R, n)$ given by

$$f(a_0 + a_1 x + \ldots + a_{n-1} x^{n-1}) = (a_0, a_1, \ldots, a_{n-1}).$$

Now we use well known theorem of isomorphism of rings to obtain desired result.

We end this section with remark that a class of quasi rigid rings isn't closed for matrix extensions.

For a ring R and endomorphism $\sigma: R \to R$ we consider

$$S_4 = \begin{bmatrix} a & a_{12} & a_{13} & a_{14} \\ 0 & a & a_{23} & a_{24} \\ 0 & 0 & a & a_{34} \\ 0 & 0 & 0 & a \end{bmatrix}, a, a_{ij} \in R,$$

which is subring of $T_4(R)$. If R is a σ rigid and

$$a = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

easy calculation shows that $aS_4\bar{\sigma}(a) = 0$, inspite the fact that $a \neq 0$, which shows that S_4 is not quasi σ-rigid (see also [6]).

4 CONCLUSION

This article deals with a class of rigid rings and its generalizations. Some of classes are closed for direct products, localizations as well as for direct limits. Main results are related to the possibility of extending property of rigidity from ring to some of its extensions. We proved that the property of central rigidity can be transferred from the ring to a corresponding matrix ring, under some conditions of factor ring. Further investigation on rigid rings can be found in recent work [6]. The results on skew rigid rings are extensively used in [7].

REFERENCES

[1] M. Mehrabadi and S. Sahebi, "Central σ-rigid rings", *Palestine Journal of Mathematics*, **6** (2), 569-572 (2017).

[2] T. Ozdin and M. Kosan, "Quasi σ-rigid rings", *Int. J. Contemp. Math. Sciences*, **3** (27), 1331-1335 (2011).

[3] D. Jokanović, "Properties of Armendariz rings and weak Armendariz rings", *Publications de l'Institut Mathematique*, **85**, 131-137 (2009).

[4] D. Jokanović, "A note to quasi rigid rings", *Mathematica Montisnigri*, **3**, 14-17 (2008).
[5] H. Pourtaherian and I. Rakhimov, "On Hilbert Property of Rings", I*nternational Journal of Algebra*, **5** (7), 301-308 (2011).
[6] C. Abdiôglu, S. Sahinkaya and A. Kör, "Rigid, quasi-rigid and matrix rings with $(\overline{\sigma}, 0)$-multiplication", *Algebra and Discrete Mathematics*, **17**, 1-11 (2014).
[7] L. Ouyang, "Extensions of generalized α-rigid rings", *International Journal of Algebra*, **3**, 105-116 (2008).

Received July 10, 2020

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

SOME CONGRUENCES INVOLVING CATALAN, PELL AND FIBONACCI NUMBERS

S. KOPARAL[1*] AND N. ÖMÜR[1]

[1] Kocaeli University Mathematics Department 41380 İzmit Kocaeli Turkey.
*Corresponding author. E-mail: sibel.koparal@kocaeli.edu.tr

DOI: 10.20948/mathmontis-2020-48-2

Summary. In this paper, using some special numbers and combinatorial identities, we show some interesting congruences: for a prime $p>5$,

$$\sum_{k=0}^{(p-1)/2}\frac{C_k}{8^k(k+1)}\equiv 2^{(7-p)/2}\frac{P_p}{p}-4\left(\frac{2}{p}\right)+8-\frac{2^{p+2}}{p}\pmod p,$$

$$\sum_{k=1}^{(p+1)/2}\frac{C_{k-1}}{4^k k(k+1)}\equiv -q_p(2)+\frac{5}{6}\pmod p,$$

$$\sum_{k=1}^{(p+1)/2}\frac{C_{k-1}}{(-16)^k k(k+1)(k+2)}\equiv\frac{L_{3p}}{2^{p+2}p}-\frac{2^{p-2}}{p}+\frac{5}{2}\left(\frac{5}{p}\right)-\frac{91}{40}\pmod p,$$

where C_n, L_n and P_n are the n th Catalan number, the n th Lucas number and the n th Pell number, respectively. $\left(\frac{\cdot}{p}\right)$ denotes the Legendre symbol.

1 INTRODUCTION

The Fibonacci sequence $\{F_n\}$ and the Lucas sequence $\{L_n\}$ are defined by the following recursions:

$$F_{n+1}=F_n+F_{n-1} \quad \text{and} \quad L_{n+1}=L_n+L_{n-1},\ n>0,$$

where $F_0=0$, $F_1=1$ and $L_0=2$, $L_1=1$, respectively. If α and β are the roots of equation $x^2-x-1=0$, the Binet formulas of the sequences $\{F_n\}$ and $\{L_n\}$ have the forms

$$F_n=\frac{\alpha^n-\beta^n}{\alpha-\beta} \quad \text{and} \quad L_n=\alpha^n+\beta^n,$$

respectively.

The Pell sequence $\{P_n\}$ and the Pell-Lucas sequence $\{Q_n\}$ are defined recursively by

$$P_{n+1}=2P_n+P_{n-1} \quad \text{and} \quad Q_{n+1}=2Q_n+Q_{n-1},\ n>0,$$

in which $P_0=0$, $P_1=1$ and $Q_0=Q_1=2$, respectively. If γ and δ are the roots of equation

2010 Mathematics Subject Classification: 11B39, 05A10, 05A19.
Key words and Phrases: Central binomial coefficients, congruences, Fibonacci numbers, Pell numbers.

$x^2 - 2x - 1 = 0$, the Binet formulas of the sequences $\{P_n\}$ and $\{Q_n\}$ have the forms

$$P_n = \frac{\gamma^n - \delta^n}{\gamma - \delta} \quad \text{and} \quad Q_n = \gamma^n + \delta^n,$$

respectively.

The harmonic numbers have interesting applications in many fields of mathematics, such as number theory, combinatorics, analysis and computer science. Harmonic numbers H_n are defined as for a positive integer n

$$H_n = \sum_{k=1}^{n} \frac{1}{k},$$

where $H_0 = 0$. The first few harmonic numbers are $1, \frac{3}{2}, \frac{11}{6}, \frac{25}{12}, \ldots$

Some elementary combinatorial properties of the Catalan numbers are given in [2, 6, 8]. The Catalan numbers are given by

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \binom{2n}{n} - \binom{2n}{n+1}, \; n \in N.$$

For a prime p and an integer a with $p \nmid a$, we write the Fermat quotient $q_p(a) = \left(a^{p-1} - 1\right)/p$. For an odd prime p and an integer a, $\left(\frac{a}{p}\right)$ denotes the Legendre symbol defined by

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } p \mid a, \\ 1 & \text{if } a \text{ is a quadratic residue mod } p, \\ -1 & \text{if } a \text{ is a quadratic nonresidue mod } p. \end{cases}$$

In [3], E. Lehmer showed that for prime $p > 3$,

$$H_{(p-1)/2} \equiv -2q_p(2) \pmod{p}. \tag{1.1}$$

In [9], Z.W. Sun obtained that

$$\sum_{k=0}^{(p-1)/2} \frac{1}{m^k}\binom{2k}{k} \equiv \left(\frac{m(m-4)}{p}\right) \pmod{p},$$

and

$$\sum_{k=0}^{(p-1)/2} \frac{1}{m^k(k+1)}\binom{2k}{k} \equiv \frac{m}{2} - \frac{m-4}{2}\left(\frac{m(m-4)}{p}\right) \pmod{p},$$

where p is an odd prime and m is any integer not divisible by p.

Let p be a fixed prime > 3. Define

$$q(x) = \frac{x^p - (x-1)^p - 1}{p} \qquad \text{and} \qquad G_n(x) = \sum_{k=1}^{p-1} \frac{x^k}{k^n},$$

where x is variable.

In [1], A. Granville showed that

$$q(x) \equiv -G_1(x) \pmod p,$$
$$G_2(x) \equiv G_2(1-x)+x^p G_2(1-1/x) \pmod p,$$
$$(q(x))^2 \equiv -2x^p G_2(x)-2(1-x^p)G_2(1-x) \pmod p.$$

It was proved in [1] (the last expression on page 3) that for any integer $n>1$,

$$\sum_{k=1}^{n-1}\binom{n-1}{k-1}\frac{(-x)^k}{k^2}=\frac{1}{n}\sum_{k=1}^{n-1}\frac{(1-x)^k-1}{k}+\frac{(1-x)^n-(-x)^n-1}{n^2}. \qquad (1.2)$$

In [7], Z.H. Sun obtained the following congruences: for an odd prime p and $G_n(x)\in Z_p[x]$,

$$G_2(x)\equiv\frac{1}{p}\left(\frac{1+(x-1)^p-x^p}{p}-\sum_{k=1}^{p-1}\frac{(1-x)^k-1}{k}\right)+p\sum_{r=2}^{p-1}\frac{x^r}{r^2}\sum_{s=1}^{r-1}\frac{1}{s} \pmod{p^2},$$

and for a prime $p>3$ and $n\in N$,

$$npG_{n+1}(x)\equiv(-1)^n x^p G_n(1/x)-G_n(x) \pmod{p^2}.$$

In [10], Z.W. Sun showed that for a prime $p>3$,

$$\sum_{k=1}^{p-1}\frac{H_k L_k}{k}\equiv 0 \pmod p \quad \text{and} \quad \sum_{k=1}^{p-1}\frac{H_k F_k}{k}\equiv\frac{2}{p}\sum_{k=1}^{p-1}\frac{F_k}{k} \pmod p.$$

In [5], H. Pan and Z.W. Sun obtained that for a prime $p>5$,

$$\sum_{k=1}^{p-1}\frac{L_k}{k^2}\equiv 0 \pmod p,$$

and for a prime $p\neq 2,5$,

$$\sum_{k=0}^{p^a-1}(-1)^k\binom{2k}{k}\equiv\left(\frac{p^a}{5}\right)\left(1-2F_{p^a-\left(\frac{p^a}{5}\right)}\right) \pmod{p^3},$$

where a is a positive integer.

In [4], S. Mattarei and R. Tauraso showed that for a prime $p>3$,

$$\sum_{k=1}^{p-1}\binom{2k}{k}\frac{x^k}{k^2}\equiv 2\left(G_2(\lambda)+G_2(\mu)\right) \pmod p,$$

where $\lambda=\frac{1}{2}\left(1+\sqrt{1-4x}\right)$ and $\mu=\frac{1}{2}\left(1-\sqrt{1-4x}\right)$.

2 SOME CONGRUENCES INVOLVING SPECIAL NUMBERS

In this section, we will give the congruences involving some special numbers. Now, we give the following lemma for further use.

Lemma 1. Let n be any positive integer. Then

$$\sum_{k=1}^{n}\binom{n-1}{k-1}\frac{(-x)^{k+1}}{k^2(k+1)}=\frac{1}{n}\sum_{k=1}^{n-1}\frac{(1-x)^{k+1}-1}{k(k+1)}+\frac{(1-x)^{n+1}-1}{n^2(n+1)}+\frac{x}{n}H_n$$

and

$$\sum_{k=1}^{n}\binom{n-1}{k-1}\frac{(-x)^{k+2}}{k^2(k+1)(k+2)}=\frac{1}{n}\sum_{k=1}^{n-1}\frac{(1-x)^{k+2}-1}{k(k+1)(k+2)}+\frac{(1-x)^{n+2}-1}{n^2(n+1)(n+2)}+\frac{x}{n+1}-\frac{x^2}{2n}H_n.$$

Proof. From Binomial Theorem, we have

$$\begin{aligned}
&\sum_{k=1}^{n-1}\frac{(1-x)^{k+1}-1}{k(k+1)}+\frac{(1-x)^{n+1}-1}{n(n+1)}+xH_n\\
&=\sum_{k=1}^{n}\frac{(1-x)^{k+1}-1}{k(k+1)}+x\sum_{k=1}^{n}\frac{1}{k}=\sum_{k=1}^{n}\frac{(1-x)^{k+1}-1+(k+1)x}{k(k+1)}\\
&=\sum_{k=2}^{n+1}\frac{(1-x)^{k}-1+kx}{k(k-1)}=\sum_{k=2}^{n+1}\frac{1}{k(k-1)}\sum_{j=2}^{k}\binom{k}{j}(-x)^{j}=\sum_{k=2}^{n+1}\frac{1}{k-1}\sum_{j=2}^{k}\frac{(-x)^{j}}{j}\binom{k-1}{j-1}\\
&=\sum_{k=2}^{n+1}\sum_{j=2}^{k}\frac{(-x)^{j}}{j(j-1)}\binom{k-2}{j-2}=\sum_{j=2}^{n+1}\frac{(-x)^{j}}{j(j-1)}\sum_{k=j}^{n+1}\binom{k-2}{j-2}.
\end{aligned}$$

From the equality $\sum_{k=j}^{n+1}\binom{k-2}{j-2}=\binom{n}{j-1}$, we get

$$\begin{aligned}
&\sum_{k=1}^{n-1}\frac{(1-x)^{k+1}-1}{k(k+1)}+\frac{(1-x)^{n+1}-1}{n(n+1)}+xH_n\\
&=\sum_{j=2}^{n+1}\binom{n}{j-1}\frac{(-x)^{j}}{j(j-1)}=\sum_{j=1}^{n}\binom{n}{j}\frac{(-x)^{j+1}}{j(j+1)}=n\sum_{j=1}^{n}\binom{n-1}{j-1}\frac{(-x)^{j+1}}{j^2(j+1)}.
\end{aligned}$$

Thus, the first identity is obtained. Similarly, we obtain other identity.

For a prime p and a rational number $x=\frac{a}{b}$ written in reduced form, $x\equiv 0\ (mod\ p)$ if and only if a is divisible by p.

Theorem 1. Let p be an odd prime. For any rational number x such that $x^{-1}\not\equiv 0(mod\ p)$,

$$\sum_{k=0}^{(p-1)/2}\frac{C_k}{k+1}x^{k+1}\equiv 1-(\lambda-\mu)^{p+1}+\frac{2^{p-1}}{p}\left(\lambda^p+\mu^p-1\right)\ (mod\ p), \tag{2.1}$$

where $\lambda=\frac{1}{2}\left(1+\sqrt{1-4x}\right)$ and $\mu=\frac{1}{2}\left(1-\sqrt{1-4x}\right)$.

Proof. Replacing n and x by $(p+1)/2$ and $4x$ in (1.2), respectively, we have

$$\frac{p+1}{2}\sum_{k=0}^{(p-1)/2}\binom{(p-1)/2}{k}\frac{(-4x)^{k+1}}{(k+1)^2}=\sum_{k=1}^{(p-1)/2}\frac{(1-4x)^k}{k}+2\frac{(1-4x)^{(p+1)/2}-1}{p+1}-H_{(p-1)/2}.$$

From the congruences $\binom{(p-1)/2}{k}(-4)^k\equiv\binom{2k}{k}\pmod p$ for $k=1,\ldots,(p-1)/2$,

$\frac{1}{p+k}\equiv\frac{1}{k}\pmod p$ for $k=1,\ldots,p-1$ and (1.1), we write

$$-2\sum_{k=0}^{(p-1)/2}\binom{2k}{k}\frac{x^{k+1}}{(k+1)^2}\equiv 2\left((1-4x)^{(p+1)/2}-1\right)+\sum_{k=1}^{(p-1)/2}\frac{(1-4x)^k}{k}+2q_p(2)\pmod p. \tag{2.2}$$

By congruence $\binom{p-1}{k}\equiv(-1)^k\pmod p$ for $k=1,\ldots,p-1$, we get

$$\sum_{k=1}^{(p-1)/2}\frac{(1-4x)^k}{k}\equiv-\sum_{k=1}^{(p-1)/2}\binom{p-1}{2k-1}\frac{(1-4x)^k}{k} \tag{2.3}$$

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

Substituting (2.3) into (2.2), we have

$$\sum_{k=0}^{(p-1)/2}\frac{C_k}{k+1}x^{k+1}=\sum_{k=0}^{(p-1)/2}\binom{2k}{k}\frac{x^{k+1}}{(k+1)^2}$$

$$\equiv 1-(1-4x)^{(p+1)/2}-q_p(2)-\frac{\left(\sqrt{1-4x}-1\right)^p-\left(\sqrt{1-4x}+1\right)^p+2}{2p}$$

$$=1-(\lambda-\mu)^{p+1}+\frac{2^{p-1}}{p}\left(\lambda^p+\mu^p-1\right)\pmod p.$$

Thus, this concludes the proof.

From Theorem 1, we immediately deduce the following results.

Corollary 1. Let p be an odd prime. Then

$$\sum_{k=0}^{(p-1)/2}\frac{C_k}{4^k(k+1)}\equiv 4\left(1-q_p(2)\right)\pmod p,$$

$$\sum_{k=0}^{(p-1)/2}\frac{(-1)^k C_k}{k+1}\equiv 2^{p-1}\frac{1-L_p}{p}+5\left(\frac{5}{p}\right)-1\pmod p, \tag{2.4}$$

$$\sum_{k=0}^{(p-1)/2}\frac{C_k}{8^k(k+1)}\equiv 2^{(7-p)/2}\frac{P_p}{p}-4\left(\frac{2}{p}\right)+8-\frac{2^{p+2}}{p}\pmod p,$$

$$\sum_{k=0}^{(p-1)/2} \frac{C_k}{(-4)^k (k+1)} \equiv -\frac{2Q_p}{p} + \frac{2^{p+1}}{p} + 8\left(\frac{2}{p}\right) - 4 \pmod p,$$

$$\sum_{k=0}^{(p-1)/2} \frac{C_k}{(-16)^k (k+1)} \equiv -\frac{2^{3-p}}{p} L_{3p} + \frac{2^{p+3}}{p} + 20\left(\frac{5}{p}\right) - 16 \pmod p,$$

and for $p > 3$

$$\sum_{k=0}^{(p-1)/2} \frac{C_k}{9^k (k+1)} \equiv \frac{2^{p-1}}{p}\left(\frac{L_{2p}}{3^{p-2}} - 9\right) - 5\left(\frac{5}{p}\right) + 9 \pmod p.$$

Proof. For the proof (2.4), considering $x = -1$ in Theorem 1, we have

$$\sum_{k=0}^{(p-1)/2} \frac{C_k}{(-1)^{k+1}(k+1)}$$

$$\equiv 1 - 5^{(p+1)/2} + \frac{2^{p-1}}{p}\left(\left(\frac{1+\sqrt{5}}{2}\right)^p + \left(\frac{1-\sqrt{5}}{2}\right)^p - 1\right) \equiv 1 - 5\left(\frac{5}{p}\right) + \frac{2^{p-1}}{p}\left(L_p - 1\right) \pmod p.$$

Similarly, we obtain other five congruences of Corollary 1.

Theorem 2. Let $p > 3$ be a prime. For any rational number x such that $x^{-1} \not\equiv 0 \pmod p$,

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{k(k+1)} x^{k+1} \equiv \frac{2^{p-1}x}{p}\left(\lambda^p + \mu^p - 1\right) - \frac{8x+1}{12}(\lambda - \mu)^{p+1} + \frac{6x+1}{12} \pmod p,$$

where λ and μ are defined as before.

Proof. By the first identity of Lemma 1, we have

$$\sum_{k=0}^{n-1} \binom{n-1}{k} \frac{(-x)^{k+2}}{(k+1)^2 (k+2)} = \frac{1}{n}\sum_{k=1}^{n-1} \frac{(1-x)^{k+1} - 1}{k(k+1)} + \frac{(1-x)^{n+1} - 1}{n^2 (n+1)} + \frac{x}{n} H_n.$$

Replacing n and x by $(p+1)/2$ and $4x$ in the aboving sum, respectively, we have

$$\sum_{k=0}^{(p-1)/2} \binom{(p-1)/2}{k} \frac{(-4x)^{k+2}}{(k+1)^2 (k+2)} = \frac{2}{p+1} \sum_{k=1}^{(p-1)/2} \frac{(1-4x)^{k+1} - 1}{k(k+1)} + 8\frac{(1-4x)^{(p+3)/2} - 1}{(p+1)^2 (p+3)} + \frac{8x}{p+1} H_{(p+1)/2}.$$

From the congruences $\binom{(p-1)/2}{k}(-4)^k \equiv \binom{2k}{k} \pmod p$ for $k = 1, \ldots, (p-1)/2$,

$\frac{1}{p+k} \equiv \frac{1}{k} \pmod p$ for $k = 1, \ldots, p-1$ and (1.1), we write

$$\sum_{k=0}^{(p-1)/2} \binom{2k}{k} \frac{x^{k+2}}{(k+1)^2(k+2)}$$

$$\equiv \frac{(1-4x)-1}{8} \sum_{k=1}^{(p-1)/2} \frac{(1-4x)^k}{k} + \frac{1}{4(p+1)} - \frac{(1-4x)^{(p+1)/2}}{4(p+1)} + \frac{(1-4x)^{(p+3)/2}}{6} + x - \frac{x}{2} - xq_p(2) - \frac{1}{6} \pmod p.$$

From the congruences (2.3) and $\frac{1}{p+1} \equiv 1 \pmod p$, we have

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{k(k+1)} x^{k+1}$$

$$\equiv -\frac{x}{2p}\left(2 - \left(1+\sqrt{1-4x}\right)^p - \left(1-\sqrt{1-4x}\right)^p\right) - \frac{(1-4x)^{(p+1)/2}}{4} + \frac{(1-4x)^{(p+3)/2}}{6} + \frac{x}{2} - xq_p(2) - \frac{5}{2}$$

$$= -\frac{x}{2p}\left(2^p - \left(1+\sqrt{1-4x}\right)^p - \left(1-\sqrt{1-4x}\right)^p\right) - \frac{(1-4x)^{(p+1)/2}}{4} + \frac{(1-4x)^{(p+3)/2}}{6} + \frac{x}{2} - \frac{5}{2} \pmod p.$$

Thus, the desired result is obtained.

From the above equation, we get the assertion of the theorem.

As an immediate consequence of Theorem 2, we obtain the following result.

Corollary 2. Let $p > 3$ be a prime. Then

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{4^k k(k+1)} \equiv -q_p(2) + \frac{5}{6} \pmod p,$$

$$\sum_{k=1}^{(p+1)/2} \frac{(-1)^k C_{k-1}}{k(k+1)} \equiv 2^{p-1} \frac{L_p - 1}{p} - \frac{35}{12}\left(\frac{5}{p}\right) + \frac{5}{12} \pmod p,$$

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{8^k k(k+1)} \equiv \frac{P_p}{2^{(p-1)/2} p} - \frac{2^{p-1}}{p} - \frac{4}{3}\left(\frac{1}{2}\right)^{(p+1)/2} + \frac{7}{6} \pmod p,$$

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{9^k k(k+1)} \equiv \frac{2^{p-1}}{p}\left(\frac{L_{2p}}{3^p} - 1\right) - \frac{85}{108}\left(\frac{5}{p}\right) + \frac{5}{4} \pmod p,$$

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{(-4)^k k(k+1)} \equiv \frac{Q_p - 2^p}{2p} - \frac{2}{3}\left(\frac{2}{p}\right) + \frac{1}{6} \pmod p,$$

and

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{(-16)^k k(k+1)} \equiv \frac{L_{3p}}{2^{p+1} p} - \frac{2^{p-1}}{p} + \frac{5}{6}\left(\frac{5}{p}\right) - \frac{5}{6} \pmod p.$$

Theorem 3. Let $p > 5$ be a prime. For any rational number x such that $x^{-1} \not\equiv 0 \pmod p$,

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{k(k+1)(k+2)} x^{k+2} \equiv \frac{2^{p-2}}{p} x^2 \left(\lambda^p + \mu^p - 1\right) - \frac{1}{60}(\lambda - \mu)^{p+5} + \frac{5}{96}(\lambda - \mu)^{p+3} - \frac{1}{32}(\lambda - \mu)^{p+1} + \frac{1}{8}x^2 + \frac{1}{12}x - \frac{1}{240} \pmod p,$$

where λ and μ are defined as before.

As an direct consequence of Theorem 3, we obtain the following result.

Corollary 3. Let $p > 5$ be a prime. Then

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{4^k k(k+1)(k+2)} \equiv -\frac{1}{2} q_p(2) + \frac{47}{120} \pmod p,$$

$$\sum_{k=1}^{(p+1)/2} \frac{(-1)^k C_{k-1}}{k(k+1)(k+2)} \equiv 2^{p-2} \frac{L_p - 1}{p} - \frac{15}{16}\left(\frac{5}{p}\right) + \frac{3}{80} \pmod p,$$

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{8^k k(k+1)(k+2)} \equiv \frac{P_p}{2^{(p+1)/2} p} - \frac{2^{p-2}}{p} - \frac{3}{5}\left(\frac{1}{2}\right)^{(p+1)/2} + \frac{21}{40} \pmod p,$$

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{9^k k(k+1)(k+2)} \equiv \frac{2^{p-2}}{p}\left(\frac{L_{2p}}{3^p} - 1\right) - \frac{29}{48}\left(\frac{5}{9}\right)^{(p+1)/2} + \frac{43}{80} \pmod p,$$

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{(-4)^k k(k+1)(k+2)} \equiv \frac{Q_p - 2^p}{4p} + \frac{1}{5}\left(\frac{2}{p}\right) - \frac{11}{40} \pmod p,$$

and

$$\sum_{k=1}^{(p+1)/2} \frac{C_{k-1}}{(-16)^k k(k+1)(k+2)} \equiv \frac{L_{3p}}{2^{p+2} p} - \frac{2^{p-2}}{p} + \frac{5}{2}\left(\frac{5}{p}\right) - \frac{91}{40} \pmod p.$$

3. CONCLUSION

Using combinatorial identities, some interesting congruences are investigated on the sums which include Catalan numbers. It is conceivable to extend our results using the useful technics or methods. A first question on the extension of these congruences can be viewed as generalizations on using the q-Binomial coefficients instead of the binomial coefficients. A second question on such extensions of these congruences can be observed that for an odd prime p and any rational number x such that $x^{-1} \not\equiv 0 \pmod p$, the congruences of the sums with decreasing factorials including Catalan numbers $\sum_{k=1}^{(p-1)/2} \frac{C_{k-1}}{k^{\underline{m}}} x^k \pmod p$.

REFERENCES

[1] A. Granville, "The square of the Fermat quotient", *Integers: Electron. J. Combin. Number*

Theory, **4,** A22, (2004).
[2] P. Hilton and J. Pedersen, "Catalan numbers, their generalization, and their uses", *The Math. Intelligencer*, **13**, 64-75 (1991).
[3] E. Lehmer, "On congruences involving Bernoulli numbers and the quotients of Fermat and Wilson", *Ann. of Math.*, **39**, 350-360 (1938).
[4] S. Mattarei and R. Tauraso, "Congruences for central binomial sums and finite polylogarithms", *J. Number Theory*, **133**, 131-157 (2013).
[5] H. Pan and Z.W. Sun, "Proof of three conjectures on congruences", *Sci. China Math.*, **57**(10), 2091-2102 (2014).
[6] L.W. Shapiro, "A Catalan triangle", *Discrete Math.*, **14**, 83-90 (1976).
[7] Z.H. Sun, "Congruences involving Bernoulli and Euler numbers", *J. Number Theory*, **128**, 280-312 (2008).
[8] Z.W. Sun, "Binomial coefficients, Catalan numbers and Lucas quotients", *Sci. China Math.*, **53**(9), 2473-2488 (2010).
[9] Z.W. Sun, "Congruences involving generalized central trinomial coefficients", *Sci. China Math.*, **57** (7), 1375-1400 (2014).
[10] Z.W. Sun, "On harmonic numbers and Lucas sequences", *Publ. Math. Debrecen*, **80**(1-2), 25-41 (2012).

Received April 11, 2020

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

ON THE CONSTRUCTION OF A GENERALIZED COMPUTATIONAL EXPERIMENT IN VERIFICATION PROBLEMS

A. K. ALEKSEEV, A.E. BONDAREV*, V.A. GALAKTIONOV, A.E. KUVSHINNIKOV

Keldysh Institute of Applied Mathematics Russian Academy of Sciences
Moscow, Russia
*Corresponding author. E-mail: bond@keldysh.ru

DOI: 10.20948/mathmontis-2020-48-3

Summary. This work is devoted to the application of a generalized computational experiment for a comparative assessment of numerical methods accuracy. A generalized computational experiment allows one to obtain a numerical solution for a class of problems determined by the ranges of defining parameters variation. The applications of such approach in the presence of a reference solution and in its absence are discussed. An example of error surfaces constructing is given when the solvers of the OpenFOAM software package are compared. The classic inviscid problem of oblique shock wave is used as a basic task. Variations of the key parameters of the problem — the Mach number and angle of attack — are considered. In addition to the OpenFOAM solvers, the comparison included the WW method, which has a second order of accuracy in time and space and an adjustable artificial viscosity. The problem of flow around a cone at an angle of attack with varying Mach number, cone angle and angle of attack is also considered. The concept of an error index is introduced as an integral characteristic of deviations from the exact solution for each solver.

1 INTRODUCTION

Throughout the history of the development of computational mathematics and mathematical modeling, problems of verification of numerical methods have occupied a special place. When creating a new numerical method or modifying an existing one, the authors had to show the efficiency of their developments and evaluate their accuracy before proceeding with solving practical problems. A huge number of works are devoted to these studies. As an example, we can point to the works [1-12]. Verification of the obtained results and assessment of the accuracy of the applied numerical method was an obligatory part of research in the field of mathematical modeling of physical processes.

As a rule, a comparison of the numerical results was carried out with some reference solution, in the role of which the exact solution was used if available or the available experimental data. A separate problem is the estimation of the accuracy of numerical methods in the absence of a reference solution.

The relevance of the problems of verification of numerical methods and calculations based on them is also evidenced by the presence of federal standards, both foreign [13,14] and recently appeared Russian [15]. Such standards determine the direction of research in this area. However, all these methodological documents are focused on verification in relation to a specific task with fixed values of key parameters.

2010 Mathematics Subject Classification: 49Q99, 76M27.
Key words and Phrases: generalized computational experiment, numerical methods, verification problems, flow around a cone, oblique shock wave, error surfaces, error index.

It should be noted that at present the relevance of verification problems is steadily increasing due to the widespread use of open and commercial packages for solving various problems of mathematical modeling. As a rule, such packages provide the user with a certain set of numerical methods presented in the form of solvers integrated into the software package. In this case, the user is faced with the problem of choosing a solver. And here a number of difficulties arise. Not all solvers undergo comprehensive testing before being implemented into a software package. Commercial packages do not provide complete open information about the implemented numerical methods and their properties. Various development teams can add solvers to open source packages, but they often cannot provide full testing. Therefore, research in the field of verification and comparative evaluation of numerical methods is becoming more and more necessary.

Historically, verification in problems of computational aerogasdynamics consisted of two parts. The first is modeling a qualitative flow pattern containing discontinuities, separated flows, vortices, etc. The second is to ensure the accuracy of the calculation of quantitative characteristics. Here it was necessary to rely on a reference solution - experimental, accurate, or obtained by calculations using other methods.

Verification was usually carried out for one separate task. By default, it was assumed that with a small variation in the governing parameters of the problem (velocity, viscosity, time scales, thermophysical characteristics of the medium, geometric parameters), the numerical method under consideration will be applicable and provide a similar accuracy.

At the present stage, researchers need more comprehensive estimates of the accuracy of numerical methods. For example, the researchers need to have an assessment of accuracy, not for a single task, but for a class of tasks. By a class of tasks we mean a basic task considered in the ranges of change in the set of key parameters. In computational aerodynamics characteristic numbers that determine flow velocity, viscosity, thermophysical properties of the medium, geometric parameters, etc., can serve as such parameters. An opportunity of getting solution for a class of problems is provided by the construction of a generalized computational experiment. Also, a generalized computational experiment can be very useful in assessing the accuracy in the absence of a reference solution. In this case, it is possible to estimate the accuracy using an ensemble of solutions obtained by various numerical methods. The variation of the solver is considered as a parameter and the parametric problem is solved using a generalized computational experiment.

The concept, basic methods and approaches of a generalized computational experiment, as well as a number of software tools for its implementation were developed in Keldysh Institute of Applied Mathematics RAS. The main aspects of constructing a generalized computational experiment and examples of its implementation are described in detail in [16–23, 28-30].

2 GENERALIZED COMPUTATIONAL EXPERIMENT

The emergence of the concept of a generalized computing experiment is associated with the development of high-performance computing clusters and parallel technologies. In problems of computational aerodynamics, parallel technologies usually provide the ability to quickly calculate on detailed grids. However, parallel technologies provide us with another important opportunity. This is the ability to simultaneously calculate on different nodes the same task with different input data. As a rule, such a calculation is performed in multitasking mode.

This opens up the possibility of implementing a generalized computational experiment. The key parameters of the problem under consideration are divided in certain ranges with a certain step, forming a grid partition of a multidimensional box in a multidimensional space of key parameters. The basic problem is solved using parallel technologies at each point of the grid partition. The obtained results represent multidimensional data volumes. Processing, analysis and visual presentation of this data is carried out using methods of visual analytics and scientific visualization. This computing technology is the most general description of a generalized computing experiment.

Obviously, such a concept can be applied to a wide range of tasks. This range includes parametric studies, optimization problems. A generalized computational experiment is an effective tool for solving inverse problems.

A large number of different applications of a generalized computational experiment are described in detail in [16-23, 28-30]. The concept of a generalized computational experiment was applied to a wide range of both model and practical problems.

These tasks include the analysis of the interaction of a viscous supersonic flow with a jet barrier, the flows in the wake of the body, the problems of the interaction of jets, the problem of flowing around a cone at an angle of attack, the problem of oblique shock waves, and many others. The approach to constructing a generalized computational experiment was applied to the problem of finding the optimal three-dimensional shape of the blades assembly for a power plant in terms of power loads.

Also, this approach was applied to the problems of verification of numerical methods. A comprehensive comparative analysis of a number of solvers of the OpenFOAM open software package [24] was carried out in [20-23, 29, 30]. As basic tasks, we used problems that have a reference solution (exact solution or experimental data). These tasks include the problem of a supersonic inviscid flow around a cone at an angle of attack and the problem of an oblique shock wave formation. In both cases, a class of problems was considered, formed by key parameters variations of the problem in question.

3 THE APPROACH OF ACCURACY ESTIMATION ON THE ENSEMBLE OF SOLUTIONS

The estimation of the accuracy of numerical methods in the absence of a reference solution is a separate problem.

Undoubtedly, at present, the understanding of the need to estimate the calculation error is present in the field of CFD and is even formulated as standards [13-15]. However, the methods proposed there are based mainly on the convergence of the solution over the grid (according to [2], this approach goes back to C. Runge) and on Richardson's extrapolation. Both of these approaches are based on the asymptotic behavior of the lowest (in the expansion in terms of the grid step) term of the approximation error and, accordingly, do not provide strict inequalities in the error estimation. For convergence “by adhesion” (Runge), the difference of two solutions (on coarse and fine meshes) is used as an estimate of the error. In Richardson's method, this difference between the solutions is refined using a coefficient that depends on the order of approximation. An additional problem in the field of CFD, which complicates the application of the Richardson method, is the space-variable order of convergence of different algorithms. In particular, on the shock wave for schemes of any approximation order, the convergence order demonstrates values around unity. To take this

effect into account, the generalized Richardson method is used, which allows one to estimate the local order of convergence. Unfortunately, this method is significantly unstable and requires at least four successive mesh refinements, which creates huge computational problems.

One of the alternatives in this case is the estimation of accuracy on the ensemble of solutions. The ensemble of solutions obtained by various numerical methods on the same grid allows us to estimate the location of the exact solution and to divide the obtained numerical solutions into clusters of different levels of accuracy. This direction is being actively developed at present and is presented in [25-27]. A natural drawback of this approach is the need for researcher to have at his disposal a certain number of solvers that implement numerical methods with different computational properties.

In general, a fairly large volume of numerical experiments [25-27] confirms the possibility of estimating the approximation error on an ensemble of independent numerical calculations, which cannot but arouse interest in the analysis of this approach. It seems likely that the transition from a single numerical solution to an ensemble of independent solutions opens up opportunities for the implementation of non-standard concepts of a numerical solution. These topics need deeper analysis and development. However, if successful, one can hope for the creation of computationally efficient algorithms that ensure the verification of numerical solutions even in the absence of reference solutions. An important role in this can be played by the construction of a generalized computational experiment, where the parameter is the choice of the solver, and the numerical solution is implemented in parallel mode simultaneously for the solvers participating in the calculation. In this case, in the presence of a certain number of independent solvers providing an ensemble of solutions, the construction of a generalized computational experiment can dramatically speed up the estimate.

4 COMPARATIVE ACCURACY ESTIMATION USING REFERENCE SOLUTION

This section provides two examples of constructing a generalized computational experiment for a comparative assessment of numerical methods accuracy. As examples, we use the numerical results described in detail in the authors' works [20, 23, 29, 30]. In these papers, two classes of computational gas dynamics problems were considered.

The first class of problem describes a supersonic inviscid flow around a cone at an angle of attack.

The second class describes the incidence of an inviscid supersonic gas flow onto a flat plate at an angle of attack. Both of these problems are fairly well known. The first problem has a tabular solution [36], used as a reference solution. The second problem has an exact solution. We consider the first class of problem for 3D statement. The second class is considered as 2D problem.

Let’s consider the first class of problems. We solve the problem of a supersonic gas flow around a cone at an angle of attack. Variable parameters are angle of attack $\alpha = 0°, 5°, 10°$, Mach number $M = 3, 5, 7$ and cone half-angle $\beta = 10°, 15°, 20°$. The flow scheme is shown in Figure 1. The conditions of the incoming stream at the input are indicated by the index “∞”, and at the output, by the index ξ, since the solution is self-similar and depends on the dimensionless variable.

For calculation, the Euler system of equations is used. The system is supplemented by the ideal gas equation of state.

Three solvers were selected from the OpenFOAM software package: rhoCentralFoam (rCF), sonicFoam (sF), and pisoCentralFoam (pCF). Solver rhoCentralFoam is based on a central-upwind scheme which is a combination of central difference and upwind schemes [31,32]. Solver sonicFoam is based on the PISO algorithm (Pressure Implicit with Splitting of Operator) [33]. Solver pisoCentralFoam is a combination of a central-upwind scheme with the PISO algorithm [34]. This solver is not included in the standard set of OpenFOAM solvers. It was created by independent team of developers at the Ivannikov Institute for System Programming RAS. All the calculations were performed using the OpenFOAM version 2.3.0.

We solved the problem with each solver for the entire set of variable parameters. Thus, we obtained a set of numerical solutions. The exact solution was obtained by interpolating the table solution from [3]. Then we found the error of the solution in the norms L1 and L2. Since different solvers implement different numerical methods, the errors were markedly different from each other. The initial and boundary conditions, as well as the settings of the solvers, were set similarly to [26, 28].

Fig. 2 shows the steady-state solution for the pressure field obtained by interpolating the tabular solution from [36], cone half-angle $\beta = 20°$, angle of attack $\alpha = 10°$, Mach number $M = 3$.

Figure 3 shows the error surface in L2 norm for this problem with the variation of the solver and the half-cone angle at fixed Mach number 3 and fixed angle of attack 5°. It can be seen that the deviation from the exact solution increases with the growth of the half-solution angle. One can also see that the rhoCentralFoam and pisoCentralFoam solvers are approximately equally accurate, while the sonicFoam solver accuracy is much lower.

Figure 4 shows the error surface for the same problem with variation of solvers and angle of attack at a Mach number of 5 and a cone half-angle of 15°.

Figure 5 shows the error surface (deviation from reference solution) for the angle of attack 5° and a cone half-angle 20° with variation of solvers and Mach number.

Thus, analyzing Figures 3, 4, 5 we see that for all solvers, the error increases with increasing the angle of attack, the angle of the cone half-angle and Mach number.

So, we have here the accuracy assessment for all three solvers participating in this research. This is the result of constructed generalized numerical experiment for the class of problems in question.

It is easy to see that in this case the numerical result of the generalized computational experiment is the error function in the L2 norm of 4 variables (Mach number, angle of attack, cone half angle, solver number):

$$Err = F(M, \alpha, \beta, Ns)$$

A complex visual representation of such a function is a separate task and is not considered in this article.

For the considered class of problems, the construction of a generalized computational experiment provided a full-fledged comparative estimate of the accuracy of the three selected solvers of the OpenFOAM package in the range of variation of the determining parameters.

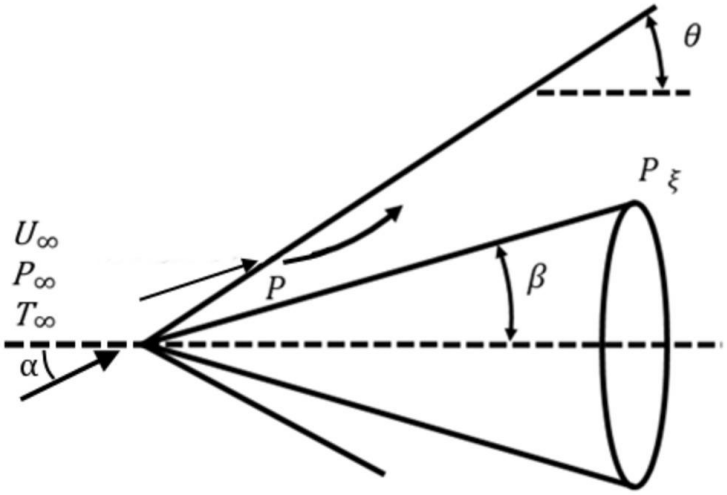


Fig. 1. Flow scheme.

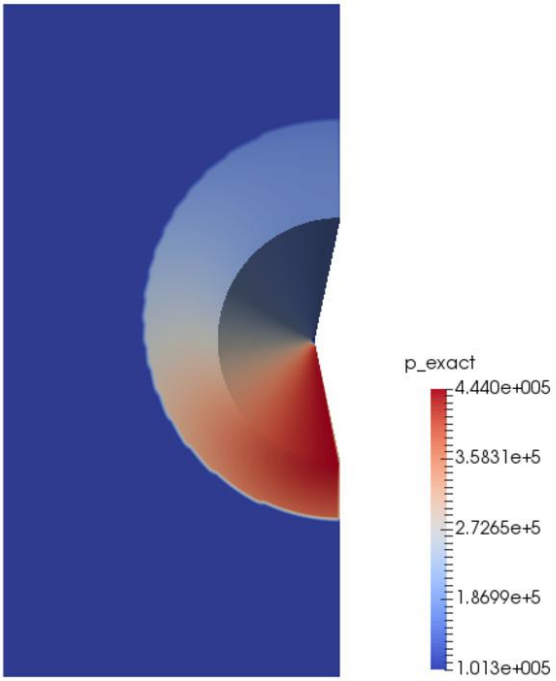


Fig. 2. Pressure field for steady flow.

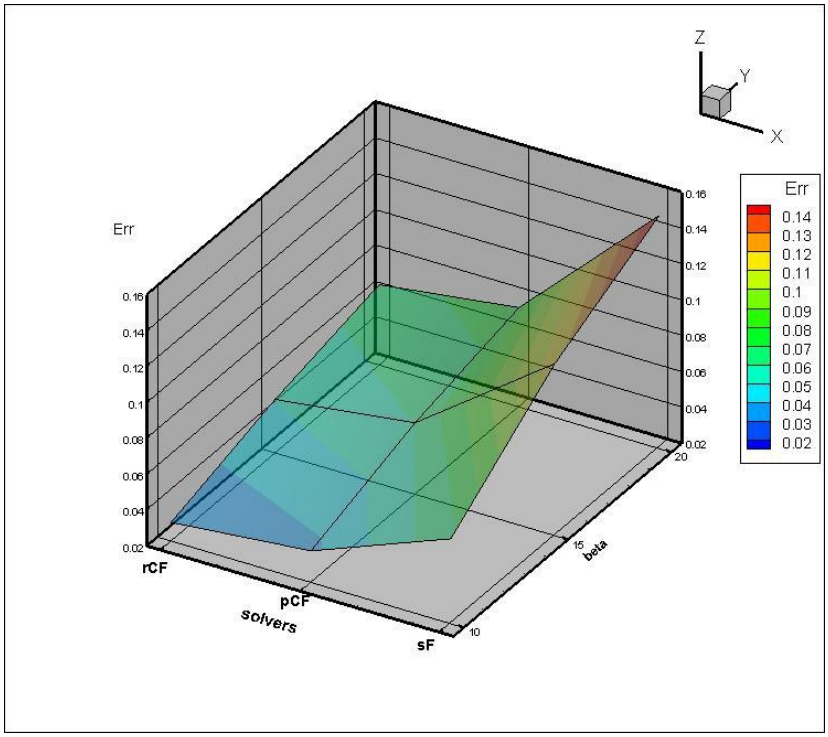


Fig. 3. Image of the error surface for the Mach number 3 and the angle of attack 5° with variation of solvers and half-cone angle.

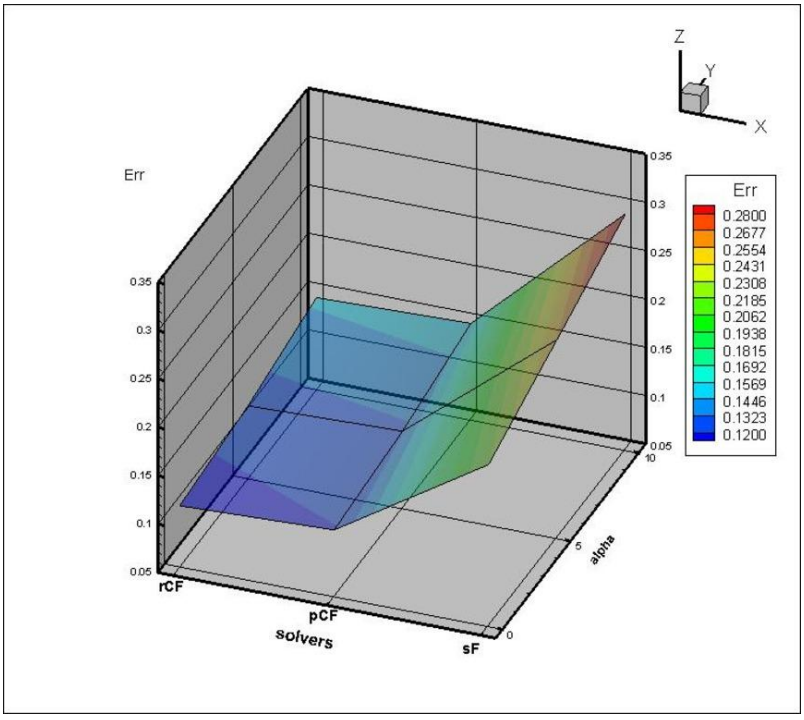


Fig. 4. Image of the error surface for the Mach number 5 and a cone half-angle 15° with variation of solvers and angle of attack.

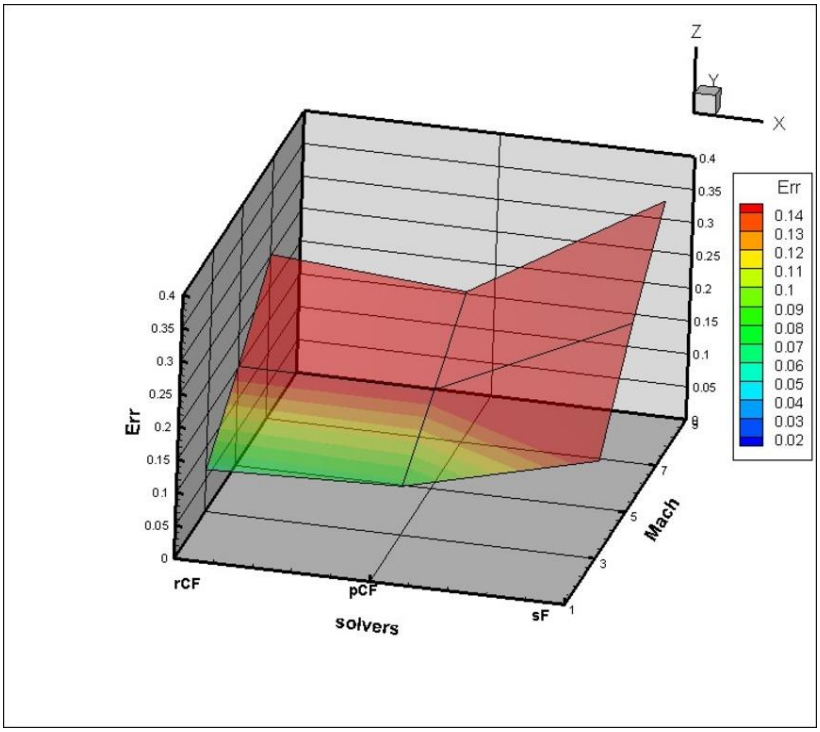


Fig. 5. Image of the error surface for the angle of attack 5° and a cone half-angle 20° with variation of solvers and Mach number.

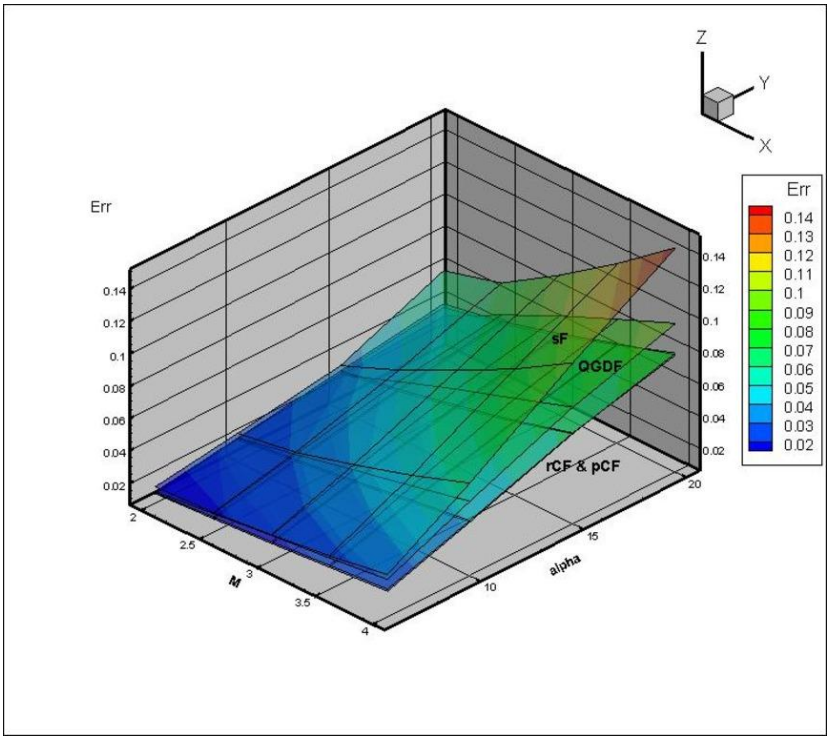


Fig. 6. Error surfaces with variation of the Mach number and angle of attack for the oblique shock wave [29].

Let's consider another class of problems. In this case we use well-known problem of oblique shock wave formation. We consider this problem in 2D statement.

A supersonic gas flow falls on the plate at an angle. Reflecting from the plate, the flow forms an oblique shock wave. The problem has exact solution. In the problem, the Mach number M and the angle of incidence of the supersonic flow β varied, similarly to [29,30]. Figure 6 shows the error surfaces for this problem for 4 solvers: rhoCentralFoam (rCF), sonicFoam (sF), pisoCentralFoam (pCF), QGDFoam (QGDF). Now we include into consideration a new solver QGDFoam (QGDF). This solver is based on a system of quasi-hydrodynamic equations. The solver was also created by independent developers [35].

Carrying out similar calculations for several numerical methods implemented in the solvers of the open software package OpenFOAM, makes it possible to build several such surfaces on one drawing. This opens up the possibility of a deep and clear comparative analysis of the accuracy of the studied numerical methods. The construction of such a generalized computational experiment involves the creation of computational technology from solving a direct problem up to visual analysis of the results.

This technique allows carrying out a detailed visual comparison of deviations from the exact solution. It can be seen that in our case, all error surfaces change in the same way. The error increases with the growth of key parameters. The best accuracy in this class of problems is provided by the rCF and pCF solvers, for which the error surfaces are almost identical. Thus, the construction of a generalized computational experiment allows us to conduct a full-fledged comparative accuracy assessment for four solvers of the OpenFOAM software package in the class of problems. The class of tasks in this particular case is determined by the basic task (oblique shock wave) and the ranges of variation of the key parameters of the problem — the Mach number and angle of attack.

The use of a generalized computational experiment makes it possible to involve new numerical methods in research on the comparative assessment of accuracy. This problem was also solved for the numerical WW-method [37], which does not belong to OpenFOAM solvers, but is implemented as a separate software package. It's ADI-method [38] modification using hybrid implicit finite difference scheme. The review describing hybrid schemes is presented in [39]. The scheme [37] has second order accuracy in space and time. Also the scheme (we'll call this scheme as WW-scheme) is unconditionally stable and simple for programming. Except these properties WW-scheme has one interesting and useful feature. When non-linear problem with strong shocks is solved, one has to reduce undesirable solution oscillations. There are two ways for this. The first way is concerned with procedure of smoothing between time-steps. The second way consists in adding some terms with artificial viscosity to basic equations. Both ways require more calculations and complicate algorhytm. The present numerical method doesn't need these ways. Needed for stabilization of solution artificial viscosity is an internal property of WW-scheme. One can regulate the artificial viscosity by the choice of weight parameters. This property is quite suitable for practical applications. It should be noted that QGDFoam solver [35] also has regulated artificial viscosity. This opens up prospects in the future for studying the effect of artificial viscosity on the comparative assessment of accuracy for similar numerical methods.

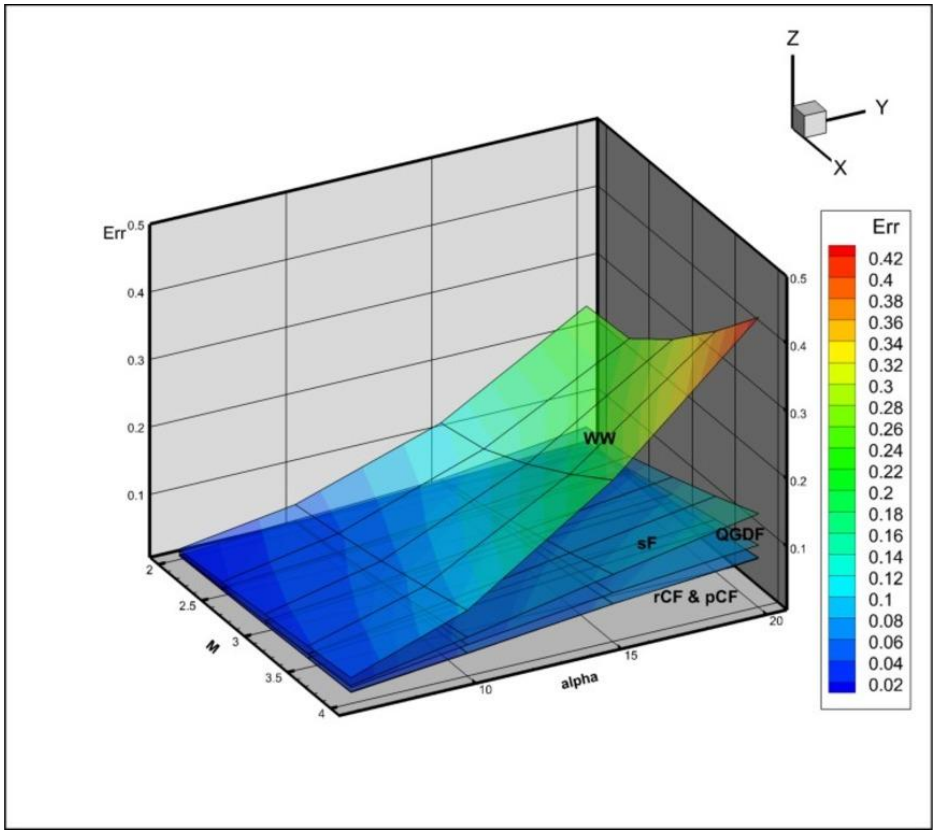


Fig. 7. Error surfaces with variation of the Mach number and angle of attack for the oblique shock wave for OpenFOAM solvers and WW-method.

Figure 7 presents the error surfaces for OpenFOAM solvers with addition such surface for WW-method. One can see that for WW-method, the error increases with an increase in the Mach number and the angle. You can also notice that the error surface for the WW-method is located much higher than the similar surfaces for OpenFOAM solvers. However, this discrepancy can presumably be reduced using artificial viscosity parameters.

The images of error surfaces presented in Figures 6, 7 give a fairly clear idea of the comparative accuracy of numerical methods in the class of problems. However, for a more complete assessment, we enter an integral characteristic for each surface. We will call this characteristic the Error Index (EI). The error index is defined as follows.

Let we have K key parameters. Each of them has its own grid partition, and A - the deviation from the exact solution at each point of the grid partition. We denote the total number of points in the resulting multidimensional space as N. Then the error index is defined as:

$$EI = (\sum A)/N \tag{1}$$

First, we calculate the Error Index for the problem of flow around a cone at an angle of attack.

Solver	rCF	pCF	sF
Error Index	0.13336	0.13366	0.24086

Table 1. Error Index values for the problem of flow around a cone at an angle of attack

Next, we calculate the values of the error index for the problem of oblique shock formation.

Solver	rCF	pCF	QGDF	sF	WW
Error Index	0.037734	0.038751	0.0453406	0.058216	0.14888

Table 2. Error Index values for the problem of oblique shock formation

Tables 1 and 2 show that the values of the error index EI fully correspond to the relative positions of the numerical results presented in figures above. Therefore, the calculated error index can serve as a characteristic of the accuracy of numerical methods in the selected class of problems.

5 CONCLUSIONS

The application of a generalized computational experiment to the problems of comparative estimation of the accuracy of numerical methods is considered. A generalized computational experiment allows simultaneous calculations of the same problem with different input data based on parallel technologies in a multitasking mode. The obtained multidimensional results are examined using visual analysis tools.

Two examples of constructing a generalized computational experiment for classes of problems are presented - flow around a cone at an angle of attack and the formation of an oblique shock wave. For both cases the class of problems is formed on the basis of the basic problem and variations of the determining parameters of the problem. For both classes of problems, a comparative assessment of the accuracy of the solvers of the software package OpenFOAM. For the case of oblique shock wave WW-method is added to comparison. This method does not belong to OpenFOAM solvers and is implemented as a separate software package. An example of constructing error surfaces is given. The concept of a numerical method error index for a class of problems is introduced. Such indexes are computed for all cases in this research. The calculated error indexes can serve as a characteristic of the accuracy of numerical methods in the selected class of problems.

The construction of a generalized computational experiment can serve as an effective tool for verification problems.

REFERENCES

[1] R.D. Skeel, “Thirteen ways to estimate global error”, *Numer. Math*, **48**, 1-20 (1986).

[2] S.I. Repin, *A posteriori estimates for partial differential equations*, Walter de Gruyter, Vol. 4 (2008).

[3] J.T. Oden and S. Prudhomme, “Goal-oriented error estimation and adaptivity for the finite element method”, *Comput. Math. Appl.*, **41**, 735-756 (2001).

[4] S. Prudhomme and J.T. Oden, “On goal-oriented error estimation for elliptic problems: application to the control of pointwise errors”, *Comput. Method. Appl. M.*, **176**, 313-331 (1999).

[5] M Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, New York: Wiley (2000).

[6] I. Babuska, J. Osborn, “Can a finite element method perform arbitrarily badly?”, *Math. Comput.*, **69**, 443-462 (2000).

[7] M.H. Carpenter and J.H. Casper, Accuracy of shock capturing in two spatial dimensions, *AIAA J.*, **37**, 1072-1079 (1999).

[8] J.W. Banks, J.A.F. Hittinger and C.S. Woodward, “Numerical error estimation for nonlinear hyperbolic PDEs via nonlinear error transport”, *Comput. Method. Appl. M.*, **213**, 1-15 (2012).

[9] F. Rauser, J. Marotzke and P. Korn, “Ensemble-type numerical uncertainty quantification from single model integrations”, *J. Comp. Phys.*, **292**, 30-42 (2015).

[10] C. Johnson, “On computability and error control in CFD”, *Int. J. Numer. Meth. Fl.*, **20**, 777-788 (1995).

[11] I. Babuska and W. Rheinboldt, “A posteriori error estimates for the finite element method”, *Int. J. Numer. Meth. Eng.*, **12**, 1597-1615 (1978).

[12] Ch.J. Roy and A. Raju, “Estimation of discretization errors using the method of nearby problems”, *AIAA J.*, **45**, 1232-1243 (2007).

[13] Guide for the verification and validation of computational fluid dynamics simulations, American Institute of Aeronautics and Astronautics, AIAA-G-077-1998, Reston, VA (1998).

[14] Standard for verification and validation in computational fluid dynamics and heat transfer, ASME V&V 20-2009 (2009).

[15] Federal standard P 57700.12–2018. Numerical simulation of supersonic flows for an inviscid gas. Software verification - National standard of the Russian Federation for numerical modeling of physical processes (2018).

[16] A.E. Bondarev, “Analysis of space-time flow structures by optimization and visualization methods”, *Lect. Notes Comput. Sc.*, **7870**, 158-168 (2013).

[17] A.E. Bondarev and V.A. Galaktionov, “Parametric optimizing analysis of unsteady structures and visualization of multidimensional data”, *Int. J. Model. Simul. Sci. Comput.*, **04 No. supp01** (2013).

[18] A.E. Bondarev, “On the Construction of the Generalized Numerical Experiment in Fluid Dynamics”, *Mathematica Montisnigri*, **42,** 52-64 (2018).

[19] A.E. Bondarev, “On visualization problems in a generalized computational experiment”, *Scientific Visualization*, **11**,156-162 (2019).

[20] A.E. Bondarev and A.E. Kuvshinnikov, “Analysis of the accuracy of OpenFOAM solvers for the problem of supersonic flow around a cone”, *Lect. Notes Comput. Sc.*, **10862**, 221-230 (2018).

[21] A.E. Bondarev, “On the estimation of the accuracy of numerical solutions in CFD problems”, *Lect. Notes Comput. Sc.*, **11540**, 325-333 (2019).

[22] A.E. Bondarev and V.A. Galaktionov, “Generalized computational experiment and visual analysis of multidimensional data”, *Scientific Visualization*, **11**, 102-114 (2019).

[23] A.E. Bondarev, A.E. Kuvshinnikov, “Analysis of the accuracy of OpenFOAM solvers for the problem of supersonic flow around a cone”, *Lect. Notes Comput. Sc.*, **10862**, 221-230 (2018).
[24] OpenFOAM Foundation. https://openfoam.org (Accessed August 5, 2020).
[25] A.K. Alekseev and A.E. Bondarev, “On exact solution enclosure on ensemble of numerical simulations”, *Mathematica Montisnigri*, **38**, 63-77 (2017).
[26] A.K. Alekseev, A.E. Bondarev and A.E. Kuvshinnikov, “Verification on the ensemble of independent numerical solutions”, *Lect. Notes Comput. Sc.*, **11540**, 315-324 (2019).
[27] A.K. Alekseev and A.E. Bondarev, “Estimation of the distance between true and numerical solutions”, *Comput. Math. Math. Phys.*, **59**, 857-863 (2019).
[28] A.K. Alekseev, A.E. Bondarev and A.E. Kuvshinnikov, “On uncertainty quantification via the ensemble of independent numerical solutions”, *J. Comput. Sci.*, **42**, 101114 (2020).
[29] A.K. Alekseev, A.E. Bondarev and A.E. Kuvshinnikov, “Comparative analysis of the accuracy of OpenFOAM solvers for the oblique shock wave problem”, *Matematica Montisnigri*, vol. **45**, 95-105 (2019).
[30] A. Bondarev, A. Kuvshinnikov, “Parametric study of the accuracy of OpenFOAM solvers for the oblique shock wave problem”, I*EEE The Proceedings of the 2019 Ivannikov ISPRAS Open Conference*, 108-112 (2019).
[31] A. Kurganov and E. Tadmor, “New high-resolution central schemes for nonlinear conservation laws and convection-diffusion equations”, *J. Comput. Phys.*, **160**, 241-282 (2000).
[32] C. Greenshields, H. Wellerr, L. Gasparini, and J. Reese, “ Implementation of semi-discrete, non-staggered central schemes in a colocated, polyhedral, finite volume framework, for high-speed viscous flows”, *Int. J. Numer. Meth. Fluids*, **63**, 1-21 (2010).
[33] R. Issa, “Solution of the implicit discretized fluid flow equations by operator splitting”, *J. Comput. Phys.*, **62**, 40-65 (1986).
[34] M. Kraposhin, A. Bovtrikova and S. Strijhak, “Adaptation of Kurganov-Tadmor numerical scheme for applying in combination with the PISO method in numerical simulation of flows in a wide range of Mach numbers”, *Procedia Computer Science*, **66**, 43-52 (2015).
[35] M.V. Kraposhin, E.V. Smirnova, T.G. Elizarova, and M.A. Istomina, “Development of a new OpenFOAM solver using regularized gas dynamic equations”, *Computers & Fluids*, **166**, 163-175 (2018).
[36] K.I. Babenko, G.P. Voskresenskii, A.N. Lyubimov, and V.V. Rusanov, *Three-dimensional ideal gas flow past smooth bodies*, Moscow: Nauka (1964).
[37] A.E. Bondarev, “On hybrid numerical method for 2d viscous flows”, *Mathematica Montisnigri*, **29**, 59-67 (2014).
[38] D.W. Peaceman and H.H. Rachford Jr., “The numerical solution of parabolic and elliptic differential equations”, *Journal of the Society for Industrial and Applied Mathematics*, **3**, 28-41 (1955).
[39] A.G. Kulikovsky, N.V. Pogorelov, and A.Yu. Semenov, *Matematicheskie voprosy chislennogo resheniya giperbolicheskih sistem uravneniy* Moscow, PhysMathLit (2001).

Received June 12, 2020

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

COMPUTATIONAL MODEL FOR HIGH-SPEED MULTICOMPONENT FLOWS

V.E. BORISOV, O.B. FEODORITOVA, N.D. NOVIKOVA, YU.G. RYKOV, V.T. ZHUKOV*

Keldysh Institute of Applied Mathematics, RAS. Moscow, Russia
*Corresponding author. E-mail: vic.zhukov@gmail.com

DOI: 10.20948/mathmontis-2020-48-4

Summary. The article describes a robust technology for numerical simulation of multicomponent flows in the presence of strong shock waves, for example, in the flight of a high-speed vehicle with a ramjet/scramjet engine. This takes into account the phenomena of turbulence, multicomponent diffusion and heat transfer. High-temperature phenomena that occur when moving at a high supersonic speed and are associated with dissociation and ionization of the gas environment are not considered, although they fit well into the framework of the general model of a multicomponent reacting gas. The advantage of the algorithm proposed in this article is its adaptability to both the simulation of strong shock waves and to the simulations on small scales, as well as a high potential for parallelization. Illustrative examples of simulations based on the developed method are given.

1 INTRODUCTION

Currently, there is a rise in a new wave of interest in the implementation of controlled flight at a high supersonic speed in the stratosphere using a ramjet/scramjet engine. The processes occurring in a ramjet/scramjet engine during high-speed flight are a complex combination of gas-dynamic phenomena, such as the development of turbulent boundary layers, their interaction with shock waves, the formation of a pseudo-shock, and phenomena associated with subsonic/supersonic combustion of fuel to create thrust. Due to the high speeds, complex and highly transient structure of the processes and the resulting complexity of problem setting and high cost of natural experiments, the ability to perform predictive numerical simulation plays an important role. At the same time, it is necessary to calculate the parameters of the external flow of a high-speed aircraft in order to correctly determine the flow characteristics at the entrance to the ramjet/scramjet tract. In addition high supersonic speeds lead to various high-temperature phenomena begin to appear, which require considering this medium as a multicomponent with the possibility of chemical reactions between components. A similar situation occurs in simulation of fuel combustion in the ramjet/scramjet tract: there are components of gas, fuel and combustion products and there is an additional energy release due to chemical reactions.

All this suggests that the predictive numerical model for the flight of a high-speed aircraft with a ramjet/scramjet engine is multi-scale and, accordingly, requires a high level of detail. The solution of such problems in a reasonable time for practice is possible only on high-performance parallel computing systems.

2010 Mathematics Subject Classification: 35Q30, 65Q10, 39A70.
Key words and Phrases: Multicomponent gas mixture, Thermal conductivity, Viscosity and Diffusion, Scramjet, Hypersonic flight.

Accordingly, the numerical algorithm should take into account a variety of physical processes on the one hand, and not be too complex to be able to control the results responsibly on the other hand. The algorithm should also be robust and well parallelized, preferably regardless of the specific architecture of the supercomputer, for example, multiprocessor or hybrid.

The purpose of this paper is to describe an algorithm of this type for numerical simulation of multi-scale multicomponent gas dynamics problems with the possibility of reactions between components. In the world literature, there is a huge number of works devoted to the actual simulations of high-speed flight in the stratosphere using the thrust of a ramjet/scramjet, as well as verification and validation of numerical modeling. These procedures are necessary elements for creating predictive numerical models. Most of the work focuses either on the external flow around the aircraft, see, for example, [1, 2], or on the processes occurring in the ramjet/scramjet tract, in particular, on the propagation and stabilization of the flame front depending on the geometric features of the tract, see, for example, [3, 4]. The algorithm proposed in this article is especially adaptable to both the simulation of strong shock waves and the simulations on small scales, where combustion occurs, for example, as well as it has a high parallelization potential. This is achieved by using three main components. First, the method of splitting by physical processes is used; generally convective flows and flows that have a diffusive character are split. Second, the calculation of the hyperbolic part (convective flows) uses a multi-component modification of the Godunov scheme with the exact solution of the Riemann problem. Third, for solving a subproblem involving diffusive flows (the parabolic part), the original iterative method is used, a scheme of the LI-M type, see, for example, [5]. The use of a LI-M scheme allows, instead of implicit approximation of parabolic operators, to make the transition to the next time layer using explicit Chebyshev iterations, the number of which is inversely proportional to the grid parameter and is determined without using empirical parameters. The simulation is performed with the Courant number typical for hyperbolic problems. In addition to significantly reducing the simulation time of the complete problem, this methodology is well parallelized on arbitrary unstructured grids, which distinguishes it from other methodologies based on the use of implicit schemes.

The algorithm is illustrated by an example of numerical simulation of multicomponent flow in a model ramjet/scramjet tract with injection of a hydrocarbon gas-phase fuel. As a result, a multi-scale picture of the interaction of the shock wave system with the boundary layer appears and a complex flow of a multicomponent gas mixture is formed. The problem under consideration is of real practical interest, since it prepares a natural transition to three-dimensional problem statements, the simulation of real layouts, combined with the modeling of external flow.

All the algorithmic constructions described above are based on the templates of the author's package of Keldysh Institute of Applied Mathematics NOISEtte [6, 7]. As a result, the new computer code MCFL (MultiComponent FLows) inherits the high parallel efficiency inherent in the NOISEtte code. This latter code is not chosen by chance, it is among the most advanced Russian CFD codes, it includes a large set of ready-made numerical models, and there are many alternative solutions. Note that there are different approaches to creating software packages for numerical modeling, see, for example, [8].

The MCFL code is in the process of development, verification tests are performed in various tasks, starting with simple ones, see, for example, [9]. In this paper, verification is

performed on the complex problem of simulation of multicomponent mixture flows without combustion in a model combustion chamber, for which the results of experiments and simulations using different methods are known, i.e. it is essentially a specialized benchmark. Nevertheless, despite the complexity and non-linearity of the problem, the numerical results demonstrate the refinement of approximate solutions when reducing the size of the mesh cells. The simulations were performed using supercomputer resources of Keldysh Institute of Applied Mathematics (K-100, K-10) with MPI communications between various computing nodes.

The results show that the new methodology can provide a structure of complex processes. In this case, the strategy for simulation of complex problems is to solve a series of simpler problems sequentially and perform sequential verification by establishing convergence over a set of grids and comparing numerical solutions obtained using various numerical methods.

2 BASIC EQUATIONS OF THE MATHEMATICAL MODEL FOR A MULTICOMPONENT REACTING MIXTURE

We will assume that there is a gas mixture consisting of a basic gas, such as air, and a set of chemically reacting components, such as components of a hydrocarbon fuel. This takes into account the phenomena of turbulence, multicomponent diffusion and heat transfer, including the interaction of these processes with chemical processes. High-temperature phenomena that occur when moving at a high supersonic speed and are associated with dissociation and ionization of the basic gas will not be considered here, although they fit well into the framework of the general model of a multicomponent reacting gas.

Simulations of gas mixture flows are based on the system of Unsteady Reynolds averaged Navier-Stokes equations with the introduction of additional terms and equations for accounting for the effects of turbulence and combustion (it is assumed to sum up the repeated indexes below; $i, j = 1, 2, 3$), see, for example, [10, 11].

Continuity equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u_i)}{\partial x_i} = 0. \tag{1}$$

Momentum conservation:

$$\frac{\partial (\rho u_i)}{\partial t} + \frac{\partial (\rho u_i u_j)}{\partial x_j} = -\frac{\partial p^*}{\partial x_i} + \frac{\partial}{\partial x_j}\left[\tau_{ij}\right]. \tag{2}$$

Energy conservation:

$$\frac{\partial (\rho E)}{\partial t} + \frac{\partial (\rho u_j E + u_j p^*)}{\partial x_j} = -\frac{\partial q_j^T}{\partial x_j} + \frac{\partial}{\partial x_j}\left[u_i \tau_{ij}\right]. \tag{3}$$

Equations for describing turbulence:

$$\frac{\partial (\rho k)}{\partial t} + \frac{\partial}{\partial x_j}(\rho k u_j) = \frac{\partial}{\partial x_j}\left((\mu + \sigma_k \mu_t)\ \frac{\partial}{\partial x_j} k\right) + P_k\ -\ \rho \beta^* k \omega, \tag{4}$$

$$\frac{\partial(\rho\omega)}{\partial t}+\frac{\partial}{\partial x_j}\left(\rho\omega \mathrm{u}_j\right)=\frac{\partial}{\partial x_j}\left(\left(\mu+\sigma_\omega\mu_t\right)\ \frac{\partial}{\partial x_j}\omega\right)+$$

$$\alpha\rho S^2-\ \rho\beta\omega^2+(1\text{-}F_1)2\rho\sigma_{\omega 2}\frac{1}{\omega}\frac{\partial\omega}{\partial x_j}\frac{\partial k}{\partial x_j}\,.$$

The transfer equations of chemical components with mass fractions Y_m, $m=1, ..., N_{sp}$ have the form:

$$\frac{\partial(\rho Y_m)}{\partial t}+\frac{\partial\left(\rho u_j Y_m\right)}{\partial x_j}=-\frac{\partial J_{j,m}}{\partial x_j}+\dot{\omega}_m\,, \tag{5}$$

and the following conditions are set

$$\sum Y_m=1, \qquad \sum J_{j,m}=\ \sum\dot{\omega}_m\ =0\,.$$

Summation in these and similar formulas is based on the number of components from $m=1$ to $m=N_{sp}$. Let's explain the notations, variables, and functions used in the equations (1) – (5): $\vec{u}=\{u_i\}$ is the velocity vector of the averaged flow of a gas mixture; ρ is the mixture density; $p^*=p+2/3\ \rho k$, p is the thermodynamic pressure of the mixture, k is the energy of turbulent pulsations, ω is the specific rate of turbulent energy dissipation; $E=e+0.5\cdot u_k u_k+k\equiv e+k+K$ is the total energy of multicomponent flow, e is specific internal energy, K is the kinetic energy; $\vec{q}^T=\left\{q_i^T\right\}$ is vector of flow caused by temperature changes; the tensor of viscous stresses τ_{ij} is put in the form

$$\tau_{ij}=\left(\mu+\mu_t\right)\left[\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right)-\frac{2}{3}\delta_{ij}\frac{\partial u_k}{\partial x_k}\right], \tag{6}$$

where μ and μ_t are the coefficients of the molecular and turbulent viscosity of the gas mixture respectively. Specific form of diffusion and heat fluxes $J_{j,m}$, q_j^T, as well as source term $\dot{\omega}_m$, will be described in the next section.

To describe turbulence, the SST Menter model is used, see [12, 13], which includes two differential equations (4) for k and ω. But another model can also be used. An original holistic view regarding the nature of turbulence can also be found in [14], and a view of critical phenomena from the standpoint of non-equilibrium phase transitions can be found in [15]. In addition non-traditional approach to conservation laws theory is contained in [16].

The system of equations (1) – (5) is supplemented by the Mendeleev-Clapeyron equation of state for a mixture of ideal gases

$$p=\rho\frac{R}{M}T\,, \tag{7}$$

where R is the universal gas constant, M is the molar mass of the total mixture and $1/M = \sum Y_m / M_m$, M_m is the molar mass of the mth component.

For a reacting multicomponent mixture, the specific internal energy e is determined via the enthalpy h of the mixture using the formula $e = h - p/\rho$. At the same time

$$h = h_s + h_{ch} \equiv \int_{T_0}^{T} C_p(\theta)d\theta + \sum \Delta h_m^0 Y_m = \sum h_m Y_m \equiv \sum Y_m \int_T^T C_{p,m}(\theta)d\theta + \sum \Delta h_m^0 Y_m \quad (8)$$

where $C_{p,m}(T)$ is the heat capacity of the mth component at constant pressure, Δh_m^0 is the enthalpy required to form 1 kg of the component at standard temperature $T_0 = 298.15\,K$.

Finally, we note that instead of the energy equation (3), we often use the equation for enthalpy, which is a consequence of (1) – (3). The advantage of the equation for enthalpy is that it describes the contribution of the chemical energy of reactions to the increase in the flow temperature.

3 COEFFICIENTS OF MULTICOMPONENT DIFFUSION, VISCOSITY, THERMAL CONDUCTIVITY, AND CHEMICAL REACTIONS

In the framework of the model considered in this paper, the flux $\vec{q}^T$ consists of two parts: the actual heat flux and the diffusion flux of the mixture:

$$q_j^T = -\lambda \frac{\partial T}{\partial x_j} + \sum h_m J_{m,j} . \quad (9)$$

Diffusion flux $J_{m,j}$ of each component is defined according to the formula $J_{m,j} = -\rho D_m \cdot \partial Y_m / \partial x_j$, where the effective diffusion coefficients D_m can be found by Wilke's rule, see [17–19], using binary diffusion coefficients of the individual components. The coefficients of the mixture's viscosity and thermal conductivity can be written using, for example, the Wilke approximations [17] and the Mason-Saxena approximations [20]. A more accurate approximation according to [21] of the effective coefficients of viscosity, thermal conductivity, and diffusion is written using the so called reduced Q–integrals as the first approximations of the Chapman-Enskog method.

We will use simpler approximations at this stage. To calculate the molecular viscosity of each mixture component, the Sutherland formula is used $\mu_m(T) = C_{1,m} T^{\Delta_m} / \left(T + C_{0,m}\right)$ with specially chosen constants. The thermal conductivity and diffusion coefficients of each component are taken as $\lambda_m = C_{p,m}(T)\mu_m / \mathrm{Pr}$ and $D_m = \mu_m / (\rho Sc)$ assuming Prandtl number Pr and Schmidt number Sc. To average the coefficients of the thermal conductivity and the viscosity, we use the following formulas

$$\mu = \frac{1}{2}\left[\sum \mu_k X_k + \left(\sum (X_k / \mu_k)\right)^{-1}\right], \quad \lambda = \frac{1}{2}\left[\sum \lambda_k X_k + \left(\sum (X_k / \lambda_k)\right)^{-1}\right],$$

where $X_k = Y_k M / M_k$. The mixture turbulent viscosity μ_t when using Menter model (4) is taken according to the formula $\mu_t = \rho k / \omega$.

The term responsible for chemical reactions is described in the standard way

$$\dot{\omega}_m = M_m \sum_{j=1}^{N_r} \dot{\omega}_{mj}, \qquad \dot{\omega}_{mj} = \left(\upsilon''_{mj} - \upsilon'_{mj}\right)\left(k_{fj} \prod_{l=1}^{N_{sp}} \left[\frac{\rho Y_l}{M_l} \right]^{\upsilon'_{lj}} - k_{bj} \prod_{l=1}^{N_{sp}} \left[\frac{\rho Y_l}{M_l} \right]^{\upsilon''_{lj}} \right). \quad (10)$$

Here N_r is the number of reactions, $\upsilon'_{mj}, \upsilon''_{mj}$ are the molar stoichiometric coefficients, k_{fj}, k_{bj} are respectively, the forward and backward reaction rate coefficients, depending on the temperature.

4 MAIN FEATURES OF THE NUMERICAL ALGORITHM

Let us introduce the state vector of conservative variables $\boldsymbol{U} \equiv \rho\left(1, u_1, u_2, u_3, E, k, \omega, \left\{Y_m, m = 1, ..., N_{sp}\right\}\right)$. Then the difference approximation of the system of equations (1) – (5) can be written as follows

$$\frac{\partial}{\partial t}\boldsymbol{U} + C_h\left(\boldsymbol{U}\right) = D_h\left(\boldsymbol{U}\right), \quad (11)$$

where $C_h\left(\boldsymbol{U}\right)$ is nonlinear convective difference operator on a grid with a characteristic cell diameter h, and $D_h\left(\boldsymbol{U}\right)$ is, generally speaking, a nonlinear diffusive operator. By linearizing (11) and taking a time step τ, an explicit scheme can be written in the following form

$$\frac{\boldsymbol{U}_{j+1} - \boldsymbol{U}_j}{\tau} + \tilde{C}_h \boldsymbol{U}_j = \tilde{D}_h \boldsymbol{U}_j. \quad (12)$$

The main goal is to offer a specific implementation of (11) that preserves the time step constraints that correspond only to the hyperbolic part. To achieve this goal, we use the method of splitting the system into hyperbolic and diffusion parts and solving the appropriate diffusion problem using a special iterative process. Namely, we will solve equations (11) at each time step using two stages. At the first stage we put $\tilde{D}_h \equiv 0$ and find the solution of an intermediate difference hyperbolic equations using an explicit scheme

$$\frac{\bar{\boldsymbol{U}}_{j+1} - \boldsymbol{U}_j}{\tau} + \tilde{\boldsymbol{C}}_h \boldsymbol{U}_j = 0. \quad (13)$$

In this case, the Godunov method for calculating flows with an exact solution of the Riemann problem for a multicomponent mixture is used [22]. Note that the Riemann problem for turbulent equations (4) is solved separately. Using an exact solution to the Riemann problem allows us to better take into account the specifics of the flows with strong shock waves. As is well known, multicomponent flow generates a number of computational instabilities, to mitigate which the Godunov scheme must be modified. The modification [22] uses the so-called "double flow" method, see, for example, [23].Thus, intermediate values $\bar{\boldsymbol{V}}_j, \bar{\boldsymbol{U}}_j = \rho \bar{\boldsymbol{V}}_j$ are obtained at the hyperbolic stage. At the second stage we put $\tilde{C}_h \equiv 0$ and write difference parabolic problem with new initial data

$$\frac{\boldsymbol{U}_{j+1}-\bar{\boldsymbol{U}}_j}{\tau} = \tilde{D}_h \hat{\boldsymbol{U}}_j . \tag{14}$$

If in (14) $\boldsymbol{U}_j$ is taken instead of $\hat{\boldsymbol{U}}_j$, then the summation of (14) and (13) leads to the conservation equations (12), but the resulting scheme requires diffusive limitation on the time step size, $\tau \sim h^2$. To avoid such a strict limitation we find $\hat{\boldsymbol{U}}_j$ in the course of iterations of the LI-M scheme detailed in [5]. The construction of this scheme is based on the use of properties of Chebyshev polynomials. Approximation and stability are inherent in the construction of the scheme. The equations (14) are the result of final iteration of the LI-M scheme, and the number of iterations p is inversely proportional to the grid parameter h and is determined without using empirical parameters by an exact formula $p = \left\lceil 0.25\pi \sqrt{\tau\, \lambda_{max} + 1} \right\rceil$, where λ_{max} is the upper bound of the spectrum of the discrte operator $\tilde{D}_h$. The resulting algorithm remains stable under the hyperbolic time step constraint $\tau \sim const \cdot h$ and lends itself well to the parallelization process.

Above it is described only the principle scheme of splitting. In the concrete implementation at the current stage of development the turbulent equations (4) do not participate.

5 SIMULATION RESULTS

Below there are demonstrative numerical simulation results of multicomponent turbulent flows with strong shock waves performed on a multiprocessor supercomputer K-100 of the hybrid architecture using the methodology described above. The simulation strategy is developed using the example of a two-dimensional problem of modeling the flow of a chemically reacting multicomponent medium in a channel with a backward step. This channel is a model of the combustion chamber of a high-speed aircraft. The task is set as follows, see [24]. The combustion chamber is a duct with a step that consists of a short part with an expansion angle of 0.5° and a long constant-area part, Figure 1.

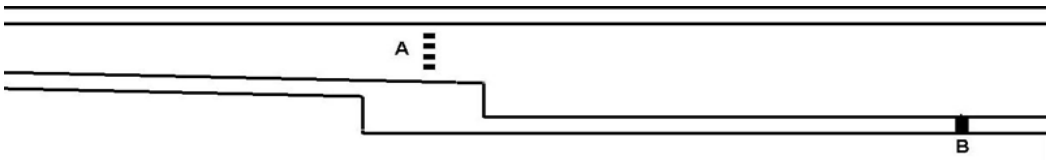


Figure 1. The schematic of the combustion chamber

At the entrance to the chamber, inflow supersonic flow (Mach number 2.5) is set, containing the products of hydrocarbon fuel combustion. The fuel injectors are located in the front of the chamber in the area marked with the letter A. To ensure the ignition of the mixture, a short-term locking of the channel with a stream of compressed air is used, the supply of which is carried out across the main flow using a pneumatic throttle B. Calculations are performed on a refined sequence of nested meshes, indicated below 1x, 2x, 4x. Each of the meshes is obtained from the previous one by doubling the number of cells in both directions. The 1x grid consists of ~ 110 thousand nodes, the 2x grid ~ 440 nodes, etc. In [25], similar numerical experiments were performed using the OpenFOAM software package. Note

that the results of calculations on the 2x and 4x grids do not visually differ much, so the graphs below are given for the 1x and 2x grids. The time-marching scheme is implemented to obtain a steady-state solution or oscillating mode in the combustion process in the chamber. The second order spatial Galerkin approximation is used in the NOISEtte for all viscous, heat conduction and diffusion fluxes.

In Figure 2 the flow structure of a multicomponent mixture is shown. Figure 3 shows the flow details in the fuel injection zone in the gas phase, and in additional, two coordinate lines $x = const$ and $y = const$ are marked, along which Figure 4 shows one-dimensional profiles of fuel concentration and pressure for grids 1x and 2x. Figures 2 and 4 demonstrate the convergence of approximate solutions when reducing the size of grid's cells.

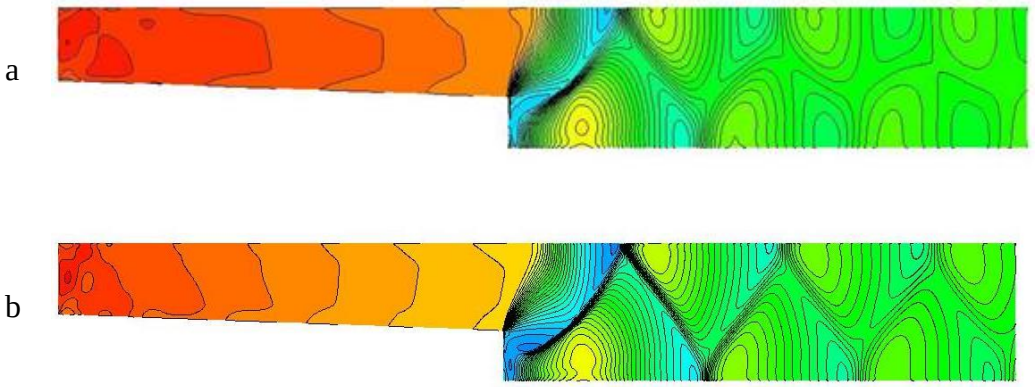


Figure 2. Normalized pressure field on the grids 1x (a) and 2x (b)

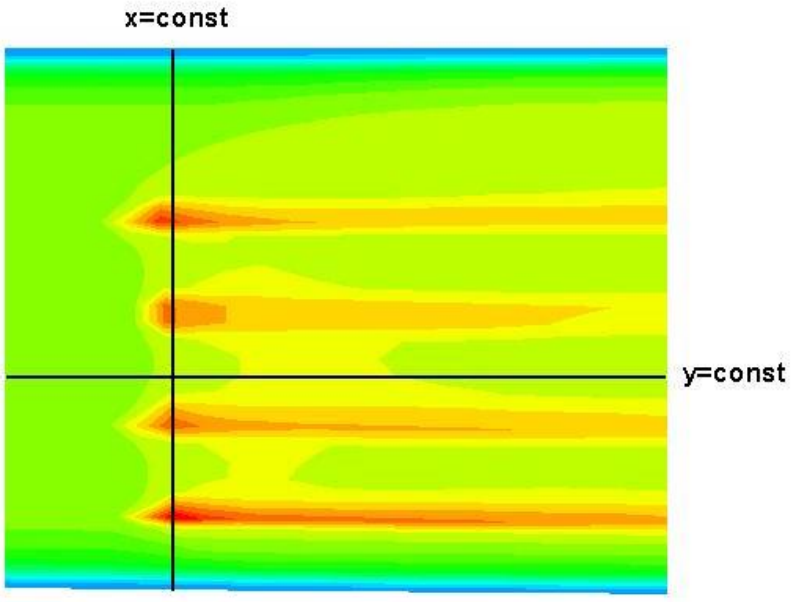


Figure 3. Fuel injection zone; lines $x = const$ and $y = const$ for the imaging 1D profiles

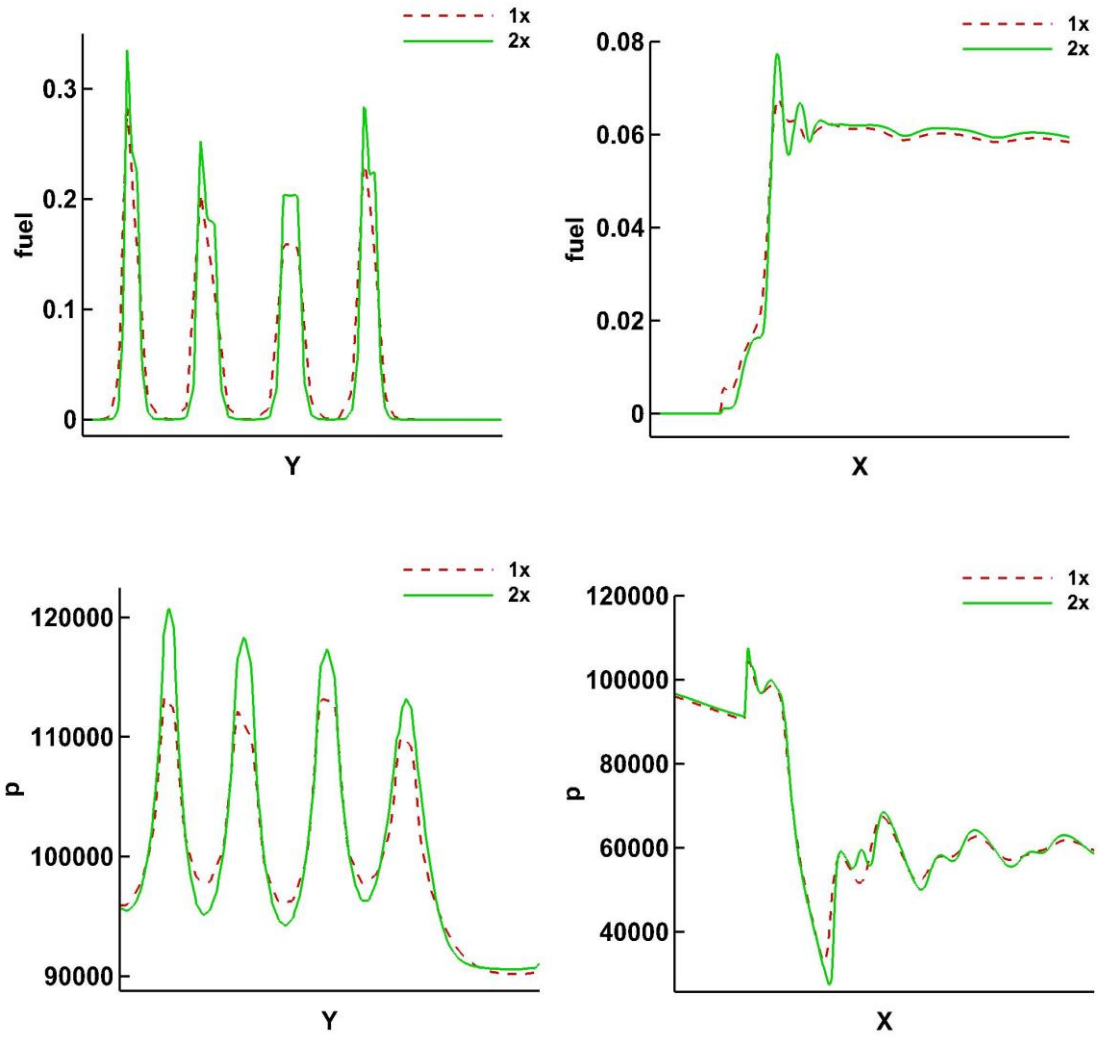


Figure 4. Profiles of fuel fraction (top row) and pressure (bottom row) along the chosen lines for the grids 1x and 2x

Majority of numerical-experimental investigations of the flow in a chamber with the step were implemented for the validation of numerical technologies, see for example [24]. Similarly, the results presented here show that the developed numerical technique describes high-speed flow in the combustor duct correctly.

6 CONCLUSION

The paper presents an original numerical method for integrating multi-component Navier-Stokes equations. The difference method used allows us to more accurately account for the presence of strong shock waves and maintain stability under hyperbolic restrictions on the value of the time step. The results of simulations are qualitatively consistent with experimental data and calculated data from other studies.

The proposed algorithm is well parallelized, which makes it a robust tool for numerical modeling complex flows on grids with a large number of nodes, for example, the external flow of a high-speed aircraft and the internal flow in a ramjet/scramjet tract. The presented methodology can serve as a development basis for creating a fully functional tool for numerical modeling of complex processes on high-performance computing systems.

Acknowledgements: This work was supported by Russian Science Foundation, project № 19-71-30004.

REFERENCES

[1] T. Murugan, S. De, V. Thiagarajan, "Validation of three-dimensional simulation of flow through hypersonic air-breathing engine", *Defence Science Journal*, **65**(4), 272–278 (2015).
[2] T.V. Markova et all, "Simulating flow around scaled model of a supersonic vehicle in wind tunnel", *J. Phys.: Conf. Ser.*, **774**, 012095 (2016).
[3] F.H.E. Ribeiro, R. Boukharfane, A. Mura, "Highly-resolved large-eddy simulations of combustion stabilization in a scramjet engine model with cavity flameholder", *Computers and Fluids*, **197,** 104344 (2020).
[4] E.Jeong, S.O'Byrne, I-S. Jeung, A.F.P. Houwing, "The effect of fuel injection location on supersonic hydrogen combustion in a cavity-based model scramjet combustor", *Energies*, **13**(1), 1–16 (2020).
[5] V.T. Zhukov, "On explicit methods for the time integration of parabolic equations", *Math. Models Comput. Simul.*, **3**(3), 311–332 (2011).
[6] P.A. Bakhvalov, I.V. Abalakin, T.K. Kozubskaya, "Edge-based reconstruction schemes for unstructured tetrahedral meshes", *Int. J. Numer. Methods Fluids*. **81**(6), 331–356 (2016).
[7] I.V. Abalakin. A.V. Gorobets, A.P. Duben, T.K. Kozubskaya, P.A. Bakhvalov, "Parallel Research Code NOISEtte for Large-Scale CFD and CAA Simulations", *Numerical methods and programming*. **13**, 110–125 (2012).
[8] I.N. Konshin, K.M. Terekhov, Yu.V. Vassilevski, "Numerical modelling via INMOST software platform", *Mathematica Montisnigri*, **47**, 75-86 (2020).
[9] V.T. Zhukov, N.D. Novikova, O.B Feodoritova, "An Approach to Time Integration of the Navier–Stokes Equations", *Comp. Math. and Math. Physics*, **60**(2), 272–285 (2020).
[10] T. Poinsot, D. Veynante, "*Theoretical and numerical combustion"*, Edwards, 3rd Edition, 522 p. (2011).
[11] S.T. Surzhikov, " *Radiation gas dynamics of the spacecraft. Multi-temperature models"*, IPMech RAS, 706 p. (2013) (in Russian).
[12] F.R. Menter, "Two-equation eddy-viscosity turbulence models for engineering applications", *AIAA-Journal*, **32**(8), 269–289, (1994).
[13] W. Vieser, T. Esch, F. Menter. "Heat transfer predictions using advanced two-equation turbulence models", *CFX Validation Report 10/0602, AEA Technology*, 1–69 (2002).
[14] M.Ya. Marov, A.V. Kolesnichenko, "*Turbulence and self-organization*", Springer, (2013), 657 p.
[15] E.V. Radkevich, E.A. Lukushov, N.N. Yakovlev, O.A. Vasilieva, M.I. Sidorov, *"Introduction to the generalized theory of non-equilibrium phase transitions and thermodynamic analysis of solid environment mechanics tasks"*, Lomonosov Moscow State University, (2019), 342 c. (in Russian)
[16] Yu.G. Rykov, "Extremal properties of the functionals connected with the systems of conservation laws", *Mathematica Montisnigri*, **46**, 21–30 (2019).
[17] C.R. Wilke, "A viscosity equation for gas mixtures", *J. Chem. Phys.*, **18**(4), 517–522 (1950).
[18] C.R. Wilke, "Diffusional Properties of Multicomponent Gases", *Chemical engineering progress*, **46**(2), 95–104 (1950).
[19] D. F. Fairbanks, C.R. Wilke, "Diffusion Coefficients in Multicomponent Gas Mixtures", *Ind.*

Eng. Chem., **42**(3), 471–475 (1950).
[20] E.A. Mason, S.C. Saxena, "Approximate Formula for the Thermal Conductivity of Gas Mixtures", *Physics of Fluids*, **1**(5), 361–369 (1958).
[21] J.H. Ferziger, H.G. Kaper,"Mathematical Theory of Transport Processes in Gases", North-Holland Publishing Company, 1972.
[22] V.E. Borisov, Yu.G. Rykov, "Modified Godunov method for multicomponent flow simulation", *J. Phys.: Conf. Ser.*, **1250** (2019), 012006.
[23] G. Billet, R. Abgrall, "An Adaptive Shock-Capturing Algorithm for Solving Unsteady Reactive Flows", *Computers & Fluids*, **32** (10), 1473–1495 (2003).
[24] M. Ivankin, A. Nikolaev, V. Sabelnikov, A. Shiryaeva, V. Talyzin, V. Vlasenko, "Complex Numerical-Experimental Investigations of Combustion in Model High-Speed Combustor Ducts", *Acta Astronaut.*, **158**, 425–437(2019).
[25] V.T. Zhukov, N.D. Novikova, O.B Feodoritova, "On the Numerical Simulation of Combustion in a Scramjet Combustor Using OpenFOAM", *Math Models Comput Simul*, **11**(2)**,** 266–276 (2019).

Received June 6, 2019

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

MATHEMATICAL MODELING OF THE CONTACT INTERACTION OF FUEL ELEMENTS USING THE MORTAR METHOD

P.S. ARONOV[1,2*], M.P. GALANIN[1,2], AND A.S. RODIN[1,2]

[1] Keldysh Institute of Applied Mathematics (Russian Academy of Sciences). Moscow, Russia

[2] Bauman Moscow State Technical University. Moscow, Russia

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

DOI: 10.20948/mathmontis-2020-48-5

Summary. The article discusses the implementation of the algorithm for solving axisymmetric contact problems of the thermoelasticity theory using the mortar method. This algorithm is used for numerical simulation of the contact interaction of several bodies under thermomechanical loading. The ill-conditioned system of linear algebraic equations obtained as a result of finite element discretization is numerically solved using the modified symmetric successive over-relaxation method (MSSOR), generalized to the case of contact of several bodies. The results of the algorithm application are demonstrated on a problem simulating some processes in a fuel element with a different number of bodies. The effect of the contacting bodies number and mesh steps on the number of iterations necessary to achieve a given accuracy while solving the system of equations is investigated.

1 INTRODUCTION

Accounting the contact interaction of various functional equipment units allows to obtain a more accurate estimation of the construction stress-strain analysis. The most promising and often used technique of the contact interaction studying is numerical methods, the leading place among which is occupied by the finite element method.

It is often not possible to use matched meshes while modeling the contact of a large number of bodies. The numerical solution of such problems is carried out using various algorithms, among which we can distinguish the domain decomposition method [1,2], the penalty method [3,4], various versions of the Lagrange multiplier method [5,6], in particular, the mortar method [7,8], based on the construction of a sufficiently detailed auxiliary mesh for determining the Lagrange multipliers in the case when the meshes are unmatched. Algorithms designed for modelling the interaction of bodies in dynamic problems can also be noted [9].

This paper discusses a sufficiently general statement of the contact interaction problem of several bodies is given and an implementation of an algorithm for numerically solving contact problems using the mortar method is presented. The block system of linear algebraic equations with a saddle point obtained as a result of the problem discretization is ill-conditioned, and a modified symmetric successive over-relaxation method is used to solve it, generalized to the case of contact interaction of several bodies.

The article considers the problem modeling some processes in a fuel element with a different number of contacting bodies on unmatched meshes. The dependence of iterations number on the selected finite element mesh and the number of bodies is analyzed.

2010 Mathematics Subject Classification: 74M15, 74S05, 74B05.
Key words and Phrases: Contact Problem of the Elasticity Theory, Finite Element Method, Mortar-method, Succesive Over-Relaxation Method.

2 MATHEMATICAL FORMULATION OF THE PROBLEM

Let a group of axisymmetric thermoelastic contacting bodies (Fig. 1) be located in three-dimensional space $\mathbb{R}^3$, occupying a domain $G = \bigcup_{\alpha} G_\alpha$ (α is a body number index), bounded by a piecewise smooth border $\partial G = \bigcup_{\alpha} \partial G_\alpha$.

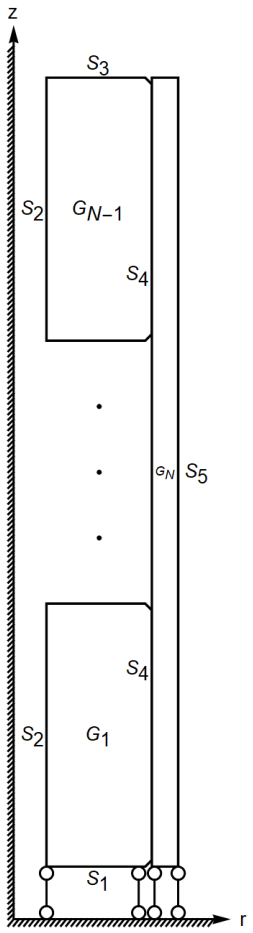


Figure 1. Scheme of contact interaction of several bodies

Consider the following problem: inside a cylindrical cladding G_N there is a column of several placed on top of each other identical cylindrical pellets $G_1,\ldots,G_{N-1}$, having an inner hole and chamfers at both ends. Each pellet (except G_1 и G_{N-1}) comes into contact with two adjacent pellets and the cladding (it is believed that there is no initial gap between them). S_1 is the lower end of the lower pellet, S_2 is the inner surface of the pellets, S_3 is the top end of the top pellet, S_4 is the boundary between the inner surfaces of the pellets and the cladding, S_5 is the outer surface of the cladding. The lower ends of the cladding and lower pellets are fixed vertically. There is no friction on the contact surfaces. This problem simulates some thermomechanical processes taking place in a fuel element.

Suppose that the coupling effect (the dependence of temperature on the deformation of the body) can be neglected; therefore, we will solve the heat conduction problem separately, and use the obtained temperature field to solve the contact problem of the thermoelasticity theory.

Consider the following initial-boundary value problem for the nonlinear heat equation:

$$\begin{gathered} c(T)\rho\frac{\partial T}{\partial t} = (k_{ij}(T)T_{,j})_{,i} + q(\mathbf{x}), \quad \mathbf{x} \in G; \\ T(\mathbf{x},0) = T_0(\mathbf{x}), \quad x \in G; \\ -n_i k_{ij}(T)T_{,j}\,|_{S_1\cup S_2\cup S_3} = q_w(\mathbf{x},t), \quad \mathbf{x} \in S_1 \cup S_2 \cup S_3, \quad t > 0; \\ -n_i k_{ij}(T)T_{,j}\,|_{S_4} = \alpha(T)[T(\mathbf{x},t) - T_f(\mathbf{x},t)], \quad \mathbf{x} \in S_4, \quad t > 0, \end{gathered} \tag{1}$$

where $c(T)$ is the specific heat capacity of the medium, ρ is the medium density, t is time, k_{ij} is the thermal conductivity tensor components, $T_{,j} = \dfrac{\partial T}{\partial x_j}$, $q(\mathbf{x})$ is the power of internal sources (drains) of the body, $T_0(\mathbf{x})$ is the initial temperature, $T(\mathbf{x},t)$ is temperature at time t, n_i are the components of the unit vector of the external normal to the boundary ∂G, $q_w(\mathbf{x},t)$ is the heat flux density on the surfaces S_1, S_2, S_3, $\alpha(T)$ is the heat transfer coefficient on the surface S_4, $T_f(\mathbf{x})$ is the temperature at a similar point lying on the opposite side of the contact pair.

The mathematical formulation of the contact problem of the elasticity theory for the case when there are no bulk forces includes the following relations [10] for each body $G_\alpha \subset \mathbb{R}^3$, participating in the contact ($i, j = \overline{1,3}$):

- equilibrium equations

$$\sigma_{ji,j}(\mathbf{u}) = 0, \quad \mathbf{x} \in G_\alpha; \tag{2}$$

- kinematic boundary conditions

$$\mathbf{u}(\mathbf{x}) = \mathbf{u_0}(\mathbf{x}), \quad \mathbf{x} \in S_1; \tag{3}$$

- force boundary conditions

$$\sigma_{ji}(\mathbf{u})n_j = g_i(\mathbf{x}), \quad x \in S_3 \cup S_4; \tag{4}$$

- Cauchy relations

$$\varepsilon_{ij}(\mathbf{x}) = \frac{1}{2}(u_{i,j}(\mathbf{x}) + u_{j,i}(\mathbf{x})), \quad \mathbf{x} \in G_\alpha; \tag{5}$$

- governing equations (Hooke's law)

$$\sigma_{ij}(\mathbf{x}) = C_{ijkl}(\varepsilon_{kl}(\mathbf{x}) - \varepsilon^0_{kl}(\mathbf{x})), \quad \mathbf{x} \in G_\alpha; \tag{6}$$

- kinematic contact condition

$$u_n^{\alpha_1}(\mathbf{x}) = -u_n^{\alpha_2}(\mathbf{x}), \quad x \in S_k^{\alpha_{12}}; \tag{7}$$

- force contact condition

$$\sigma_n^{\alpha_1}(\mathbf{x}) = \sigma_n^{\alpha_2}(\mathbf{x}) \leqslant 0, \quad \mathbf{x} \in S_k^{\alpha_{12}}, \tag{8}$$

where x_i are the coordinates of the vector $x \in G_\alpha$; σ_{ij} are the stress tensor components; ε_{kl} are the strain tensor components; ε_{kl}^0 are the components of the initial strain tensor (for a thermoelastic body such are thermal strains); u_i are the displacement vector components; C_{ijkl} are the components of the elastic constants tensor; g_i are the surface force vector components; n_j are the components of the external normal vector to the corresponding surface S_j; $u_n^{\alpha_i}$ are the projections of displacement vectors of boundary points on the direction of the external normal n to the body α_i boundary; $\sigma_n^{\alpha_i}$ are the projections of stress vectors on the directions of external normals n_i.

The conditions of contact interaction with respect to displacements and stresses must be fulfilled when solving the problem on the contact surfaces of bodies.

For the case under consideration of the axisymmetric formulation of the problem, the vectors of stresses $\boldsymbol{\sigma}$, strains $\boldsymbol{\varepsilon}$, и and displacements $\mathbf{u}$ cylindrical coordinate system are written as follows:

$$\boldsymbol{\sigma} = \begin{Bmatrix} \sigma_r \\ \sigma_z \\ \sigma_\theta \\ \tau_{r\theta} \end{Bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{Bmatrix} \varepsilon_r \\ \varepsilon_z \\ \varepsilon_\theta \\ \gamma_{r\theta} \end{Bmatrix}, \quad \mathbf{u} = \begin{Bmatrix} u_r \\ u_z \end{Bmatrix}.$$

The solution of the problem (2)-(8) is equivalent to [11] minimizing the functional

$$\Pi = \frac{1}{2}\int_G \boldsymbol{\sigma}^T \varepsilon \, dG - \int_S \mathbf{u}^T \mathbf{g} \, dS + \int_{S_k} \lambda_n (u_n^2(\mathbf{x}) - u_n^1(\mathbf{x})) \, dS \tag{9}$$

when fulfilling the kinematic boundary conditions (3), where λ_n are the Lagrange multipliers, which are projections of the stress vectors on the directions of the external normals, $u_n = u_r n_r + u_z n_z$.

3 BASIC MATRIX RELATIONS OF THE FINITE ELEMENT METHOD

For the numerical solution of the problem (2)-(8) we will use the finite element method. The finite element mesh consists of second-order quadrangular elements.

The components $u_r^{(e)}$, $u_z^{(e)}$ of the displacement vector $\mathbf{u}$ inside the finite element with the index e are determined using the dependence

$$\begin{Bmatrix} u_r^{(e)} \\ u_z^{(e)} \end{Bmatrix} = \left[N\right]^{(e)} \left\{u\right\}^{(e)},$$

where $\left[N\right]^{(e)}$ is the matrix of the shape functions [12] of the finite element with the index e, and $\left\{u\right\}^{(e)}$ is the the combined vector of displacement components in all nodes of the finite element with the number e.

Relations between deformations and displacements are written as follows [12]:

$$\left\{\varepsilon\right\}^{(e)} = \left[B\right]^{(e)} \left\{u\right\}^{e}, \tag{10}$$

where $\left[B\right]^{(e)}$ is the gradient matrix [11] of the finite element with the index e.

Stresses are expressed through deformations using the Hooke law:

$$\left\{\sigma\right\}^{(e)} = \left[D_{\alpha}\right]^{(e)} \left\{\varepsilon\right\}^{(e)},$$

or, taking into account (10),

$$\left\{\sigma\right\}^{(e)} = \left[D_{\alpha}\right]^{(e)} \left[B\right]^{(e)} \left\{u\right\}^{(e)},$$

where $\left[D_{\alpha}\right]^{(e)}$ is the local elasticity matrix of the finite element [12] with the index e for the body with the index α. For the axisymmetric formulation of the problem the matrix $\left[D_{\alpha}\right]^{(e)}$ is written as follows:

$$\left[D_{\alpha}\right]^{(e)} = \frac{E}{(1+\nu)(1-2\nu)} \begin{pmatrix} 1-\nu & \nu & \nu & 0 \\ \nu & 1-\nu & \nu & 0 \\ \nu & \nu & 1-\nu & 0 \\ 0 & 0 & 0 & \dfrac{1-2\nu}{2} \end{pmatrix},$$

where E is the Young's modulus and ν is the Poisson's ratio.

4 APPLICATION OF THE MORTAR METHOD FOR SOLVING CONTACT PROBLEMS

The mortar method for solving contact problems of the elasticity theory is based on the independent finite element discretization of disjoint subdomains. The meshes on these subdomains are, generally speaking, unmatched on the contact line, and the continuity of the solution is achieved through the use of Lagrange multipliers [13]. Among the main advantages of the mortar method, the possibility of independent selection of various types of finite elements and shape functions both at the boundaries of contacting bodies and during integration along the contact line can be noted.

To simplify the recording, we restrict ourselves to the case of two bodies with one pair of contact surfaces. Let the body G_m be master and the body G_s be slave. The contact line from the side of the body G_m is denoted by Γ_m and from the side of the body G_s is denoted by Γ_s. We consider one-dimensional second order finite elements on the contact lines Γ_m and Γ_s. From the nodes of these elements on the contact line Γ_m we draw normals to contact line Γ_s.

For the finite elements formed at the intersection of the normals and Γ_s we will carry out further integration, considering them also as one-dimensional quadratic elements with similar shape functions. The division of bodies into master/slave is largely conditional and non-obvious, but ultimately this choice determines the discretization of Lagrange multipliers [14].

Consider the following integral:

$$\int_{\Gamma} \boldsymbol{\lambda}^T (\mathbf{u_m} - \mathbf{u_s})\, d\gamma = \sum_{i=1}^{k_m} \int_{\Gamma_{mi}} \boldsymbol{\lambda}^T \mathbf{u_m}\, d\gamma - \sum_{i=1}^{k_s} \int_{\Gamma_{si}} \boldsymbol{\lambda}^T \mathbf{u}_s\, d\gamma, \tag{11}$$

where $\Gamma = \Gamma_m \cup \Gamma_s$, k_m and k_s are the total number of finite elements into which the contact lines Γ_m and Γ_s are divided, respectively, the vectors $\mathbf{u_m}$ and $\mathbf{u_s}$ consist of the normal components of the displacement vectors of the finite element nodes on the contact lines Γ_m and Γ_s, the vector $\boldsymbol{\lambda}$ consists of Lagrange multipliers corresponding to the projections of stress vectors on the directions of external normals on the contact line Γ_s. Inside the finite element with the index (e) values of λ_n, u_s and u_m are expressed as follows:

$$\lambda_n = \left[N_\lambda\right]^{(e)} \left\{\lambda\right\}^{(e)}, \quad u_s = \left[N_s\right]^{(e)} \left\{u_s\right\}^{(e)}, \quad u_m = \left[N_m\right]^{(e)} \left\{u_m\right\}^{(e)},$$

where $\left[N_\lambda\right]^{(e)}$, $\left[N_s\right]^{(e)}$, $\left[N_m\right]^{(e)}$ are the matrices of the shape functions of the one-dimensional quadratic element with the index (e).

Minimization of functional (9) together with integral (11) leads to the formation of the following system of linear algebraic equations [15]:

$$\begin{pmatrix} A_{11} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & M_1 \\ \mathbf{0} & A_{22} & \mathbf{0} & \dots & \mathbf{0} & M_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & A_{(N-1)(N-1)} & \mathbf{0} & M_{N-1} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & A_{NN} & M_N \\ M_1^T & M_2^T & \dots & M_{N-1}^T & M_N^T & \mathbf{0} \end{pmatrix} \begin{Bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \\ \vdots \\ \mathbf{u_{N-1}} \\ \mathbf{u_N} \\ \boldsymbol{\lambda} \end{Bmatrix} = \begin{Bmatrix} \mathbf{R_1} \\ \mathbf{R_2} \\ \vdots \\ \mathbf{R_{N-1}} \\ \mathbf{R_N} \\ \mathbf{0} \end{Bmatrix}, \tag{12}$$

where

$$\left[A_{ii}\right] = \sum_{e=1}^{k_i} \left[a_G\right]^{(e)T} \left(\int_{G_i} \left[B\right]^{(e)T} \left[D_i\right]^{(e)} \left[B\right]^{(e)}\, dG \right) \left[a_G\right]^{(e)},$$

$$\left\{R_i\right\} = \sum_{e=1}^{k_i} \left[a_S\right]^{(e)T} \left(\left[N\right]^{(e)T} \left[g_i\right]^{(e)}\, dV \right),$$

$$\lambda = \begin{pmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_{q-1} & \lambda_q \end{pmatrix}^T,$$

$$M_i = \begin{pmatrix} 0 & \dots & 0 & M_{i1} & 0 & \dots & 0 & M_{ij} & 0 & \dots & 0 & M_{ip} & 0 & \dots & 0 \end{pmatrix},$$

where N is the number of bodies, q is the number of contact pairs and in the columns of the

matrices M_i corresponding to the numbers of the bodies in contact with the body with the number i, are the following matrices

$$\left[M_{ij}\right]=\sum_{e=1}^{k_i}\left[a_S\right]^{(e)T}\left(\int\limits_{\Gamma_{si}}\left[N\right]^{(e)T}\left[N\right]^{(e)}\,d\gamma\right)\left[a_S\right]^{(e)}.$$

Here $\left[a_G\right]^{(e)}$, $\left[a_S\right]^{(e)}$ are the matrices of geometric interactions of the finite element with the index (e), $\left[D_i\right]^{(e)}$ are the local elasticity matrices of the finite element with the index (e), $\{g_i\}$ are the local surface force vectors, k_i are the number of finite elements into which the body G_i is divided.

Matrices A_{ii} have the dimension $n_i\times n_i$, $i=1,\ldots,N$, n_i is the the number of unknown displacements in the i-th body, $n=n_1+\ldots+n_N$ is the total number of unknown displacements, matrices M_i have the dimension $n_i\times m$, $m=m_1+\ldots+m_q$ is the total unknown Lagrange multipliers, matrices M_{ij} have the dimension $n_i\times m_j$, m_j is the the number of unknown Lagrange multipliers for the j-th contact pair, which includes the i-th body.

5 ALGORITHM FOR SOLVING THE SYSTEM OF LINEAR ALGEBRAIC EQUATIONS

The system of linear algebraic equations (12) with a saddle point is ill-conditioned and has a zero block on the main diagonal, therefore, for its numerical solution we will use modified symmetric successive over-relaxation method (MSSOR) [16,17], generalized to the case of contact of several bodies:

$$\begin{cases}
\frac{1}{\tau}A_{11}(\mathbf{u_1}^{\mathbf{k+\frac{1}{2}}}-\mathbf{u_1^k})+A_{11}\mathbf{u_1^k}+M_1\boldsymbol{\lambda}^{\mathbf{k}}=\mathbf{R_1},\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\frac{1}{\tau}A_{NN}(\mathbf{u_N}^{\mathbf{k+\frac{1}{2}}}-\mathbf{u_N^k})+A_{NN}\mathbf{u_N^k}+M_N\boldsymbol{\lambda}^{\mathbf{k}}=\mathbf{R_N},\\
-\frac{\alpha}{2\tau}B_1(\boldsymbol{\lambda}_{\mathbf{1}}^{\mathbf{k+1}}-\boldsymbol{\lambda}_{\mathbf{1}}^{\mathbf{k}})+\sum_{t=1}^{N}M_{t1}^T\mathbf{u_t}^{\mathbf{k+\frac{1}{2}}}=\mathbf{0},\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
-\frac{\alpha}{2\tau}B_q(\boldsymbol{\lambda}_{\mathbf{q}}^{\mathbf{k+1}}-\boldsymbol{\lambda}_{\mathbf{q}}^{\mathbf{k}})+\sum_{t=1}^{N}M_{tq}^T\mathbf{u_t}^{\mathbf{k+\frac{1}{2}}}=\mathbf{0},\\
\frac{1}{\tau}A_{11}(\mathbf{u_1}^{\mathbf{k+1}}-\mathbf{u_1}^{\mathbf{k+\frac{1}{2}}})+A_{11}\mathbf{u_1}^{\mathbf{k+\frac{1}{2}}}+M_1\boldsymbol{\lambda}^{\mathbf{k+1}}=\mathbf{R_1},\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\frac{1}{\tau}A_{NN}(\mathbf{u_N}^{\mathbf{k+1}}-\mathbf{u_N}^{\mathbf{k+\frac{1}{2}}})+A_{NN}\mathbf{u_N}^{\mathbf{k+\frac{1}{2}}}+M_N\boldsymbol{\lambda}^{\mathbf{k+1}}=\mathbf{R_N},
\end{cases}\qquad(13)$$

where k is an iteration number, α and τ are the iterative parameters.

Before the first iteration, it is necessary to set the initial (zero) value to the vector of Lagrange multipliers $\boldsymbol{\lambda}$ and then calculate the displacement vectors $\mathbf{u}_i$ from the first m equations of system (12).

Using the scheme (13) allows us to reduce the solution of the general ill-conditioned system of equations for all contacting bodies to the sequential solution of three blocks of systems of equations: N systems for calculating $\mathbf{u}_i^{k+\frac{1}{2}}$, q systems for calculating $\boldsymbol{\lambda}_i^{k+1}$ и N systems for calculating $\mathbf{u}_i^{k+1}$. Within each of these blocks, systems of equations can be solved independently of each other, including in parallel. All these systems of equations are solved using the conjugate gradient method. Matrices selected as preconditioners are $B_i = \sum_{j=1}^{N} M_{ji}^T (\mathrm{diag}\{A_{jj}\})^{-1} M_{ji}$, and the values of the iterative parameters are set as follows: $\alpha = 0,05$, $\tau = 0,5$.

6 RESULTS OF THE NUMERICAL SOLUTION

First, we will solve the initial-boundary value problem of the heat equation (1), and we will use the obtained temperature field to solve the contact problem of the elasticity theory (2)-(8). In the pellets, a constant heat release is set, and the temperature in all nodes of the cladding is assumed to be constant (623 K). Constant pressure $p_1 = 10$ MPa is set on the outer surface of the cladding, and constant pressure $p_2 = 50$ MPa is set on the upper surface of the upper pellet. The pellets are made of uranium dioxide, the cladding is made of an alloy of zirconium. The elastic moduli, thermal expansion coefficients, specific heat capacity and thermal conductivity of both materials are temperature dependent, and the Poisson's ratios and density are constant.

We will carry out a series of calculations with a different number of pellets. For the case of five pellets, we present the distribution of radial and axial displacements and stresses at the contact boundary between the pellets and the cladding. We will consider unmatched meshes: the pellets $G_1, \ldots, G_{N-1}$ are divided into 40 elements in the r direction and into 80 elements in the z direction, and the cladding G_N is divided into 10 elements in the r direction and into 400 elements in the z direction.

The considered problem has the following specific features:

- at the contact boundaries between the pellets, almost all finite elements (except for a few elements near the inner surface) exit the contact;
- axial displacements reach significant values: for example, for the case of 100 pellets, the upper pellet displacements relative to its initial position by an amount comparable to the size of several pellets.

Fig. 2-3 show the graphs of the distributions of radial and axial displacements at the boundary between the pellets and the cladding. Fig. 4-5 show the graphs of the distributions of radial and axial stresses at the boundary between the pellets and the cladding. The graphs of radial displacements stresses are visually indistinguishable and coincide everywhere, except for the vicinity of the chamfers of the pellets; no oscillations are observed.

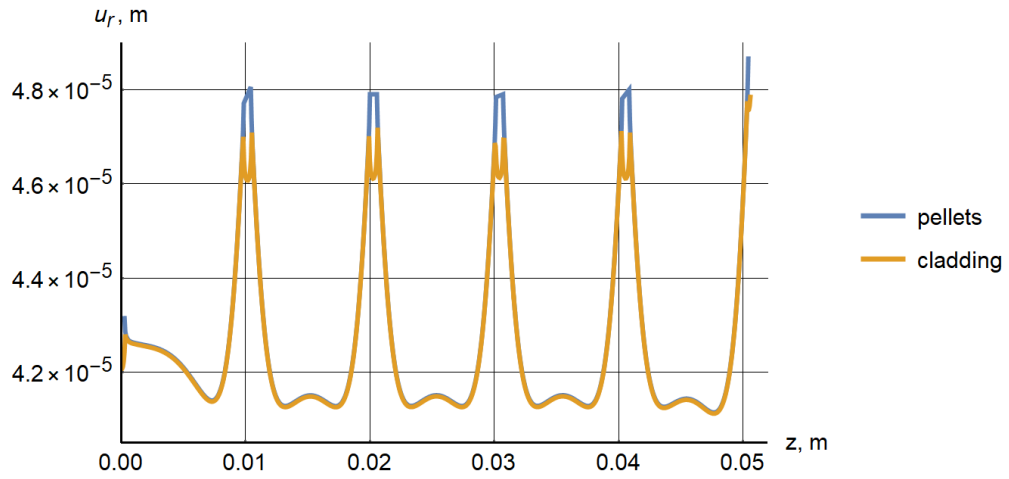


Figure 2. Radial displacements $u_r(z)$ in the elements nodes

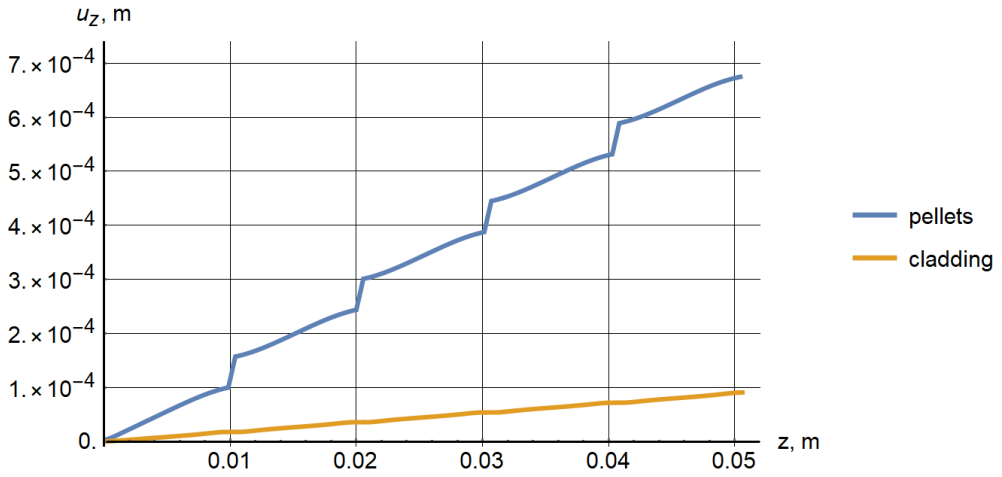


Figure 3. Axial displacements $u_z(z)$ in the elements nodes

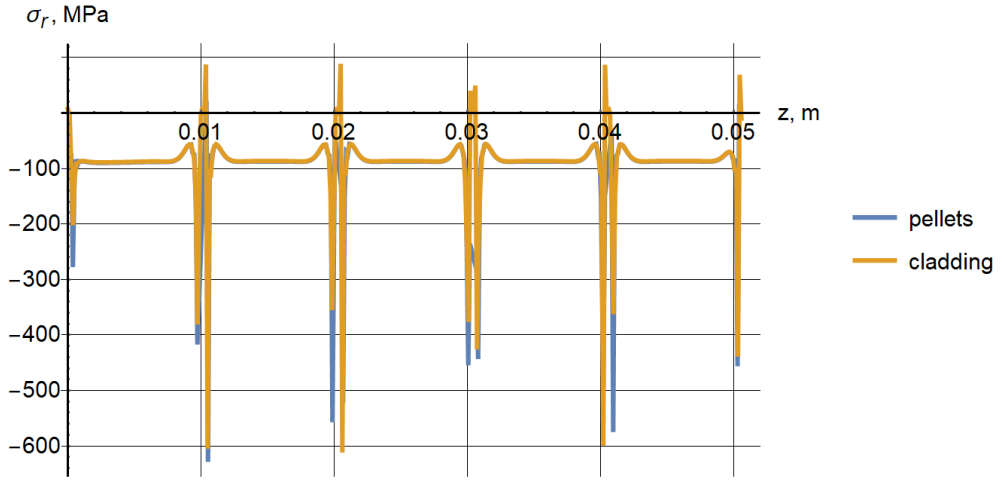


Figure 4. Radial stresses $\sigma_r(z)$ in the elements nodes

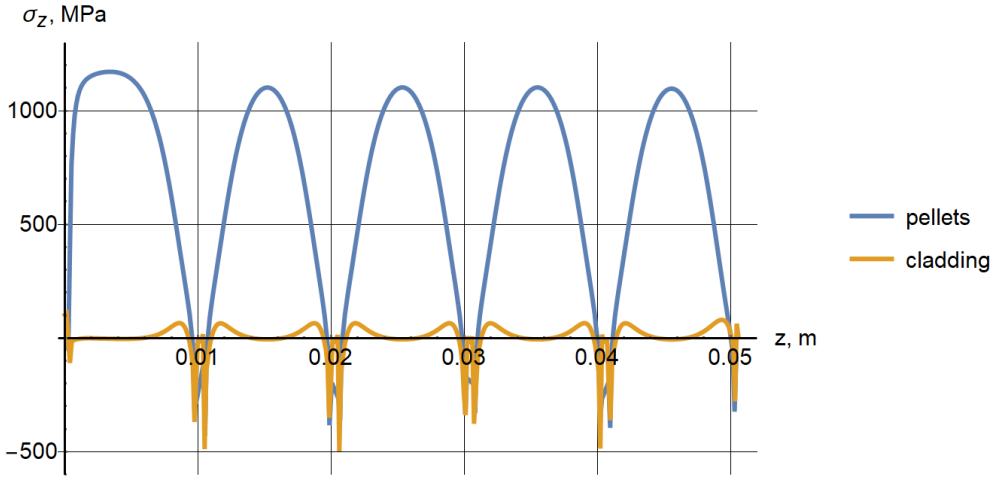
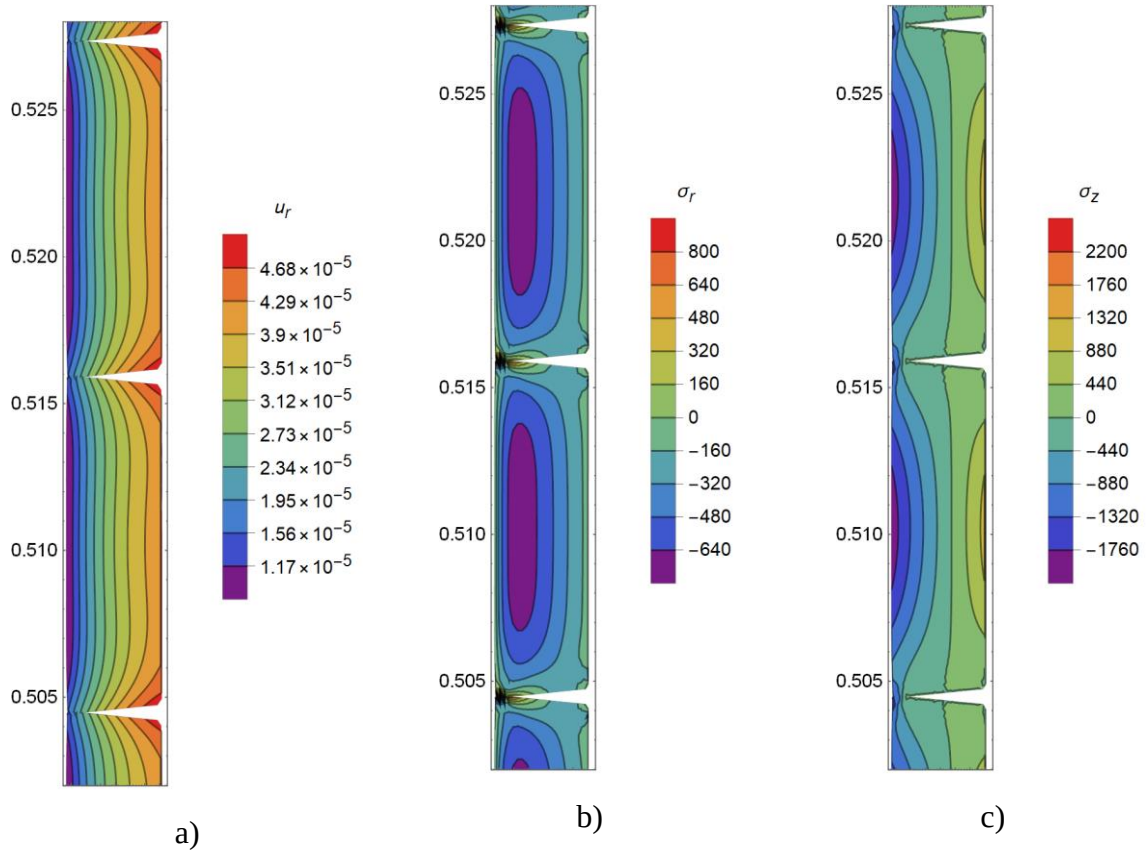


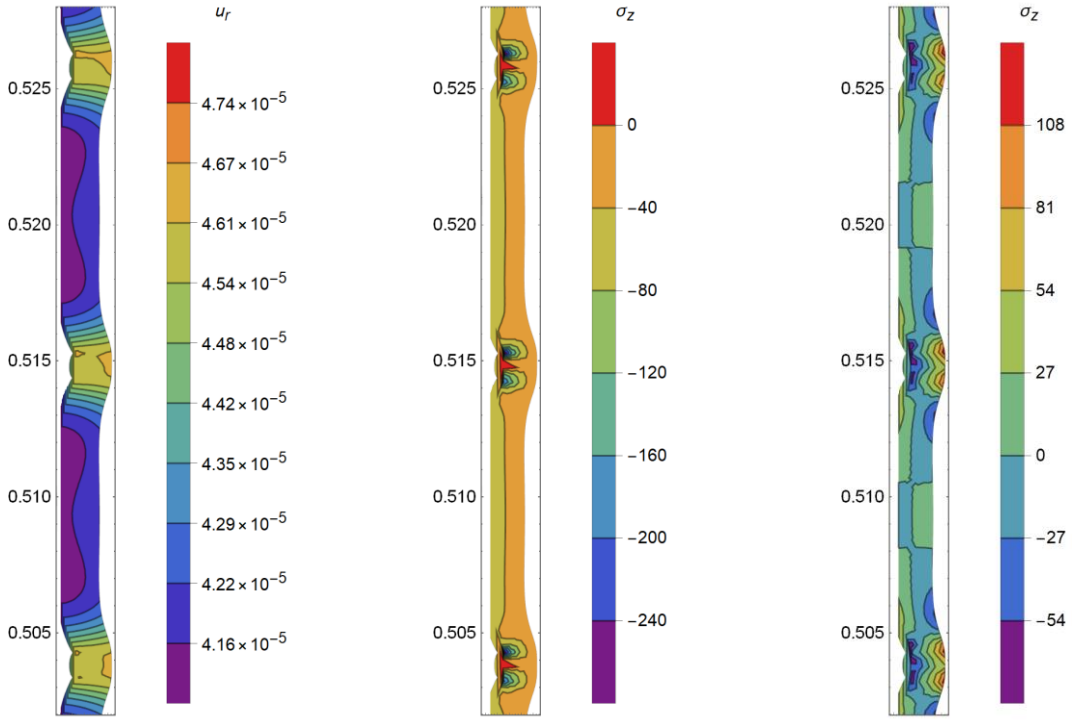
Figure 5. Axial stresses $\sigma_z(z)$ in the elements nodes

For the case of 100 pellets, we present fragments of two-dimensional distributions of displacements and stresses in the middle of a pillar column.



a) b) c)

Figure 6. Two-dimensional distributions in the elements nodes of the 49th and 50th pellets: a) — radial displacements, b) — radial stresses, c) — axial stresses



a) b) c)

Figure 7. Two-dimensional distributions in the elements nodes of the cladding: a) — radial displacements, b) — radial stresses, c) — axial stresses

Fig. 6-7a show two-dimensional distributions of radial displacements, and Fig. 6-7b and 6-7c show two-dimensional distributions of radial and axial stresses for the case of 100 pellets. Fragments of the distributions corresponding to the 49th and 50th pellets are shown, and when constructing deformed bodies, the applied displacements were increased by 10 times for pellets and 50 times for the cladding for clarity.

We will carry out a series of calculations with three different unmatched meshes:

- Mesh 1: the pellets are divided into 20 elements in the r direction and into 40 elements in the z direction, the cladding is divided into 5 elements in the r direction,
- Mesh 2: the pellets are divided into 40 elements in the r direction and into 80 elements in the z direction, the cladding is divided into 10 elements in the r direction,
- Mesh 3: the pellets are divided into 80 elements in the r direction and into 160 elements in the z direction, the cladding is divided into 20 elements in the r direction.

In the z direction the cladding is divided into the required number of elements in proportion to the number of pellets. For example, for five pellets, this will be 200, 400, and 800 elements for the three variants of the mesh. Fig. 8-9 show the graphs of the distribution of axial stress at the contact boundary between the 3rd pellet and the cladding. It is seen that when using a coarser mesh, oscillations arise near the chamfers, with an increase in the number of elements, the amplitude of the oscillations decreases and on the smallest mesh they disappear.

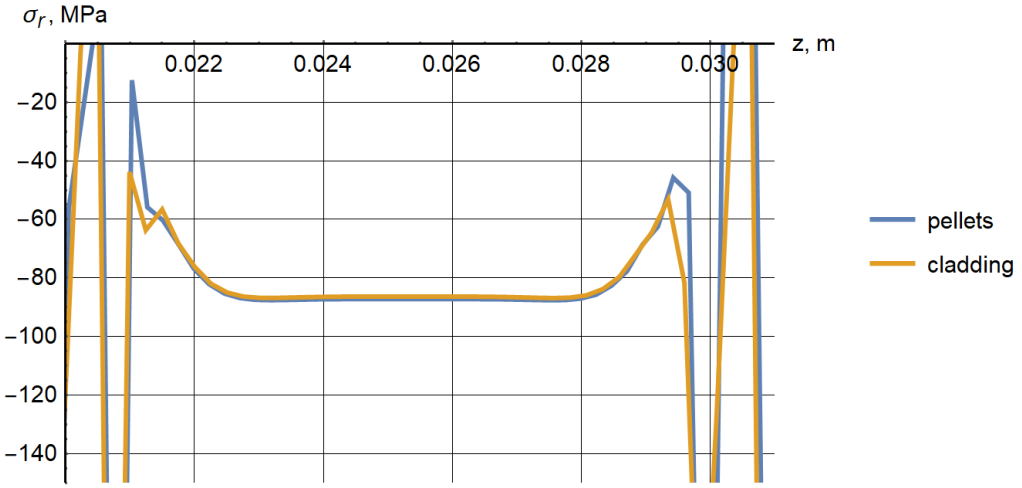


Figure 8. Radial stresses $\sigma_r(z)$ in the elements nodes of the 3rd pellet (mesh 1)

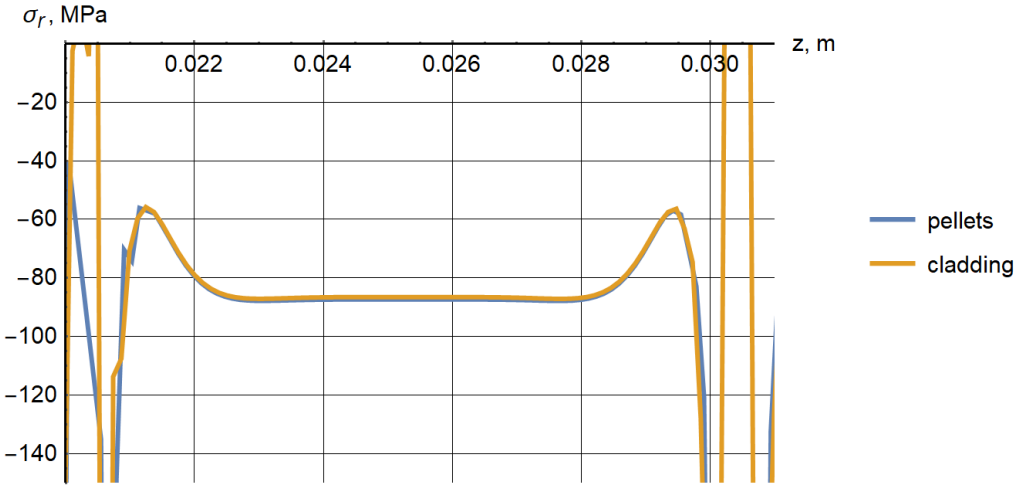


Figure 9. Radial stresses $\sigma_r(z)$ in the elements nodes of the 3rd pellet (mesh 2)

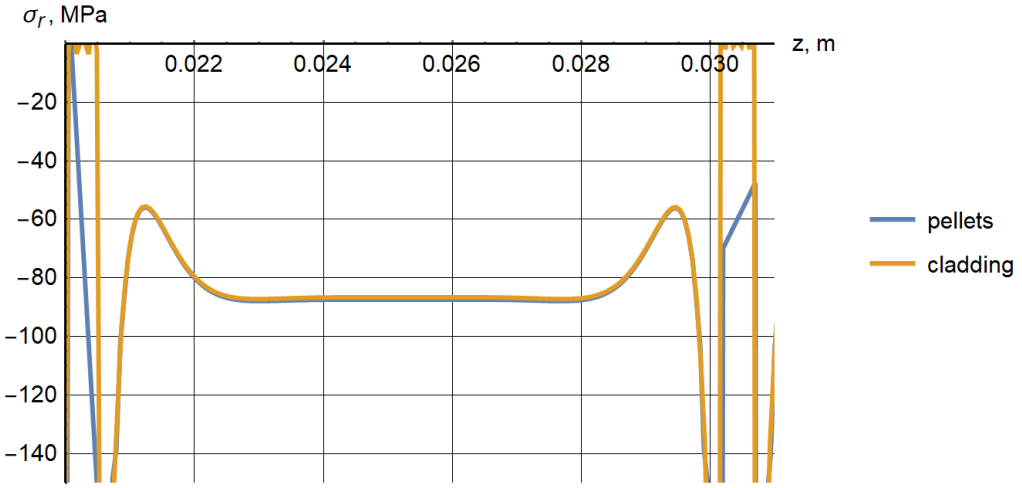


Figure 10. Radial stresses $\sigma_r(z)$ in the elements nodes of the 3rd pellet (mesh 3)

Let us compare the number of iterations necessary to achieve the given accuracy of solving the system of equations (12), with a different number of pellets. In this case, the relative accuracy is calculated as follows:

$$\varepsilon = \sqrt{\frac{\sum_i S_i \dfrac{(\hat{u}_{r_i} - u_{r_i})^2 + (\hat{u}_{z_i} - u_{z_i})^2}{u_{r_i}^2 + u_{z_i}^2}}{\sum_i S_i}},$$

where $\hat{u}_{r_i}$, $\hat{u}_{z_i}$ are the values of radial and axial displacements at a new iteration, S_i is the sum of the areas of the finite elements, which includes the considered node, divided by the number of nodes in the finite element.

ε	Mesh 1	Mesh 2	Mesh 3
10^{-2}	5	5	6
10^{-3}	6	7	8
10^{-4}	12	15	21
10^{-5}	29	38	56

Table 1. The number of iterations required to achieve accuracy of ε (5 pellets)

ε	Mesh 1	Mesh 2	Mesh 3
10^{-2}	8	10	12
10^{-3}	14	18	22
10^{-4}	30	39	52
10^{-5}	64	80	94

Table 2. The number of iterations required to achieve accuracy of ε (25 pellets)

ε	Mesh 1	Mesh 2	Mesh 3
10^{-2}	16	19	24
10^{-3}	25	31	40
10^{-4}	51	66	90
$5 \cdot 10^{-5}$	85	102	148

Table 3. The number of iterations required to achieve accuracy of ε (100 pellets)

Tables 1-3 show that with increasing accuracy, the number of bodies and decreasing mesh step, the number of iterations increases. For 100 pellets, maximum accuracy of $\varepsilon = 5 \cdot 10^{-5}$ is achieved, there is no convergence for the accuracy of $\varepsilon = 10^{-5}$.

Let us determine some empirical patterns of growth in the number of iterations depending on the number of pellets and nodes of the finite element model. With an increase in the number of nodes, the number of iterations increases in proportion to n^y, where $y \in (0,225; 0.275)$. The above range was obtained for calculations with a different number of bodies (5, 10, 25, 50, 100 pellets). It is known that when using the conjugate gradient method without preconditioning for

matrices obtained by discretizing the Laplace operator, the number of iterations increases in proportion to $\sqrt{n}$ [18]. If unit matrices are used as preconditioners B_i, then there is no convergence. The selected preconditioner allows faster convergence compared to the conjugate gradient method for conventional matrices.

As the number of pellets increases, the number of iterations increases in proportion to N^x, where $x \in (0,37; 0,41)$. This range was obtained for calculations with three different meshes (Fig. 11). The number of iterations for the calculation with five pellets is somewhat inconsistent with the proposed regularity, therefore, $N = 10$ was taken as a starting point for constructing the dependence graph of $CN^{0,4}$. In real fuel elements, the number of pellets reaches several hundred; therefore, the proposed algorithm allows us to obtain a numerical solution with sufficient accuracy for a moderate number of iterations.

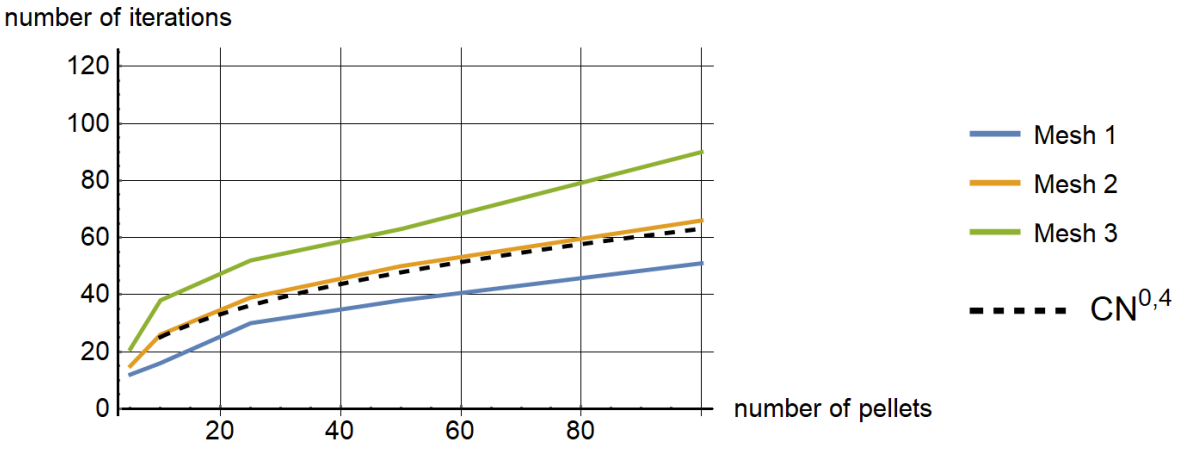


Figure 11. Dependence of the number of iterations on the number of pellets for $\varepsilon = 10^{-4}$

7 CONCLUSIONS

The statement of the problem of contact interaction of a system of axisymmetric thermoelastic bodies under thermomechanical loading is presented. The results of the numerical implementation of the algorithm for solving the problem using the mortar method are presented on the example of a demonstration task that simulates some processes in a fuel element. A generalization of the algorithm for solving a system of linear algebraic equations arising as a result of discretization of the problem by the finite element method, in the case of contact of several bodies. A comparison is made of the number of iterations required to achieve a given accuracy of solving the system of equations, depending on the number of pellets and mesh steps. It is shown that the proposed method for the numerical solution of an ill-conditioned system of equations allows convergence to be achieved with sufficient accuracy for a moderate number of iterations. It is also demonstrated that with an increase in the number of elements, the amplitude of the axial stress oscillations arising near the chamfers of the pellets decreases.

Acknowledgements: The authors with deep gratitude and appreciation remember untimely departed D.Sc. in Engineering Stankevich Igor Vasilyevich.

REFERENCES

[1] M.P. Galanin, A.V. Krupkin, V.I. Kuzneczov, V.V. Lukin, V.V. Novikov, A.S. Rodin, I.V. Stankevich, "Matematicheskoe modelirovanie termouprugoplasticheskogo kontaktnogo vzaimodejstviya sistemy tel", *Mathematica Montisnigri*, **30**, 99-114 (2014).

[2] I.V. Stankevich, M.E. Yakovlev, Si Tu Xtet, "Razrabotka algoritma kontaktnogo vzaimodejstviya na osnove alterniruyushhego metoda Shvarcza", *Vestnik MGTU im. N.E. Baumana. Ser. Estestvennye nauki. Specz. vyp. Prikladnaya matematika*, 134-141 (2011).

[3] I. Babuska, "The finite element method with penalty", *Mathematics of Computation*, 27, 221-228 (1973).

[4] M.V. Mikhaylyuk, E.V. Strashnov, "Simulation of articulated rigid bodies with penalty method", *Mathematica Montisnigri*, **27**, 91-106 (2013).

[5] P. Le Tallec, T. Sassi, "Domain decomposition with nonmatching grids: augmented Lagrangian approach", *Mathematics of Computation*, 64, 1367-1396 (1995).

[6] M.P. Galanin, P.V. Gliznucina, V.V. Lukin, A.S. Rodin, *Varianty realizacii metoda mnozhitelej Lagranzha dlya resheniya dvumernyx kontaktnyx zadach*, Preprint IPM, No. 89, (Moscow: KIAM), (2015). URL: http://library.keldysh.ru/preprint.asp?id=2015-89.

[7] P. Wriggers, *Computational Contact Mechanics*, Speinger-Verlag (2006).

[8] B.P. Lamichhane, *Higher Order Mortar Finite Elements with Dual Largange Multiplier Spaces and Applications*, Universitat Stuttgart (2006).

[9] E.I. Kraus, I.I. Shabalin, "The tool for high-velocity interaction and damage of solids", *Mathematica Montisnigri*, **39**, 18-29 (2017).

[10] V.S. Zarubin, G.N. Kuvyrkin, *Matematicheskie modeli mexaniki i elektrodinamiki sploshnoj sredy*, Izd-vo MGTU im. N.E. Baumana (2008).

[11] L.A. Rozin, *Variacionnye postanovki zadach dlya uprugix sistem*, Izd-vo Leningradskogo un-ta (1978).

[12] O. Zenkevich, K. Morgan, *Konechnye elementy i approksimaciya*. Mir (1986).

[13] B.I. Wohlmuth, "A mortar finite element method using dual spaces for the Lagrange multiplier", *SIAM Journal on Numerical Analysis*, **38** (3), 989-1012 (2000).

[14] P.S. Aronov, A.S. Rodin, *Matematicheskoe modelirovanie kontaktnogo vzaimodejstviya dvux uprugix tel s krivolinejnymi graniczami na nesoglasovannyx setkax*, Preprint IPM, No. 87, (Moscow: KIAM), (2019). doi:10.20948/prepr-2019-87.

[15] I.V. Stankevich, P.S. Aronov, "Matematicheskoe modelirovanie kontaktnogo vzaimodejstviya dvux uprugix tel s pomoshhyu mortar-metoda", Matematika i matematicheskoe modelirovanie, 3, 26-44 (2018).

[16] Yu.V. Bychenkov, E.V. Chizhonkov, *Iteracionnye metody resheniya sedlovyx zadach*, BINOM (2010).

[17] P.S. Aronov, M.P. Galanin, A.S. Rodin, I.V. Stankevich, "Reshenie zadachi kontakta dvux uprugix tel mortar-metodom i metodom Shvarcza na nesoglasovannyx setkax", *Tavricheskij vestnik informatiki i matematiki*, 1, 24-42, (2019).

[18] Dzh. Demmel, *Vychislitelnaya linejnaya algebra. Teoriya i prilozheniya*, Mir (2001).

Received May 10, 2020

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

MODELING OF TRANSDUCER CALIBRATION FOR PRESSURE MEASUREMENT IN NANOSECOND LASER ABLATION

A.A. SAMOKHIN*, P.A. PIVOVAROV, A.L. GALKIN

Prokhorov General Physics Institute of the Russian Academy of Sciences
119991, Moscow, Vavilov str., 38
*Corresponding author. E-mail: asam40@mail.ru

DOI:10.20948/mathmontis-2020-48-6

Summary. Pressure transducer calibration is modeled in the regime where calculated thermo-acoustic signal and experimental information about normal boiling temperature attainment due to nanosecond laser pulses irradiation are used. It is shown that the regime demonstrated recently for 30 ns laser action is not straightforwardly applicable for shorter pulses due to strong thermo-acoustic signal enhancement compared with vaporization signal near normal boiling point. For subnanosecond laser irradiation two pulses action is suggested where shorter and longer pulses are used simultaneously.

1. INTRODUCTION.

In [1], a method for calibrating a pressure transducer using a change in the thermoacoustic signal due to the appearance of evaporative pressure as a marker for reaching the target normal boiling point was proposed and implemented. Additional information about the actual temperature of the irradiated target with its known thermophysical parameters allows us to accurately calculate the value of thermoacoustic pressure and thereby calibrate the acoustic recording scheme used.

The fact of simultaneous recording of thermoacoustic and evaporative signals has been recorded in irradiated dielectrics, metals and semiconductors for a long time [2-5]. However, the use of this fact for the calibration of piezoelectric sensors has not been reported until recently, although piezoelectric sensors have been used in the diagnosis of pulsed exposure for more than half a century [6].

A 7 mm thick piezoelectric sensor operating in the current source mode was calibrated in [1] under the influence of a laser pulse (25 ns, 1.06 μm) on the surface of a metal target made of liquid mercury, which was in contact with the sensor through a 3 mm thick glass layer. The value of thermoacoustic pressure at the time the boiling point was reached was comparable to atmospheric pressure, which facilitated its simultaneous observation with the evaporative signal.

Since the value of the thermoacoustic signal at a fixed pulse intensity grows inversely with its duration, the question arises of how much this calibration technique can be used for shorter pulses when, against the background of increased thermoacoustic pressure, the evaporation signal in the temperature range close to normal boiling becomes relatively less noticeable.

In this paper, we analyze the possibility of extending this method to shorter laser pulses.

2. THERMOACOUSTIC PRESSURE SIGNAL IN LINEAR APPROXIMATION.

The one-dimensional hydrodynamic problem for calculating the thermoacoustic pressure pulse when the absorbing target is heated by a laser pulse is described by the following system of equations:

$$\frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial z} + \rho\frac{\partial u}{\partial z} = 0 \qquad (1)$$

2010 Mathematics Subject Classification: 78A60, 80A20, 81T80.
Key words and Phrases: Laser Ablation, Thermoacoustic Signal, Pressure Sensor Calibration, Mercury.

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial z} + \frac{1}{\rho}\frac{\partial P}{\partial z} = 0 \quad (2)$$

$$\rho C_p\left(\frac{\partial T}{\partial t} + u\,\frac{\partial T}{\partial z}\right) = \frac{\partial}{\partial z}\,\kappa\frac{\partial T}{\partial z} + Q \quad (3)$$

$$Q(z,t) = \alpha I(t) e^{-\alpha z} \quad (4)$$

$$I(t) = I_m e^{-b\left(\frac{t}{\tau_p}-2.5\right)^2},\ b = 2.772 \quad (5)$$

Here ρ, u, T, P - density, speed, temperature, pressure are space-time dependent functions (z, t). C_p, κ - heat capacity and thermal conductivity (known temperature functions). Q, α, I_m, τ_p- source power density, absorption coefficient, peak laser intensity and duration of a laser pulse of a Gaussian shape.

The heat equation (3) should be supplemented with initial and boundary conditions:

$$T(z,0) = T_o, \quad \kappa\frac{\partial T}{\partial z}(0,t) = 0, \quad T(L,t) = T_o \quad (6)$$

The right-hand boundary condition (6) should be set at a distance L, where the temperature change can be considered negligible in comparison with the quantities of interest to us, which thereby virtually cease to depend on L.

Let us now consider various approximate approaches to the solutions of these equations, supplemented by the equation of state $\rho(T, P)$, which are used to determine pressure pulses excited in absorbing condensed media under the action of laser radiation.

If we do not take into account the dependence of density on pressure P, which is true for fairly slow processes, then from (2) in the linear approximation we obtain the well-known expression for thermoacoustic pressure in terms of time derivatives of surface temperature and absorbed intensity [3]:

$$P = -\rho\int_0^x \frac{\partial u_{lin}}{\partial t}dx_1 \approx \frac{1}{C_p}\,\frac{\partial\rho}{\rho\partial T}\left[\kappa\,\frac{\partial T}{\partial t} + \frac{i}{\alpha}\right] \quad (7)$$

It is seen from (7) that, in the case of a fixed intensity maximum, the pressure increases inversely with the duration of the laser pulse. For small values of the absorption coefficient in (7), the second term related to the volume character of the absorption is dominant, and in the opposite limiting case, the main role is played by the mechanism of diffusion thermal heating of the target. Such an explicit mechanism separation is possible only in the linear approximation (7).

These expressions are valid at distances exceeding the total heating length, but small in comparison with the length of the corresponding sound wave. Taking into account the dependence of density on pressure in a linear approximation for pressure, we obtain the well-known wave equation with a source:

$$\frac{1}{c^2}\frac{\partial^2 P}{\partial t^2} - \frac{\partial^2 P}{\partial z^2} = \frac{\partial\rho}{\partial T}\frac{\partial^2 T}{\partial t^2} \quad (9)$$

Formula (7) is obtained from (9) if the term with the square of the speed of sound in the denominator is neglected on the left side and the remaining term is double integrated over the coordinate of the heating source on the right side.

In the general case, consideration of the nonlinear response requires solving the complete system (1) - (3). When using the incompressible fluid approximation, an approach is possible

when the system of continuity and thermal conductivity equations is considered first, and then its solution is used to determine the pressure from the Euler equation (2).

To numerically solve the heat equation, which is necessary when determining the thermoacoustic pressure (7), you can use finite-difference schemes or the Green's function method with subsequent calculation of the corresponding integrals.

$$T(z,t) = \mathrm{T}_0 + \int_0^t \int_0^\infty G(z, z_1, t-\tau)\, q_V(z,t)\mathrm{d}z_1\mathrm{d}\tau,$$

$$G(z,z_1,t) = \frac{1}{\sqrt{4\pi\chi t}}\left\{e^{-\frac{(z-z_1)^2}{4\chi t}} + e^{-\frac{(z+z_1)^2}{4\chi t}}\right\},\ \chi = \frac{\kappa}{\rho C_p},\quad q_V(z,t) = \frac{Q(z,t)}{\rho C_p} \qquad (10)$$

$$\kappa T(z,t) = \int_{-\infty}^{t} \sqrt{\frac{\chi}{\pi(t-\tau)}}\, q_s(\tau) exp\left\{-\frac{z^2}{4\chi(t-\tau)}\right\} d\tau,\quad \chi = \frac{\kappa}{\rho C_p},\quad q_s(t) = I(t) \qquad (11)$$

where χ is the thermal diffusivity.

Formulas (10) and (11) give expressions for the behavior of the temperature profile in the case of bulk and surface absorption, respectively. In the latter case, the formula for surface temperature takes a simple form:

$$\kappa T(0,t) = \int_{-\infty}^{t} \sqrt{\frac{\chi}{\pi(t-\tau)}}\, q(\tau) d\tau \qquad (12)$$

The results and analysis of specific calculations of the behavior of *P(t)* are given in the next section for mercury using the known values of its thermophysical parameters [7–9]:

***T*, K**	**C_p, 10^{-3}, J/(g·K)**	**κ, 10^{-2}, W/(cm·K)**	**χ, 10^{-2}, cm^2/s**	**КТР, 10^{-6}, 1/K**	***P* vap.satur., bar**	**ρ, g/cm^3**
300	139	8	4.45	181	$3.68\cdot10^{-6}$	13.521
350	138	8.5	4.78	182	$1.23\cdot10^{-4}$	13.4
400	137	9.3	5.02	182	$1.61\cdot10^{-3}$	13.279
450	137	9.8	5.4	183	0.01178	13.16
500	137	10.4	6	183	0.05758	13.04
550	137	11.5	6.48	185	0.2993	12.921
600	137	12.4	6.84	185	0.6146	12,8
650	137	13	7.12	186	1.521	12.68
700	137	13.5	7.41	186	33	12.56
750	137	13.8	7.67	185	64.5	12.436
800	136	14.2	7.95	185	11.58	12.31

Table 1: Thermophysical parameters of mercury.

3 RESULTS AND DISCUSSION.

3.1 The dependence of the thermoacoustic signal on the pulse duration.

First, we consider the change in the thermoacoustic pressure (7) with a decrease in the pulse duration in the target heating from the initial temperature $T_o = 0.45T_b$ to a predetermined value corresponding to the normal boiling point of mercury T_b = 630 K. Because of this additional condition the pressure amplitude increases faster than the reciprocal of the pulse duration, since a shortening of the laser pulse requires an additional increase in its intensity. The considered problem for this reason becomes effectively nonlinear. A similar note also applies to the dependence of the pressure amplitude on the magnitude of the absorption coefficient.

As thermophysical constants in calculating the temperature and pressure, we will use three sets of values from Table 1 corresponding to the target temperatures T / T_b = 0.45, 1, and 1.1.

$$T_o = 0.45T_b, \qquad \rho = 13.59\frac{g}{cm^3}, \qquad \rho C_p = 1.865\frac{J}{cm^3K},$$
$$\frac{\partial\rho}{\rho\partial T} = 1.815\cdot 10^{-4}K^{-1}, \qquad \chi = 0.032\frac{cm^2}{s} \tag{13}$$

$$T_o = T_b, \qquad \rho = 12.70\frac{g}{cm^3}, \qquad \rho C_p = 1.74\frac{J}{cm^3K},$$
$$\frac{\partial\rho}{\rho\partial T} = 1.856\cdot 10^{-4}K^{-1}, \qquad \chi = 0.076\ \frac{cm^2}{s} \tag{14}$$

$$T_o = 1.11T_b, \qquad \rho = 12.52\frac{g}{cm^3}, \qquad \rho C_p = 1.716\frac{J}{cm^3K},$$
$$\frac{\partial\rho}{\rho\partial T} = 1.86\cdot 10^{-4}K^{-1}, \qquad \chi = 0.082\ \frac{cm^2}{s} \tag{15}$$

. Figures 1 and 2 show the behavior of the temperature of the irradiated surface and the thermoacoustic pressure signal in the linear approximation of heating from the initial temperature of $0.45T_b$ to T_b at various durations of the absorbed laser pulse with a decrease from 30 to 0.1 ns. The thermophysical parameters indicated in (13) - (15) were used at T_0 / T_b = 0.45 and 1 and two absorption coefficients of 10^5 cm^{-1} and 10^6 cm^{-1}.

From a comparison of Figures 1 and 2, it can be seen that the transition from the use of data (13) to data (14) practically does not affect the shape of the pressure response, with the exception of the ratio between the amplitudes of the positive and negative pressure waves, the values of which approach unity with decreasing pulse duration and absorption coefficient. This behavior is associated with a relative change in the length of the thermal effect due to thermal conductivity and the radiation absorption length .

However, the amplitude values of pressure and absorbed intensity change more significantly with a change in the pulse duration, as is also shown in Table 2, where these values are given for a set of constants (13) and (14) for two absorption coefficients of 10^5 cm^{-1} and 10^6 cm^{-1}.

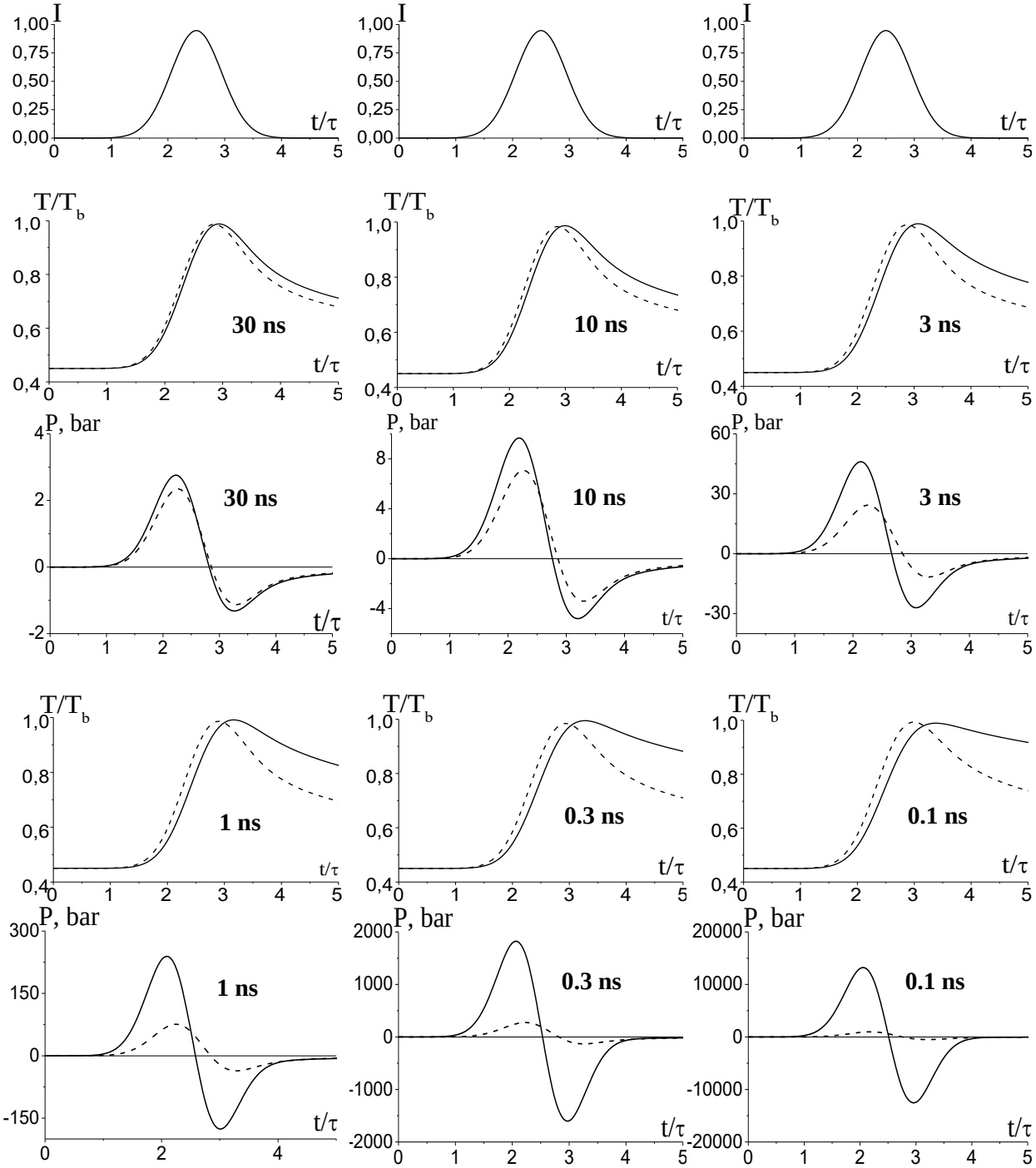


Fig. 1 Temporal profile of intensity *(I)*, relative temperature *(T / T_b)* and pressure *(P)* for different pulse durations (they are shown in the graphs) and $\alpha = 10^5$ cm^{-1} (solid lines), $\alpha = 10^6$ cm^{-1} (dashed lines). The values of maximum intensities for two absorbtion cooficients and difeerent pulse duration are: 30 ns - I_m=1.2 MW/cm^2 and I_m=1.04 MW/cm^2; 10 ns - I_m=2.3 MW/cm^2 and I_m=1.8 MW/cm^2; 3 ns - I_m=5 MW/cm^2 and I_m=3.4 MW/cm^2; 1 ns - I_m=11 MW/cm^2 and I_m=6 MW/cm^2; 0.3 ns - I_m=29 MW/cm^2 and I_m=12 MW/cm^2; 0.1 ns - I_m=74 MW/cm^2 and I_m=23.3 MW/cm^2.

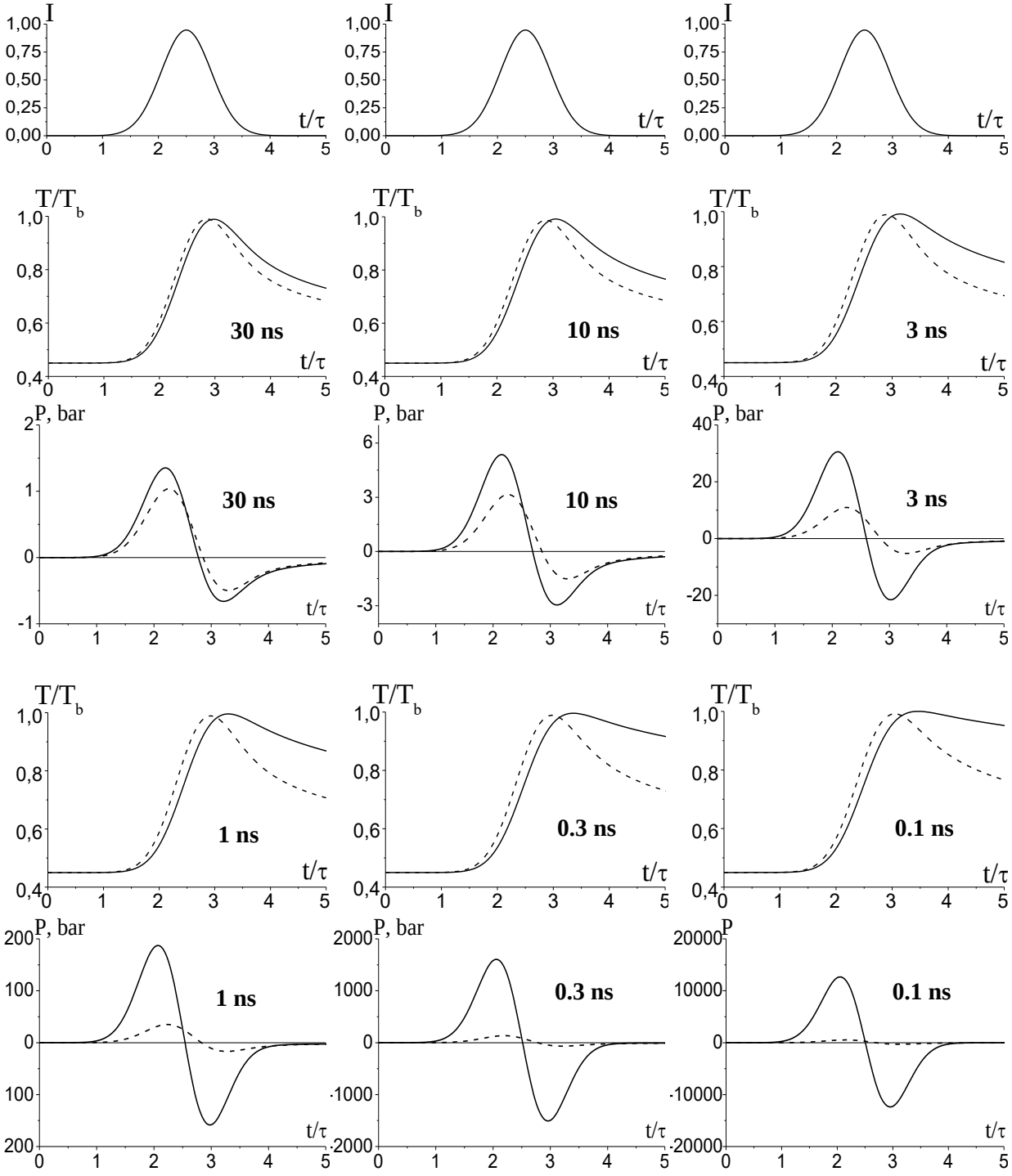


Fig. 2 Temporal profile of intensity *(I)*, relative temperature *(T / T_b)* and pressure *(P)* for different pulse durations (they are shown in the graphs) and $\alpha = 10^5$ cm^{-1} (solid lines), $\alpha = 10^6$ cm^{-1} (dashed lines). The values of maximum intensities for two absorbtion cooficients and difeerent pulse duration are: 30 ns - I_m=1 MW/cm^2 and I_m=0.7MW/cm^2; 10 ns - I_m=1.8 MW/cm^2 and I_m=1.3 MW/cm^2; 3 ns - I_m=4.2 MW/cm^2 and I_m=2.4 MW/cm^2; 1 ns - I_m=10 MW/cm^2 and I_m=4.5 MW/cm^2; 0.3 ns - I_m=28 MW/cm^2 and I_m=12 MW/cm^2; 0.1 ns - I_m=74 MW/cm^2 and I_m=18 MW/cm^2.

	$\alpha = 10^5$ cm^{-1}				$\alpha = 10^6$ cm^{-1}			
$\tau_{p.}$, ns	I_{m1}, MW/cm^2	P_{m1}, bar	I_{m2}, MW/cm^2	P_{m2}, bar	I_{m1}, MW/cm^2	P_{m1}, bar	I_{m2}, MW/cm^2	P_{m2}, bar
0,1	74	140000	74	140000	18.2	600	23.3	1100
0,3	27.5	1750	29.2	2000	9	150	11.9	300
1	10	210	11.2	260	4.45	38	6.1	80
3	4.24	33	5.07	50	2.43	12.5	3.4	27
10	1.82	5.75	2.3	10	1.28	3.5	1.81	7.6
30	0.9	1.45	1.2	3	0.73	1.15	1.04	2.6

Table 2: The values of the maximum pressure and the corresponding radiation intensity depending on the duration of the absorbed laser pulse for two values of the absorption coefficient and heating to T_b. from initial temperature $T_o = 0.45T_b$.

The ratio of the pressure amplitudes calculated for different sets of constants (13), (14) at various values of the absorption coefficient is shown in Fig. 3 (two upper curves) depending on the duration of the laser pulse. The solid curve ($\alpha = 10^5$ cm^{-1}) tends to unity for short pulses where the heat conduction effect is smaller than at $\alpha = 10^6$ cm^{-1}.

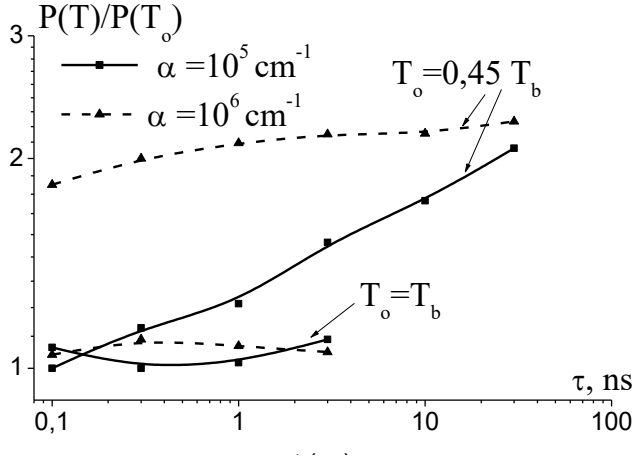


Fig. 3 Ratio of maximum pressures at various thermophysical parameters for two values of the absorption coefficient depending on the pulse duration.

The pressure in Fig. 1 and Fig. 2 for 30 ns is approximately consistent with the calculation result in [1]. It also follows from the figures and Table 2 that the amplitude of the thermoacoustic pressure signal grows much faster than $(\tau_p)^{-1}$ when τ_p diminishes. This difference, clearly demonstrated in Fig. 3, as mentioned above, is associated with an additional condition for reaching a given temperature $T_o = T_b$ for each pulse duration.

Using simple estimates, it is easy to find that in this case, with the volume or surface nature of the absorption, the amplitude of the linear response with decreasing pulse duration will increase as $(\tau_p)^{-2}$ or $(\tau_p)^{-3/2}$. With a comparable effect from these two heating mechanisms, the exponent will lie in the range 1.5 - 2, as can be seen from Fig. 4.

Due to such a rapid increase, the amplitude of the thermoacoustic signal exceeds the atmospheric pressure by more than an order of magnitude even with a duration of 3 ns. Such an excess, which increases with a further decrease in the pulse duration, greatly complicates

the direct application of the method used in [1] for calibrating a piezoelectric sensor in the case of short laser pulses.

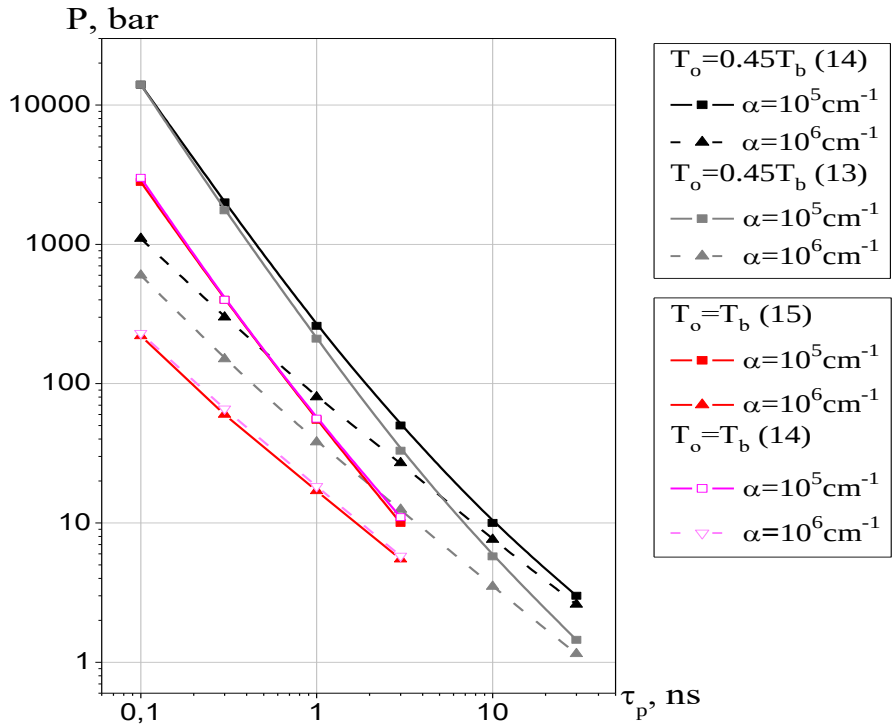


Fig. 4 Change in pressure maximum depending on the pulse duration for two sets of thermophysical constants (13)-(14) (grey and black) with initial temperature $T_o = 0.45T_b$ and (14)-(15) (violet and red) with initial temperature $T_o = T_b$. Triangles and squares correspond to two values of the absorption coefficient $\alpha = 10^6$ cm^{-1} and $\alpha = 10^5$ cm^{-1}, respectively.

3.2. Modification of the approach for short pulses.

A decrease in pressure for short pulses can be achieved by preheating the target to a temperature close to the value of T_b with an additional longer laser pulse. The two-pulse technique, when the first pulse transfers the irradiated system to the desired state, and the second carries out its monitoring is widely used in various situations (see, for example, [10]). In our case, the short pulse should be combined with the moment of reaching the temperature maximum from the first long pulse, when the thermoacoustic signal generated by it is close to its passage through zero.

The action of a short laser pulse should provide heating of the target from T_b to values at which the saturated vapor pressure exceeds 3 bar. For mercury, in accordance with Table 1 (saturation pressure versus temperature), this gives T value about $1.1\ T_b$ The ablation regime then approaches evaporation into vacuum with a recoil pressure of $P_r \cong 1$ bar. However, a consistent quantitative analysis of the problem of the unsteady process of formation of a pulse of evaporative pressure under similar conditions, when the saturated vapor pressure as a result exceeds the external pressure by several times, as far as we know, has not yet been carried out. Note that in the conditions of interest to us, the evaporation and thermoacoustic signals are simply summed, since the cooling effect of evaporation on the temperature profile is small.

Figure 5 shows the temporal dependences of the thermoacoustic pressure upon heating the target from $T_0 = T_b$ to $T = 1.1\ T_b$ with laser pulses of various durations (3, 1, 0.3, and 0.1 ns), and Table 3 shows the maximum achievable values of thermoacoustic pressure under such conditions.

It can be seen that the magnitude of the thermoacoustic signal is much smaller than in Fig. 1 and Table 2; however, at 0.1 ns its amplitude nevertheless considerably exceeds 1 bar even for the absorption coefficient 10^6 cm $^{-1}$. The total thermoacoustic signal when exposed to two pulses with durations of 30 ns and 1 ns is shown in Fig. 6.

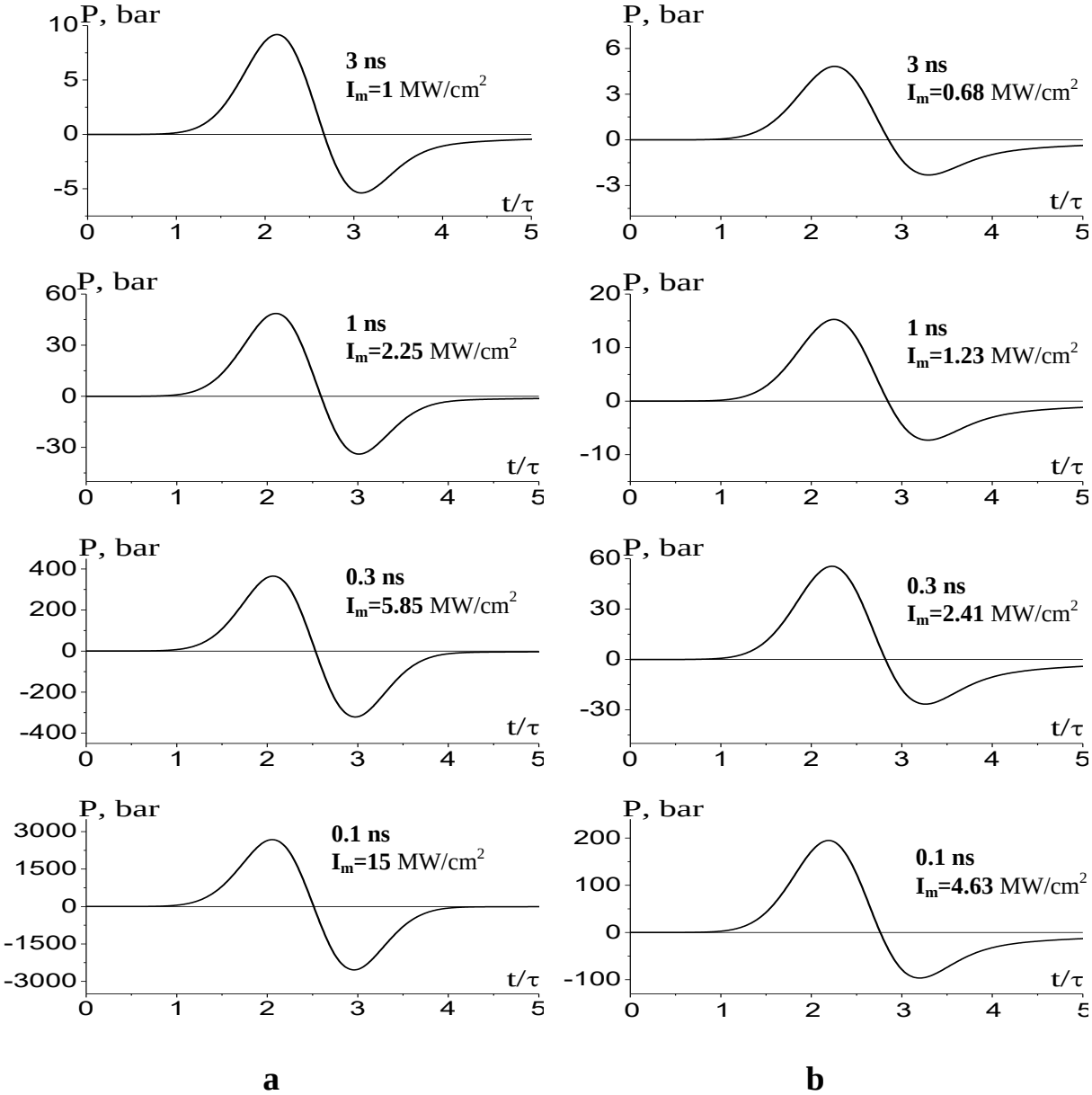


Fig.5. Temporal profile of thermoacoustic pressure depending on the laser pulse duration and $\alpha = 10^5$ cm^{-1} (**a**), $\alpha = 10^6$ cm^{-1} (**b**) upon heating of the target from $T = T_b$ to $T = 1.11\ T_b$ with the corresponding intensities.

	$\alpha = 10^5$ cm^{-1}				$\alpha = 10^6$ cm^{-1}			
τ, нс	I_{m1}, MW/cm^2	P_{m1}, bar	I_{m2}, MW/cm^2	P_{m2}, bar	I_{m1}, MW/cm^2	P_{m1}, bar	I_{m2}, MW/cm^2	P_{m2}, bar
0.1	15	2800	15	3000	4.63	220	4.73	230
0.3	5.85	400	5.87	400	2.41	60	2.47	66
1	2.25	55	2,27	56	1.23	17	1.25	18.3
3	1.01	10	1.03	11	0.675	5.5	0.7	5.8

Table 3: The values of the maximum pressure and the corresponding radiation intensity depending on the duration of the absorbed laser pulse for two values of the absorption coefficient and heating up to $1.11T_b$, from the initial temperature $T_o = T_b$.

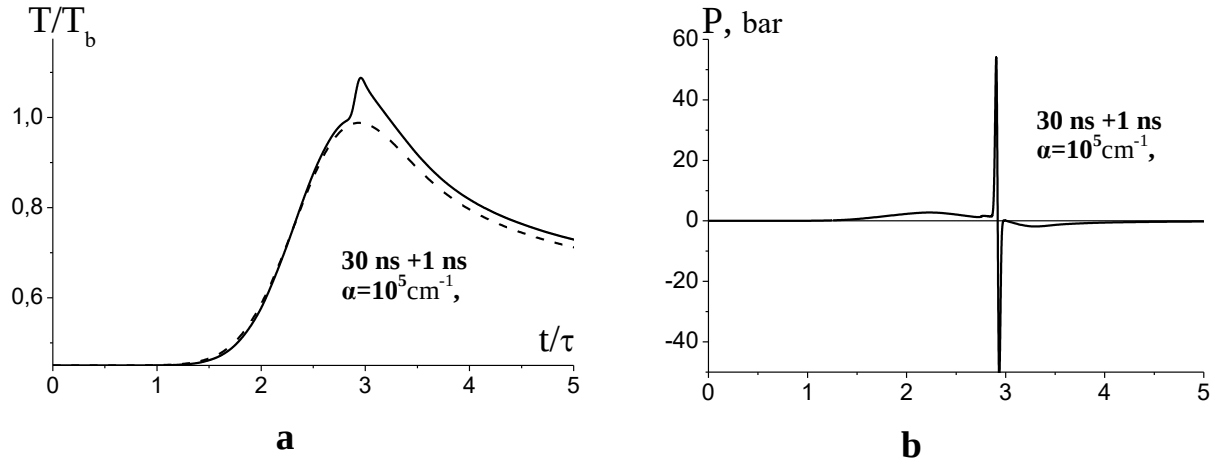


Fig. 6. The temporal profile of temperature (a) and pressure (b) for a laser pulse with a duration of 30 ns I_{m1} = 1.2 MW/cm^2 followed by heating of the target from $T = T_b$. up to $T = 1.11T_b$. 1 ns pulse with I_{m2}=3.2 MW/cm^2; $\alpha = 10^5$ cm^{-1}.

Note that for $\alpha = 10^6$ cm^{-1} and a pulse duration of 0.3 ns, the maximum of the acoustic pressure is only slightly higher than the maximum for the duration of 1 ns, shown in Fig. 6 This means that calibration in the case under consideration is possible for the subnanosecond mode if the absorption coefficient is large enough.

The question of choosing the wavelength of laser radiation to ensure such a value of α is beyond the scope of this paper. At the same time, for mercury at higher temperatures approaching the critical point, there appears to be a sharp decrease in α, characteristic of the metal-insulator transition (see, for example, [11-14] and referencies therein).

3.3. Applicability of the approximations used.

The use of the equation of state in this work, which takes into account the dependence of the density of the liquid only on temperature, suggests, in particular, the slowness of the processes under consideration on a scale specified by the speed of sound c_s. This means that the convective velocity υ associated with thermal expansion should be small compared to c_s >> υ and the characteristic spatial size l of the processes considered over time t should be small relative to the mean free path of sound over the same time. This inequality holds in all of the exposure modes discussed.

The effect of the calculated pressure P on the density change can be approximately estimated using a simple relation that takes into account the proportionality of this change to the ratio P/c_s^2. This contribution is significant only for the shortest pulses during direct target heating from the initial room temperature to the boiling point. However, these cases are not of interest for calibration precisely because of the large magnitude of the resulting thermoacoustic pressure.

The applicability of the linear approximation is evidenced by a comparison of the results obtained when two sets of constants corresponding to the initial and final values of the target temperature are used in the calculations. The results are shown in Fig. 3 and in Tables 2 and 3. For the heating interval $T = (0.45 - 1)\, T_b$, the difference between the sets (13) and (14) is not negligible due to a noticeable variation in thermal diffusivity in this range. However, this difference does not play a significant role, if we take into account the use of a two-pulse technique. At the same time, for short pulses with a heating interval $T = (1 - 1.1)\, T_b$, the applicability of the linear approximation is quite satisfactory, as can be seen from Fig. 3 (two lower curves), table 3 and Fig. 4 where in the case $T_o = T_b$ there is no visible difference between the curves with the sets (14) and (15) for the same α value.

4. CONCLUSION.

Thus, in the present work, as a result of numerical simulation of various modes of the method of calibrating pressure sensors, originally proposed and implemented [1] for a laser pulse duration of 30 ns, it is shown that the direct extension of this method to shorter durations is difficult. This is due to the rapid increase in thermoacoustic pressure with decreasing pulse duration, which turns out to be sharper than the inverse dependence characteristic of a linear response to a given laser intensity, due to the fixation of the temperature range in the calibration mode

A natural way to overcome this difficulty is to use a double-pulse heating of the target, when the first long pulse brings the temperature of the target closer to the boiling point, and the second, shorter one, provides the appearance of an evaporative signal with a reduced (compared to single-pulse mode) value of the thermoacoustic signal. This approach allows us to extend the calibration method implemented in [1] to subnanosecond pulses, thereby expanding its range of applicability by almost two orders of magnitude in reducing the laser pulse duration.

REFERENCES.

[1] A.A.Samokhin , E.V.Shashkov, N.S., Vorobiev, A.E Zubko., “Nanosecond Calibration of a Piezo Transducer by Comparing Thermoacoustic and Vaporization Pressure Signals at Pulsed Laser Irradiation of a Metal Target”, *Physics of Wave Phenomena*, **27** (4), 268–270 (2019)

[2] M.W. Sigrist and F.K. Kneubuhl, ”Laser-generated stress waves in liquids”, *Journal Acoustical Society of America*, **64**(6), 1652-1663 (1978).

[3] A.A.Samokhin “First-order phase transitions induced by laser radiation in absorbing condensed matter”, *Proceedings of the Institute of General Physics Academy of Science of the USSR*, Commack, New York, **13**, 1-161, (1990).

[4] I.A. Veselovskii, B.M. Zhiryakov, N.I. Popov, A.A. Samokhin, “”The photoacoustic effect andphase transitions in semiconductors and metals irradiated by laser pulses”, *Proceeding of the Institute of General Physics Academy of Sciences of the USSR*, **13**, 179-198 (1990).

[5] A.E. Zubko, A.A. Samokhin, “Modeling of thermoacoustic and evaporation pressure signals in absorbing liquids irradiated with nanosecond laser pulses”, *Mathematica Montesnigri*, **36,**78-85 (2016).
[6] A.E.Graham, F.W.Neilson, W.B.Benedick, “Piezoelectric current from schock-loaded quartz – a submicrosecond stress gauge”, *J. Appl. Phys*., **36** (5), 1775-1783 (1965)
[7] V.E. Zinov'ev. *Teplofizitcheskie svoistva metallov pri visokich temperaturach*, Moskva: Metallurgiya (1989).
[8] A.P. Babichev, N.A. Babushkina, A.M. Bratkovskiy, *Fizitcheskie velitchini*, Moskva: Energoatomizdat (1991).
[9] A.I. Volkov, I.M. Jarskiy, *Bol'shoy chimitcheskiy spravotchnik*, Moskva: Sovetskaya shkola (2005).
[10] A.A. Samokhin, S.M. Klimentov, P.A. Pivovarov, "Acoustic diagnostics of the explosive boiling up of a transparent liquid on an absorbing substrate induced by two nanosecond laser pulses", *Quantum Electron*., **37** (10), 967–970 (2007).
[11] S.N. Andreev, V.I. Mazhukin, N.M. Nikiforova, A.A. Samokhin, “On possible manifestations of the induced transparency during laser evaporation of metals”, *Quantum Electronics*, 33, 771–776 (2003).
[12] A.A. Samokhin, A.E. Zubko, “On metal-dielectric transition in laser ablation modeling”, *Mathematica Montesnigri*, **46**, 114-122 (2019).
[13] Yu Zhang, Daixian Zhang, Jianjun Wu, Zhen He, and Xiong Deng, “A thermal model for nanosecond pulsed laser ablation of aluminum”, *AIP Advances*, **7**, 075010 (2017)
[14] A. A..Samokhin, E.V..Shashkov, N. S..Vorobiev, A.E..Zubko, "On acoustical registration of irradiated surface displacement during nanosecond laser-metal interaction and metal–nonmetal transition effect", *Appl. Surf. Scienc*e, **502**, 144261 (2020).

Received May 10, 2020

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

ATOMISTIC MODELING OF CRYSTAL-MELT INTERFACE MOBILITY OF FCC (AL, CU) AND BCC (FE) METALS IN STRONG SUPERHEATING/UNDERCOOLING STATES

V.I. MAZHUKIN*, A.V. SHAPRANOV, O.N. KOROLEVA

Keldysh Institute of Applied Mathematics, Russian Academy of Science
*Corresponding author. E-mail: vim@modhef.ru

DOI: 10.20948/mathmontis-2020-48-7

Summary. A detailed study of the mobility and kinetic properties of solid - liquid interfaces (SLI) with different types of crystal lattices (fcc - Al, Cu) and (bcc - Fe) metals in a wide range of temperatures and pressures was carried out using atomistic modeling. The ranges of maximum permissible values of superheated/undercooled states for each metal have been determined. The ultimate goal of the study was to determine the temperature dependences of the stationary front velocity $v_{s\ell}(\Delta T)$ describing the SLI mobility in each of the metals in an analytical form. The analytical dependence $v_{s\ell}(\Delta T)$ was constructed by comparing the results of atomistic modeling in the area of maximum permissible superheating/undercooling values with the data of the main kinetic models of Wilson - Frenkel (WF) and Broughton, Gilmer and Jackson (BGJ). An acceptable agreement was achieved by introducing appropriate correction parameters into the kinetic models using the least squares method. The influence of the crystallographic orientation of metals and external pressure on the SLI mobility is investigated.

1 INTRODUCTION

The development and evolution of nanostructures of materials [1] in many cases is determined by the motion of interphase interfaces, and therefore it becomes urgent to determine the main equilibrium and nonequilibrium properties of interfaces.

Equilibrium properties of interfaces are largely related to grain boundaries, which are interfaces between differently oriented crystals of the same material [2]. The importance of studying the structure and thermodynamics of heterophase interfaces of solid materials [3] is determined by the role they play in the stability and morphology of materials. In particular, in situations where the primary solidification process is fully completed, the solid phase, as a rule, can pass into new phases and secondary microstructures [4].

The non-equilibrium properties of mobile solid-liquid interfaces (SLI) are manifested in fast melting and solidification processes, where they play an important role in the establishment of various structural and kinetic properties, the morphology of the growth of a new phase, and non-equilibrium transfer of matter across the phase boundaries [4], [5] .

Initially, the greatest efforts in the atomistic research were devoted to the crystal-melt interfaces in one-component systems. These studies included a detailed analysis of the structure at the interface [6], direct calculations of the interfacial free energies [7], kinetic coefficients, [8,9] and associated crystalline anisotropies [10], as well as other structural and thermodynamic parameters of the solid-liquid interface [11]. At that, the main attention was paid to the determination of the most important property of the mobility of the interphase

2010 Mathematics Subject Classification: 74A20, 74A50, 74A25.
Key words and Phrases: Kinetic models, Kinetic properties, Solid - liquid interface, Superheated/undercooled states, Analytical dependence of front velocity.

boundary – the speed of heterogeneous phase transformations $\upsilon_{s\ell} = \upsilon_{s\ell}(\Delta T)$, which depends on the value of superheating/undercooling $\Delta T = T_{s\ell} - T_m$ of the interphase boundary $\Gamma(t)$.

According to thermodynamic theory, if the constituent parts of a thermodynamic system are not in equilibrium with each other, then thermodynamic flows arise through their interfaces, accompanied by the process of transformation of matter from one state of aggregation to another one (phase transition). Assuming that the processes occurring in the system are quasi-static, and the fluxes are infinitely small, one can use the methods of equilibrium thermodynamics to describe such a nonequilibrium system. In this case, an infinitesimal difference in the thermodynamic parameters in different parts of the system is assumed. The driving force of first-order phase transitions is the difference in the free energy of two phases at the interface $\Gamma(t)$. Using the theory of thermodynamic potentials, it can be shown that the difference in free energy is linearly proportional to undercooling (superheating) [8, 12]:

$$\Delta G = \Delta S \cdot \Delta T = \frac{L\Delta T}{T_{eq}}$$

where ΔS, ΔT are the change of the entropy and temperature, L, T_{eq} are the specific heat and equilibrium phase transition temperature.

Identifying the difference in free energy ΔG as the speed of the phase transformation υ, one obtains the expression for the speed of the transformation, which in the thermodynamic approach at constant pressure and small deviations from the equilibrium is proportional to superheating/undercooling: ΔT: $\upsilon \approx \mu \Delta T$, where for the case of the solid-liquid interface $\Gamma_{s\ell}(t)$, μ is the proportionality constant between the normal boundary speed and its superheating/undercooling.

The use of the equilibrium theory of thermodynamic potentials to describe phase transformations (nonequilibrium processes) makes it possible to take into account only a small entry into the metastable superheated/undercooled region and to study phase transformations near the equilibrium line, where the temperature dependence of $\upsilon_{s\ell}(\Delta T)$ is mainly controlled by the difference of the free energy of the crystal and liquid phase. This determines the main disadvantages of the thermodynamic approach. Since thermodynamics does not take into account the internal structure of the bodies under consideration, a number of its conclusions and provisions do not have sufficient accuracy and physical clarity.

Structural particles of matter in continuous motion are displayed in the main statements of the molecular kinetic theory, in which all processes are considered at the atomic or molecular level, and the particles have a Maxwellian velocity distribution.

The simplest kinetic dependence in the form of the linear growth rate of a crystal from a melt, which outwardly agrees well with the above thermodynamic relationship, was obtained on the basis of classical molecular kinetic models. In this dependence, the proportionality constant is called the kinetic coefficient [14, 15]. The main application of the kinetic coefficient and its various modifications [16, 17] was found in the description of melting - solidification processes in the vicinity of the equilibrium melting temperature $T_{s\ell} \approx T_m$, where the coefficient μ is the main parameter characterizing the mobility of the SLI. In the kinetic approach in the most general form, the coefficient μ is written as [9, 18]:

$$\upsilon_{\langle hk\ell \rangle}(\Delta T) = \mu_{\langle hk\ell \rangle} \Delta T \qquad (1)$$

where $\langle hk\ell \rangle$ are the indices characterizing the anisotropy of the kinetic coefficient and indicating the dependence of the SLI rate on the crystallographic orientation of the interface. The anisotropy of the kinetic coefficient and its value are the most important factors affecting the growth rate and the associated morphology of solidification at small undercooling.

The need to consider an extended range of superheated/undercooled states was caused by a large number of theoretical [19 - 24] and experimental [25-27] works that showed that the kinetics of melting/solidification processes far from the equilibrium melting temperature T_m is significantly different from the kinetics in the vicinity of T_m.

In this work, using atomistic modeling, a detailed study of the mobility and kinetic properties of SLI with different lattices (fcc - Al, Cu) and (bcc - Fe) and crystallographic orientations for metals is carried out in a wide range of temperatures and pressures. The ultimate goal of the study is to construct the temperature dependences of the stationary front velocity describing the SLI mobility in metals in a wide range of maximum permissible values of superheating/undercooling. The analytical dependences of $v_{s\ell}(\Delta T)$ were determined by comparing the results of the atomistic modeling in the area of maximum permissible values of superheating/undercooling with the data of the kinetic models. An acceptable match is achieved by introducing the corresponding correction parameters into the model [28], [29].

2 MAIN KINETIC MODELS

Traditionally, the kinetic theories have developed most intensively in the direction of the processes of crystallization - solidification of melts, which is explained by the large content in them (along with the fundamental aspects) of a large number of technological applications, in which crystallization of metastable phases in deeply supercooled melts leads to significant structural changes associated with thermodynamic, kinetic and mechanical properties of materials.

Kinetic theories depicting temperature dependences $v_{s\ell}(\Delta T)$ of SLI mobility are based on various physical phenomena. The most well-known kinetic theories include the classical Wilson - Frenkel (WF) theory [30-32] with a diffusion mechanism for controlling the interface kinetics, the kinetic model of Broughton, Gilmer and Jackson (BGJ) [33] with a collisional thermal mechanism, and the kinetic theory [7,34], with the mechanism of influence of density changes, the so-called density functional theory (DFT).

In the kinetic theory of WF, [30-32], the interface speed is related to the diffusion of atoms in the liquid phase. This theory is often called the transition state theory, since it is based on the assumption that melting or solidification occurs through some intermediate or transition state. In this theory, the SLI rate is controlled by a diffusion limiting mechanism. This mechanism is based on the assumption that atoms (molecules) must overcome the diffusion barrier during the transition from liquid to solid phase. The transition is accompanied by a significant restructuring of the interface. In this case, the rate of the crystallization process is assumed to be proportional to the diffusion coefficient, which is usually presented in the form of the Arrhenius equation

$$D = D_0 \exp\left(-\frac{Q}{k_B T_{s\ell}} \right) \tag{2}$$

where Q is the energy of activation for diffusive motion in liquid, k_B is the Boltzmann constant, k_BT is the average thermal energy for a single atom, D_0 is the prefactor determining the speed of the process.

Temperature dependence of the crystallization / melting front velocity $\upsilon_{s\ell}(\Delta T)$ in the model with diffusive limitation is expressed in the generalized form as:

$$\left[\upsilon_{s\ell}(\Delta T)\right]^{\mathrm{WF}}_{\langle hk\ell\rangle} = \frac{a\,f}{\lambda^2} D\left[\exp\left(\frac{\Delta G}{k_B T_{s\ell}}\right)-1\right] = C^{\mathrm{WF}}_{\langle hk\ell\rangle}\frac{D}{a}\left[\exp\left(\frac{L_m}{k_B T_m}\frac{\Delta T}{T_{s\ell}}\right)-1\right] \qquad (3)$$

where $C^{\mathrm{WF}}_{\langle hk\ell\rangle} = \dfrac{a^2}{\lambda^2} f$, a is the interatomic distance, λ is the man free path for the atoms of this process, it is assumed to be proportional to the lattice constant, a: $\lambda < a$, f is the efficiency coefficient (a constant of the order of unity, $f < 1$), characterizing the fraction of collisions of liquid atoms with solid, which leads to crystallization. These quantities do not have a rigorous definition, are difficult to measure, and, moreover, depend on the crystallographic orientation of the interface [18, 35].

The BGJ theory [33], originally proposed as an improvement on the earlier WF theory, uses the frequency of thermal collisions of atoms with the interface [36] as a constraint. The modification of the WF transition state theory was motivated by the results of MD simulation [33] performed with the Lennard-Jones potential, which was assumed to be metal-like, which showed that the growth of crystals of monatomic systems may not in all cases be limited by diffusion. In particular, in the region of very low temperatures, the diffusion coefficient tends to zero, but according to the simulation results, the SLI rate is still finite and the WF model turns out to be wrong. Following the hypothesis [36] that the solidification of monatomic metals is limited only by the frequency of collisions of melt atoms with the crystal surface, the authors of the BGJ model [33] replaced the diffusion term in (3) with the average thermal velocity of atoms $\upsilon_T = \sqrt{3k_B T_{s\ell}/m}$

$$\left[\upsilon_{s\ell}(\Delta T)\right]^{\mathrm{BGJ}}_{\langle hk\ell\rangle} = \frac{a}{\lambda} f_0 \upsilon_T\left[\exp\left(\frac{L_m}{k_B T_m}\frac{\Delta T}{T_{s\ell}}\right)-1\right] = C^{\mathrm{BGJ}}_{\langle hk\ell\rangle}\sqrt{\frac{3k_B T_{s\ell}}{m}}\left[\exp\left(\frac{L_m}{k_B T_m}\frac{\Delta T}{T_{s\ell}}\right)-1\right] \qquad (4)$$

where $C^{BGJ}_{(hk\ell)} = \dfrac{a}{\lambda} f_0$ is a dimensionless coefficient, m is the atom mass.

In the kinetic theory developed on the basis of DFT [7, 34], the interface kinetics is controlled by the relaxation of short-wave density waves. Density functional theory also explicitly explains the anisotropy of the coefficient $\mu_{\langle hk\ell\rangle}$, confirming that it is determined only by $\langle hk\ell\rangle$ factors.

Within the framework of the kinetic approach it is easy to formulate an analytical form of the kinetic coefficient $\mu_{\langle hk\ell\rangle}$ (1). For the temperatures in the vicinity of the equilibrium melting temperature T_m, ($T_{s\ell} \approx T_m$) from the Eqs. (3) и (4) one can obtain the coefficient $\mu_{\langle hk\ell\rangle}$ in the form:

$$\mu^{\mathrm{WF}}_{\langle hk\ell\rangle} = C^{\mathrm{WF}}_{\langle hk\ell\rangle}\frac{D}{a}\frac{L_m}{k_B T_m^2} \qquad (5)$$

$$\mu_{\langle hk\ell\rangle}^{BGJ} = C_{\langle hk\ell\rangle}^{\mathrm{BGJ}} \sqrt{\frac{3k_B T_m}{m}} \frac{L_m}{k_B T_m^2} \tag{6}$$

It is easy to see that the equations (5), (6) represent the first term in the expansion of equations (3), (4) in the neighborhood of T_m. The dimensionless coefficients $C_{\langle hk\ell\rangle}^{\mathrm{WF}}$, $C_{\langle hk\ell\rangle}^{BGJ}$ do not have a strict definition and depend on the crystallographic orientation of the interface [6].

The presence of the values without strict definition Q, D_0, λ and dimensionless coefficients f, $C_{\langle hk\ell\rangle}^{\mathrm{WF}}$, $C_{\langle hk\ell\rangle}^{BGJ}$) the models (3) - (6) required further investigation of the kinetics of the phase transformations. It was performed using atomistic modeling. Modeling based on molecular dynamic methods using both pair Lennard-Jones [37] and many-particles EAM [4, 5, 23,35,38 - 41] potentials for metallic interfaces with different crystallographic orientation ⟨100⟩, ⟨110⟩ and ⟨111⟩ allowed to find that both theories (models WF (3) и BGJ (4)) near the equilibrium temperature T_m give the results that are in reasonable agreement between themselves and the results of theory (DFT) [7, 34].

With the development of more accurate many-particle potentials for metals, a number of molecular dynamics studies of crystal growth in pure metals were carried out. Atomistic modeling [23, 27] showed that in the range of values close to the melting point, the crystallization process can be represented with acceptable accuracy by kinetic models with diffusion (3) and collisional thermal constraints (4), as well as by models of the density functional theory [7, 34]. For deep undercooling, it is preferable [23, 27] to use the kinetic model with collisional thermal constraint [33] and the Arrhenius-type model [35]. In the region of intermediate undercooling at a level of ~ $0.7T_m$, the advantage is retained by the model of the transition state with diffusion limitation of WF. In the paper [42], a specially developed semi-empirical potential was used to simulate the phase transformation in a disordered one-component system. The modeling showed that the WF theory satisfactorily describes the results of MD simulation of interface migration in the temperature range from $0.55T_m$ to T_m, while the BGJ theory is less accurate in describing the temperature dependence of the SLI speed in the same temperature range. Below 0.55 T_m, none of the existing theories is able to reproduce the temperature dependence of the interface speed.

Recent use of ultrashort fs, ps - laser pulses to achieve deep undercooling in melts of thin (10–50 nm) metal films [23–27] has brought to the fore the study of bulk and surface mechanisms of melting of solids [43, 44]. The degree of undercooling during solidification can be easily controlled by varying the thickness of the thin films. The problem of determining the mobility of the crystal-melt interface during melting / solidification of metals [49] as a function of temperature $\upsilon_{s\ell}(\Delta T)$ in the entire range of deep superheating/undercooling became urgent especially with the subsequent transition to the problems of crystal growth from deeply undercooled melts [45, 46], as well as kinetic glass transition of undercooled liquids [47, 48].

Since in the overwhelming majority of works [5, 9-11, 18, 23, 27, 35, 42] the study of the temperature dependence of the stationary speed $\upsilon(T_{s\ell})$ was carried out in the temperature range of crystallization with deep undercooling, the important question of the possibility of using the analyzed kinetic models remained open: is it possible to determine the speed of movement of a SLI with acceptable agreement in the temperature range of melting with strong superheating of the solid phase? In the literature there is a small number of works [28,

29, 49-54], in which the kinetic models (3), (4) were tested in a wide temperature range, covering the processes of crystallization and melting.

3 STATEMENT OF THE PROBLEM AND COMPUTATIONAL ALGORITHM

The determination of the stationary temperature dependence of the kinetic rate of SLI in the range of maximum permissible values of superheating/undercooling was carried out using a computational experiment consisting of a large series of molecular dynamics calculations.

Metals of two types were involved in the modeling: with fcc lattice - aluminum (Al), copper (Cu); and bcc lattice - iron (Fe) with different crystallographic orientations. Atomistic models are based on the model concept of a polyatomic molecular system in which all atoms are represented by material points, the motion of which is described in the classical case by Newton's equations. As a result, the evolution of an ensemble of N point particles is described by a system of $2N$ ordinary differential equations. The interaction between particles was described by various many-particle EAM potentials: for aluminum [38], copper [39], and for iron [40] with parameterization [41]. The integration of this system of equations for all N particles requires the knowledge of the coordinates and velocities $(\vec{r}_i, \vec{\upsilon}_i)\big|_{t=0}$ at the initial time $t = 0$. For Al (and in a similar way for other metals), the computational domain was set in the form of a parallelepiped with a size of 5×5×41 nm and filled with 57 600 particles. All the atoms of the parallelepiped were formed in the form of a set of 20×10×10 lattice unit cells. The periodic boundary conditions were set in all three spatial directions at the boundaries of the computational domain, i.e. the simulated object was an infinite single crystal of metal.

The initial state of the computational domain for modeling of the process of heterogeneous melting of a metal is a solid-state structure with a liquid layer in the middle of the computational domain, in which the crystalline and liquid phases are separated by two flat interfaces. To study the melting process, the liquid phase occupies about 18% of the volume of the computational domain, and to study the crystallization process ~ 80%. Subsequently, the interface speed was measured directly as a function of its temperature.

In the course of calculations using a thermostat, a fixed temperature value was established and maintained throughout the entire computational domain during the entire numerical experiment. At the same time, the barostat kept a constant value of the external pressure. This excluded the reverse effect of the release / absorption of the latent heat of fusion on the local temperature at the fronts. As a result, the process of heterogeneous melting / crystallization quickly reaches a stationary regime, and the change in the amount of a new phase occurs almost linearly. The position of the melting-crystallization fronts was tracked using the order parameter.

The maximum permissible values of overheating/overcooling are understood as the temperatures at which, in the event of overheating, the initial crystal still retains mechanical stability, the loss of which is associated with the onset of homogeneous melting. For stationary action conditions, the value of the limiting superheating is $Ts\ell \approx 1.25T_m$ which is in good agreement with the results of [43]. Under unsteady action, the limiting value of overheating reached the value of $1.5T_m$ or more, which coincides with the estimates [55].

In the case of undercooling, the limiting temperature is the temperature at which the undercooled melt is still pure liquid. The limitation of deep penetration into the metastable undercooled region is associated with the formation of an intermediate (interstitial) phase, for

which the order parameter turns out to be much larger than that of a liquid, but much less than that of a normal crystal. The appearance of the interstitial phase indicates the beginning of the vitrification process. The glass transition temperature for most metals is in the region $T_c \approx (0.3 \div 0.5)T_m$ [35]. The variants of calculations in which a noticeable proportion of the interstitial phase appeared were excluded from consideration.

Taking into account the above estimates in the calculations, the temperature range of the superheating/undercooling limit values was chosen within $T_{s\ell} \approx (0.3 \div 1.25)T_m$. For the considered metals Cu, and Fe, the influence of the crystallographic orientation of the interface in the plane ⟨100⟩ was considered and for Al - ⟨100⟩, ⟨111⟩. The influence of external pressure for two values, 0 and 80 kbar, was also considered for each of the metals.

4 THE RESULTS OF THE ATOMISTIC MODELING. THE CONSTRUCTION OF AN ANALYTICAL DEPENDENCE $v_{\langle hk\ell\rangle}(\Delta T)$

The result of the molecular dynamics simulation was obtaining a discrete set of values of the phase front velocity depending on the crystallographic orientation of the interface: for Al: $v_{\langle 100\rangle}(\Delta T)$, for Cu and Fe: $v_{\langle 100\rangle}(\Delta T)$ in the range of extreme values of superheating/undercooling. Discrete values $v_{\langle hk\ell\rangle}(\Delta T)$ in Fig. 1, 2, 3, 4 are marked with markers (circles and triangles – P = 80 kbar).

The analytical dependencies were constructed using the least squares method.

The discrete set of values $v_{\langle hk\ell\rangle}(T_{s\ell})$ obtained from the molecular dynamics modeling was compared with the results of the kinetic models of BGJ [33], (equation (4)) and WF [30-32], (equation (3)). When comparing, it was taken into account that equations (3), (4) contain two thermophysical parameters - the equilibrium melting temperature T_m and the latent heat of melting L_m, the values of which, due to the peculiarities of the interaction potentials used in molecular dynamics calculations, may slightly differ from the reference ones. To correctly compare the MD data with the kinetic data in equations (3), (4), we used the values of T_m and L_m determined for all considered metals Al, Cu, Fe from additional MD calculations performed by the method [56] with the same potentials [38], [39], [40, 41]. The calculation results for 2 values of external pressure are shown in Table 1.

Metal	Pressure [kbar]	T_m, [K]	L_m, [kJ/mol]
Al	0	949	8.90
	80	1332	11.30
Cu	0	1315	11.48
	80	1602	13.21
Fe	0	1775	15.57
	80	2062	17.14

Table 1. The calculation results of T_m and L_m for 2 values of external pressure

The model of BGJ. Full alignment of a discrete set of values $\upsilon_{\langle hk\ell\rangle}(\Delta T)$ with equation (4) was achieved using a 2-parametric $C_{\langle hk\ell\rangle}$, $\beta_{\langle hk\ell\rangle}$ approximation:

$$\left[\upsilon_{s\ell}(\Delta T)\right]^{\mathrm{BGJ}}_{\langle hk\ell\rangle} = C^{BGJ}_{\langle hk\ell\rangle}\sqrt{\frac{3k_B}{m}T_{s\ell}}\left[\exp\left(\beta^{BGJ}_{\langle hk\ell\rangle}\,\frac{L_m(P)}{k_B T_m(P)}\cdot\frac{T_{s\ell}-T_m(P)}{T_{s\ell}}\right)-1\right] \qquad (7)$$

where $C^{BGJ}_{\langle hk\ell\rangle}$, $\beta^{BGJ}_{\langle hk\ell\rangle}$ are the approximation parameters. By introducing additional parameters $C^{BGJ}_{\langle hk\ell\rangle}$ before the whole expression and $\beta^{BGJ}_{\langle hk\ell\rangle}$ in the exponent, it is possible to achieve the required precision of the temperature dependence $\upsilon^{\mathrm{BGJ}}_{\langle hk\ell\rangle}(\Delta T)$ in a wide temperature range. The values of the parameters $C^{BGJ}_{\langle hk\ell\rangle}$ found from MD calculations allow to automatically estimate the value of $\frac{a}{\lambda}f$, which is not very precisely determined in the BGJ model.

Metal	Pressure [kbar]	$C^{BGJ}_{\langle 100\rangle}$	$\beta^{\mathrm{BGJ}}_{\langle 100\rangle}$	σ [m/s]
Al $\upsilon^{\mathrm{BGJ}}_{\langle 100\rangle}(\Delta T)$	P = 0	0.344	5.01	5.97
	P = 80	0.374	5.37	8.00
Cu $\upsilon^{\mathrm{BGJ}}_{\langle 100\rangle}(\Delta T)$	P = 0	0.416	5.74	11.75
	P = 80	0.441	5.51	12.36
Fe $\upsilon^{\mathrm{BGJ}}_{\langle 100\rangle}(\Delta T)$	P = 0	0.365	6.34	3.89
	P = 80	0.416	6.31	8.56
Al $\upsilon^{\mathrm{BGJ}}_{\langle 111\rangle}(\Delta T)$	P = 0	$C^{\mathrm{BGJ}}_{\langle 111\rangle}$	$\beta^{\mathrm{BGJ}}_{\langle 111\rangle}$	σ [m/s]
		0.165	7.27	5.11

Table 2. The values of the approximation parameters and the mean square deviation σ of the approximating function $\upsilon^{\mathrm{BGJ}}_{\langle hk\ell\rangle}(\Delta T)$ from the discrete set of values $\upsilon_{\langle hk\ell\rangle}(\Delta T)$.

The best agreement, with an error not exceeding a few percent, over the entire temperature range was achieved with the values of the approximating parameters $C^{BGJ}_{\langle hk\ell\rangle}$, $\beta^{BGJ}_{\langle hk\ell\rangle}$ presented in Table 2.

In Figures 1, 2, 3, solid red and dashed blue lines show the plotted dependences of the SLI speed in the crystallographic planes ⟨100⟩ and ⟨111⟩ at the external pressure P = 0 and P = 80 kbar for the elements Al, Cu, Fe.

In Fig. 4 for Al, the solid red and dashed black lines show the plotted stationary dependences of the SLI speed in the crystallographic planes ⟨100⟩ and ⟨111⟩ and at external pressure P = 0.

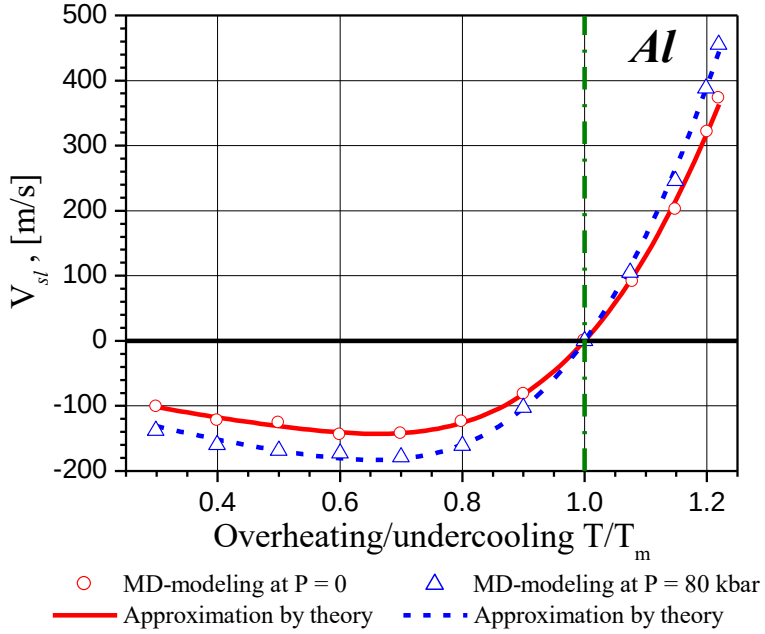


Fig. 1. The dependence of the SLI rate in the crystallographic plane ⟨100⟩ on the superheating/undercooling value for Al at P = 0 and P = 80 kbar.

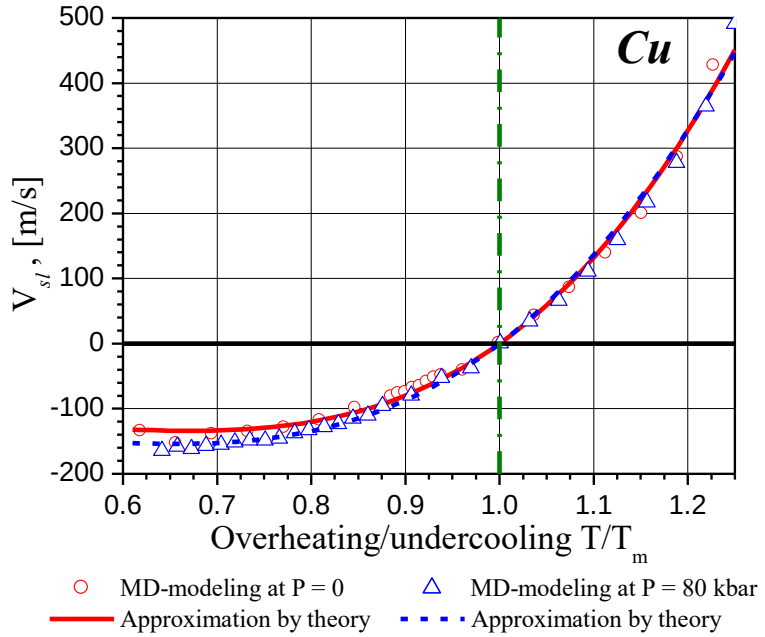


Fig. 2. The dependence of the SLI rate in the crystallographic plane ⟨100⟩ on the superheating/undercooling value for Cu at P = 0 and P = 80 kbar.

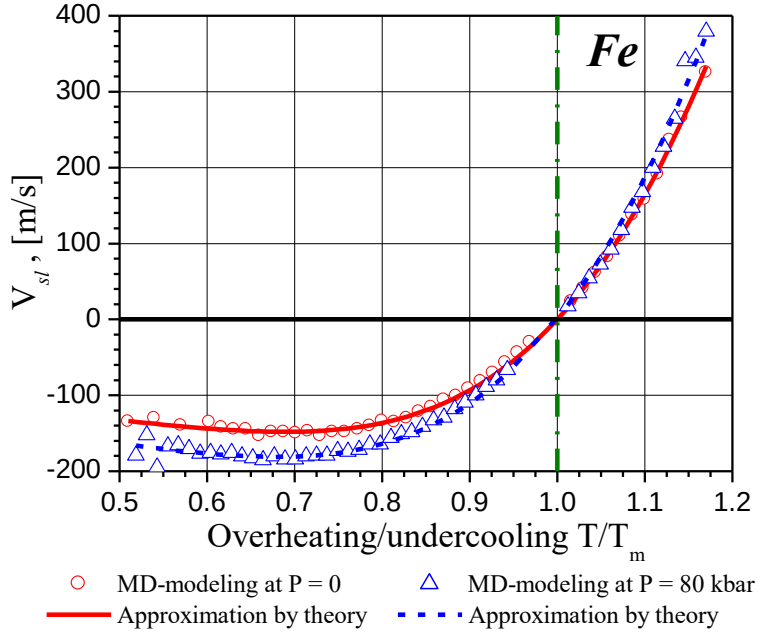


Fig.3 The dependence of the SLI rate in the crystallographic plane ⟨100⟩ on the superheating/undercooling value for Fe at P = 0 and P = 80 kbar.

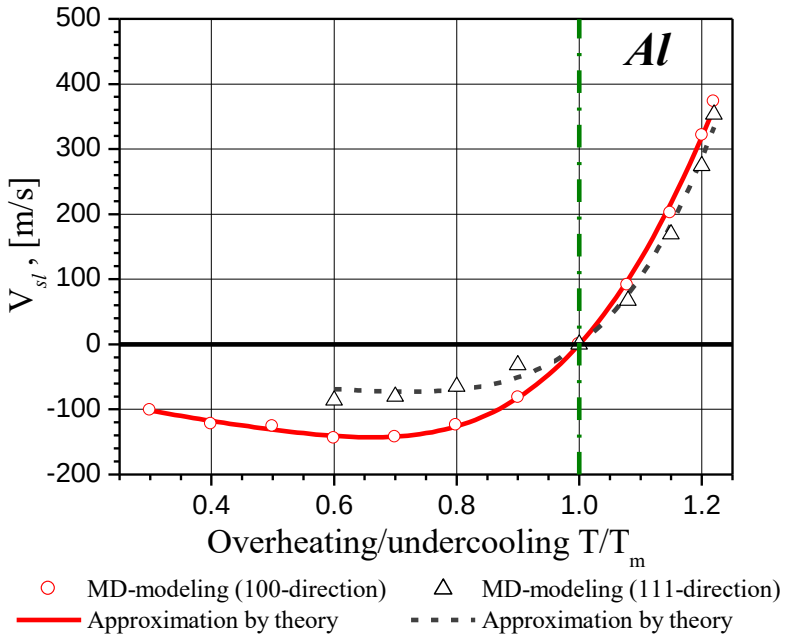


Fig. 4. The dependence of the SLI rate in the crystallographic planes ⟨100⟩ and ⟨111⟩ on the superheating/undercooling value for Al at P = 0.

Note that when simulating the crystallization of aluminum in the region of significant undercooling of the interface with the orientation ⟨111⟩, the probability of the appearance of stacking faults significantly increases. Their accumulation leads to a curvature of the initial crystallographic plane and the further propagation of the front is going in a direction different from ⟨111⟩. For this reason in the crystallographic direction ⟨111⟩, it was not possible to enter the undercooling region below $0.6\times T_m$.

The model of WF. The diffusion limited model [30 - 32] was also used to approximate the discrete set of values $\upsilon_{\langle hk\ell\rangle}(T_{s\ell})$ from molecular dynamics modeling for aluminum at two values of external pressure (P = 0, P = 80 kbar).

A complete fit of the discrete values $\upsilon_{\langle hk\ell\rangle}(\Delta T)$ to the equation (5) was reached using a three-parameter $C^{\mathrm{WF}}_{\langle hk\ell\rangle}$, $\beta^{WF}_{\langle hk\ell\rangle}$, $Q_{\langle hk\ell\rangle}$ approximation:

$$\left[\upsilon_{s\ell}(\Delta T)\right]^{WF}_{\langle hk\ell\rangle} = C^{WF}_{\langle hk\ell\rangle}\, dim\left[\frac{D_0}{a}\right] exp\left(-\frac{Q_{\langle hk\ell\rangle}}{k_B T_{s\ell}}\right)\left[\beta^{WF}_{\langle hk\ell\rangle}\, exp\left(\frac{L_m(P)}{k_B T_m(P)}\frac{T_{s\ell}-T_m(P)}{T_{s\ell}}\right)-1\right] \qquad (8)$$

In the considered range of the limiting values of superheating/undercooling in the crystallographic plane ⟨100⟩, the best fit with an error less than few percent was reached with the values of the approximation parameters $C^{\mathrm{WF}}_{\langle 100\rangle}$, $\beta^{\mathrm{WF}}_{\langle 100\rangle}$, $Q_{\langle 100\rangle}$ listed in the Table 3. The value of $C^{\mathrm{WF}}_{\langle 100\rangle}$ was determined keeping in mind that $\dim\left[\frac{D_0}{a}\right]$ = (m/s) is only a dimensionality of the relation D_0/a.

Metal	Pressure [kbar]	$C^{\mathrm{WF}}_{\langle 100\rangle}$ [m/s]	$Q_{\langle 100\rangle}$ [kJ/mol]	$\beta^{\mathrm{WF}}_{\langle 100\rangle}$	σ [m/s]
Al $\upsilon^{\mathrm{WF}}_{\langle 100\rangle}(\Delta T)$	P = 0	198.9	1.58	5.65	5.82
	P = 80	233.8	1.73	6.28	6.90
Cu $\upsilon^{\mathrm{WF}}_{\langle 100\rangle}(\Delta T)$	P = 0	140.64	0	6.88	10.70
	P = 80	163.18	0	6.67	11.48
Fe $\upsilon^{\mathrm{WF}}_{\langle 100\rangle}(\Delta T)$	P = 0	251.02	0.051	6.63	3.87
	P = 80	244.11	0.033	7.10	8.31

Table 3. The values of the approximation parameters and the mean square deviation σ of the approximating function $\upsilon^{\mathrm{WF}}_{\langle 100\rangle}(\Delta T)$.

A comparison of $\upsilon^{\mathrm{WF}}_{\langle hk\ell\rangle}(\Delta T)$ (8) with the curves $\upsilon^{\mathrm{BGJ}}_{\langle hk\ell\rangle}(\Delta T)$, (7), at Figs. (1 – 3) showed almost complete fit with an error less than 1%. For this reason in this paper, a separate plot of $\upsilon^{\mathrm{WF}}_{\langle hk\ell\rangle}(T)$ is not provided.

5 SHORT ANALYSIS

The obtained curves $\upsilon_{\langle hk\ell\rangle}^{\mathrm{WF,\,BGJ}}(\Delta T)$, characterizing the mobility of SLI, for the considered (fcc) and (bcc) metals in the crystallographic plane $\langle 100\rangle$ have a great generality. The melting branches in the range $T_m \le T_{s\ell} \le 1.25T_m$ have exponential behavior as the superheating rises, reaching the maximum values $\upsilon_{s\ell}^{WF,\,BGJ}(\Delta T) \sim 300 \div 350 m/s$. The crystallization process in the undercooling region $0.5T_m < T_{s\ell} \le T_m$ takes place in a more complicated manner. The speed of crystallization $\upsilon_{s\ell}^{WF,\,BGJ}(\Delta T)$ at all curves, Figs.1-3, in the vicinity of $T_{s\ell} \sim 0.7T_m$ have a well-noticeable maximum of ~ 140÷160 *m/s.* It should be noted that the maximum crystallization rate for Fe coincides with similar data obtained in [35]. The appearance of the maximum in the crystallization rate is associated with the beginning of the formation of the interstitial phase, which slows down the speed of the phase front. In this work, only crystallization processes in a undercooled pure liquid are considered. Glass transition processes occurring near and below the temperature T_g, $0 < T_{s\ell} \le T_g$, are not included in the consideration, since the complexity and importance of this process, in particular, for technological applications deserves a separate consideration.

External pressure (P = 80 kbar) has no significant effect on the behavior of the kinetic speed in the entire considered range of superheating/undercooling (blue curves). But the maximum values of the rate of melting and crystallization in this case increase to 350 ÷ 450 m/s and 160 ÷ 180 m/s, respectively. To a much greater extent, SLI mobility is affected by its crystallographic orientation. For example, for aluminum, the ratio of the maximum crystallization rates reaches $\upsilon_{\langle 100\rangle}^{\mathrm{BGJ}}(\Delta T)/\upsilon_{\langle 111\rangle}^{\mathrm{BGJ}}(\Delta T) \approx 2$ times, and that of melting, 1.2 times, Fig. 4.

6 CONCLUSIONS

The molecular dynamics method was used to study the kinetics of melting/crystallization of (fcc) and (bcc) metals (Al, Cu, Fe) in the range of limiting values of superheating/undercooling. The limiting superheating/undercooling for each of the metals is determined by the temperature values at which the superheated crystal retains the properties of the crystal, and the undercooled melt still remains liquid.

a) The modeling results showed that for the considered metals, the range of the limiting values of superheating/undercooling is in the range $T_{s\ell} \approx (0.3 \div 1.25)T_m$ for $\mathrm{Al}_{\langle 100\rangle}$, $T_{s\ell} \approx (0.6 \div 1.2)T_m$ for $\mathrm{Al}_{\langle 111\rangle}$, $T_{s\ell} \approx (0.6 \div 1.3)T_m$ for Cu, $T_{s\ell} \approx (0.65 \div 1.18)T_m$ for Fe.
b) A discrete set of velocity values $\upsilon_{\langle hk\ell\rangle}(T_{s\ell},P)$ obtained from the atomistic modeling together with the kinetic models of WF (3) and BGJ (4) were used to construct analytical dependences of the stationary velocity of motion of the SLI $\upsilon_{\langle hk\ell\rangle}^{WF,\,BGJ}(\Delta T,P)$ (Eqs. (7, 8)) over the entire range of limiting values of superheating/undercooling.
c) In the considered range of limiting superheating/undercooling, both kinetic models of WF and BGJ allow, with practically the same error of ~ 1%, to construct analytical expressions for the velocities in the form of two- and three-parametric curves in the case of using models of BGJ (4) and WF (3) respectively.

d) The temperature dependences of the solid/liquid interface velocity for Al, Cu, and Fe, determined from the results of simulations using different crystallographic planes, demonstrate a clear asymmetry with respect to the melting point T_m, which is explained by the strong difference between the melting kinetics in a highly superheated state and the kinetics of solidification in a highly undercooled state.
e) For all metals, the change in the temperature dependence of the velocity $v_{s\ell}(\Delta T)$ when passing through the melting point T_m occurs smoothly without a bend in the slope.
f) The crystallographic orientation of the metal, rather than its crystal lattice type, has the greatest impact on SLI mobility.
g) External pressure has no significant effect on the kinetic velocity behavior over the entire superheating/undercooling range under consideration, but leads to an increase in the maximum melting and crystallization rates.
h) The obtained results of atomistic modeling indicate that the capabilities of the classical Wilson - Frenkel model [30–32] were greatly underestimated.

Acknowledgements: This study was supported by the Russian Science Foundation (project no. 18-11-00318).

REFERENCES

1. *Handbook of Materials Modeling, Vol. 1, 2,* Yip Sidney (Ed.), Springer, Dordrecht, Berlin, Heidelberg, New York (2005).
2. K. Thornton, J. Ågren, P.W. Voorhees, "Modelling the evolution of phase boundaries in solids at the meso- and nano-scales", *Acta Materialia*, **51** (19), 5675–5710 (2003).
3. A.P. Sutton, R.W. Balluffi, *Interfaces in crystalline materials*, Oxford: Clarendon Press (1995).
4. Y. Mishin, M. Asta, Ju Li. "Atomistic modeling of interfaces and their impact on microstructure and properties. Overview № 148", *Acta Materialia*, **58**(4), 1117–115**1**, (2010). https://doi.org/10.1016/j.actamat.2009.10.049
5. J.J. Hoyt, M. Asta, A. Karma, "Atomistic and continuum modeling of dendritic solidification", *Materials Science and Engineering R*, **41**, 121–163 (2003).
6. 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 No. 146", *Acta Materialia*, **57**, 941–971 (2009).
7. Yu C. Shen and David W. Oxtoby, "Density functional theory of crystal growth: Lennard-Jones fluids", *J. Chem. Phys.*, **104** (11), 4233 – 4242 (1996). doi: 10.1063/1.471234.
8. G.H. Rodway, J.D. Hunt, "Thermoelectric investigation of solidification of lead. I. Pure lead", *J. Cryst. Growth*, **112** (2-3), 554-562 (1991).
9. J. Monk, Y. Yang, M. I. Mendelev, M. Asta, J. J. Hoyt and D. Y. Sun, "Determination of the crystal-melt interface kinetic coefficient from molecular dynamics simulations", *Modelling Simul. Mater. Sci. Eng.*, **18**, 015004(1-18) (2010). doi:10.1088/0965-0393/18/1/015004.
10. D.Y. Sun, M. Asta, and J. J. Hoyt, "Kinetic coefficient of Ni solid-liquid interfaces from molecular-dynamics simulations", *Phys. Rev. B*, **69**, 024108 (2004). doi:10.1103/PhysRevB.69.024108
11. Y.F. Gao, Y.Yang, D.Y.Sun, M.Asta, J.J.Hoyt, "Molecular dynamics simulations of the crystal–melt interface mobility in HCP Mg and BCC Fe", *J.Crystal Growth*, **312** (21) 3238–3242 (2010). doi.org/10.1016/j.jcrysgro.2010.07.051
12. K.A. Jackson, *Kinetic processes:crystal growth, diffusion, and phase transitions in materials*, Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA, (2004).

13. V.I. Mazhukin, "Kinetics and Dynamics of Phase Transformations in Metals Under Action of Ultra-Short High-Power Laser Pulses. Ch.8", 219 -276, *Laser Pulses – Theory, Technology, and Applications*, I. Peshko (Ed.), InTech, Croatia, (2012). http://dx.doi.org/10.5772/50731
14. B. Chalmers, *Principles of Solidification*, N. Y. John Wiley &Sons, (1964).
15. S.R. Coriell, D. Turnbull, "Relative roles of heat transport and interface rearrangement rates in the rapid growth of crystals in undercooled melts", *Acta Metallurgica*, **30** (12), 2135-2139 (1982).
16. M. Amini, B.B. Laird, "Kinetic coefficient for hard-sphere crystal growth from the melt", *Phys. Rev. Lett.*, **97**, 216102(1-4) (2006).
17. J.J. Hoyt, M. Asta, A. Karma, "Atomistic simulation methods for computing the kinetic coefficient in solid-liquid systems", *Interface Science*, **10** (2-3), 181-189 (2002).
18. M.I. Mendelev, M.J. Rahman, J.J. Hoyt, M. Asta, "Molecular-dynamics study of solid–liquid interface migration in fcc metals", *Modelling Simul. Mater. Sci. Eng.*, **18,** 074002(1-18) (2010). http://dx.doi.org/10.1088/0965-0393/18/7/074002
19. W. Kurz and D. J. Fisher, *Fundamentals of Solidification*, *3rd ed*., Trans Tech, Aedermannsdorf, Switzerland, (1992).
20. R.Trivedi, W.Kurz, "Solidification microstructures: A conceptual approach", *Acta Metallurgica et Materialia*. **42** (1), 15-23 (1994). https://doi.org/10.1016/0956-7151(94)90044-2
21. P. Galenko, S. Sobolev. "Local nonequilibrium effect on undercooling in rapid solidification of alloys", *Phys. Rev. E*, **55** (1), 343 – 352 (1997).
22. D. M.Herlach, "Solidification from undercooled melts", *Materials Science and Engineering: A*, **226–228**, 348-356 (1997). https://doi.org/10.1016/S0921-5093(96)10644-4
23. Y. Ashkenazy, R. S. Averback, "Atomic mechanisms controlling crystallization behaviour in metals at deep undercoolings", *Europhysics Letters (EPL)*, **79** (2), 26005(1-6) (2007). doi: 10.1209/0295-5075/79/26005.
24. V.I. Mazhukin, "Nanosecond laser ablation: mathematical models, computational algorithms, modeling Chapter 2", 31 -55, *Laser Ablation - From Fundamentals to Applications*, Tatiana Itina (Ed.), InTech, Croatia, (2017).
25. C. A. MacDonald, A. M. Malvezzi, & F. Spaepen, "Picosecond time-resolved measurements of crystallization in noble metals", *JAP*, **65** (1), 129–136, (1989). doi:10.1063/1.342586
26. M. B. Agranat, S. I. Ashitkov, V. E. Fortov, A. V. Kirillin, A. V. Kostanovskii, S. I. Anisimov, & P. S. Kondratenko, "Use of optical anisotropy for study of ultrafast phase transformations at solid surfaces", *Appl. Phys. A: Materials Science & Processing*, **69**(6), 637–640, (1999). doi:10.1007/s003390051045
27. W. L. Chan, R. S. Averback, D. G. Cahill, Y. Ashkenazy, "Solidification velocities in deeply undercooked silver", *Phys. Rev. Lett.*, **102**(9), 095701(1-4) (2009). https://doi.org/10.1103/PhysRevLett.102.095701
28. 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).
29. 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).
30. H.A. Wilson, "On the velocity of solidification and viscosity of supercooled liquids", *Philos. Mag.*, **50**, 238-250 (1900).
31. J.I. Frenkel, "Note on the relation between the speed of crystallization and viscosity", *Phys. Z. Sowjet Union*, **1**, 498 – 499 (1932).
32. J. Frenkel, *Kinetic Theory of Solids*, Oxford University Press, N. Y., (1946).
33. J.Q. Broughton, G.H. Gilmer, K.A. Jackson, "Crystallization Rates of a Lennard-Jones Liquid", *Phys. Rev. Let.*, **49**, 1496 -1500 (1982).

34. L.V. Mikheev, A. A. Chernov, "Mobility of a diffuse simple crystal—melt interface", *J. Crystal Growth*, **112** (2-3), 591–596 (1991). doi:10.1016/0022-0248(91)90340-b
35. Y. Ashkenazy, R.S. Averback, "Kinetic stages in the crystallization of deeply undercooled body-centered-cubic and face-centered-cubic metals", *Acta Materialia*, **58**, 524–530 (2010).
36. D. Turnbull, "On the relation between crystallization rate and liquid structure", *J. Phys. Chem.*, **62** (4), 609 – 613 (1962).
37. J.E. Jones, "On the determination of molecular fields. I. From the Variation of the Viscosity of a Gas with Temperature", *Proc. Royal Society of London. Series A*, **106 (**738**)**, 441-462 (1924).
38. V.V. Zhakhovskii, N.A. Inogamov, Yu.V. Petrov, S.I. Ashitkov, K. Nishihara, "Molecular dynamics simulation of femtosecond ablation and spallation with different interatomic potentials", *Appl. Surf. Sci.*, **255**(24), 9592-9596 (2009).
39. S. M. Foiles, M. I. Baskes, M. S. Daw, "Embedded-Atom-Method Functions for the Fcc Metals Cu, Ag, Au, Ni, Pd, Pt and their alloys", *Phys. Rev. B*, **33**, 7983 -7991 (1986).
40. M.I. Mendelev, S. Han, D.J. Srolovitz, G.J. Ackland, D.Y. Sun, M. Asta, "Development of new interatomic potentials appropriate for crystalline and liquid iron", *Philosophical Magazine*, **83** (35), 3977–3994 (2003).
41. G.J. Ackland, M.I. Mendelev, D.J. Srolovitz, S. Han, A.V. Barashev, "Development of an interatomic potential for phosphorus impurities in α-iron", *J. Phys. Condens. Matter*, **16**, s2629 (1-14) (2004).
42. M.I. Mendelev, "Molecular dynamics simulation of solidification and devitrification in a one-component system", *Modelling Simul. Mater. Sci. Eng.*, **20**(4), 045014(1-17) (2012). https://doi.org/10.1088/0965-0393/20/4/045014
43. Q.S. Mei, K. Lu, "Melting and superheating of crystalline solids: From bulk to nanocrystals", *Progress in Materials Science*, **52** (8), 1175–1262 (2007).
44. V. I. Mazhukin, M. M. Demin, A. V. Shapranov, "High-speed laser ablation of metal with pico- and subpicosecond pulses", *Appl. Surf. Sci.*, **302**, 6-10 (2014). https://doi.org/10.1016/j.apsusc.2014.01.111
45. Gang Sun, Jenny Xu & Peter Harrowell, "The mechanism of the ultrafast crystal growth of pure metals from their melts", *Nature Materials*, **17**, 881–886 (2018). doi: https://doi.org/10.1038/s41563-018-0174-6
46. Peter K. Galenko, Dmitri V. Alexandrov, "From atomistic interfaces to dendritic patterns", *Medicine, Physics. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **376** (2113) (2018). doi:10.1098/rsta.2017.0210.
47. J. Schmelzer, T. Tropin, "Glass Transition, Crystallization of Glass-Forming Melts, and Entropy", *Entropy*, **20**(2), 103(1-32) (2018). doi:10.3390/e20020103
48. T. V. Tropin, J. W. Schmelzer & V. L. Aksenov, "Modern aspects of the kinetic theory of glass transition", *Physics-Uspekhi*, **59**(1), 42–66 (2016). doi:10.3367/ufne.0186.201601c.0047
49. 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).
50. 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).
51. M.D. Kluge, J.R. Ray, "Pulsed laser melting of silicon: A molecular dynamics study", *J.Chem. Phys.*, **87** (4), 2336 – 2339 (1987).
52. 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).
53. C.J. Tymczak, J.R. Ray, "Interface response function for a model of sodium. A molecular dynamics study", *J. Chem. Phys.*, **92** (12), 7520 – 7530 (1990).
54. V.I. Mazhukin, A.V. Shapranov, A.V. Mazhukin, P.V. Breslavsky, "Atomistic modeling of the dynamics of the solid/liquid interface of Si melting and crystallization taking into account deeply

superheated/supercooled states", *Mathematica Montisnigri*, **47**, 87-99 (2020). doi: http://doi.org/10.20948/mathmontis-2020-47-8

55. B. Rethfeld, K. Sokolowski-Tinten, D. von der Linde, S.I. Anisimov, "Ultrafast thermal melting of laser-excited solids by homogeneous nucleation", *Phys. Rev. B*, **65**, 092103(1-4) (2002).
56. V.I. Mazhukin, A.V. Shapranov, V.E. Perezhigin, "Matematicheskoe modelirovanie teplofizicheskix svojstv, processov nagreva i plavleniya metallov metodom molekulyarnoj dinamiki", *Mathematica Montisnigri*, **24**, 47 – 66 (2012).

Received June 18, 2020

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

CALCULATION OF MIS WEIGHTS FOR BIDIRECTIONAL PATH TRACING WITH PHOTON MAPS IN PRESENCE OF DIRECT ILLUMINATION

S. V. ERSHOV, A. G. VOLOBOY*

Keldysh Institute of Applied Mathematics of RAS,
Miusskaya Sq. 4, Moscow, Russia, 125047
*Corresponding author. E-mail: voloboy@gin.keldysh.ru

DOI: 10.20948/mathmontis-2020-48-8

Summary. Multiple importance sampling (MIS) is a well-known method for noise reduction in Monte-Carlo ray tracing. It weights contributions from merging camera and light paths in different vertices. Since noise strongly depends on these weights, the problem of the optimal choice of weight to reach the minimal noise is very important. For bi-directional Monte-Carlo ray tracing with photon maps (BDPM), different join paths are not statistically independent because several light paths are checked against the same camera path and vice versa. As a result, the optimal weights which minimize the noise functional in the classic Monte Carlo ray tracing and in BDPM are different. In this paper we calculate weights for the simple reduced case of just two strategies, i.e. merging at just two vertices of camera ray. We show that these weights obey an integral equation which is qualitatively different from the well-known MIS formulae for uncorrelated samples. The integral equation is solved analytically in a closed form and one can see that the MIS weights for BDPM algorithm depend on the number of rays and scene geometry. In this paper we also correctly take into account the direct illumination to pixel luminance.

1 INTRODUCTION

Currently simulation of light propagation is widely used not only in optical engineering bust also in design of new materials. It is intensively applied in architectural, automotive and aircraft design tasks. If wave effects can be neglected, various kinds of stochastic ray tracing are a good choice. This realm mainly includes Metropolis light propagation and Monte-Carlo ray tracing (MCRT). When an image should be calculated, the classic forward ray tracing from light source is inefficient and various bi-directional modifications [1] are applied. The weak side of all stochastic methods is that their results are noisy. The amplitude of noise depends on the method of ray generation and scattering.

Therefore the problem of the optimal probability distribution of ray scattering has being addressed since long. An important foundation which many modern approaches originated from is the Veach's work [2]. Its theorems about Multiple Importance Sampling (MIS) in Monte-Carlo based methods are still a base for current research. The idea is

2010 Mathematics Subject Classification: 78-04, 65C05, 65C20.
Key words and Phrases: Monte-Carlo ray tracing, photon maps, reduction of noise, multiple importance sampling, weights..

to generate several random numbers, one for each "strategy" (i.e. probability density which admits an efficient generation of samples), and then sum their contributions to the accumulated average with weight (which usually depends on the phase space point). With the optimal choice of weights for each strategy, the noise can be substantially reduced, and [2] suggested the famous "balanced heuristic" and "power heuristic" methods of calculation of weights. A proof had been produced that albeit these weights are only sub-optimal, the resulting noise does not much exceed the absolute minimum. Recently this latter point had been attacked in e.g. [3].

Veach results are based on the theorem which assumes independent samples, but then were applied to the more advanced methods like bi-directional Monte Carlo path tracing (BDPT), bi-directional Monte-Carlo ray tracing with photon maps (BDPM) [4, 5] or their combination [6]. Meanwhile now the "samples" i.e. light paths connecting the source and the camera, happen to be not independent because e.g. in BDPM or "vertex merging" the same light path is merged with many camera paths and vice versa. So the resulting joined full trajectories have common part and thus they are not independent.

In this work we investigate the weights optimal to BDPM and demonstrate the equations for them are qualitatively different from the famous Veach heuristics. This is because for the bidirectional Monte Carlo path tracing method or for BMCRT with photon maps the noise does not follow the rule like in the classic MCRT [7]. Therefore the optimal weights which minimize the noise functional in the classic MCRT and in BDPM are different. Since the BDPM noise is a quadratic functional over the ray contribution as it is shown in [7], it is also quadratic in weights. So while calculation of weights that minimize that noise functional looks mathematically trivial it is not so in practice.

There is an infinite set of weights in bidirectional MCRT with own set of weights for each join path length. The weights from different sets are defined in different functional spaces (they have different number of arguments, i.e. vertices and so on), while all they are "coupled" in the noise functional. There are also problems with ray absorption etc. As a result the optimal weights obey an infinite system of linear integral equations which are extremely complex. And their kernels must be calculated from solving yet other integral equations similar to the "rendering equation" [8] and so on.

We derive the formulae for the optimal MIS weights in case of photon maps, i.e. when each camera ray is merged with several light rays. Unlike the algebraic equations of the Veach's "balance (or power) heuristic", these are integral equations which happily admit solution in closed form (i.e. it is an analytic formula which includes several integrals of the input functions). One can also see that they depend not only on the BDFs and light emission distribution, like in the Veach case, but also they depend on the number of rays and scene geometry.

2 BACKGROUND

The issue of optimal MIS weights choosing is very important. It influences on quality of resultant realistic image or, what is the same, on a convergence speed of algorithm. This is why lot of studies is devoted to the issue.

The MIS formulae and heuristic methods for optimal weights calculation were proposed in [2, 9]. The formulae derived there depend on the scattering characteristics (BDFs) and

distribution of light source emission only. Further this approach was applied in many papers of which we should recommend [10].

Some extension was suggested in [11] which operates probability density of the join paths (i.e. rays from the source to the camera obtained by "merging vertices" of the forward and backward MCRT). The number of vertices in the join path and two halves which constitute it is different, because two close ones merge into one. Thus the phase spaces are different, but the author unify them and calculate the density in the space of join paths. Naturally, it is proportional to the importance of the join path, and also contains scale factor proportional to the squared merging radius. What is important here is that this radius can be different in different parts of the scene. As a result, after substitution the density of the join path into Veach-like formulae the authors can investigate dependence on the merging radius. The remaining problem is that this is still a one-sample density which is insufficient to treat "correlation terms" in BDPM [7], so accounting for the integration radius does not help much. Notice that our approach does not use the density of join paths, but only those of the light and camera paths, because the noise is an integral with narrow kernels due to "vertex merging", and these weakly converge to the delta function so that not causing much problems.

The general idea of the papers [12, 13] is that the forward MCRT ray hits within the integrating sphere around the point of the camera ray. This method is named photon mapping or "vertex merging". In the same time it is possible to connect the segment point of the camera ray with the forward MCRT point ("vertex connection"). Of course here it is necessary check the shading, etc. In general these are two different bidirectional ray tracing methods. And the authors united those using weights to produce the promising "vertex connection and merging" (VCM) method. But the weights are all the same fractionally rational as those of Veach. Non-locality and non-linearity (outside the Veach's "power heuristic") is absent there.

S. Popov and others in [14] proposed original algorithm for generation of full paths connecting camera and light source. They stochastically generated several paths with different weights. After that they chose weights to minimize stochastic noise in resultant image. Their algorithm differs from ours because they neglect statistical dependence between "camera rays of different length". This sounds reasonable, but even if they are really independent, they are merged with the same light path. The authors claim they have a full derivation which does account for correlations and leads to integral equations, but it is absent in the paper. The initial equations they start from are correct but too complex to treat and the authors simplified the problem by replacing the actual noise by its upper bound.

Naturally, this does not guarantee the weights found do lead to a smaller noise: although the upper bound decreases, the noise itself may even increase.

Several works [15, 16, 17] are devoted to the issue of efficient computation of the MIS weights.

In [18] the problem of calculations of the optimal weights had been considered for a *limited* MIS when we mix contributions from just *two* vertices, M-th and $(M+1)$-th of the camera path, when M is fixed. This consideration assumed that there is no direct illumination at the camera vertices up to the last (i.e. the $(M+1)$-th), see Section 7 of this paper. While possible in reality, this is a severe limitation, especially when M is large

enough. Current paper eliminates this restriction.

3 LIMITED MIS IN BDPM

The general idea is described in [18]: we trace camera ray until it underwent $M+1$ *diffuse* events, it never can go further. While the direct and caustic illumination is collected in all of them, *diffuse* one is collected only in the M-th and $M+1$-th hits, with weights w_0 and $w_1 = 1 - w_0$ correspondingly. These weights are deterministic functions of $M+1$ vertices of the *join* trajectory (merged camera and light paths), counting from camera. These weights affect the noise while the limiting accumulated image luminance is the same, and so can be chosen so as to reduce the noise. Details and reasoning produced in [18] up to Section 7 remain still valid. We also assume that $M > 0$ so that the density of camera hits starting from the M-th is *diffuse*. The difference is in Eq. (3) of [18], where the contribution to pixel luminance from merging the camera and light paths neglected the direct illumination while here it will be taken into account.

4 RAY CONTRIBUTION

Here and below all calculations are for ***one*** *pixel. The total flux of all scene lights is assumed 1 to not bother about scaling between the density of photons and irradiance.*

Retaining the direct illumination, the contribution[1] to pixel luminance from merging the camera path $\{x_0, x_1, ...\}$ (counting from the first camera hit x_0) and light path $\{x_0^{(F)}, x_1^{(F)}, ...\}$ (counting vertices from the light source) is

$$\begin{aligned} C \;=\; & \sum_{m=0}^{M} E(x_0, ..., x_m) L_0(x_{m-1}, x_m) \\ & + \sum_{n\geq 2} K(x_M - x_n^{(F)}) w_0(\boldsymbol{\mathcal{X}}, x_{n-1}^{(F)}) E(\boldsymbol{\mathcal{X}}) f(x_{n-1}^{(F)} \to x_n^{(F)}, x_M \to x_{M-1}, x_M) \\ & + \sum_{n\geq 2} K(x_{M+1} - x_n^{(F)}) w_1(\boldsymbol{\mathcal{X}}, x_{M+1}) E(\boldsymbol{\mathcal{X}}, x_{M+1}) f(x_{n-1}^{(F)} \to x_n^{(F)}, x_{M+1} \to x_M, x_{M+1}) \\ & + w_1(\boldsymbol{\mathcal{X}}, x_{M+1}) E(\boldsymbol{\mathcal{X}}, x_{M+1}) L_0(x_M, x_{M+1}) \end{aligned}$$

where

$$\boldsymbol{\mathcal{X}}_m \equiv (x_0, ..., x_m)$$

is the initial part of the camera path,

$$x \to y \equiv \frac{y - x}{|y - x|}$$

denotes the *unit* vector from x to y, K is the integration kernel, w_0 and $w_1 = 1 - w_0$ are the weights, $L_0(x, y)$ is the surface luminance in point y under *direct* illumination coming

[1] That is, accumulated pixel value increases by C.

from the point x, $f(v,u,x)$ is *diffuse* BDF in luminance factor units at surface point x for illumination direction (towards the surface) v and direction of scattering (away from the surface) u, $E(x_0,...,x_m)$ is the energy (or transmission factor in [10] terms) of the camera ray *before* hitting x_m, while for the light ray the energy is always 1. Notice that since the direct component is taken explicitly (via the terms L_0), it is not taken from photon maps to avoid double counting, i.e. the light segments $[x_0^{(F)}, x_1^{(F)}]$ which represent the *direct* illumination do not interact with integration spheres and thus the sums start from $n=2$.

The "energy" or camera ray, which is also named *transmission factor*, is defined as usual: it is 1 just after leaving the camera, i.e. $E(x_0)=1$ and then

$$E(x_0,...,x_{m+1}) = \mu(x_{m-1},x_m)E(x_0,...,x_m) \tag{1}$$

where

$$\mu(y,x) \equiv \int f(v, x\to y, x)\,|(v\cdot n(x))|\,d^2v, \tag{2}$$

n being the local normal.

Below we shall denote

$$\mathcal{L}_0(x_0,...,x_M) \equiv \sum_{m=0}^{M} E(x_0,...,x_m)L_0(x_{m-1},x_m)$$

which is what the pure path tracing (without using FMCRT) would accumulate. Ray contribution then becomes

$$\begin{aligned} C &= \mathcal{L}_0(\boldsymbol{\mathcal{X}}_M) + w_1(\boldsymbol{\mathcal{X}}_{M+1})E(\boldsymbol{\mathcal{X}}_{M+1})L_0(x_M,x_{M+1}) \\ &+ \sum_{n\ge 2} K(x_M - x_n^{(F)})w_0(\boldsymbol{\mathcal{X}}_M, x_{n-1}^{(F)})E(\boldsymbol{\mathcal{X}}_M)f(x_{n-1}^{(F)}\to x_n^{(F)}, x_M\to x_{M-1}, x_M) \\ &+ \sum_{n\ge 2} K(x_{M+1} - x_n^{(F)})w_1(\boldsymbol{\mathcal{X}}_M, x_{M+1})E(\boldsymbol{\mathcal{X}}_{M+1})f(x_{n-1}^{(F)}\to x_n^{(F)}, x_{M+1}\to x_M, x_{M+1}) \end{aligned} \tag{3}$$

5 NOISE

In BDPM (with or without weights), the variance of the pixel luminance calculated from N_F forward rays and N_B backward rays (started from the same pixel) obeys the general law [7], see also [18]:

$$\begin{aligned} V &= \frac{1}{N_F N_B}\left(\langle\langle C^2\rangle\rangle - \langle\langle C\rangle\rangle^2\right) + \frac{1-N_F^{-1}}{N_B}\left(\langle\langle C\rangle_F^2\rangle_B - \langle\langle C\rangle\rangle^2\right) \\ &+ \frac{1-N_B^{-1}}{N_F}\left(\langle\langle C\rangle_B^2\rangle_F - \langle\langle C\rangle\rangle^2\right) \end{aligned}$$

Here $\langle\cdot\rangle_B$ is the averaging over the BMCRT ensemble for the fixed FMCRT ray and $\langle\cdot\rangle_F$ is the averaging over the FMCRT ensemble for the fixed camera ray. Notice the linear

term $\langle\langle C\rangle\rangle$ is independent from the order of averaging so we drop subscripts here. It is also independent from weights, while $\langle\langle C^2\rangle\rangle$ and $\langle\langle C\rangle_F^2\rangle_B$ depend on them.

Averaging over the ensemble of light paths resp. camera paths is

$$\langle\cdot\rangle_F = \int (\cdot) p_F(x_0^{(F)}, x_1^{(F)}, ..., x_n^{(F)}, ...) dx_0^{(F)} \cdots dx_n^{(F)} \cdots \tag{4}$$

$$\langle\cdot\rangle_B = \int (\cdot) p_B(x_0, x_1, ..., x_{M+1}) dx_0 dx_1 \cdots dx_{M+1} \tag{5}$$

where p_F and p_B are the probability densities of the light and camera paths. Since we assume that FMCRT uses Russian roulette to kill rays while keeps ray energy, p_F is *not* normalized. Below we shall sometimes use the spatial and sometimes angular probability densities keeping in mind the obvious relation between differentials

$$d^2(x \to y) = s(x,y) d^2 y \tag{6}$$

$$s(y,x) \equiv |((x \to y) \cdot n(x))| \tag{7}$$

where n is the local normal.

While neither explicit form nor the properties of $p_F(\cdots)$ are used, for the camera path density we need the recurrence relation. Namely, in BMCRT where direction of the next segment of the camera ray is chosen proportionally to BDF,

$$p_B(x_0, ..., x_{M+1}) = \tilde{f}(v, u, x_M) \, |(v \cdot n(x_M))| \, s(x_M, x_{M+1}) \times p_B(x_0, ..., x_M), \tag{8}$$

where

$$\begin{aligned} \tilde{f}(v,u,x_M) &\equiv \frac{f(v,u,x_M)}{\mu(x_M, x_{M+1})} \\ v &\equiv x_{M+1} \to x_M \\ u &\equiv x_M \to x_{M-1} \end{aligned} \tag{9}$$

is the "normalized" BDF for BMCRT, and the total backward scattering μ was defined in (2).

Notice that when $N_F \to \infty$ while N_B is fixed the noise does not vanish. This remaining noise $N_B^{-1}\left(\langle\langle C\rangle_F^2\rangle_B - \langle\langle C\rangle\rangle^2\right)$ can be naturally termed the "BMCRT noise", and $\langle\langle C\rangle_F^2\rangle_B$ named "BMCRT term". Similarly, when $N_B \to \infty$ while N_F is fixed, the remaining noise $N_F^{-1}\left(\langle\langle C\rangle_B^2\rangle_F - \langle\langle C\rangle\rangle^2\right)$ is termed "FMCRT noise" and $\langle\langle C\rangle_B^2\rangle_F$ is named the "FMCRT term". The last quadratic average $\langle\langle C^2\rangle\rangle$ will be named the "cross term".

Again, like in [18], we neglect the "FMCRT noise", because for most practical cases it is much smaller than the other two terms. Assuming $N_F \gg 1$, we arrive at the approximate noise law:

$$N_B V \approx N_F^{-1}\left(\langle\langle C^2\rangle\rangle - \langle\langle C\rangle\rangle^2\right) + \langle\langle C\rangle_F^2\rangle_B - \langle\langle C\rangle\rangle^2 \tag{10}$$

Now let us calculate the weight-dependent quadratic averages $\langle\langle C^2\rangle\rangle$ and $\langle\langle C\rangle_F^2\rangle_B$.

5.1 Cross term $\langle\langle C^2\rangle\rangle$

Averages over the ensemble of light paths (4) of the terms proportional to the 0th or 1st power of kernel are all limited for $S \to 0$. So in $\langle\langle C^2\rangle\rangle$ dominating are terms quadratic in kernel. Let us assume that the integration kernel K is uniform within the integration sphere, so that $K^2 = S^{-1}K$ where S is its area. Then

$$\begin{aligned} C^2 &= \frac{1}{S}\sum_{n\geq 2} K(x_M - x_n^{(F)}) w_0^2(\boldsymbol{\mathcal{X}}_M, x_{n-1}^{(F)}) g_M^2(\boldsymbol{\mathcal{X}}_M, x_{n-1}^{(F)}, x_n^{(F)}) \\ &\quad + \frac{1}{S}\sum_{n\geq 2} K(x_{M+1} - x_n^{(F)}) w_1^2(\boldsymbol{\mathcal{X}}_{M+1}) g_{M+1}^2(\boldsymbol{\mathcal{X}}_{M+1}, x_{n-1}^{(F)}, x_n^{(F)}) \\ &\quad + \sum_{n\neq n'} K(x_M - x_n^{(F)}) K(x_{M+1} - x_{n'}^{(F)}) \times (\cdots) \\ &\quad + \sum_{n\geq 2} K(x_M - x_n^{(F)}) K(x_{M+1} - x_n^{(F)}) \times (\cdots) \\ &\quad + O(1), \\ g_m(\boldsymbol{\mathcal{X}}_m, x_{n-1}^{(F)}, x_n^{(F)}) &\equiv E(\boldsymbol{\mathcal{X}}_m) f(x_{n-1}^{(F)} \to x_n^{(F)}, x_m \to x_{m-1}, x_m) \end{aligned}$$

where we dropped the terms proportional to the product $K(x_M - x_n^{(F)})K(x_M - x_{n'}^{(F)})$ which for $n \neq n'$ and $S \to 0$ vanishes because in a "normal" scene the light path segment can not have zero length. Therefore

For *different* light path vertices $n \neq n'$ the average of $K(x_M - x_n^{(F)})K(x_{M+1} - x_{n'}^{(F)})$ times a limited function over the light paths is limited for $S \to 0$. The product $K(x_M - x_n^{(F)})K(x_{M+1} - x_n^{(F)})$ of kernels with different centres is 0 for a.e. camera path, thus the average of a limited function times it over the combined camera and light paths ensemble thus goes to 0 as $S \to 0$. Therefore, the two first lines strongly dominate for $S \to 0$.

Now let us applying to them averaging (4) then (5) and taking into account that by virtue of FMCRT

$$p_F(x_{n-1}^{(F)}, x) d^2 x_{n-1}^{(F)} = |(v \cdot n(x))|\, I_n(v, x) d^2 v = |(v \cdot n(x))|\, I_n(v, x) s(x, x_{n-1}^{(F)}) d^2 x_{n-1}^{(F)} \quad (11)$$

where n is the local normal, I_n is *angular density* of illumination of the n-th order (i.e. after $n-1$ scattering events, $n = 1$ is for direct) in x from direction $v \equiv x_{n-1}^{(F)} \to x$. We then obtain

$$\begin{aligned} \langle\langle C^2\rangle\rangle &\approx S^{-1} \int w_0^2(\boldsymbol{\mathcal{X}}_{M+1}) E^2(\boldsymbol{\mathcal{X}}_M) f^2(x_{M+1} \to x_M, x_M \to x_{M-1}, x_M) \\ &\qquad \times J(x_{M+1} \to x_M, x_M) p_B(\boldsymbol{\mathcal{X}}_M) s(x_M, x_{M+1}) d^2 \boldsymbol{\mathcal{X}}_{M+1} \\ &\quad + S^{-1} \int w_1^2(\boldsymbol{\mathcal{X}}_{M+1}) b(x_M, x_{M+1}) E^2(\boldsymbol{\mathcal{X}}_{M+1}) p_B(\boldsymbol{\mathcal{X}}_{M+1}) d^2 \boldsymbol{\mathcal{X}}_{M+1} \end{aligned} \quad (12)$$

where

$$I \equiv \sum_{n\geq 2} I_n \tag{13}$$

is the angular density of the *full* diffuse illumination,

$$J(v,x) \equiv |(v \cdot n(x))|\, I(v,x) \tag{14}$$

is *diffuse* irradiance and

$$b(x_M, x_{M+1}) \equiv \int f^2(v', x_{M+1} \to x_M, x_{M+1}) J(v', x_{M+1}) d^2 v' \tag{15}$$

Notice this integral is over the *diffuse* illumination only.

Introducing

$$Q(\boldsymbol{\mathcal{X}}_{M+1}) \equiv E(\boldsymbol{\mathcal{X}}_M) E(\boldsymbol{\mathcal{X}}_{M+1}) p_B(\boldsymbol{\mathcal{X}}_{M+1}) \tag{16}$$

and applying to it the recurrence relation (1) one obtains two useful identities:

$$\begin{aligned} E(\boldsymbol{\mathcal{X}}_{M+1}) p_B(\boldsymbol{\mathcal{X}}_{M+1}) &= E^{-1}(\boldsymbol{\mathcal{X}}_M) Q(\boldsymbol{\mathcal{X}}_{M+1}) & (17)\\ E^2(\boldsymbol{\mathcal{X}}_{M+1}) p_B(\boldsymbol{\mathcal{X}}_{M+1}) &= \mu(x_{M-1}, x_M) Q(\boldsymbol{\mathcal{X}}_{M+1}) & (18) \end{aligned}$$

Substituting (18) into (12) we arrive at

$$\begin{aligned} \langle\langle C^2 \rangle\rangle \approx\ & S^{-1} \int w_0^2(\boldsymbol{\mathcal{X}}_{M+1}) f(x_{M+1} \to x_M, x_M \to x_{M-1}, x_M) \\ & \times I(x_{M+1} \to x_M, x_M) Q(\boldsymbol{\mathcal{X}}_{M+1}) d^2 \boldsymbol{\mathcal{X}}_{M+1} \\ & + S^{-1} \int w_1^2(\boldsymbol{\mathcal{X}}_{M+1}) b(x_M, x_{M+1}) \mu(x_{M-1}, x_M) Q(\boldsymbol{\mathcal{X}}_{M+1}) d^2 \boldsymbol{\mathcal{X}}_{M+1} \end{aligned} \tag{19}$$

Then, since $w_0 = 1 - w_1$,

$$\begin{aligned} \langle\langle C^2 \rangle\rangle \approx\ & S^{-1} \int f(v, u, x_M) I(v, x_M) Q(\boldsymbol{\mathcal{X}}_{M+1}) d^2 \boldsymbol{\mathcal{X}}_{M+1} \\ & + S^{-1} \int w_1^2(\boldsymbol{\mathcal{X}}_{M+1}) \beta(x_{M-1}, x_M, x_{M+1}) I(v, x_M) Q(\boldsymbol{\mathcal{X}}_{M+1}) d^2 \boldsymbol{\mathcal{X}}_{M+1} \\ & - 2S^{-1} \int w_1(\boldsymbol{\mathcal{X}}_{M+1}) f(v, u, x_M) I(v, x_M) Q(\boldsymbol{\mathcal{X}}_{M+1}) d^2 \boldsymbol{\mathcal{X}}_{M+1} \end{aligned} \tag{20}$$

where

$$\beta(x_{M-1}, x_M, x_{M+1}) \equiv f(x_{M+1} \to x_M, x_M \to x_{M-1}, x_M) + \mu(x_{M-1}, x_M) \frac{b(x_M, x_{M+1})}{I(v, x_M)} \tag{21}$$

5.2 BMCRT term $\langle\langle C\rangle_F^2\rangle_B$

Averaging (3) over the light paths ensemble and using (11), (13), (6) and (14), we have for $S \to 0$ when the kernel is nearly a delta-function,

$$\langle C\rangle_F \approx \mathcal{L}_0(\boldsymbol{\mathcal{X}}_M) + E(\boldsymbol{\mathcal{X}}_M)G_0(\boldsymbol{\mathcal{X}}_M) + E(\boldsymbol{\mathcal{X}}_{M+1})w_1(\boldsymbol{\mathcal{X}}_{M+1})L(x_M, x_{M+1}) \tag{22}$$

where:

$$L_d(x_M, x_{M+1}) \equiv \int f(v, x_{M+1} \to x_M, x_{M+1})J(v, x_{M+1})d^2v \tag{23}$$

is luminance under *diffuse* illumination, $L = L_0 + L_d$ is the *full* luminance and

$$\begin{aligned} G_k(\boldsymbol{\mathcal{X}}_M) \;\equiv\; & \int w_k(\boldsymbol{\mathcal{X}}_{M+1})f(x_{M+1} \to x_M, x_M \to x_{M-1}, x_M) \\ & \times J(x_{M+1} \to x_M, x_M)s(x_M, x_{M+1})d^2x_{M+1} \end{aligned} \tag{24}$$

Squaring (22) and averaging over the BMCRT paths gives

$$\begin{aligned} \langle\langle C\rangle_F^2\rangle_B \;=\; & \int \mathcal{L}_0^2(\boldsymbol{\mathcal{X}}_M)p_B(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\ & + \int G_0^2(\boldsymbol{\mathcal{X}}_M)\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\ & + \int w_1^2(\boldsymbol{\mathcal{X}}_{M+1})L^2(x_M, x_{M+1})E^2(\boldsymbol{\mathcal{X}}_{M+1})p_B(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\ & + 2\int \mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)G_0(\boldsymbol{\mathcal{X}}_M)E(\boldsymbol{\mathcal{X}}_M)p_B(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\ & + 2\int w_1(\boldsymbol{\mathcal{X}}_{M+1})\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)L(x_M, x_{M+1})E(\boldsymbol{\mathcal{X}}_{M+1})p_B(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\ & + 2\int w_1(\boldsymbol{\mathcal{X}}_{M+1})G_0(\boldsymbol{\mathcal{X}}_M)L(x_M, x_{M+1})\mu^{-1}(x_{M-1}, x_M) \\ & \qquad \times E^2(\boldsymbol{\mathcal{X}}_{M+1})p_B(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \end{aligned}$$

where

$$\rho(\boldsymbol{\mathcal{X}}_M) \equiv E^2(\boldsymbol{\mathcal{X}}_M)p_B(\boldsymbol{\mathcal{X}}_M) \tag{25}$$

Applying (17) and (18) this becomes

$$\begin{aligned}
\langle\langle C\rangle_F^2\rangle_B &= \int \mathcal{L}_0^2(\boldsymbol{\mathcal{X}}_M)p_B(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\
&+ \int G_0^2(\boldsymbol{\mathcal{X}}_M)\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\
&+ \int w_1^2(\boldsymbol{\mathcal{X}}_{M+1})L^2(x_M,x_{M+1})\mu(x_{M-1},x_M)Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\
&+2\int G_0(\boldsymbol{\mathcal{X}}_M)E^{-1}(\boldsymbol{\mathcal{X}}_M)\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\
&+2\int w_1(\boldsymbol{\mathcal{X}}_{M+1})E^{-1}(\boldsymbol{\mathcal{X}}_M)\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)L(x_M,x_{M+1})Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\
&+2\int w_1(\boldsymbol{\mathcal{X}}_{M+1})G_0(\boldsymbol{\mathcal{X}}_M)L(x_M,x_{M+1})Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1}
\end{aligned}$$

Then,

$$\begin{aligned}
&\int G_0(\boldsymbol{\mathcal{X}}_M)E^{-1}(\boldsymbol{\mathcal{X}}_M)\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\
&\quad+\int w_1(\boldsymbol{\mathcal{X}}_{M+1})E^{-1}(\boldsymbol{\mathcal{X}}_M)\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)L(x_M,x_{M+1})Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\
&=\int E(\boldsymbol{\mathcal{X}}_M)\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)L_d(u,x_M)p_B(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M
\end{aligned}$$

(here we replaced L with I using (30) and recalled normalization $w_0+w_1=1$) we finally arrive at

$$\begin{aligned}
\langle\langle C\rangle_F^2\rangle_B &= \int \mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)\left(\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)+2E(\boldsymbol{\mathcal{X}}_M)L_d(u,x_M)\right)p_B(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\
&+ \int G_0^2(\boldsymbol{\mathcal{X}}_M)\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\
&+ \int w_1^2(\boldsymbol{\mathcal{X}}_{M+1})\mu(x_{M-1},x_M)I^2(v,x_M)Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\
&+2\int G_0(\boldsymbol{\mathcal{X}}_M)\left(\int w_1(\boldsymbol{\mathcal{X}}_{M+1})I(v,x_M)Q(\boldsymbol{\mathcal{X}}_{M+1})d^2x_{M+1}\right)d^2\boldsymbol{\mathcal{X}}_M
\end{aligned}$$

Applying the definition (24) to the inner integral in the 4th line we have

$$\begin{aligned}
\langle\langle C\rangle_F^2\rangle_B &= \int \mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)\left(\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)+2E(\boldsymbol{\mathcal{X}}_M)L_d(u,x_M)\right)p_B(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\
&+ \int \left(G_0(\boldsymbol{\mathcal{X}}_M)+G_1(\boldsymbol{\mathcal{X}}_M)\right)^2\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\
&+ \int w_1^2(\boldsymbol{\mathcal{X}}_{M+1})\mu(x_{M-1},x_M)I^2(v,x_M)Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\
&- \int G_1^2(\boldsymbol{\mathcal{X}}_M)\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M
\end{aligned}$$

From (24), and normalization $w_0 + w_1 = 1$ it follows that $G_0 + G_1 = L_d(x_M \to x_{M-1}, x_M)$ thus

$$\begin{aligned}\langle\langle C\rangle_F^2\rangle_B &= \int \mathcal{L}_0(\boldsymbol{\mathcal{X}}_M)\left(\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M) + 2E(\boldsymbol{\mathcal{X}}_M)L_d(u, x_M)\right) p_B(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\ &+ \int L_d^2(u, x_M)\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\ &+ \int w_1^2(\boldsymbol{\mathcal{X}}_{M+1})\mu(x_{M-1}, x_M)I^2(v, x_M)Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\ &- \int G_1^2(\boldsymbol{\mathcal{X}}_M)\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \end{aligned} \tag{26}$$

or

$$\begin{aligned}\langle\langle C\rangle_F^2\rangle_B &= \int \left(\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M) + E(\boldsymbol{\mathcal{X}}_M)L_d(u, x_M)\right)^2 p_B(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\ &+ \int w_1^2(\boldsymbol{\mathcal{X}}_{M+1})\mu(x_{M-1}, x_M)I^2(v, x_M)Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\ &- \int G_1^2(\boldsymbol{\mathcal{X}}_M)\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \end{aligned} \tag{27}$$

5.3 Resulting noise

Substituting (20) and (26) into (10) we have

$$\begin{aligned}N_B V &\approx n_F^{-1}\int f(v, u, x_M)I(v, x_M)Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\ &+ \int \left(\mathcal{L}_0(\boldsymbol{\mathcal{X}}_M) + E(\boldsymbol{\mathcal{X}}_M)L_d(u, x_M)\right)^2 p_B(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\ &- \left(n_F^{-1} + 1\right)\langle\langle C\rangle\rangle^2 \\ &+ \int w_1^2(\boldsymbol{\mathcal{X}}_{M+1})a(x_{M-1}, x_M, x_{M+1})\mu(x_{M-1}, x_M)I(v, x_M)Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \\ &- \int G_1^2(\boldsymbol{\mathcal{X}}_M)\rho(\boldsymbol{\mathcal{X}}_M)d^2\boldsymbol{\mathcal{X}}_M \\ &-2n_F^{-1}\int w_1(\boldsymbol{\mathcal{X}}_{M+1})f(v, u, x_M)I(v, x_M)Q(\boldsymbol{\mathcal{X}}_{M+1})d^2\boldsymbol{\mathcal{X}}_{M+1} \end{aligned} \tag{28}$$

where

$$\begin{aligned}n_F &\equiv SN_F \\ a(x_{M-1}, x_M, x_{M+1}) &\equiv n_F^{-1}\tilde{f}(v, u, x_M) + n_F^{-1}\frac{b(x_M, x_{M+1})}{L(x_M, x_{M+1})} + L(x_M, x_{M+1}) \end{aligned} \tag{29}$$

and $\tilde{f}$ was defined in (9). We also used the obvious relation that in absence of specular objects and volumetric absorption angular density of *secondary* (*diffuse*) illumination equals the *full* radiance of the surface point seen from that direction:

$$I(x_{M+1} \to x_M, x_M) = L(x_M, x_{M+1}) \tag{30}$$

6 OPTIMAL WEIGHTS

By definition, this are the weights which minimize the noise functional. In our case there is only one independent weight, let it be w_1. Variation of noise (28) with respect to it is

$$\begin{aligned}\frac{N_B}{2}\delta V \;\approx\; & \int \delta w_1 w_1(\boldsymbol{\mathcal{X}}_{M+1}) a(x_{M-1}, x_M, x_{M+1}) \mu(x_{M-1}, x_M) \\ & \qquad \times I(v, x_M) Q(\boldsymbol{\mathcal{X}}_{M+1}) d^2\boldsymbol{\mathcal{X}}_{M+1} \\ & - \int \delta G_1(\boldsymbol{\mathcal{X}}_M) G_1(\boldsymbol{\mathcal{X}}_M) \rho(\boldsymbol{\mathcal{X}}_M) d^2\boldsymbol{\mathcal{X}}_M \\ & - n_F^{-1} \int \delta w_1(\boldsymbol{\mathcal{X}}_{M+1}) f(v, u, x_M) I(v, x_M) Q(\boldsymbol{\mathcal{X}}_{M+1}) d^2\boldsymbol{\mathcal{X}}_{M+1}\end{aligned}$$

where here and below

$$\begin{aligned} u &\equiv x_M \to x_{M-1} \\ v &\equiv x_{M+1} \to x_M \end{aligned}$$

and

$$\delta G_1(\boldsymbol{\mathcal{X}}_M) \equiv \int \delta w_1(\boldsymbol{\mathcal{X}}_{M+1}) f(v, u, x_M) J(v, x_M) s(x_M, x_{M+1}) d^2 x_{M+1}$$

Expanding,

$$\begin{aligned}\frac{N_B}{2}\delta V \;\approx\; & \int \delta w_1(\boldsymbol{\mathcal{X}}_{M+1}) \left(w_1(\boldsymbol{\mathcal{X}}_{M+1}) a(x_{M-1}, x_M, x_{M+1}) \mu(x_{M-1}, x_M) \right. \\ & \qquad \left. \left| -n_F^{-1} f(v, u, x_M) - G_1(\boldsymbol{\mathcal{X}}_M) \right) \right. \\ & \qquad \times I(v, x_M) Q(\boldsymbol{\mathcal{X}}_{M+1}) d^2\boldsymbol{\mathcal{X}}_{M+1}\end{aligned}$$

The optimal weight is the one for which the functional V reaches its minimum, i.e. $\delta V = 0$ for an arbitrary δw_1. This happens if and only if

$$w_1(\boldsymbol{\mathcal{X}}_{M+1}) \left(n_F^{-1} \beta(x_{M-1}, x_M, x_{M+1}) + \mu(x_{M-1}, x_M) I(v, x_M) \right) = n_F^{-1} f(v, u, x_M) + G_1(\boldsymbol{\mathcal{X}}_M)$$

or

$$w_1(\boldsymbol{\mathcal{X}}_{M+1})\left(n_F^{-1}\tilde{f}(v,u,x_M)+n_F^{-1}\frac{b(x_M,x_{M+1})}{I(v,x_M)}+I(v,x_M)\right) = n_F^{-1}\tilde{f}(v,u,x_M) + \frac{G_1(\boldsymbol{\mathcal{X}}_M)}{\mu(x_{M-1},x_M)}$$

Let

$$\begin{aligned} a(x_{M-1},x_M,x_{M+1}) &\equiv n_F^{-1}\tilde{f}(v,u,x_M)+n_F^{-1}\frac{b(x_M,x_{M+1})}{L(x_M,x_{M+1})}+L(x_M,x_{M+1}) \\ W_1(v,u,x_M) &\equiv \frac{n_F^{-1}\tilde{f}(v,u,x_M)}{a(x_{M-1},x_M,x_{M+1})} \end{aligned} \tag{31}$$

then

$$w_1(x_0,...,x_{M+1}) = W_1(v,u,x_M)+\frac{\tilde{G}_1(x_0,...,x_M)}{a(x_{M-1},x_M,x_{M+1})}$$

where

$$\begin{aligned} \tilde{G}_1(x_0,...,x_M) &\equiv \int w_1(x_0,...,x_{M+1})\tilde{f}(v,u,x_M)J(v,x_M)s(x_M,x_{M+1})d^2x_{M+1} \\ &= \int w_1(x_0,...,x_{M+1})\tilde{f}(v,u,x_M)J(v,x_M)d^2v \\ v &\equiv x_{M+1}\to x_M \end{aligned}$$

Substituting this w_1 into the above definition of $\tilde{G}_1(x_0,...,x_M)$ we have

$$\begin{aligned} \tilde{G}_1(x_0,...,x_M) = &\int W_1(v,u,x_M)\tilde{f}(v,u,x_M)J(v,x_M)d^2v \\ &+n_F\tilde{G}_1(x_0,...,x_M)\int W_1(v,u,x_M)J(v,x_M)d^2v \end{aligned}$$

from what it follows that

$$\tilde{G}_1(x_0,...,x_M)=\frac{A(u,x_M)}{1-n_FB(u,x_M)}$$

where

$$A(u,x_M) \equiv \int W_1(v,u,x_M)\tilde{f}(v,u,x_M)J(v,x_M)d^2v \tag{32}$$

$$B(u,x_M) \equiv \int W_1(v,u,x_M)J(v,x_M)d^2v \tag{33}$$

Finally

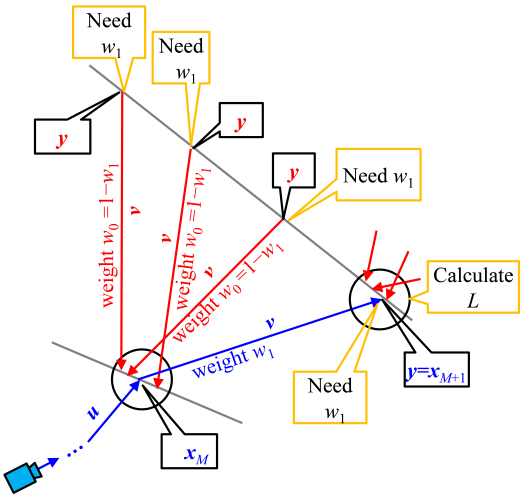


Figure 1: Use of weights in the limited MIS. Blue color relates to camera ray, red color relates to diffuse light ray.

$$w_1(x_{M-1}, x_M, x_{M+1}) = W_1(v, u, x_M) + \frac{A(u, x_M)}{a(x_{M-1}, x_M, x_{M+1})(1 - n_F B(u, x_M))} \tag{34}$$

and it depends only on three vertices (x_{M-1}, x_M, x_{M+1}) of the join path.

7 CALCULATION OF WEIGHT IN RAY TRACING

As seen from (3), for kernel which is S^{-1} inside integration sphere the camera path $(x_0, ..., x_{M+1})$ after interaction with *all* light photons increases the luminance of pixel by

$$\begin{aligned} \Delta L \;=\; & \sum_{m=0}^{M} E(x_0, ..., x_m) L_0(x_{m-1}, x_m) \\ & + E(x_0, ..., x_M) \sum_p w_0(x_{M-1}, x_M, x_p) f(v_p, x_M \to x_{M-1}, x_M) \\ & + w_1(x_{M-1}, x_M, x_{M+1}) E(x_0, ..., x_{M+1}) L(x_M, x_{M+1}) \end{aligned} \tag{35}$$

where $\sum_p$ is over the *diffuse* FMCRT ***p***hotons (x_p, v_p) inside the integration sphere around x_M and $L(x_M, x_{M+1})$ is estimation of the *full* radiance at x_{M+1} towards x_M (diffuse component is from photon maps, the direct one being calculated deterministically), calculated as it would be without weights. This use of weights is schematically shown in Figure 1.

We therefore need weight $w_1(x_{M-1}, x_M, y)$ for the the following y:

1. Hit point of the *camera* ray after scattering at x_M
2. *Previous* hitpoint of all *diffuse* FMCRT photons which hit the integration sphere about x_M

For each y the weight is given by (34) which includes *integrals* $b(x_M, y)$ and $L(x_M, y)$. As seen from (15) and (23), they can be estimated with Monte-Carlo method from photon maps:[2]

$$b(x_M, y) \approx \frac{1}{SN_F} \sum_p f^2(v_p, y \to x_M; y) \tag{36}$$

$$L(x_M, y) \approx L_0(x_M, y) + \frac{1}{SN_F} \sum_p f(v_p, y \to x_M; y) \tag{37}$$

where the sums are over the *diffuse* FMCRT $\boldsymbol{p}$hotons (x_p, v_p) inside the integration sphere around y. These calculations are very much similar to the usual estimation of surface radiance from photon maps. Then (29), (31) give us the base component of weight $W_1(v, u, x_M)$ for this set of directions $v = y \to x_M$.

Regrettably their result is inevitably *noisy* while (as said in the very beginning) the weights must be *deterministic* functions of the join path so that the weight calculated for given (x_{M-1}, x_M, y) must be *the same* during all the MCRT process. Meanwhile in BDPM the photon maps change from iteration to iteration, thus had that same trajectory (x_{M-1}, x_M, y) been encountered latter, the weight calculated for it would be different.

A simple remedy is to *freeze* the photon map used for calculation of integrals in the weight formula so that it is the same for all iterations. For example, we can always use photon map from the 1st iteration. In this case $L(x_M, x_{M+1})$ is calculated differently: for the 3rd line of (35) it is calculated from the "main" photon maps of current iteration while for the weight it is calculated from the *frozen* map.

The scheme of calculations is shown in Figure 2 (left).

Besides, we need A and B which are independent from y. The integrals (32) and (33) with respect to the measure $J(v, x_M)d^2v$ proportional to the number of the incident photons per surface area also can be estimated with the Monte-Carlo method as:

$$A \approx \frac{1}{SN_F} \sum_p W_1(v_p, u, x_M) \tilde{f}(v_p, u, x_M) \tag{38}$$

$$B \approx \frac{1}{SN_F} \sum_p W_1(v_p, u, x_M) \tag{39}$$

where the sums are again over the FMCRT photons (x_p, v_p) which hit the integration sphere around x_M.

Again to make the result deterministic we use photons from the frozen map. Notice we need W_1 another set of directions than those in (36), (37) because there v_p was from the "main" photon map while now it is from the frozen photon map.

We therefore need to calculate W_1 there, too. This is by the same (36), (37), only for y which is now the previous hit from not the main but the frozen map.

The scheme of calculations is shown in Figure 2 (right).

[2]The full radiance L, can be calculated either completely from photon maps (using both *direct* and diffuse photons), or its *direct* component L_0 can be calculated separately while and only the diffuse component (23) is taken from the photon maps

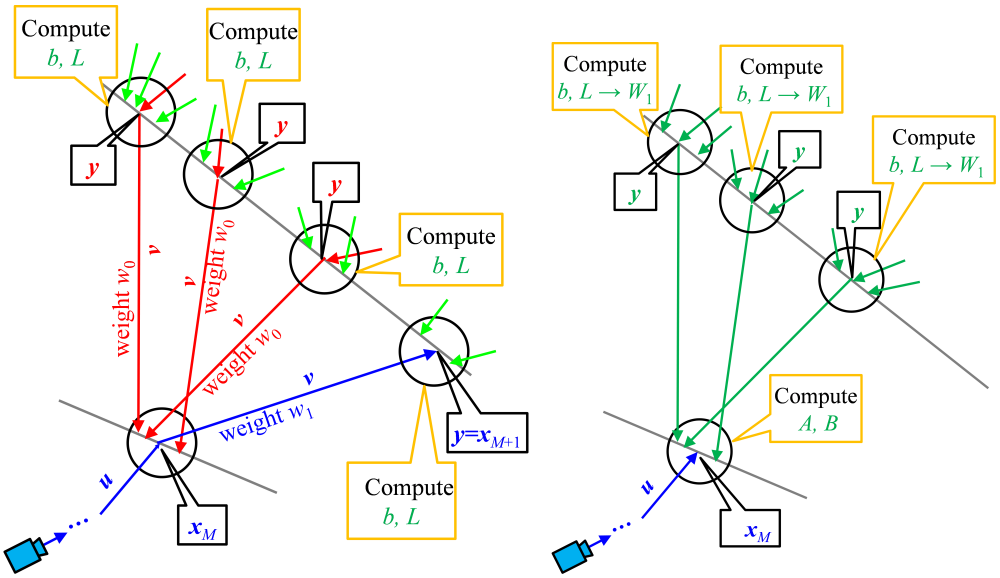


Figure 2: **Left**: Calculation of weights assuming A and B had been already calculated. Blue color relates to camera ray, red color relates to light ray from current photon map and green color relates to light ray from frozen photon map. In each integration sphere we compute b, L from the *frozen* photon map and then calculate w_1 from them. **Right**: Calculation of A and B from the frozen photon map. In each integration sphere we compute b, L from the *frozen* photon map, then calculate W_1 from them and eventually average into A and B.

8 CONCLUSION

The "full" MIS requires that all vertices in the join path be used for intersection of light and camera rays. It happens that even for very simple scenes the problem is very sophisticated. We therefore proposed a compromise sub-optimal approach when only two weights can be not 0. These strategies are: either terminate camera ray at the given vertex or continue it by yet one segment. Then, we treat the weights of these two strategies as functions of the camera path and the last light path vertex but not of the "early" part of light path. In this case the optimal weights obey a linear integral equation that admits solution in close form, i.e. the solution is an analytic formula which though depends on some integrals that must be calculated numerically. We demonstrate how they can be calculated using FMCRT in presence of the direct illumination to pixel luminance. The algorithm does not require too sophisticated calculations and is applicable in practice.

Unlike the widely used Veach heuristic [2], [10] the optimal MIS weights for BDPM algorithm are not local and depend on the scene properties in points outside the current path. Besides, they depend not only on BDFs and distribution of light source emission but also on the number of forward (light) paths traced in one BDPM iteration and on the area of the integration sphere.

REFERENCES

[1] M. Pharr and G. Humphreys, *Physically Based Rendering: From Theory To Implementation*, 2nd ed., Morgan Kaufmann Publishers Inc., CA, USA (2010).

[2] E. Veach, *Robust Monte-Carlo methods for light transport simulation*, Dissertation, Stanford Univ. (1997).

[3] M. Sbert, V. Havran, and L. Szirmay-Kalos, "Multiple importance sampling revisited: breaking the bounds", *EURASIP Journal on Advances in Signal Processing*, 15(1–15) (2018).

[4] H. W. Jensen, "Global illumination using photon maps", *Proceedings of the Eurographics Workshop on Rendering Techniques '96, London, UK*, 21–30 (1996).

[5] H. W. Jensen and P. Christensen, "High quality rendering using ray tracing and photon mapping", *ACM SIGGRAPH 2007 Courses, New York, NY, USA*, Course 8 (1–130) (2007).

[6] N. Dodik, "Implementing probabilistic connections for bidirectional path tracing in the Mitsuba Renderer", https://www.cg.tuwien.ac.at/research/publications/2017/dodik-2017-pcbpt/ (Accessed July 17, 2020).

[7] S. V. Ershov, D. D. Zhdanov, and A. G. Voloboy, "Estimation of noise in calculation of scattering medium luminance by MCRT", *Math. Montisnigri*, **45**, 60–73 (2019).

[8] J. T. Kajiya, "The rendering equation", *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '86*, 143–150 (1986).

[9] E. Veach and L. J. Guibas, "Optimally combining sampling techniques for Monte Carlo rendering", *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '95*, 419–428 (1995).

[10] J. Vorba, "Bidirectional photon mapping", *Proceedings of CESCG 2011: The 15th Central European Seminar on Computer Graphics, Prague*, 25–32 (2011).

[11] T. Hachisuka, J. Pantaleoni, and H. W. Jensen, "A path space extension for robust light transport simulation", *ACM Trans. Graph.*, **31**(6), 191(1-10) (2012). DOI: 10.1145/2366145.2366210.

[12] I. Georgiev, *Implementing vertex connection and merging*, Tech. rep., Saarland Univ. (2012).

[13] I. Georgiev, J. Křivánek, T. Davidovič, and P. Slusallek, "Light transport simulation with vertex connection and merging", *ACM Trans. Graph.*, **31**(6), 192(1–10) (2012).

[14] S. Popov, R. Ramamoorthi, F. Durand, and G. Drettakis, "Probabilistic Connections for Bidirectional Path Tracing", *Computer Graphics Forum*, **34**(4), 75–86 (2015).

[15] D. van Antwerpen, *Recursive MIS computation for streaming bdpt on the GPU*, Tech. rep., Delft Univ. of Technology (2012).

[16] Q. Liu and Y. Zhang, "Light transport simulation via generalized multiple importance sampling", arXiv:1803.04305v2 [cs.GR] (2018).

[17] V. Frolov, A. Kharlamov, V. Galaktionov, and K. Vostryakov, "Multiple reference octrees for a GPU photon mapping and irradiance caching", *Programming and Computer Software*, **40**(4), 208–214 (2014).

[18] S. V. Ershov, D. D. Zhdanov, and A. G. Voloboy, "Calculation of MIS weights for bidirectional path tracing with photon maps", *Proceeding of conference on Computing for Physics and Technology*, Moscow, 20(1–5) (2020).

Received June 20, 2020

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

FUZZY METRIC SPACE AND APPLICATIONS IN IMAGE PROCESSING

NEBOJŠA M. RALEVIĆ[1*], MARIJA PAUNOVIĆ[2], AND BRATISLAV IRIČANIN[3,4]

[1] University of Novi Sad, Faculty of Technical Sciences,
Department of Fundamental Sciences. Novi Sad, Serbia
[2] University of Kragujevac, Faculty of Hotel Management and Tourism in Vrnjačka Banja.
Vrnjačka Banja, Serbia
[3] University of Beograd, School of Electrical Engineering. Belgrade, Serbia
[4] University of Kragujevac, Faculty of Mechanical and Civil Engineering in Kraljevo, Kraljevo, Serbia
*Corresponding author. E-mail: nralevic@uns.ac.rs

DOI:10.20948/mathmontis-2020-48-9

Summary. In this paper, the notions of fuzzy T-metric and fuzzy S-metric have been introduced, and then, examples of known fuzzy metrics are provided, as well as theorems that enable algorithms for putting up new metrics. Recently, there has been renewed interest in some of their properties, which are further shown, these being polygonal inequality, and two new classes of functions have been shown to be regarded as fuzzy metric. By applying these fuzzy metrics, an algorithm has been used in order to remove image noise. The goal was to improve the sharpness and the quality of the image, which is expressed and measured by means of the image quality index UIQI. It has been shown that the general image, which is filtered by this algorithm, has greater sharpness than the image filtered by the median filter, which is probably the most commonly used vector filter.

1 INTRODUCTION

There is evidence that the notion of a probabilistic metric space plays a crucial role in regulating a generalization of metric space in which the distance between x and y is replaced by the distribution function $F_{x,y}$ where the value of that distribution function at some point a is interpreted as the probability in which case, the distance between x and y is smaller of a. A key aspect of and the introduction of the concept of fuzzy set was proposed by L. Zadeh in 1965.

In recent years, there has been an increasing interest in fuzzy logic and recent developments in fuzzy sets have heightened the need for initiating wider development of the theory of fuzzy logic and fuzzy sets, as well as other mathematical fields where they were meant to replace standard sets. Characterization of fuzzy sets is important for our increased understanding of fuzzy logic and its consequences, and in this manner the theory of fuzzy metric spaces has evolved, where the distance between elements x and y is a fuzzy set with certain properties. There are several important areas where this theory makes an original contribution to finding increasing application particularly to specific practical problems one of which is image processing.

The whole concept of thinking that a distance between two objects is no longer a classical, but a fuzzy set, has led to further development of theoretical parts of mathematics in which the has term appeared, as well as the application of a part of this knowledge in various engineering sciences, even in medicine' sciences. Perhaps the best example is provided by the functional analysis, where the counterparts are defined by the classical notions of a sphere,

2010 Mathematics Subject Classification: 03E72, 54E25, 62H35, 68U10.
Key words and Phrases: Fuzzy metric, Image filtering, Triangular conorms and Triangular norms.

neighborhood, convergence, etc., in fuzzy metric spaces. It is in this manner that the theory of a fixed point in the complete fuzzy metric spaces is considered as a natural continuation of research in this field in probabilistic metric spaces (see e.g. [2]).

Practical problems in which distance was used and their solution with fuzzy distance have obtained a new meaning of problem interpretation and the way of solving it. This is perhaps best seen in shape recognition, shape analysis and image processing, which have direct applications in practice. For example, classifying data into clusters, in which case each of them contains objects having the maximum of similarity, i.e. the minimum distance between each other, with as many objects from other clusters as possible.

Image processing and the problems considered in this area, such as, for example, filtering and segmentation of the image, abundantly use the distance in terms of similarities, i.e. differences of images, i.e. parts of images, and the "ambiguity" itself directly indicates the need for fuzzy distances.

The paper is organized as follows. The second section presents the basic notions of t−norm, t−conorms, fuzzy complements and their properties which will be required when proving the properties of fuzzy metrics (see [9], [10]). In the third section, the fuzzy S−metric and the fuzzy T−metric are defined and the examples relevant to the application are provided. Polygonal inequality has also been proven and two new classes of set functions have been defined, and the findings should make an important contribution to the field of fuzzy metrics. The properties by which new fuzzy metrics can be defined from existing classes are presented. Evidence suggests that fuzzy metric parameter values are among the most important factors for enabling selection of the metric that further offers the best performance in image filtering. In the section 4, the values of the described parameters are further experimentally determined. Ultimately, the main aim of this study is to investigate the differences between the quality of image filtering, utilizing our algorithm and filtering using VMF both of which, were calculated by means of the UIQI metric, which has been defined in [21], and they are subsequently compared.

2 PRELIMINARIES

This paper sets out to assess the effects of triangular norms and conorms and thus, definitions of triangular norms and conorms and some of their properties are taken into consideration and provided in the references (see [9], [10]).

Definition 2.1. For binary operation $N : [0,1]^2 \to [0,1]$ which satisfies the following axioms for all $a,b,c,d \in [0,1]$:

n_1) $N(a,e) = a$ (boundary condition);

n_2) $c \le d \Rightarrow N(a,c) \le N(a,d)$ (monotonicity);

n_3) $N(a,b) = N(b,a)$ (commutativity);

n_4) $N(a,N(b,c)) = N(N(a,b),c)$ (associativity).

we say that N is *norm*.

If $e = 1$, then N is the *triangular norm* (shorter t−norm), and instead of N we write T. If $e = 0$, then N is the *triangular conorm* (shorter t−conorm), and instead of N we write S.

A norm N, is an *Archimedean* norm if N is continuous and for all $a \in (0,1)$, $N(a,a) < a$ for t−norm N and for all $a \in (0,1)$, $N(a,a) > a$ for t−conorm N.

As $b \le 1$, $T(a,b) \le T(a,1) = a$, and similarly $T(a,b) \le b$, imply

$$T(a,b) \le min\{a,b\}. \tag{1}$$

As $b \ge 0$, $S(a,b) \ge S(a,0) = a$, and similarly $S(a,b) \ge b$, imply

$$S(a,b) \ge max\{a,b\}. \tag{2}$$

Remark 1. From the conditions given in the definition of the norm follows the monotonicity by coordinates, i.e. for all $a_1,a_2,b_1,b_2 \in [0,1]$

$$a_1 \le a_2 \wedge b_1 \le b_2 \Rightarrow N(a_1,b_1) \le N(a_2,b_2). \tag{3}$$

It is necessary here to clarify exactly what is meant by the monotonic condition replacement. Namely, replacing the monotonic condition in Definition 2.1 by condition (3), an equivalent definition of the norm is obtained.
If, in the definition of the norm, instead of the axiom of monotonicity, a strict monotonicity is valid, i.e.

$$a_1 < a_2 \wedge b_1 < b_2 \Rightarrow N(a_1,b_1) < N(a_2,b_2),$$

for all $a_1,a_2,b_1,b_2 \in [0,1]$, then the norm is *strict*.

Definition 2.2. A *decreasing generator g* is a continuous strictly decreasing function from [0,1] to $\mathbb{R}$, such that $g(1) = 0$.

An *increasing generator g* is a continuous strictly increasing function from [0,1] to $\mathbb{R}$, such that $g(0) = 0$.

Definition 2.3. *The power of the norm* is given by formulas:

$$N^1(a_1,a_2) = N(a_1,a_2),\ N^n(a_1,...,a_n,a_{n+1}) = N\left(N^{n-1}(a_1,...,a_n),a_{n+1}\right)\ (n \ge 2).$$

Remark 2. If T is a t−norm, then:

$$T(a_1,a_2) = 1 \Leftrightarrow a_1 = a_2 = 1,$$
$$T^n(a_1,a_2,...,a_{n+1}) = 1 \Leftrightarrow a_1 = ... = a_{n+1} = 1.$$

Remark 3. If T is a strict t−norm, then:

$$T(a_1,a_2) = 0 \Leftrightarrow a_1 = 0 \vee a_2 = 0,$$
$$T^n(a_1,a_2,...,a_{n+1}) = 0 \Leftrightarrow a_1 = 0 \vee ... \vee a_{n+1} = 0.$$

Remark 4. If S is a t−conorm, then:

$$S(a_1,a_2) = 0 \Leftrightarrow a_1 = a_2 = 0.$$
$$S^n(a_1,a_2,...,a_{n+1}) = 0 \Leftrightarrow a_1 = ... = a_{n+1} = 0.$$

Remark 5. If S is a strict t−conorm, then:

$$S(a_1,a_2) = 1 \Leftrightarrow a_1 = 1 \vee a_2 = 1.$$
$$S^n(a_1,a_2,...,a_{n+1}) = 1 \Leftrightarrow a_1 = 1 \vee ... \vee a_{n+1} = 1.$$

Definition 2.4. The function $c : [0,1] \to [0,1]$ is a *fuzzy complement*, if following conditions are satisfied:

c_1) $c(0) = 1$ i $c(1) = 0$, (boundary conditions)

c_2) $(\forall a,b \in [0,1])\ a \leq b \Rightarrow c(a) \geq c(b)$ (monotonicity).

If $c(c(a)) = a$ holds for all $a \in [0,1]$, then a function c is *involutive*.

If c is a continuous function, then we say that *c is a continuous fuzzy complement*.

If $c : [0,1] \to [0,1]$ is an involutive monotonic non-increasing function, it follows that c is a continuous bijective function for which boundary conditions are valid (see [10]).

The triangular norm T and the triangular conorm S are *dual* with respect to the fuzzy complement c if and only if

$$c\left(T\left(a,b\right)\right) = S\left(c\left(a\right),c\left(b\right)\right) \text{ and } c\left(S\left(a,b\right)\right) = T\left(c\left(a\right),c\left(b\right)\right).$$

(T,S,c) is called a *dual triple*.

For the triangular norm T and the involutive fuzzy complement c, the binary operation S on $[0,1]$ defined with

$$S\left(a,b\right) = c\left(T\left(c\left(a\right),c\left(b\right)\right)\right)$$

for all $a,b \in [0,1]$, is a triangular conorm S such that (T,S,c) is a dual triple.

3 FUZZY METRICS

This section presents the findings of the research, focusing on the fuzzy S−metric and the fuzzy T−metric, which are being considered. Some of the new characteristics of the T−fuzzy metrics are presented. The main issues addressed in this section of the paper refer to the well known notions and characteristics and in the literature, these terms tend to be used to refer to the notions given and proven in the papers [6], [7], and [19].

Definition 3.1. [19] Let $X \neq \emptyset$,

i) S be a continuous t- conorm,

ii) T be a continuous t- norm,

and $\mathbf{d}$ is a fuzzy set defined on $X \times X \times (0,+\infty)$, that satisfies the following conditions for all $x,y,z \in X, \alpha,\beta > 0$:

(1) i) $\mathbf{d}(x,y,\alpha) \in [0,1)$, ii) $\mathbf{d}(x,y,\alpha) \in (0,1]$;

(2) i) $\mathbf{d}(x,y,\alpha) = 0 \Leftrightarrow x = y$, ii) $\mathbf{d}(x,y,\alpha) = 1 \Leftrightarrow x = y$;

(3) i), ii) $\mathbf{d}(x,y,\alpha) = \mathbf{d}(y,x,\alpha)$;

(4) i) $S(\mathbf{d}(x,y,\alpha),\mathbf{d}(y,z,\beta)) \geq \mathbf{d}(x,z,\alpha + \beta)$, ii) $T(\mathbf{d}(x,y,\alpha),\mathbf{d}(y,z,\beta)) \leq \mathbf{d}(x,z,\alpha + \beta)$;

(5) i), ii) $\mathbf{d}(x,y,\_) : (0,+\infty) \to [0,1]$ is a continuous function.

The fuzzy set $\mathbf{d}$ is called

i) a *fuzzy S−metric* and a triple $(X,\mathbf{d},S)$ is *fuzzy S−metric space* (where $\mathbf{d}$ satisfies (1-5)i)) ;

ii) a *fuzzy T−metric* and a triple $(X,\mathbf{d},T)$ is *fuzzy T−metric space* (where $\mathbf{d}$ satisfies (1-5) ii)).

If instead of (1), it holds that $\mathbf{d}(x,y,\alpha) \in [0,1]$, the fuzzy set $\mathbf{d}$ is a *fuzzy S−metric* (*fuzzy T−metric*) *in the broader sense*, and $(X,\mathbf{d},S)$ ($(X,\mathbf{d},T)$) is a *fuzzy S−metric* (*fuzzy T−metric*) *space in the broader sense*.

We will mark the fuzzy S−metric with $\mathbf{s}$ and the fuzzy T−metric with $\mathbf{t}$, and we will write

the mark **d** if some statement is valid in both cases and use the term fuzzy metric.

Definition 3.2. [19] Fuzzy metric **d** is *stationary* on X if **d** does not depend of α, i.e. if for all fixed $x,y \in X$, the function $\mathbf{d}_{x,y}(\alpha) = \mathbf{d}(x,y,\alpha)$ is a constant.

Remark 6. Fuzzy S−metric $\mathbf{s}(x,y,\_) : (0,+\infty) \to [0,1]$ is non-decreasing function, and fuzzy T−metric $\mathbf{t}(x,y,\_) : (0,+\infty) \to [0,1]$ is non-increasing function.

Putting that $y = z$ in the triangle inequality

$$S\left(\mathbf{s}(x,y,\alpha),\mathbf{s}(y,y,\beta)\right) = S\left(\mathbf{s}(x,y,\alpha),0\right) = \mathbf{s}(x,y,\alpha) \geq \mathbf{s}(x,y,\alpha+\beta),$$

supposing $0 < \alpha_1 < \alpha_2$, for $\alpha = \alpha_1$, and $\beta = \alpha_2 - \alpha_1$, imply $\mathbf{s}(x,y,\alpha_1) \geq \mathbf{s}(x,y,\alpha_2)$, i.e. $\mathbf{s}(x,y,\_)$ is non-increasing function.

Putting that $y = z$ in the triangle inequality

$$T\left(\mathbf{t}(x,y,\alpha),\mathbf{t}(y,y,\beta)\right) = T\left(\mathbf{t}(x,y,\alpha),1\right) = \mathbf{t}(x,y,\alpha) \leq \mathbf{t}(x,y,\alpha+\beta),$$

supposing $0 < \alpha_1 < \alpha_2$, for $\alpha = \alpha_1$, and $\beta = \alpha_2 - \alpha_1$, imply $\mathbf{t}(x,y,\alpha_1) \leq \mathbf{t}(x,y,\alpha_2)$, i.e. $\mathbf{t}(x,y,\_)$ is non-decreasing function.

Remark 7. The triangle inequality, is trivial satisfied if $x = y$ or $y = z$ or $x = z$.
Indeed, for example (4) i)

$$x = z \Rightarrow S\left(\mathbf{d}(x,y,\alpha),\mathbf{d}(y,x,\beta)\right) \geq \mathbf{d}(x,x,\alpha+\beta) = 0 \Leftrightarrow \top,$$
$$x = y \Rightarrow S\left(\mathbf{d}(x,x,\alpha),\mathbf{d}(x,z,\beta)\right) = S\left(0,\mathbf{d}(x,z,\beta)\right) = \mathbf{d}(x,z,\beta) \geq \mathbf{d}(x,z,\alpha+\beta) \Leftrightarrow \top,$$
$$y = z \Rightarrow S\left(\mathbf{d}(x,y,\alpha),\mathbf{d}(y,y,\beta)\right) = S\left(\mathbf{d}(x,y,\alpha),0\right) = \mathbf{d}(x,y,\alpha) \geq \mathbf{d}(x,y,\alpha+\beta) \Leftrightarrow \top.$$

Theorem 3.1. *If* $\mathbf{d} : X \times X \times (0,+\infty) \to \mathsf{R}$ *is a fuzzy T−metric with respect to the norm T, then polygonal inequality, for* $n \geq 2$*, hold:*

$$T^{n-1}\left(\mathbf{d}(x_1,x_2,\alpha_1),\mathbf{d}(x_2,x_3,\alpha_2),...,\mathbf{d}(x_n,x_{n+1},\alpha_n)\right) \leq \mathbf{d}(x_1,x_{n+1},\alpha_1+\alpha_2+...+\alpha_n). \quad (4)$$

For case S−metric with respect to the conorm S, for $n \geq 2$*, hold:*

$$S^{n-1}\left(\mathbf{d}(x_1,x_2,\alpha_1),\mathbf{d}(x_2,x_3,\alpha_2),...,\mathbf{d}(x_n,x_{n+1},\alpha_n)\right) \geq \mathbf{d}(x_1,x_{n+1},\alpha_1+\alpha_2+...+\alpha_n). \quad (5)$$

Proof. Let us show the polygonal inequality for the case of the fuzzy T−metric. It is analogous to the fuzzy S−metric as well. If $n = 2$, i.e.,

$$T\left(\mathbf{d}(x_1,x_2,\alpha_1),\mathbf{d}(x_2,x_3,\alpha_2)\right) \leq \mathbf{d}(x_1,x_3,\alpha_1+\alpha_2),$$

statement is valid, because axiom (4)ii). Suppose the claim holds for $n = k$ and prove it for $n = k + 1$:

$$T^{k-1}\left(\mathbf{d}(x_1,x_2,\alpha_1),\mathbf{d}(x_2,x_3,\alpha_2),...,\mathbf{d}(x_k,x_{k+1},\alpha_k)\right) \leq \mathbf{d}(x_1,x_{k+1},\alpha_1+\alpha_2...+\alpha_k)$$
$$\Rightarrow T^{k}\left(\mathbf{d}(x_1,x_2,\alpha_1),\mathbf{d}(x_2,x_3,\alpha_2),...,\mathbf{d}(x_k,x_{k+1},\alpha_k),\mathbf{d}(x_{k+1},x_{k+2},\alpha_{k+1})\right)$$
$$= T\left(T^{k-1}\left(\mathbf{d}(x_1,x_2,\alpha_1),\mathbf{d}(x_2,x_3,\alpha_2),...,\mathbf{d}(x_k,x_{k+1},\alpha_k)\right),\mathbf{d}(x_{k+1},x_{k+2},\alpha_{k+1})\right)$$
$$\leq T\left(\mathbf{d}(x_1,x_{k+1},\alpha_1+\alpha_2+...+\alpha_k),\mathbf{d}(x_{k+1},x_{k+2},\alpha_{k+1})\right)$$
$$\leq d\left(x_1,x_{k+2},\alpha_1+\alpha_2+...+\alpha_k+\alpha_{k+1}\right). \qquad \square$$

Theorem 3.2. [19] *If* $(X,\mathbf{s},S)$ *is a fuzzy S−metric space and the T is a t−norm dual to the t−conorm S with respect to the continuous involutive fuzzy complement c, then* $(X,c\circ\mathbf{s},T)$ *is a fuzzy T−metric space.*

If $(X,\mathbf{t},T)$ *is a fuzzy T−metric space and S is a t−conorm dual to the norm T with respect to a continuous involutive fuzzy complement c, then* $(X,c\circ\mathbf{t},S)$ *is a fuzzy S−metric space.*

The theorem is also valid for fuzzy metric spaces in the broader sense.

Example 1. [6, 19] The mapping $\mathbf{t}_K:\mathbb{R}^+\times\mathbb{R}^+\to\mathbb{R}$ defined by $\mathbf{t}_K(x,y)=\dfrac{\min\{x,y\}+K}{\max\{x,y\}+K}$, where $K>0$, is a fuzzy T-metric with respect to multiplication, and $\mathbf{s}_K(x,y)=\dfrac{|x-y|}{\max(x,y)+K}$ is a fuzzy S-metric with respect to the algebraic sum, $S(x,y)=1-(1-x)(1-y)=x+y-xy$, dual to T with respect to the standard fuzzy complement.

Example 2. Let $f:X\to\mathbb{R}^+$ be one-to-one function and let $g:\mathbb{R}^+\to[0,\infty]$ be an increasing continuous function. The mapping $\mathbf{t}:X\times X\times\mathbb{R}^+\to[0,1]$ defined by

$$\mathbf{t}(x,y,\alpha)=\left(\frac{\frac{(f(x))^p+(f(y))^p}{2}+g(\alpha)}{(\max\{f(x),f(y)\})^p+g(\alpha)}\right)^q, \tag{6}$$

where $p,q>0$ fixed, is a fuzzy T-metric with respect to multiplication.

Proof. $f(x),f(y)\in\mathbb{R}^+, g(\alpha)\geq 0$. Without loss of generality, let $f(x)\leq f(y)$. Then $(f(x))^p\leq(f(y))^p$, i.e.,

$$(f(x))^p+(f(y))^p\leq(f(y))^p+(f(y))^p=2(f(y))^p=2\max\{(f(x))^p,(f(y))^p\}$$

$$\Rightarrow\frac{(f(x))^p+(f(y))^p}{2}\leq\max\{(f(x))^p,(f(y))^p\}$$

$$\Rightarrow 0<\frac{(f(x))^p+(f(y))^p}{2}+g(\alpha)\leq\max\{(f(x))^p,(f(y))^p\}+g(\alpha)$$

$$\Rightarrow 1\geq\mathbf{t}(x,y,\alpha)=\left(\frac{\frac{(f(x))^p+(f(y))^p}{2}+g(\alpha)}{\max\{(f(x))^p,(f(y))^p\}+g(\alpha)}\right)^q>0.$$

$$(\Leftarrow)\ x=y\Rightarrow\mathbf{t}(x,y,\alpha)=\left(\frac{\frac{(f(x))^p+(f(y))^p}{2}+g(\alpha)}{\max\{(f(x))^p,(f(y))^p\}+g(\alpha)}\right)^q=\left(\frac{(f(x))^p+g(\alpha)}{(f(x))^p+g(\alpha)}\right)^q=1$$

$$(\Rightarrow)\ \mathbf{t}(x,y,\alpha)=\left(\frac{\frac{(f(x))^p+(f(y))^p}{2}+g(\alpha)}{\max\{(f(x))^p,(f(y))^p\}+g(\alpha)}\right)^q=1\Leftrightarrow\frac{(f(x))^p+(f(y))^p}{2}+g(\alpha)=$$

$$\max\{(f(x))^p,(f(y))^p\}+g(\alpha)\Leftrightarrow(f(x))^p+(f(y))^p=2\max\{(f(x))^p,(f(y))^p\}:$$

$f(x) \geq f(y) \Rightarrow (f(x))^p + (f(y))^p = 2(f(x))^p \Rightarrow (f(y))^p = (f(x))^p \Rightarrow f(y) = f(x) \Rightarrow y = x,$

$f(x) \leq f(y) \Rightarrow (f(x))^p + (f(y))^p = 2(f(y))^p \Rightarrow (f(x))^p = (f(y))^p \Rightarrow f(x) = f(y) \Rightarrow x = y$ (f is one-to-one function).

$$\mathbf{t}(x,y,\alpha) = \left(\frac{\frac{(f(x))^p + (f(y))^p}{2} + g(\alpha)}{(\max\{f(x), f(y)\})^p + g(\alpha)}\right)^q = \left(\frac{\frac{(f(y))^p + (f(x))^p}{2} + g(\alpha)}{(\max\{f(y), f(x)\})^p + g(\alpha)}\right)^q = \mathbf{t}(y,x,\alpha).$$

Let's prove inequality

$$\mathbf{t}(x,y,\alpha) \cdot \mathbf{t}(y,z,\alpha) \leq \mathbf{t}(x,z,\alpha). \tag{7}$$

$$(7) \Leftrightarrow \left(\frac{\frac{(f(x))^p + (f(y))^p}{2} + g(\alpha)}{(\max\{f(x), f(y)\})^p + g(\alpha)}\right)^q \cdot \left(\frac{\frac{(f(y))^p + (f(z))^p}{2} + g(\alpha)}{(\max\{f(y), f(z)\})^p + g(\alpha)}\right)^q$$

$$\leq \left(\frac{\frac{(f(x))^p + (f(z))^p}{2} + g(\alpha)}{(\max\{f(x), f(z)\})^p + g(\alpha)}\right)^q$$

$$\Leftrightarrow \frac{\frac{(f(x))^p + g(\alpha) + (f(y))^p + g(\alpha)}{2}}{\max\{(f(x))^p + g(\alpha), (f(y))^p + g(\alpha)\}} \cdot \frac{\frac{(f(y))^p + g(\alpha) + (f(z))^p + g(\alpha)}{2} + g(\alpha)}{\max\{(f(y))^p + g(\alpha), (f(z))^p + g(\alpha)\}}$$

$$\leq \frac{\frac{(f(x))^p + g(\alpha) + (f(z))^p + g(\alpha)}{2}}{\max\{(f(x))^p + g(\alpha), (f(z))^p + g(\alpha)\}} \Leftrightarrow$$

$$\frac{X+Y}{\max\{X,Y\}} \cdot \frac{Y+Z}{\max\{Y,Z\}} \leq 2 \cdot \frac{X+Z}{\max\{X,Z\}}, \tag{8}$$

where $X = (f(x))^p + g(\alpha), Y = (f(y))^p + g(\alpha), Z = (f(z))^p + g(\alpha).$

There are three cases: 1) $f(x) \leq f(y) \leq f(z)$, 2) $f(x) \leq f(z) \leq f(y)$, and 3) $f(y) \leq f(x) \leq f(z)$, i.e., 1) $X \leq Y \leq Z$, 2) $X \leq Z \leq Y$, and 3) $Y \leq Z \leq Z$.

1) $(8) \Leftrightarrow \frac{X+Y}{Y} \cdot \frac{Y+Z}{Z} \leq 2\frac{X+Z}{Z}$

$\Leftrightarrow (X+Y)(Z+Y) \leq 2(X+Z)Y$

$\Leftrightarrow Y^2 + XY + ZY + XZ \leq 2XY + 2ZY$

$\Leftrightarrow Y^2 - (X+Z)Y + XZ \leq 0 \Leftrightarrow (Y-X)(Y-Z) \leq 0 \Leftrightarrow \top.$

The inequality is correct because two following inequalities hold: $X \leq Y$ and $Y \leq Z$.

2) $(8) \Leftrightarrow \frac{X+Y}{Y} \cdot \frac{Y+Z}{Y} \leq 2\frac{X+Z}{Z}$

$\Leftrightarrow \left(Y^2 + (X+Z)Y + XZ\right)Z \leq 2(X+Z)Y^2$

$\Leftrightarrow (X+Z)ZY + XZ^2 \le 2XY^2 + ZY^2$

$\Leftrightarrow XZY + Z^2Y + XZ^2 \le XY^2 + XY^2 + ZY^2.$

This inequality is valid because:

$Z \le Y \Rightarrow XZY \le XY^2,$

$Z \le Y \Rightarrow XZ^2 \le XY^2,$

$Z \le Y \Rightarrow Z^2Y \le ZY^2.$

3) $(8) \Leftrightarrow \frac{X+Y}{X} \cdot \frac{Y+Z}{Z} \le 2\frac{X+Z}{Z} \Leftrightarrow (X+Y)(Y+Z) \le 2X \cdot (X+Z).$

This inequality is true because the following inequalities hold:

$Y \le X \Rightarrow X+Y \le 2X,\ Y \le X \Rightarrow Y+Z \le X+Z.$

The function $F(\alpha) = (\frac{a+g(\alpha)}{b+g(\alpha)})^q,$ where $a,b,g(\alpha) > 0$, $a < b$, is monotonously increasing, because g is monotonously increasing, so

$$\mathbf{t}(x,y,\alpha) \le \mathbf{t}(x,y,\alpha+\beta), \mathbf{t}(y,z,\beta) \le \mathbf{t}(y,z,\alpha+\beta), \text{ i.e., (7) implies:}$$

$$\mathbf{t}(x,y,\alpha) \cdot \mathbf{t}(y,z,\beta) \le \mathbf{t}(x,y,\alpha+\beta) \cdot \mathbf{t}(y,z,\alpha+\beta) \le \mathbf{t}(x,z,\alpha+\beta).$$

The mapping **t** is obviously a continuous function, because g is continuous function. □

Special case of Example 2 is next example:

Example 3. [19] The mapping $\mathbf{t}_K : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ defined by $\mathbf{t}_K(x,y) = \frac{\frac{x+y}{2}+K}{\max\{x,y\}+K}$, where $K > 0$, is a fuzzy *T*-metric with respect to multiplication, and $\mathbf{s}_K(x,y) = \frac{|x-y|}{2(\max(x,y)+K)}$ is the fuzzy *S*-metric with respect to the algebraic sum, dual to *T* with respect to standard fuzzy complement.

Example 4. [19] The mapping $\mathbf{t}_p : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R},\ p > 0$ defined by

$$\mathbf{t}_p(x,y) = \frac{\sqrt[p]{\frac{x^p+y^p}{2}}}{\max\{x,y\}}, \tag{9}$$

is a fuzzy *T*-metric with respect to multiplication.

Example 5. [6] Let $g : \mathbb{R}^+ \to [0,\infty]$ be an increasing continuous function. If (X,d) is a metric space then the mapping $\mathbf{t} : X \times X \times \mathbb{R}^+ \to \mathbb{R}$ defined by

$$\mathbf{t}(x,y,t) = \frac{g(t)}{g(t)+d(x,y)} \tag{10}$$

is a fuzzy *T*-metric with respect to the multiplication. Specially, for $f(\alpha) = \alpha$:

$$\mathbf{t}(x,y,t)=\frac{t}{t+d(x,y)} \tag{11}$$

and its dual (with respect to the standard fuzzy complement)

$\mathbf{s}(x,y,t)=1-\mathbf{t}(x,y,t)=\dfrac{d(x,y)}{t+d(x,y)}$ is a fuzzy *S*-metric with respect to the algebraic sum.

Theorem 3.3. *Let* $f:X\to(0,1]$ *be one-to-one function and let* $g:\mathbb{R}^+\to(0,1]$ *be an increasing continuous function. The mapping* $d:X\times X\times(0,+\infty)\to\mathbb{R}$ *defined by*

$$\mathbf{d}(x,y,\alpha)=\begin{cases}1, & x=y\\ T(T(f(x),f(y)),g(\alpha)), & x\neq y\end{cases}=\begin{cases}1, & x=y\\ T^2(f(x),f(y),g(\alpha)), & x\neq y\end{cases},$$

is a fuzzy T−metric with respect to the continuous t−norm T.

Proof. From Remark 3 follows

$\mathbf{d}(x,y,\alpha)=0\Leftrightarrow T(f(x),f(y))=0\vee g(\alpha)=0\Leftrightarrow f(x)=0\vee f(y)=0\vee g(\alpha)=0\Leftrightarrow\perp,$

i.e. $\mathbf{d}(x,y,\alpha)=T\left(T(f(x),f(y)),g(\alpha)\right)\in(0,1]$;

Obviously, $x=y\Rightarrow\mathbf{d}(x,y,\alpha)=1$. Suppose that $x\neq y$ and $\mathbf{d}(x,y,\alpha)=1$, then from Remark **2** follows

$$\mathbf{d}(x,y),\alpha)=T\left(T(f(x),f(y)),g(\alpha)\right)=1\Leftrightarrow T(f(x),f(y))=1\wedge g(\alpha)=1$$
$$\Leftrightarrow f(x)=1\wedge f(y)=1\wedge g(\alpha)=1\Rightarrow f(x)=f(y)\Rightarrow x=y.\ \text{Contradiction!}$$
$$\mathbf{d}(x,y),\alpha)=T(T(f(x),f(y)),g(\alpha))=T\left(T(f(y),f(x)),g(\alpha)\right)=\mathbf{d}(y,x,\alpha);$$

From associativity of *t*−norm *T*, for $x\neq y\neq z\neq x$:

$$\begin{aligned}T\left(\mathbf{d}(x,z,\alpha),\mathbf{d}(z,y,\beta)\right)&=T\left(T\left(T(f(x),f(z)),g(\alpha)\right),T\left(T(f(z),f(y)),g(\beta)\right)\right)\\&=T\left(T(f(x),f(z)),T\left(g(\alpha),T\left(T(f(z),f(y)),g(\beta)\right)\right)\right)\\&=T\left(T(f(x),f(z)),T\left(T\left(g(\alpha),T(f(z),f(y))\right),g(\beta)\right)\right)\\&=T\left(T(f(x),f(z)),T\left(T\left(T(f(z),f(y)),g(\alpha)\right),g(\beta)\right)\right)\\&=T\left(T(f(x),f(z)),T\left(T(f(z),f(y)),T\left(g(\alpha),g(\beta)\right)\right)\right)\\&=T\left(T\left(T(f(x),f(z)),T(f(z),f(y))\right),T\left(g(\alpha),g(\beta)\right)\right).\end{aligned}$$

From inequality (1), we have $T(f(x),f(z))\le f(x)$, $T(f(z),f(y))\le f(y)$,

so because of (3), it follows $T\left(T(f(x),f(z)),T(f(z),f(y))\right)\le T(f(x),f(y))$.

From (1), and monotonicity of $T:T(g(\alpha),g(\beta))\le \min\{g(\alpha),g(\beta)\}\le g(\alpha+\beta)$.

Finally,

$$T\left(\mathbf{d}(x,z,\alpha),\mathbf{d}(z,y,\beta)\right) = T\left(T\left(T\left(f(x),f(z)\right),T\left(f(z),f(y)\right)\right),T\left(g(\alpha),g(\beta)\right)\right)$$
$$\le T\left(T\left(f(x),f(y)\right),g(\alpha+\beta)\right)=\mathbf{d}(x,y,\alpha+\beta).$$

Triangular norm T and function g are continuous, then $\mathbf{d}(x,y,\_):(0,+\infty)\to[0,1]$ is a continuous function. □

The following properties hold (see [19]) which enable construction of new fuzzy metrics.

Theorem 3.4. *Let* $\mathbf{d}$ *be a stationary fuzzy metric (in the broader sense) with respect to the norm N. If N is an Archimedean norm and g its corresponding generator, then d = g ∘* $\mathbf{d}$ *is a standard metric.*

Theorem 3.5. *If* $\mathbf{d}_1 : X\times X\times(0,+\infty)\to[0,1)$ *and* $\mathbf{d}_2 : X\times X\times(0,+\infty)\to[0,1)$ *are fuzzy metrics, with respect to the strict norm N, then the mapping* $\sigma(\mathbf{d}_1,\mathbf{d}_2): X\times X\times(0,+\infty)\to\mathbb{R}$ *defined by* $\sigma(\mathbf{d}_1,\mathbf{d}_2)(x,y,\alpha)=N\left(\mathbf{d}_1(x,y,\alpha),\mathbf{d}_2(x,y,\alpha)\right)$ *is also a fuzzy metric with respect to the norm N. If N is not a strict norm, then* $\sigma(\mathbf{d}_1,\mathbf{d}_2)$ *is a fuzzy metric in a broader sense.*

Theorem 3.6. *If* $\mathbf{d}_i : X_i\times X_i\to[0,1)$, $i=1,...,n$, $n\in N$, *are fuzzy metrics with respect to the strict norm N, then* $\mathbf{d} : X^2\to[0,1]$, $X=X_1\times\cdots\times X_n$ *defined with*

$$\mathbf{d}(x,y)=N^{n-1}\left(\mathbf{d}_1(x_1,y_1),\mathbf{d}_2(x_2,y_2),...,\mathbf{d}_n(x_n,y_n)\right), x=(x_1,...,x_n),\ y=(y_1,...,y_n),$$

is the fuzzy metric with respect to the norm N. If N is not a strict norm, then $\mathbf{d}$ *is a fuzzy metric in a broader sense.*

Example 6. If $\mathbf{t}_i : X_i\times X_i\to(0,1]$, $i=1,...,n$, $n\in\mathbb{N}$, are fuzzy T−metrics with respect to the product, then $\mathbf{t} : X^2\to(0,1]$, $X=X_1\times\cdots\times X_n$ defined with

$$\mathbf{t}(x,y)=\prod_{i=1}^{n}\mathbf{t}_i(x_i,y_i),\ \ x=(x_1,...,x_n),\ y=(y_1,...,y_n),$$

is the fuzzy T−metric with respect to the product.

4 APPLICATION

In this section the application of fuzzy metrics in filtering colour images is given. The fuzzy metrics embodies a multitude of concepts that are influenced by image pixels. Each image pixel (i,J_i) (”position”, ”colour”) can be characterized by spatial coordinates of pixel i_1, i_2 (points $i=(i_1,i_2)\in I\times I$, $I=\{0,1,...,n-1\}$ from the screen), and by vector $J_i=(J_i^1,J_i^2,J_i^3)$, the first coordinate of which represents quantity of red colour, while the second coordinate is a quantity of green colour, and finally, the third one represents quantity of blue colour, these colour components being red, green, blue (RGB), respectively.

The assignment of image filtering is to replace the pixel which represents noise by pixel without noise, which can be achieved by replacing a central pixel $(\bar{i},\bar{J}_i)$ in window

$W=\{(i,J_i)\,|\,i\in I_1\times I_2\}$, ($\overline{i}=(\overline{i}_1,\overline{i}_2)\in I_1\times I_2$, and the size from window is an odd number), with pixel that represents the other pixels from *W* in the best possible way, i.e. by a pixel which is the most similar in colour and spatial distance to all the other pixels in *W*.

Selection bias is another potential concern because it is of enormous importance to pick up a good criterion for selecting such a pixel without noise, which will replace the pixel with noise in a given window *W*, because the choice of pixels affects the image quality, i.e. affects the degree of the removed noise.

A key issue is the safe disposal of criterion selection. Namely, the choice of a criterion will be conditioned by a good selection of fuzzy *T*−metric **c**. All pixels in the some window *W* a order relation will be induced by using metric **c**. This order relation will be used to compare pixels (”position”, ”colour”) of the image and to choose a pixel that differs the least from all the other pixels in the window, i.e. which is the most similar to all other pixels in *W* (regarding colour and distance). The central pixel in the given window *W* will be replaced by the pixel found using the algorithm which is applied on each sliding window.

In the algorithm for filtering the image we use fuzzy *T*−metric $\mathbf{c}:W\times W\to\mathbb{R}$ defined with:

$$\mathbf{c}\big((i,J_i),(j,J_j)\big)=\boldsymbol{\tau}\big(J_i,J_j\big)\cdot\mathbf{t}(i,j). \tag{12}$$

Fuzzy *T*−metric which is used in order to measure similarity in colours among pixels is marked with **τ**. It is defined in the following way:

$$\boldsymbol{\tau}\big(J_i,J_j\big)=\boldsymbol{\tau}_1(J_i^1,J_j^1)\cdot\boldsymbol{\tau}_2(J_i^2,J_j^2)\cdot\boldsymbol{\tau}_3(J_i^3,J_j^3)=\prod_{l=1}^{3}\frac{\frac{J_i^l+J_j^l}{2}+K}{\max\{J_i^l,J_j^l\}+K}. \tag{13}$$

Fuzzy *T*−metric that considers spatial distance between pixels is marked with **t**. It is defined in the following way:

$$\mathbf{t}(i,j)=\frac{t}{t+\sqrt{(i_1-j_1)^2+(i_2-j_2)^2}}. \tag{14}$$

That the mappings **τ, t** and **c** are fuzzy metrics follows from the Examples 3 and 5 and Theorem 3.6.

The metric used for the comparison of the quality of images is UIQI, which is defined in [21]. Let $\mathbf{x}=\{x_i|i=1,2,\ldots,n\}$ and $\mathbf{y}=\{y_i|i=1,2,\ldots,n\}$ original and test image signals. Image quality index is defined by following formulas:

$$Q=\frac{4\sigma_{xy}\overline{x}\overline{y}}{\big(\sigma_x^2+\sigma_y^2\big)\big[(\overline{x})^2+(\overline{y})^2\big]},$$

where

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\quad \overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i,$$

$$\sigma_x^2=\frac{1}{n-1}\sum_{i=1}^{n}\big(x_i-\overline{x}\big)^2,\quad \sigma_y^2=\frac{1}{n-1}\sum_{i=1}^{n}\big(y_i-\overline{y}\big)^2,$$

$$\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

Index can be write as:

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x}\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}.$$

Of particular concern is the first component, which is the correlation coefficient between *x* and *y*. Then, the second component shows how close are the pixel brightness values of *x* and *y*. And, finally, the third component measures how similar are the contrasts between *x* and *y*.

The particular metric of quality UIQI is based on the fact that every image distortion is viewed as a combination of three factors: loss of correlation, luminance distortion and contrast distortion.

For additional literature reading about fuzzy filtering, authors recommend the following list: [5], [13], [14], [15] and [20].

In the following example of image quality UIQI, the chosen size of window is 5.

Figure 1. Onion, Original image in jpg format

As can be seen, the filtered image given below is contaminated with 10% salt and pepper noise.

Figure 2. Onion, Salt and pepper noise 10% of noise

A much debated question is whether these metrics can be defined properly. The metric c which is defined with (12), where **τ** and **t** are defined by (13) and (14), respectively. The

values of metric for the image quality UIQI for each colour for the filtered image by applying the method proposed in this paper are equal to:
UIQI: [0.317457074736562,0.316259056738023,0.225750489132398].
The sharpness for image filtered by our metric is 0.9908.
The values of metric of image quality UIQI for each colour for filtered image by median filter (see [1]) with window size three are equal to:
UIQI: [0.683856162855340,0.753420696483180,0.615343870586316].
The sharpness for image filtered by VMF is 0.4635.

Figure 3. Onion, filtered image, onion *S&P* noise window size is 5, K=2250, t=0.1 fuzzy paper denoised armaks

The result was that our image has slightly lower values for corresponding UIQI image quality, but much higher sharpness. This will enhance the understanding the link between quality and sharpness. And this is very important in cases, in which details in the image itself are necessary to be properly. We have used for measuring sharpness image quality metrics introduced in [16].

Figure 4. Onion, filtered by median filter, window size is 5; onion *S&P* noise MEDIAN denoised

In the articles [8] and [19], denoising was investigated on digital images of Lena and Baboon, using fuzzy metrics. In both cases, greater sharpness of the image was shown for the selected parameter values, and in the first case better UIQI.

5 CONCLUSIONS

Distance is a term that is very commonly used both in mathematics and other sciences. The need for the notion of distance caused it to evolve depending on the nature of the set for which it is defined, and brought about the properties required to possess that particular mapping up to what constitutes its set of values. In this paper, the distance is defined over an abstract set, and the distance between its two elements is a fuzzy set (see e.g. [17, 18]). The parameter that provides it with that particular fuzzy nature enable higher possibilities of application, as is the case here. The required properties are the following ones: boundedness, symmetry, some variants of the inequality of a triangle (in relation to the conorm, i.e. the norm), continuity according to the parameter, which all lead to the notions **s** and **t** fuzzy metrics. Historically, the idea started from K. Menger [12] and probabilistic spaces, where the distance between two elements was assessed by means of the distribution function. Subsequently, the notion of the fuzzy space was reached through the whole series of mathematicians, more specifically, in the researches of I. Kramosil, J. Michalek [11], up to A. George, P. Veramani [3, 4], and later on V. Gregory, A. Sapena [6, 7], and their associates. There is a huge number of scientific papers related to the topic of these spaces.

In this paper, the inequality of polygons for the considered fuzzy spaces has been proved, as well as new examples of those spaces, which would enable a wider scope of application of such metrics. One of these applications is image processing (image filtering, image segmentation, etc.). Accompanied by a suitable choice of phase distances, and by means of varying their appropriate parameters, we determine the best possible solution to our problem. The problem of image segmentation considered in this article is comprised of replacing the noisy pixel in the image with a better one, taking into account the influence of the surrounding pixels. The process of selecting a new pixel is influenced by both the spatial fuzzy distance of the pixels and the difference in the pixel brightness within the image in the RGB format, also measured by means of the selected fuzzy metric. The new fuzzy metric combines these two fuzzy metrics in the optimal manner into one evaluation for selecting a new pixel.

New research is aimed at searching for metrics that would perform better in their application, as well as at examining their properties and finding new possibilities of application in other areas.

Acknowledgements: First author acknowledge the financial support of the Ministry of Education, Science and Technological Development of the Republic of Serbia, in the frame of Projects applied under No. TR 34014 and No. ON 174009.

The work of third author was supported by the Serbian Ministry of Education, Science and Technological Development projects III 41025 and OI 171007.

The authors express their sincere gratitude to Prof. Miloš Đurić *(ETF Belgrade, Serbia)* for his careful reading of the paper in the stage when it was a manuscript and for his valuable comments that substantially improved the text.

REFERENCES

[1] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters", *Proceedings of the IEEE*, **78**(4), 678–689 (1990). doi 10.1109/554807

[2] Lj. Gajić, N. M. Ralević, and D. Karaklić, "Prostori sa fazi rastojanjem", *META 2019, The Fourth Conference on Mathematics in Engineering: Theory and Applications, Faculty of Technical Sciences, May 10-12th, 2019*, str. 85-90, Novi Sad, Serbia.

[3] A. George and P. Veeramani P., On some results in fuzzy metric spaces. *Fuzzy Sets Syst*., **64**(3), 395-399 (1994), doi 10.1016/0165-0114(94)90162-7
[4] A. George and P. Veeramani, On some results of analysis for fuzzy metric spaces. *Fuzzy sets and systems*, **90**(3), 365-368 (1997), doi 10.1016/S0165-0114(96)00207-2
[5] V. Gregori, S. Morillas, B. Roig, and A. Sapena, "Fuzzy averaging filter for impulse noise reduction in colour images with a correction step", *Journal of Visual Communication and Image Representation*, **55**, 518-528 (2018). doi 10.1016 /j.jvcir. 2018.06.025.
[6] V. Gregori, S. Morillas, and A. Sapena, "Examples of fuzzy metrics and applications", *Fuzzy sets and systems*, **170**(1), 95–111 (2011). doi 10.1018 /f.fss. 2010.10.019
[7] V. Gregori and S. Romaguera, "Some properties of fuzzy metric spaces", *Fuzzy Sets Syst*. **115**(3) , 485–489 (2000). doi 10.1016/S0165-0114(98)00281-4
[8] D. Karaklić, Lj. Gajić, N.M. Ralević, "Some Fixed Point Results in a Strong Probabilistic Metric Spaces", *Filomat* **33**(8), 2201-2209 (2019). doi 10.2298 /FIL1908201K
[9] E.P. Klement, R. Mesiar, E. Pap, *Triangular Norms*. Kluwer Academic Publishers, Dordrecht (2000).
[10] G.J. Klir, and B. Yuan, *Fuzzy sets and fuzzy logic: Theory and Applications*. Prentice Hall, New Jersey (1995).
[11] I. Kramosil, and J. Michálek, "Fuzzy metrics and statistical metric spaces", *Kybernetika*, **11**(5), 326-334 (1975).
[12] K. Menger, "Statistical metrics", *Proc. Nat. Acad. of Sci. U.S.A.*, **28**(12), 535-537 (1942).
[13] S. Morillas, V. Gregori, G. Peris-Fajarnes, and P. Latorre, "A fast impulsive noise color image filter using fuzzy metrics", *Real-Time Imaging*, **11**(5-6), 417–428 (2005). doi 10.1016 /j.rti. 2005.06.007
[14] S. Morillas, V. Gregori, G. Peris-Fajarnes, and P. Latorre, "A New Vector Median Filter Based on Fuzzy Metrics". in: Kamel, M., Campilho, A. (Eds.). *Image Analysis and Recognition - ICIAR2005, Lecture Notes in Computer Science*, **3656**, Springer-Verlag, Berlin, Heidelberg, 81–90 (2005). doi 10.1117/1.2767335
[15] S. Morillas, V. Gregori, G. Peris-Fajarnes, and A. Sapena, "New adaptive vector filter using fuzzy metrics, *Journal of Electronic Imaging*, **16**(3), 033,007:1–15 (2007). doi 10.1116/1.2767335
[16] N.D. Narvekar and L.J. Karam, "An Improved No-Reference Sharpness Metric Based On The Probability Of Blur Detection", in: *Conference Proceedings 2009 International Workshop on Video Processing and Quality Metrics or Consumer Electronics* (VPQM) (2010).
[17] N. Ralević, S. Dukić, and D. Karaklić, Fazi metrike i primene u otklanjanju šuma na slici, *The Fourth Mathematical Conference of the Republic of Srpska, Trebinje, Republika Srpska, 6 and 7 June 2014, Vol. II*, Fakultet za proizvodnju i menadžment Trebinje, Univerzitet u Istocnom Sarajevu University of East Sarajevo, Mathematical Society of the Republic of Srpska, str. 101-109, ISBN 978-99976-600-4-6.
[18] N. M. Ralević and D. Karaklić, "Fazi rastojanja", *META 2016, The First Conference on Mathematics in Engineering: Theory and Applications, Faculty of Technical Sciences, March 4-6th, 2016*, Novi Sad, Serbia, str. 134-141, ISBN:978-86-7892-800-0.
[19] N. M. Ralević, D. Karaklić, and N. Pištinjat, "Fuzzy metric and its applications in removing the image noise", *Soft Computing*, **23**(22), 12049-12061 (2019). doi 10.1007/s00500-019-03762-5
[20] B. Smolka, M. Szczepanski, K. N. Plataniotis, and A. N. Venetsanopoulos, "On the fast modifyed vector median filter", *Canadian Conference on Electrical and Computer Engineering*, 1315–1319 (2001). doi 10.1109 /CCECE. 2001.93.23636
[21] Z. Wang, A.C. Bovik, "A universal image quality index", *IEEE Signal Processing Letters*, **9**(3), 81–84 (2002). doi 10.1109/97.995823

Received May 20, 2020

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

К ПОСТРОЕНИЮ СТАТИСТИЧЕСКОЙ ТЕРМОДИНАМИКИ НЕЭКСТЕНСИВНЫХ СИСТЕМ НА ОСНОВЕ КАППА-ЭНТРОПИИ КАНИАДАКИСА

А.В. КОЛЕСНИЧЕНКО*

Институт прикладной математики им. М.В. Келдыша РАН. Москва, Россия
* Ответственный автор. E-mail: kolesn@keldysh.ru

DOI: 10.20948/mathmontis-2020-48-10

Ключевые слова: Энтропия Каниадакиса, неэкстенсивная статистика, дивергенция Брэгмана.

Аннотация. Как известно, в физике и в других естественных науках, использующих методы статистической механики, имеются 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 (экстенсивности) для таких важнейших термодинамических величин, как энтропия или внутренняя энергия. 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 асимптотическими степенными распределениями. В частности, неэкстенсивная статистическая механика Каниадакиса успешно применяется к космическим системам с дальним силовым взаимодействием, которое и является причиной их аномальности.

В представленной работе в рамках неэкстенсивной статистики Каниадакиса, основанной на параметрической κ-энтропии, показано, как можно получить деформированную термодинамику сложных аномальных систем и определить её свойства. Приведены основные математические свойства κ-логарифма и κ-экспоненты, а также другие связанные с ними функции, возникающие при разработке неэкстенсивной механики Каниадакиса. В результате получено обобщение нулевого закона термодинамики для двух независимых подсистем при их тепловом контакте и введена так называемая физическая температура, отличная от инверсии множителя Лагранжа β. С привлечением обобщённого первого закона термодинамики и преобразования Лежандра и на основе введённой энтропии Клаузиуса получены новые термодинамические соотношения, которые отличны от выведенных ранее традиционным для неэкстенсивной статистики способом соотношений, неудовлетворительных с точки зрения макроскопической термодинамики. На основе свойства выпуклости дивергенции Брэгмана изучены спонтанные переходы между стационарными состояниями сложной κ-системы и доказаны теорема Гиббса и *H*-теорема Больцмана. Развитый в работе подход позволяет моделировать, в частности, сложные космологические и космогонические среды от галактик и астрофизических дисков до космической плазмы и пыли.

2010 Mathematics Subject Classification: 85A35, 91B50, 82C40.
Key words and Phrases: Kaniadakis entropy, non-extensive statistics, Bregman divergence.

TOWARDS THE DEVELOPMENT OF THERMODYNAMICS OF NONEXTENSIVE SYSTEMS BASED ON KAPPA-ENTROPY ҠANIADAKIS

A.V. KOLESNICHENKO*

Keldysh Institute of Applied Mathematics, Russian Academy of Science
*Corresponding author. E-mail: kolesn@keldysh.ru

DOI:10.20948/mathmontis-2020-48-10

Summary. As is known, in physics and in other natural sciences using the methods of statistical mechanics, there are numerous examples of anomalous systems with long-range force interaction, the fractal nature of the phase space and significant correlations between their individual parts. The complex spatio-temporal structure of such systems leads to a violation of the principle of additivity (extensiveness) for such important thermodynamic quantities as entropy or internal energy. At present, theories of diverse non-extensive systems are developing at an accelerated pace, at which new ideas appear that allow a deeper understanding of their nature, capabilities and limitations. Each such theory has a wide range of important applications related to the physics of statistical systems whose probabilistic properties are described not by Gibbs (and not Gaussian), but by power-law distributions. In particular, the non-extensive statistical mechanics of Kaniadakis is successfully applied to space systems with long-range force interaction, which is the reason for their anomaly.

In the framework of non-extensive statistical mechanics of Kaniadakis based on parametric κ-entropy, it is shown how to obtain the deformed statistical thermodynamics of complex anomalous systems and determine its properties. The paper presents the basic mathematical properties of the κ-logarithm and κ-exponent, as well as other related functions that arise during the development of the statistical mechanics of Kaniadakis. As a result, a generalization is obtained for the case under consideration of the zero law of thermodynamics for two independent subsystems with their thermal contact and the so-called physical temperature is introduced, which differs from the inversion of the Lagrange multiplier. Using the generalized first law of thermodynamics and the Legendre transformation, and based on the introduced Clausius entropy, new thermodynamic relations are obtained that are different from the relationships that were previously unsatisfactory from the point of view of deformed thermodynamics, which were traditionally used for non-extensive statistics. Based on the property of convexity of Bergman divergence, spontaneous transitions between stationary states of a complex β-system are studied and the Gibbs theorem and the Boltzmann *H*-theorem are proved.

The approach developed in the work allows modeling, in particular, complex cosmological and cosmogonic environments from galaxies and astrophysical disks to cosmic plasma and dust.

ВВЕДЕНИЕ

Исследования в области механики неаддитивных (неэкстенсивных) систем стали в последнее время предметом значительного интереса, что объясняется как новизной возникающих здесь общетеоретических проблем, так и важностью практических приложений. Начало систематического изучения в этом направлении связано с работой К. Тсаллиса [1], в которой была введена так называемая q-энтропия $S_q^{Ts}(p) := \frac{k_B}{q-1}\int\left(p - p^q\right)d\Gamma$, зависящая от некоторого действительного числа q (параметра деформации) и обладающая неаддитивностью для совокупности независимых аномальных систем. Теория неэкстенсивных систем, основанная на энтропии Тсаллиса, в настоящее время интенсивно развивается. В научной литературе доступны систематизированные собрания обзоров, дающие последовательное изложение многочисленных новых результатов, полученных в ходе изучения неэкстенсивных свойств в аномальных физических явлениях (см. библиографию, представленную на сайте http:/tsallis. cat.cbpf.br/ biblio.htm, которая постоянно обновляется [2]).

Определение энтропии Тсаллиса не является единственным примером деформированной энтропии. Основой исследований в области неэкстенсивных статистик, проводимых в настоящее время, являются многочисленные нелогарифмические энтропии, введенные, например, в работах [1,3-20]. При этом каждая неэкстенсивная статистика имеет широкий спектр важных приложений, связанных с физикой аномальных систем, вероятностные свойства которых определяются не гиббсовым (не гауссовым), а асимптотическим степенным законом распределения вероятностей, который не зависит от экспоненциального поведения, обусловленного распределением Гиббса. Диапазон применения разнообразных неэкстенсивных параметрических энтропий в настоящее время постоянно расширяется, охватывая различные направления в науке, такие как космология и космогония, теория плазмы, квантовая механика и статистика, специальная и общая теории относительности, стохастическая динамика и фракталы, геофизика, биомедицина и многие другие.

Среди различных деформированных энтропий неэкстенсивных систем особый интерес представляет энтропия Каниадакиса $S_\kappa(p) := -\frac{k_B}{2\kappa}\int\left(p^{\kappa+1} - p^{1-\kappa}\right)d\Gamma$, введенная впервые в работах [21,22]. Основанная на κ - энтропия неэкстенсивная статистика сохраняет математическую и гносеологическую структуру обычной статистической механики и пригодна для описания очень большого класса экспериментально наблюдаемых явлений в физике низких и высоких энергий, а также в естественных, экономических и социальных науках. В частности, деформированная энтропия Каниадакиса объективно возникает в рамках специальной теории относительности Эйнштейна [23]. При этом параметр деформации κ зависит от скорость света c и уменьшается до нуля при $c \to \infty$, восстанавливая таким образом обычную статистическую механику и термодинамику. Статистика Каниадакиса возникает в различных прикладных областях. В качестве примера можно упомянуть работы, связанные с космическими эффектами [22-24] (в частности, с звездной астрофизикой [25], с кварк-глюонной плазмой [26], с газокинетическиими моделями, описывающими взаимодействие атомов и фотонов [27], с квантовой механикой [28].

Представленная работа посвящена конструированию на основе параметрической κ-энтропии статистической термодинамики неэкстенсивных систем. Проведенное исследование базируется на свойствах негиббсового канонического κ-распределения, полученного из принципа Джейнса [29] максимума κ-энтропии при заданности усредненной внутренней энергии системы и вероятностной нормировки для функции κ-распределения. Показано, что все важные термодинамические характеристики системы, такие как энтропия, полная и свободная энергия, внутренняя энергия, физические температура и давление, коэффициенты теплопроводности и т.п. могут быть найдены с использованием только равновесной функции κ-распределения. Получено обобщение нулевого закона термодинамики для двух независимых неэкстенсивных систем при их тепловом контакте, вводящее в рассмотрение так называемую физическую температуру $T_{ph}(p)$ отличающуюся от инверсии множителя Лагранжа β. Этот факт потребовал переопределения ряда термодинамических соотношений, получаемых естественным путем в рамках статистики Каниадакиса. Путем использования обобщенной энтропии Клаузиуса к аномальной κ-системе, был получен наиболее приемлемый набор макроскопических термодинамических соотношений для неэкстенсивной κ-системы. Кроме этого, показано, что сохраняются принцип максимума равновесной энтропии, лежандрова структура теории, термодинамическая устойчивость, теорема Гиббса и *H*- теорема.

Таким образом, в работе с единых позиций изложен круг вопросов, связанных с конструированием деформированной термодинамики на основе κ-энтропии Каниадакиса и дивергенции Брэгмана для сложных аномальных систем.

1. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ, СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ И СВОЙСТВА ЭНТРОПИИ КАНИАДАКИСА

В деформированной статистической механике Каниадакиса для непрерывных аномальных систем при вероятностной нормировке

$$\int p(\boldsymbol{r})d\Gamma = 1, \quad 0 \le p(\boldsymbol{r}) < \infty \tag{1}$$

для плотности вероятности распределения систем $p(\boldsymbol{r})$ в фазовом пространстве $\boldsymbol{r} := \{\boldsymbol{q}_1,...,\boldsymbol{q}_N;\ \boldsymbol{p}_1,...,\boldsymbol{p}_N\}$ статистического ансамбля («представляющего» макроскопическое состояние системы) деформированная κ-энтропия задаётся следующим функционалом [21,28]

$$S_\kappa(p) := -k_{\mathcal{B}}\int p(\boldsymbol{r})\frac{p^\kappa(\boldsymbol{r}) - p^{-\kappa}(\boldsymbol{r})}{2\kappa}d\Gamma = -k_{\mathcal{B}}\int \frac{p^{\kappa+1}(\boldsymbol{r}) - p^{1-\kappa}(\boldsymbol{r})}{2\kappa}d\Gamma. \tag{2}$$

Здесь и далее везде область интегрирования совпадает со всем $6N$-мерным фазовым пространством, причем безразмерный элемент фазового пространства $d\Gamma$ записывается в современной форме $d\Gamma := \left\{N!h^{3N}\right\}^{-1}d\boldsymbol{r}$, где $h = 2\pi\hbar$, $k_{\mathcal{B}}$ – постоянные Планка и Больцмана соответственно. Энтропийный индекс κ в определении κ-энтропии (2) представляет собой вещественное число, принадлежащее области $|\kappa| < 1$. Подобная деформация логарифмической функции в выражении для энтропии (по сравнению с энтропией Больцма-

на−Гиббса $S_{\mathcal{BG}}(p) := -k_{\mathcal{B}}\int p\ln(p)d\Gamma$) позволяет учитывать важную особенность поведения многих аномальных систем с длинной памятью и/или дальнодействующими силовыми взаимодействиями, когда вероятность реализации $p(\boldsymbol{r})$ малых значений параметров состояния убывает (при $p\to 0^{+}$) не экспоненциально быстро, а степенным образом (закон Парето). Благодаря этому статистика Каниадакиса описывает события, практически недостижимые в простых системах, характеризуемых статистикой Больцмана−Гиббса.

Легко показать, что в пределе слабой связи $\kappa\to 0$ κ-энтропия (2) переходит в каноническую формулу $S_{\mathcal{BG}}$ обычной статистики Больцмана−Гиббса. Действительно, в пределе $\kappa\to 0$ имеем: $p^{\pm\kappa}=\mathrm{e}^{\pm\kappa\ln p}\to\ 1\pm\kappa\ln p$, и энтропия S_{κ} сводится к

$$S_{\kappa\to 0}\left(p\right)=\lim_{\kappa\to 0}\left\{-k_{\mathcal{B}}\int p\frac{p^{\kappa}-p^{-\kappa}}{2\kappa}d\Gamma\right\}=-k_{\mathcal{B}}\int p\ln p\,d\Gamma=S_{\mathcal{BG}}.$$

Энтропия Каниадакиса (2) может быть представлена также в следующей эквивалентной форме:

$$S_{\{\kappa\}}(p) := -k_{\mathcal{B}}\int p\ln_{\{\kappa\}}p\,d\Gamma=-k_{\mathcal{B}}\langle\ln_{\{\kappa\}}p\rangle, \tag{3}$$

при написании которой использован так называемый, «деформированный логарифм» [30]

$$\ln_{\{\kappa\}}(x) := \frac{x^{\kappa}-x^{-\kappa}}{2\kappa}\equiv\frac{1}{\kappa}\sinh(\kappa\ln x)\quad(\forall x>0), \tag{4}$$

а также обычное правило получения среднего значения для любой динамической переменной $\mathcal{A}_j(\boldsymbol{r})$, а именно [31]

$$\langle\mathcal{A}_j\rangle := \int p(\boldsymbol{r})\mathcal{A}_j(\boldsymbol{r})d\Gamma, \tag{5}$$

где принята нормировка (1).

Отметим, что в дискретном случае, при условии вероятностной нормировки $\sum_{i=1}^{W}p_i=1,$ нужно в приведённых выше формулах произвести замену $\int d\Gamma\leftrightarrow\sum_{i=1}^{W}$, где $p:=\left\{p_j\right\}_{j=1,2,\ldots,W}$ − дискретная функция распределения, а число W означает количество доступных в системе микросостояний.

Свойства функции $\ln_{\{\kappa\}}(x)$. Как хорошо известно, в классической статистической механике особую роль играет логарифм функции распределения со знаком минус $(-\ln(p)))$, поскольку эта величина связана с энтропией системы. В 1990-х гг. Каниадакис предложил определение деформированного логарифма (4), в котором параметр деформации κ принадлежит интервалу $(-1,1)$, давая в пределе $\kappa\to 0$ обычный логарифм $\ln(x)$. Приведем здесь некоторые наиболее важные свойства деформированного логарифма, которые будут использованы ниже.

Легко убедиться, что функция $\ln_{\{\kappa\}}(x)$ обладает следующими свойствами [22,30]):

$$\ln_{\{\kappa\}}(x) = \ln_{\{-\kappa\}}(x), \tag{6}$$

$$\ln_{\{\kappa\}}(x^{\lambda}) = \lambda \ln_{\{\lambda\kappa\}}(x), \quad \ln_{\{\kappa\}}(1/x) = -\ln_{\{\kappa\}}(x), \tag{7}$$

$$\ln_{\{\kappa\}}(0^{+}) = -\infty, \quad \ln_{\{\kappa\}}(1) = 0, \quad \ln_{\{\kappa\}}(+\infty) = +\infty, \tag{8}$$

$$d\left[\ln_{\{\kappa\}}(x)\right]/dx = x^{-1}u_{\{\kappa\}}(x), \quad d\left[x\ln_{\{\kappa\}}(x)\right]/dx = \lambda \ln_{\{\kappa\}}(x/\alpha) = \ln_{\{\kappa\}}(x) + u_{\{\kappa\}}(x). \tag{9}$$

Имеют место также следующие свойства вогнутости функции $\ln_{\{\kappa\}}(x)$:

$$d\left[\ln_{\{\kappa\}}(x)\right]/dx > 0, \quad d^2\left[\ln_{\{\kappa\}}(x)\right]/dx^2 < 0, \quad d^2\left[x\ln_{\{\kappa\}}(x)\right]/dx^2 > 0, \tag{10}$$

а также асимптотическое поведение деформированного логарифма по степенному закону:

$$\ln_{\{\kappa\}}(x) \underset{x\to 0^{+}}{\sim} \frac{1}{|2\kappa|}x^{-|\kappa|}, \quad \ln_{\{\kappa\}}(x) \underset{x\to +\infty}{\sim} \frac{1}{|2\kappa|}x^{|\kappa|}. \tag{11}$$

Тейлоровское разложение функции $\ln_{\{\kappa\}}(1+x)$ сходится в случае, если $-1 < x \le 1$; в этом случае имеем:

$$\ln_{\{\kappa\}}(1+x) = \sum_{n=1}^{\infty} b_n(\kappa)(-1)^{n-1}\frac{x^n}{n}, \tag{12}$$

где $b_n(-\kappa) = b_n(\kappa)$, $b_1(\kappa) = 1$, $b_n(0) = 1$; при $n > 1$ коэффициенты $b_n(\kappa)$ определяются соотношением:

$$b_n(\kappa) = \frac{1}{2}(1-\kappa)\left(1-\frac{\kappa}{2}\right)\ldots\left(1-\frac{\kappa}{n-1}\right) + \frac{1}{2}(1+\kappa)\left(1+\frac{\kappa}{2}\right)\ldots\left(1+\frac{\kappa}{n-1}\right).$$

Первые три члена разложения (12) принимают вид:

$$\ln_{\{\kappa\}}(1+x) \underset{x\to 0}{\sim} = x - \frac{x^2}{2} + \left(1+\frac{\kappa^2}{2}\right)\frac{x^3}{3}. \tag{13}$$

Для деформированного логарифма $\ln_{\{\kappa\}}(x)$ справедливо также следующее интегральное представление:

$$\ln_{\{\kappa\}}(x) = \frac{1}{2}\int_{1/x}^{x}\frac{1}{t^{1+\kappa}}dt. \tag{14}$$

Наконец, если ввести так называемые к-сумму и к-произведение − обобщенную сумму и обобщенное произведение статистики Каниадакиса [32,33]

$$x \overset{\kappa}{\oplus} y := x\sqrt{1+\kappa^2 y^2} + y\sqrt{1+\kappa^2 x^2}, \quad (x, y \in \mathbb{R} \text{ и } |\kappa| \le 1),$$

$$x \underset{\kappa}{\otimes} y := \left(\kappa \ln_{\{\kappa\}} x + \kappa \ln_{\{\kappa\}} y + \sqrt{1 + (\kappa \ln_{\{\kappa\}} x + \kappa \ln_{\{\kappa\}} y)^2} \right)^{1/\kappa}, \tag{15}$$

то можно получить еще два важных свойства функции $\ln_{\{\kappa\}}(x)$:

$$\ln_{\{\kappa\}}(xy) = \ln_{\{\kappa\}}(x) \overset{\kappa}{\oplus} \ln_{\{\kappa\}}(y) = \ln_{\{\kappa\}}(x) u_{\{\kappa\}}(y) + \ln_{\{\kappa\}}(y) u_{\{\kappa\}}(x) =$$

$$= \ln_{\{\kappa\}}(x) \sqrt{1 + \kappa^2 \ln^2_{\kappa}(y)} + \ln_{\{\kappa\}}(y) \sqrt{1 + \kappa^2 \ln^2_{\kappa}(x)},$$

$$\ln_{\{\kappa\}}(x) + \ln_{\{\kappa\}}(y) = \ln_{\{\kappa\}}\left(x \underset{\kappa}{\otimes} y \right). \tag{16}$$

Функция $u_{\{\kappa\}}(x)$. В соотношениях (9) и (16), а также далее фигурирует важная в статистике Каниадакиса функция [32,33]

$$u_{\{\kappa\}}(x) := \frac{x^{\kappa} + x^{-\kappa}}{2} = \sqrt{1 + \kappa^2 \ln^2_{\{\kappa\}}(x)} \equiv \cosh(\kappa \ln(x)), \tag{17}$$

обладающая следующими свойствами (см., например, [34,35]:

$$u_{\{\kappa\}}(x) = u_{\{-\kappa\}}(x), \quad u_{\{\kappa\}}(x) = u_{\{\kappa\}}(1/x), \quad u_{\{\kappa\}}(\alpha) = 1/\lambda, \quad \ln_{\{\kappa\}}(\alpha) = -1/\lambda. \tag{18}$$

Здесь введены обозначения:

$$\lambda := \sqrt{1 - \kappa^2}, \quad \alpha := \left[(1 - \kappa)/(1 + \kappa) \right]^{1/2\kappa} \tag{19}$$

При учете свойств (18), а также преобразования

$$u_{\{\kappa\}}(x) = \frac{x^{\kappa} + x^{-\kappa}}{2} = \frac{x^{\kappa} - x^{-\kappa}}{2\kappa} - \frac{(1 - \kappa)x^{\kappa} - (1 + \kappa)x^{-\kappa}}{2\kappa} =$$

$$= \ln_{\{\kappa\}}(x) - \sqrt{1 - \kappa^2} \ln_{\{\kappa\}} \left\{ x \left(\frac{1 - \kappa}{1 + \kappa} \right)^{1/2\kappa} \right\}, \tag{20}$$

легко получить следующие соотношения:

$$u_{\{\kappa\}}(x) = x^{\kappa} - \kappa \ln_{\{\kappa\}}(x) = \ln_{\{\kappa\}}(x) - \lambda \ln_{\{\kappa\}}(\alpha x) = -\ln_{\{\kappa\}}(x) - \lambda \ln_{\{\kappa\}}(\alpha / x), \tag{21}$$

$$u_{\{\kappa\}}(xy) = u_{\{\kappa\}}(x) u_{\{\kappa\}}(y) + \kappa^2 \ln_{\{\kappa\}}(x) \ln_{\{\kappa\}}(y). \tag{22}$$

С учетом (21), закон аддитивности (16) деформированного к-логарифма может быть переписан следующим образом:

$$\ln_{\{\kappa\}}(xy) = 2 \ln_{\{\kappa\}}(x) \ln_{\{\kappa\}}(y) - \lambda \left\{ \ln_{\{\kappa\}}(y) \ln_{\{\kappa\}}(\alpha x) + \ln_{\{\kappa\}}(x) \ln_{\{\kappa\}}(\alpha y) \right\}. \tag{23}$$

Легко видеть, что в пределе слабой связи $\kappa \to 0$ это свойство κ-логарифма сводится к стандартному закону аддитивности обычного логарифма $\ln(xy) = \ln(x) + \ln(y)$.

Наконец, можно убедиться в том, что имеют место следующие дифференциальные соотношения:

$$\frac{d}{dx} u_{\{\kappa\}}(x) = \kappa^2 \frac{\ln_{\{\kappa\}}(x)}{x}, \quad \frac{d}{dx}\left\{x u_{\{\kappa\}}(x)\right\} = \lambda u_{\{\kappa\}}\left(\frac{x}{\alpha}\right) = \ln_{\{\kappa\}}(x) + u_{\{\kappa\}}(x). \tag{24}$$

Неаддитивность κ-энтропии для независимых систем. Покажем теперь, что подобно энтропии Тсаллиса (см., например, [15,36]), энтропия Каниадакиса подчиняется псевдоаддитивному закону для двух статистически независимых систем.

Рассмотрим совокупную физическую κ-систему, состояние которой описывается совместным мультипликативным распределением $p^{(12)} := p^{(1)} \cdot p^{(2)}$, где $p^{(12)} := p^{(12)}(\boldsymbol{r}_1, \boldsymbol{r}_2)$, $p^{(1)} := p^{(1)}(\boldsymbol{r}_1)$, $p^{(2)} := p^{(2)}(\boldsymbol{r}_2)$. Распределение $p^{(12)}(\boldsymbol{r}_1, \boldsymbol{r}_2)$ может зависеть также и от времени, а «точки» $\boldsymbol{r}_1$ и $\boldsymbol{r}_2$ относятся к двум статистически независимым κ-системам.

Тогда полная энтропия системы задается выражением

$$S^{(12)}_{\{\kappa\}} := -k_B \int\int p^{(12)} \ln_{\{\kappa\}}(p^{(12)}) d\Gamma_1 d\Gamma_2, \tag{25}$$

где выполняются условия нормировки

$$\iint p^{(12)}(\boldsymbol{r}_1, \boldsymbol{r}_2) d\Gamma_1 d\Gamma_2 = \int p^{(1)}(\boldsymbol{r}_1) d\Gamma_1 = \int p^{(2)}(\boldsymbol{r}_2) d\Gamma_2 = 1.$$

После подстановки мультипликативного распределения $p^{(12)} = p^{(1)} \cdot p^{(2)}$ в формулу (25) получим (при учете (22)) следующее свойство псевдоаддитивности совокупной энтропии в статистике Каниадакиса для двух независимых систем:

$$S^{(12)}_{\{\kappa\}} = S_{\{\kappa\}}(p^{(1)}) \mathcal{I}_{\{\kappa\}}(p^{(2)}) + S_{\{\kappa\}}(p^{(2)}) \mathcal{I}_{\{\kappa\}}(p^{(1)}), \tag{26}$$

где

$$\mathcal{I}_{\{\kappa\}}(p) := \langle u_{\{\kappa\}}(p) \rangle_\kappa = \int p\, u_{\{\kappa\}}(p) d\Gamma = \int p \sqrt{1 + \kappa^2 \ln^2_{\{\kappa\}}(p)} d\Gamma. \tag{27}$$

Проверим свойство квазиаддитивности (26) совокупной энтропии:

$$S^{(12)}_{\{\kappa\}} \equiv -k_B \int\int p^{(12)} \ln_{\{\kappa\}}(p^{(12)}) d\Gamma_1 d\Gamma_2 = -k_B \left\langle \ln_{\{\kappa\}}(p^{(12)}) \right\rangle_{\{\kappa\}} =$$

$$= -k_B \left\langle \ln_\kappa(p^{(1)} p^{(2)}) \right\rangle_{\{\kappa\}} = -k_B \left\langle \ln_{\{\kappa\}}(p^{(1)}) u_{\{\kappa\}}(p^{(2)}) + \ln_{\{\kappa\}}(p^{(2)}) u_{\{\kappa\}}(p(1)) \right\rangle_{\{\kappa\}} =$$

$$= -k_B \int\int p^{(1)} \cdot p^{(2)} \left\{ \ln_{\{\kappa\}}(p^{(1)}) u_{\{\kappa\}}(p^{(2)}) + \ln_{\{\kappa\}}(p^{(2)}) u_{\{\kappa\}}(p^{(1)}) \right\} d\Gamma_1 d\Gamma_2 =$$

$$= S_{\{\kappa\}}(p^{(1)}) \mathcal{I}_{\{\kappa\}}(p^{(2)}) + S_{\{\kappa\}}(p^{(2)}) \mathcal{I}_{\{\kappa\}}(p^{(1)}). \tag{28}$$

В случае если $\kappa \to 0$, то $u_{\{0\}}(p)=1$ и $\mathcal{I}_{\{0\}}\left(p\right)=1$, так что из (26) следует аддитивное правило для классической энтропии Больцмана–Гиббса.

Отметим, что с учетом соотношения (21) функцию $\mathcal{I}_{\{\kappa\}}\left(p\right)$ можно записать в другом виде:

$$\mathcal{I}_{\{\kappa\}}(p)=\langle \ln_{\kappa}(p)\rangle-\lambda\langle \ln_{\kappa}(\alpha p)\rangle=\frac{1}{k_{\mathcal{B}}}\Big[-\lambda\alpha S_{\{\kappa\}}(p/\alpha)+S_{\{\kappa\}}(p)\Big]. \tag{29}$$

Тогда для совокупной энтропии Каниадакиса двух независимых систем получим еще одно представление:

$$k_{\mathcal{B}}S^{(12)}_{\{\kappa\}}=S_{\{\kappa\}}(p^{(1)})\Big\{\lambda S_{\{\kappa\}}(\alpha p^{(2)})-S_{\{\kappa\}}(p^{(2)})\Big\}+\ S_{\{\kappa\}}(p^{(2)})\Big\{\lambda S_{\{\kappa\}}(\alpha p^{(1)})-S_{\{\kappa\}}(p^{(1)})\Big\} \tag{30}$$

2. КАНОНИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ГИББСА В ДЕФОРМИРОВАННОЙ κ-СТАТАТИСТИКЕ КАНИАДАКИСА

Основные свойства экспоненты Каниадакиса $\exp_{\{\kappa\}}(x)$. Далее нам понадобится так называемая экспонента Каниадакиса. Для определения функции, обратной деформированному логарифму (2), введем переменную $x\equiv \ln_{\{\kappa\}}(y)$, что приводит к алгебраическому уравнению $z^2-2x\kappa z-1=0$ для неизвестной $z\equiv y^{\kappa}$. Его решение $z=\kappa x\pm\sqrt{1+(\kappa x)^2}$ дает выражение $y=\left\{\kappa x\pm\sqrt{1+(\kappa x)^2}\right\}^{1/\kappa}$. Функция y обратная деформированному логарифму $x=\ln_{\{\kappa\}}(y)$, представляет собой, так называемую, экспоненту Каниадакиса:

$$\exp_{\{\kappa\}}(x):=\left(\sqrt{1+(\kappa x)^2}+\kappa x\right)^{1/\kappa}\equiv\exp\left(\frac{1}{\kappa}\operatorname{arcsinh}\kappa x\right)\in C^{\infty}(\mathbb{R}), \tag{31}$$

Из определения (31) экспоненты Каниадакиса вытекают следующие свойства [32,33]:

$$\exp_{\{0\}}(x)=\lim_{\kappa\to 0}\exp_{\{\kappa\}}(x)=\exp(x),\quad \exp_{\{\kappa\}}(0)=1,\quad \exp_{\{-\kappa\}}(x)=\exp_{\{\kappa\}}(x),$$

$$\exp_{\{\kappa\}}(-\infty)=0^{+},\quad \exp_{\{\kappa\}}(+\infty)=+\infty,\quad \ln_{\{\kappa\}}\exp_{\{\kappa\}}(x)=\exp_{\{\kappa\}}\ln_{\{\kappa\}}(x)=x,$$

$$u_{\{\kappa\}}\left\{\exp_{\{\kappa\}}(x)\right\}=\sqrt{1+\kappa^2x^2},\quad \exp_{\{\kappa\}}(x)\exp_{\{\kappa\}}(-x)=1. \tag{32}$$

Кроме того, функция $\exp_{\{\kappa\}}(x)$ обладает следующими свойствами:

$$(\exp_{\{\kappa\}}(x))^{r}=\exp_{\{\kappa/r\}}(rx) \tag{33}$$

(с $r\in\mathbb{R}$, которое в пределе слабой связи $\kappa\to 0$ воспроизводит известное свойство обыкновенной экспоненты),

$$\exp_{\{\kappa\}}(x)\exp_{\{\kappa\}}(y)=\exp_{\{\kappa\}}(x\overset{\kappa}{\oplus}y)=\exp_{\{\kappa\}}\left\{x\sqrt{1+\kappa^2y^2}+y\sqrt{1+\kappa^2x^2}\right\},$$

$$\exp_{\{\kappa\}}(x+y)=\exp_{\{\kappa\}}(x)\underset{\kappa}{\otimes}\exp_{\{\kappa\}}(y). \tag{34}$$

Из (18) следует следующее правило дифференцирования экспоненты Каниадакиса:

$$\frac{d}{dx}\exp_{\{\kappa\}}(x)=\frac{\exp_{\{\kappa\}}(x)}{u_{\{\kappa\}}\left\{\exp_{\{\kappa\}}(x)\right\}}=\frac{1}{\sqrt{1+\kappa^2x^2}}\exp_{\{\kappa\}}(x). \tag{35}$$

Имеет место следующее свойство выпуклости:

$$d^2\left[\exp_{\{\kappa\}}(x)\right]/dx^2>0,\quad x\in\mathbb{R},\quad \kappa^2<1. \tag{36}$$

Безусловно, одним из наиболее важных свойств функции $\exp_{\{\kappa\}}(x)$ является ее и асимптотическое поведение по степенному закону:

$$\exp_{\{\kappa\}}(x)\underset{x\to\pm\infty}{\sim}\left|2\kappa x\right|^{\pm1/|\kappa|}, \tag{37}$$

$$\exp_{\{\kappa\}}(x)\underset{x\to0^+}{\sim}-\left|2\kappa\right|^{-1}x^{-|\kappa|},\quad \exp_{\{\kappa\}}(x)\underset{x\to+\infty}{\sim}\left|2\kappa\right|^{-1}x^{|\kappa|}. \tag{38}$$

Разложение Тейлора экспоненты $\exp_{\{\kappa\}}(x)$ имеет следующий вид [35]:

$$\exp_{\{\kappa\}}(x)=\sum_{n=0}^{\infty}\xi_n(\kappa)\frac{x^n}{n!},\quad \kappa^2x^2<1, \tag{39}$$

где $\xi_0(\kappa)=\xi_1(\kappa)=\xi_2(\kappa)=1,\quad \xi_n(\kappa)=\prod_{j=1}^{n-1}\left[1-(2j-n)\kappa\right]$. Заметим, что первые три члена в разложении Тейлора κ-экспоненты точно такие, как и для обычной экспоненты:

$$\exp_{\{\kappa\}}(x)\underset{x\to0}{\sim}=1+x+\frac{x^2}{2}+(1-\kappa^2)\frac{x^3}{3!}. \tag{39*}$$

Экспоненту $\exp_{\{\kappa\}}(x)$ можно также записать как бесконечное произведение обыкновенных экспонент [30]:

$$\exp_{\{\kappa\}}(x)=\prod_{n=0}^{\infty}\exp(c_n\kappa^{2n}x^{2n+1}),\quad c_n:=\frac{(-1)^n(2n)!}{(2n+1)2^{2n}(n!)^2}. \tag{40}$$

Наконец, для κ-экспоненты справедливы интегральные соотношения [23]:

$$\exp_{\{\kappa\}}(-x)=\int_0^{\infty}\frac{1}{\kappa s}J_{1/\kappa}\left(\frac{s}{\kappa}\right)\exp(-sx)ds,\quad \operatorname{Re}x\ge0, \tag{41}$$

$$M_{\{\kappa\}}(r) := \int_0^\infty x^{r-1} \exp_{\{\kappa\}}(-x)dx = \frac{|2\kappa|^{-r}}{1+|\kappa|r} \cdot \frac{\Gamma\left(\frac{1}{|2\kappa|} - \frac{r}{2}\right)}{\Gamma\left(\frac{1}{|2\kappa|} + \frac{r}{2}\right)} \Gamma(r), \tag{42}$$

где $0 < r < 1/|\kappa|$; $J_\nu(s)$ – функция Бесселя, $\Gamma(x)$ – Гамма- функция. Из (42) легко получить выражение $M_{\{\kappa\}}(r+2) = \frac{r(r+1)}{1-\kappa^2(r+2)^2} \cdot M_{\{\kappa\}}(r)$.

Деформированное каноническое распределение. Равновесные состояния сложных κ -систем характеризуются распределениями, которые не меняются с течением времени. Каноническое распределение Гиббса в статистике Каниадакиса может быть получено, как и в классическом случае, из «экстремизации» κ -энтропии (4) при выполнении следующих дополнительных условий: заданности средней энергии системы

$$\mathcal{E}_{\{\kappa\}} := \langle \mathcal{H} \rangle_{\{\kappa\}} = \int p(\boldsymbol{r}) \mathcal{H}(\boldsymbol{r}) d\Gamma = const \tag{43}$$

и сохранения вероятностной нормировки (1) распределения $p(\boldsymbol{r})$. Здесь $\mathcal{H} = \mathcal{H}(\boldsymbol{r})$ – функция Гамильтона, которая определяется математической моделью изучаемых физических процессов в системе. Заметим, что в общем случае эта функция может зависеть от ряда внешних параметров $a_1, a_2, ...a_s$, макроскопически характеризующих состояние статистического равновесия рассматриваемого ансамбля аномальных динамических систем, $\mathcal{H} := \mathcal{H}(\boldsymbol{r}, \{a_j\})$ [31].

Согласно вариационному принципу Джейнса [29] определим функционал

$$\mathcal{L}(p) := -k_{\mathcal{B}} \int p(\boldsymbol{r}) \ln_\kappa p(\boldsymbol{r}) d\Gamma - \beta \int p(\boldsymbol{r}) \mathcal{H}(\boldsymbol{r}) d\Gamma - k_{\mathcal{B}} \gamma \int p(\boldsymbol{r}) d\Gamma \tag{44}$$

и найдём его безусловный экстремум. Здесь β и γ суть множители Лагранжа. В соответствии с теоремой Лагранжа вероятностное распределение $p(\boldsymbol{r})$, «экстремирующее» κ -энтропию $S_\kappa(p)$ при указанных ограничениях, определяется из условия:

$$\frac{\delta \mathcal{L}(p)}{\delta p} = -k_{\mathcal{B}} \int \left[\ln_{\{\kappa\}} \left(p(\boldsymbol{r}) \right) + u_{\{\kappa\}} \left(p(\boldsymbol{r}) \right) \right] d\Gamma - \beta \int \mathcal{H}(\boldsymbol{r}) d\Gamma - k_{\mathcal{B}} \gamma \int d\Gamma = 0. \tag{45}$$

При написании (45) использовано соотношение (9). Из (35), с учетом (19), (21) и (33), следует уравнение

$$\ln_{\{\kappa\}} \left(\alpha / p(\boldsymbol{r}) \right) - \frac{1}{\lambda} \left[\gamma + \frac{\beta}{k_{\mathcal{B}}} \mathcal{H}(\boldsymbol{r}) \right] = 0, \quad \lambda := \sqrt{1-\kappa^2}, \quad \alpha = \exp_{\{\kappa\}}(-1/\lambda), \tag{46}$$

решение которого дает следующее нормированное κ -распределение $p^{eq}(\boldsymbol{r})$ в состоянии статистического равновесия системы (аналог канонического распределения Гиббса в статистике Каниадакиса)

$$p^{eq}(\boldsymbol{r},\beta)=\alpha/\exp_{\{\kappa\}}\left\{\frac{1}{\lambda}\left(\gamma+\frac{\beta}{k_{\mathcal{B}}}\mathcal{H}(\boldsymbol{r})\right)\right\}=\alpha\exp_{\{\kappa\}}\left\{-\frac{1}{\lambda}\left(\gamma+\frac{\beta}{k_{\mathcal{B}}}\mathcal{H}(\boldsymbol{r})\right)\right\}. \tag{47}$$

Следует отметить, что в пределе слабой связи $\kappa\to 0$ это распределения сводится к каноническому распределению Гиббса классической статистики.

С учетом свойств (16) и (34) распределение (47) можно записать в другом виде [37]:

$$p^{eq}(\boldsymbol{r})=\exp_{\{\kappa\}}\left(-\frac{1}{\lambda}\right)\exp_{\{\kappa\}}\left(-\frac{\mathcal{X}(\boldsymbol{r})}{\lambda}\right)=$$

$$=\exp_{\{\kappa\}}\left[-\left(\frac{1}{\lambda}\right)^{\kappa}\oplus\left(-\frac{\mathcal{X}(\boldsymbol{r})}{\lambda}\right)\right]=\exp_{\{\kappa\}}\left[-\frac{1}{\lambda}\left(\frac{\mathcal{X}(\boldsymbol{r})}{\lambda}+\sqrt{1+\kappa^2\frac{\mathcal{X}^2(\boldsymbol{r})}{\lambda^2}}\right)\right]=$$

$$=\exp_{\{\kappa\}}\left[-u_{\{\kappa\}}\left(p^{eq}(\mathcal{X})\right)-\mathcal{X}\right], \tag{48*}$$

где, в силу (22) и (48), имеем

$$u_{\{\kappa\}}\left(p^{eq}(\boldsymbol{r})\right)=\frac{1}{\lambda}\left(\frac{\kappa^2}{\lambda}\mathcal{X}(\boldsymbol{r})+\sqrt{1+\frac{\kappa^2}{\lambda^2}\mathcal{X}^2(\boldsymbol{r})}\right),\quad \mathcal{X}(\boldsymbol{r}):=\gamma+\frac{\beta}{k_{\mathcal{B}}}\mathcal{H}(\boldsymbol{r}).$$

Заметим также, что хвост распределения (48) описывается степенной асимптотикой (это так называемый закон Парето), определяемой в силу (37) выражением $p^{eq}\underset{\mathcal{X}\to\infty}{\sim}\left|2\kappa\mathcal{X}/\sqrt{1-\kappa^2}\right|^{1/\kappa}$, которое существенно отличается от экспоненциальной асимптотики обычного распределения Больцмана−Гиббса [10,31].

Найдем теперь вторую вариацию функционала (44). В результате получим

$$\delta^2\mathcal{L}(p)=-k_{\mathcal{B}}\int\frac{1}{p}\left[\kappa^2\ln_{\{\kappa\}}(p)+u_{\{\kappa\}}(p)\right]\delta p^2\,d\Gamma=-k_{\mathcal{B}}\kappa^2\int\left[\kappa+\frac{p^{2\kappa}+1}{p^{2\kappa}-1}\right]\delta p^2\,d\Gamma.$$

Легко убедиться в том, что экстремум соответствует максимуму и минимуму рассматриваемого функционала, соответственно, при $0<\kappa<1$ $(\delta^2\mathcal{L}<0)$ и $-1<\kappa<0$ $(\delta^2\mathcal{L}>0)$. Таким образом, распределение (48) максимизирует или минимизирует энтропию Каниадакиса.

В заключение этого раздела заметим, что в неэкстенсивной статистической кинетике, также как и в классической кинетике имеет место термодинамическая эквивалентность статистических ансамблей. Все ансамбли статистической механики определяются заданием внешних условий, в которых находятся системы, их составляющие. Например, канонический ансамбль Гиббса определяется постоянством числа частиц, объема и контактом с термостатом, большой канонический ансамбль Гиббса − постоянством объема, контактом с термостатом и резервуаром частиц, изобарически-изотермический ансамбль − постоянством числа частиц, давления и контактом с термостатом. Вместе с тем, при выборе ан-

самбля обычно руководствуются удобством вычислений, а не условиями, в которых находится система, поскольку, как было доказано в ряде работ (см., например, [31]), вычисленные с их помощью термодинамические функции мало отличаются между собой и совпадают в термодинамическом пределе

3.ТЕРМОДИНАМИЧЕСКИЕ СООТНОШЕНИЯ

Приступим теперь к конструированию равновесной термодинамики, основанной на неэкстенсивной статистике Каниадакиса. Используя распределение (48*), получим следующее выражение:

$$\ln_{\{\kappa\}}(p^{eq}(\boldsymbol{r})) + u_{\{\kappa\}}(p^{eq}(\boldsymbol{r})) + \gamma + \frac{\beta}{k_{\mathcal{B}}}\mathcal{H}(\boldsymbol{r}) = 0. \tag{49}$$

(заметим, что это выражение может быть получено также путем преобразования свойства (16), в котором $x := p^{eq}(\boldsymbol{r})$ и $y := 1/\alpha$, если использовать при этом формулы (7), (19) и (48)).

Усредняя выражение (49) с помощью распределения $p^{eq}(\boldsymbol{r})$ и учитывая определения (27) и (43), в результате получим равновесное значение энтропии Каниадакиса [35]

$$S^{eq}_{\{\kappa\}} = k_{\mathcal{B}}(\mathcal{I}^{eq}_{\{\kappa\}} + \gamma) + \beta\mathcal{E}^{eq}_{\{\kappa\}}\ . \tag{50}$$

При сравнении κ-энтропии (50) с ее классической версией ($\kappa\to 0$, $\mathcal{I}^{eq}_{\{0\}} = 1$), естественно определить аналог $\mathcal{Z}_{\{\kappa\}}$ классического статистического интеграла $\mathcal{Z}$ соотношением [36]):

$$k_{\mathcal{B}} \ln_{\{\kappa\}}(\mathcal{Z}_{\{\kappa\}}) := k_{\mathcal{B}}(\mathcal{I}_{\{\kappa\}} + \gamma) + \beta\mathcal{E}^{eq}_{\{\kappa\}}. \tag{51}$$

Далее везде будем опускать метку "eq" у термодинамических параметров, когда это не вызывает неопределенности. Тогда по аналогии с обычной статистикой Больцмана−Гиббса выражение (50) принимает вид:

$$S^{eq}_{\{\kappa\}} := k_{\mathcal{B}} \ln_{\{\kappa\}} \mathcal{Z}_{\{\kappa\}}\ , \tag{52}$$

откуда следует, что $\mathcal{Z}_{\{\kappa\}} = \left(\kappa S^{eq}_{\{\kappa\}}/k_{\mathcal{B}} + \sqrt{1 + \kappa S^{eq}_{\{\kappa\}}/k_{\mathcal{B}}}\right)^{1/\kappa}$.

Продифференцируем теперь логарифм $\ln_{\{\kappa\}} \mathcal{Z}_{\{\kappa\}}$ по β. Используя определение (27) функции $\mathcal{I}_{\{\kappa\}}$, а также соотношения (9), (21) и (49), в результате получим:

$$\frac{d}{d\beta}\ln_{\{\kappa\}}(\mathcal{Z}_{\{\kappa\}}) = \frac{\beta}{k_{\mathcal{B}}}\frac{d\mathcal{E}^{eq}_{\{\kappa\}}}{d\beta} + \frac{\mathcal{E}_{\{\kappa\}}}{k_{\mathcal{B}}} + \frac{d}{d\beta}\int u_{\{\kappa\}}\left(p^{eq}(\boldsymbol{r})\right)p^{eq}(\boldsymbol{r})d\Gamma =$$

$$= \frac{\beta}{k_{\mathcal{B}}} \frac{d\mathcal{E}^{eq}_{\{\kappa\}}}{d\beta} + \frac{\mathcal{E}_{\{\kappa\}}}{k_{\mathcal{B}}} + \int \left\{ u_{\{\kappa\}}(p^{eq}) \frac{dp^{eq}}{d\beta} + p^{eq} \frac{du_{\{\kappa\}}(p^{eq})}{d\beta} \right\} d\Gamma =$$

$$= \frac{\beta}{k_{\mathcal{B}}} \frac{\mathcal{E}^{eq}_{\{\kappa\}}}{d\beta} + \frac{\mathcal{E}_{\{\kappa\}}}{k_{\mathcal{B}}} + \int \left\{ u_{\{\kappa\}}(p^{eq}) \frac{dp^{eq}}{d\beta} - p^{eq} \frac{d}{d\beta} \left(\gamma + \frac{\beta}{k_{\mathcal{B}}} \mathcal{H}(\boldsymbol{r}) \right) - u_{\{\kappa\}}(p^{eq}) \frac{dp^{eq}}{d\beta} \right\} d\Gamma =$$

$$= \frac{\beta}{k_{\mathcal{B}}} \frac{d\mathcal{E}^{eq}_{\{\kappa\}}}{d\beta} + \frac{\mathcal{E}_{\{\kappa\}}}{k_{\mathcal{B}}} - \int p^{eq} \frac{d}{d\beta} \left(\frac{\beta}{k_{\mathcal{B}}} \mathcal{H}(\boldsymbol{r}) \right) d\Gamma = \frac{\beta}{k_{\mathcal{B}}} \frac{d\mathcal{E}^{eq}_{\{\kappa\}}}{d\beta} . \tag{53}$$

Следовательно, для энергии $\mathcal{E}_{\{\kappa\}}$ справедливо термодинамическое уравнение

$$\beta \frac{d\mathcal{E}_{\{\kappa\}}}{d\beta} = k_{\mathcal{B}} \frac{d}{d\beta} \ln_{\{\kappa\}} \mathcal{Z}_{\{\kappa\}} . \tag{54}$$

Если продифференцировать экстремальное значение κ-энтропии $S_{\{\kappa\}}(p^{eq})$ по $\mathcal{E}_{\{\kappa\}}$ и учесть (9), (21) и (47), то в результате получим аналог известного соотношения равновесной термодинамики:

$$\frac{dS_{\{\kappa\}}}{d\mathcal{E}_{\{\kappa\}}} = -k_{\mathcal{B}} \int \frac{d\left[p^{eq}(\boldsymbol{r}) \ln_{\{\kappa\}} p^{eq}(\boldsymbol{r}) \right]}{dp^{eq}(\boldsymbol{r})} \frac{dp^{eq}(\boldsymbol{r})}{d\mathcal{E}_{\{\kappa\}}} d\Gamma =$$

$$= -k_{\mathcal{B}} \int \frac{d\left[\ln_{\{\kappa\}} p^{eq}(\boldsymbol{r}) + u_{\{\kappa\}}\left(p^{eq}(\boldsymbol{r}) \right) \right]}{dp^{eq(\boldsymbol{r})}} \frac{dp^{eq}(\boldsymbol{r})}{d\mathcal{E}_{\{\kappa\}}} d\Gamma = -k_{\mathcal{B}} \lambda \int \ln_{\{\kappa\}} \left(\frac{p^{eq}(\boldsymbol{r})}{\alpha} \right) \frac{dp^{eq}(\boldsymbol{r})}{d\mathcal{E}_{\{\kappa\}}} d\Gamma =$$

$$= k_{\mathcal{B}} \lambda \int \left(\frac{\gamma + \beta k_{\mathcal{B}}^{-1} \mathcal{H}(\boldsymbol{r})}{\lambda} \right) \frac{dp^{eq}}{d\mathcal{E}_{\{\kappa\}}} d\Gamma = \beta. \tag{55}$$

Здесь использованы условия $\int dp^{eq} d\Gamma = 0$ и $\int \mathcal{H}(\boldsymbol{r})\, dp^{eq}(\boldsymbol{r})\, d\Gamma = d\mathcal{E}_{\{\kappa\}}$.

Определим теперь деформированную свободную энергию Гельмгольца $\mathcal{F}_{\{\kappa\}}$ выражением

$$\mathcal{F}_{\{\kappa\}} := \mathcal{E}_{\{\kappa\}} - \frac{k_{\mathcal{B}}}{\beta} \ln_{\{\kappa\}} (\mathcal{Z}_{\{\kappa\}}) , \tag{56}$$

которое при $\kappa \to 0$ совпадает со свободной энергией обычной статистики $\mathcal{F} := \mathcal{E} - \frac{k_{\mathcal{B}}}{\beta} \ln \mathcal{Z}$.

Тогда из (52), (54) и (56) вытекают искомые дифференциальные соотношения деформированной термостатики Каниадакиса:

$$S_{\{\kappa\}} = k_{\mathcal{B}} \ln_{\{\kappa\}}(\mathcal{Z}_{\{\kappa\}}), \quad \mathcal{F}_{\{\kappa\}} \equiv \mathcal{E}_{\{\kappa\}} - \frac{1}{\beta} S_{\{\kappa\}} = \mathcal{E}_{\{\kappa\}} - \frac{k_{\mathcal{B}}}{\beta} \ln_{\{\kappa\}}(\mathcal{Z}_{\{\kappa\}}),$$

$$\beta = \frac{dS_{\{\kappa\}}}{d\mathcal{E}_{\{\kappa\}}} = k_{\mathcal{B}} \frac{d \ln_{\{\kappa\}}(\mathcal{Z}_{\{\kappa\}})}{d\mathcal{E}_{\{\kappa\}}}, \quad \beta \frac{d\mathcal{E}_{\{\kappa\}}}{d\beta} = k_{\mathcal{B}} \frac{d}{d\beta} \ln_{\{\kappa\}} \mathcal{Z}_{\{\kappa\}}, \quad \mathcal{E}_{\{\kappa\}} = \frac{d(\beta \mathcal{F}_{\{\kappa\}})}{d\beta}. \tag{57}$$

Усреднение микроскопических динамических величин. Важно отметить, что введенный выше статистический интеграл $\mathcal{Z}_{\{\kappa\}}$ определяется относительно внутренней энергии $\mathcal{E}_{\{\kappa\}}$ системы (см. (51)). Однако часто удобнее использовать другое определение $\mathcal{Z}_{\{\kappa\}}$, задаваемое выражением [35]

$$\ln_{\{\kappa\}} \ddot{\mathcal{Z}}_{\{\kappa\}} := \ln_{\{\kappa\}} \mathcal{Z}_{\{\kappa\}} - k_{\mathcal{B}}^{-1} \beta \mathcal{E}_{\{\kappa\}} = \mathcal{I}_{\{\kappa\}} + \gamma. \tag{58}$$

Здесь величина $\ddot{\mathcal{Z}}_{\{\kappa\}}$ не зависит от выбора нуля энергии, а при условии $\kappa \to 0$ формула (58) обращается в хорошо известное в статистике Больцмана–Гиббса соотношение для статистического интеграла $\ln \mathcal{Z} = 1 + \gamma$. В этом случае, уравнения равновесной термодинамики (50) принимают почти классическую форму

$$S_{\{\kappa\}} = \beta \left(\mathcal{E}_{\{\kappa\}} - \ddot{\mathcal{F}}_{\{\kappa\}} \right), \quad dS_{\{\kappa\}} = \beta \, d\mathcal{E}_{\{\kappa\}},$$

$$\ddot{\mathcal{F}}_{\{\kappa\}} = -\frac{k_{\mathcal{B}}}{\beta} \ln_{\{\kappa\}} \ddot{\mathcal{Z}}_{\{\kappa\}}, \quad \mathcal{E}_{\{\kappa\}} = \frac{d(\beta \ddot{\mathcal{F}}_{\{\kappa\}})}{d\beta}, \quad C_{\{\kappa\}} = -\beta^2 \frac{d\mathcal{E}_{\{\kappa\}}}{d\beta}. \tag{59}$$

С помощью деформированного канонического распределения Гиббса (47) можно вычислить также среднее значение любой динамической переменной $\mathcal{A}_j(\boldsymbol{r})$. Дифференцируя для этого логарифм статистического интеграла $\ddot{\mathcal{Z}}_{\{\kappa\}}$ по микроскопической энергии $\mathcal{H} = \mathcal{H}(\boldsymbol{r})$ и учитывая (49), в результате получим:

$$\frac{d}{d\mathcal{H}} \ln_{\{\kappa\}} \ddot{\mathcal{Z}}_{\{\kappa\}} = \frac{d}{d\mathcal{H}} \left(\mathcal{I}_{\{\kappa\}} + \gamma \right) = \frac{d\gamma}{d\mathcal{H}} + \int \frac{d}{dp^{eq}} \left[p^{eq} u_{\{\kappa\}}(p^{eq}) \right] \frac{dp^{eq}}{d\mathcal{H}} d\Gamma =$$

$$= \frac{d\gamma}{d\mathcal{H}} + \int \left(u_{\{\kappa\}}(p^{eq}) + p^{eq} \frac{du_{\{\kappa\}}(p^{eq})}{dp^{eq}} \right) \frac{dp^{eq}}{d\mathcal{H}} d\Gamma =$$

$$= \frac{d\gamma}{d\mathcal{H}} + \int \left(-p^{eq} \frac{d}{dp^{eq}} \left(\gamma + \frac{\beta}{k_{\mathcal{B}}} \mathcal{H}(\boldsymbol{r}) \right) \right) \frac{dp^{eq}}{d\mathcal{H}} d\Gamma =$$

$$= \frac{d\gamma}{d\mathcal{H}} + \int \left(-p^{eq} \frac{d}{d\mathcal{H}} \left(\gamma + \frac{\beta}{k_{\mathcal{B}}} \mathcal{H}(\boldsymbol{r}) \right) \right) d\Gamma = -\frac{\beta}{k_{\mathcal{B}}} p^{eq}. \tag{60}$$

При выводе этого выражения было использовано преобразование

$$\frac{du_{\{\kappa\}}(p^{eq})}{dp^{eq}} = -\frac{u_{\{\kappa\}}(p^{eq})}{p^{eq}} - \frac{d}{dp^{eq}}\left(\gamma + \frac{\beta}{k_{\mathcal{B}}}\mathcal{H}(\boldsymbol{r})\right),$$

полученное с учетом соотношения (49) и формулы $d(\ln_{\{\kappa\}} x)/dx = x^{-1}u_{\{\kappa\}}x$.

Таким образом, зная κ-свободную энергию Гельмгольца $\mathcal{F}_{\kappa}$, можно вычислить равновесную функцию распределения по формуле

$$p^{eq}(\boldsymbol{r}) = -\frac{k_{\mathcal{B}}}{\beta}\frac{d}{d\mathcal{H}}\ln_{\{\kappa\}}\ddot{\mathcal{Z}}_{\{\kappa\}} = \frac{d\ddot{\mathcal{F}}_{\{\kappa\}}}{d\mathcal{H}}. \tag{61}$$

Отсюда следует, что усредненные значения микроскопических динамических переменных $\mathcal{A}_j(\boldsymbol{r})$ могут быть найдены при использовании функций $\ddot{\mathcal{Z}}_{\{\kappa\}}$, или $\mathcal{F}_{\kappa}$ следующим образом:

$$\langle\mathcal{A}_j\rangle = \int p^{eq}(\boldsymbol{r})\mathcal{A}_j(\boldsymbol{r})d\Gamma = -\frac{k_{\mathcal{B}}}{\beta}\int\mathcal{A}_j(\boldsymbol{r})\frac{d}{d\mathcal{H}}\left[\ln_{\{\kappa\}}\left(\ddot{\mathcal{Z}}_{\{\kappa\}}\right)\right]d\Gamma = \int\mathcal{A}_j(\boldsymbol{r})\frac{d\ddot{\mathcal{F}}_{\kappa}}{d\mathcal{H}}d\Gamma \tag{62}$$

Выше было отмечено, что для систем, находящихся в состоянии статистического квазиравновесия, функция Гамильтона $\mathcal{H} = \mathcal{H}(\boldsymbol{r},\{a_j\})$ может зависеть от ряда внешних параметров $a_1, a_2, \ldots a_s$, которые можно рассматривать как обобщенные координаты данной системы. По аналогии с классической статистикой определим обобщенные силы $\mathcal{A}_j(\boldsymbol{r})$ соотношением [31]

$$\mathcal{A}_j(\boldsymbol{r}) = -d\mathcal{H}(\boldsymbol{r},\{a_j\})/da_j. \tag{63}$$

Тогда, наблюдаемые значения обобщенных сил $\mathcal{A}_j(\boldsymbol{r})$, равные среднему значению по равновесному статистическому ансамблю, могут быть вычислены следующим образом:

$$\langle\mathcal{A}_j\rangle = \int\mathcal{A}_j(\boldsymbol{r})p^{eq}(\boldsymbol{r})d\Gamma = -\int\frac{d\mathcal{H}(\boldsymbol{r},\{a_j\})}{da_j}p^{eq}(\boldsymbol{r})d\Gamma =$$

$$= \frac{k_{\mathcal{B}}}{\beta}\int\frac{d\mathcal{H}}{da_j}\frac{d\ln_{\{\kappa\}}\ddot{\mathcal{Z}}_{\{\kappa\}}}{d\mathcal{H}}d\Gamma = \frac{k_{\mathcal{B}}}{\beta}\int\frac{d\ln_{\{\kappa\}}\ddot{\mathcal{Z}}_{\{\kappa\}}}{da_j}d\Gamma = -\frac{d\ddot{\mathcal{F}}_{\kappa}}{da_j}. \tag{64}$$

4. УСЛОВИЕ РАВНОВЕСИЯ И ЗАКОН КОМПОЗИЦИИ ЭНЕРГИИ ДЛЯ СИСТЕМ С ЭНТРОПИЕЙ КАНИАДАКИСА

Нахождение условия термодинамического равновесия двух независимых систем требует привлечения законов композиции для энтропий и энергий. В статистической термодинамике Больцмана–Гиббса эти законы имеют свойство аддитивности. Для неэкстенсивной κ-системы энтропия обладает свойством псевдоаддитивности (26), которое в данной работе не распространяется на энергии. Покажем, что это обстоятельство приводит к равен-

ству так называемых физических температур для двух независимых κ-систем, при условии усреднения физических величин равновесным нормированным распределением.

Физическая температура. Начиная с работы [37]), в литературе имеет место дискуссия по способу усреднения функции Гамильтона $\mathcal{H}=\mathcal{H}(\boldsymbol{r})$ и, соответственно, по выбору закона композиции усредняемых энергий двух независимых κ-систем. Далее предполагается выполнимость закона аддитивность для функции Гамильтона:

$$\mathcal{H}^{(12)}(\boldsymbol{r}_1,\boldsymbol{r}_2)=\mathcal{H}^{(1)}(\boldsymbol{r}_1)+\mathcal{H}^{(2)}(\boldsymbol{r}_2),$$

который приводит к следующему закону аддитивности усреднённой энергии совокупной системы:

$$\mathcal{E}_{\kappa}^{(12)}=\mathcal{E}_{\kappa}^{(1)}+\mathcal{E}_{\kappa}^{(2)}, \tag{65}$$

где

$$\mathcal{E}_{\kappa}^{(12)}=\iint p^{eq}(\boldsymbol{r}_1,\boldsymbol{r}_2)\mathcal{H}^{(12)}(\boldsymbol{r}_1,\boldsymbol{r}_2)d\Gamma_1 d\Gamma_2,$$

$$\mathcal{E}_{\kappa}^{(1)}=\int p^{eq}(\boldsymbol{r}_1)\mathcal{H}^{(1)}(\boldsymbol{r}_1)d\Gamma_2 \quad \mathcal{E}_{\kappa}^{(2)}=\int p^{eq}(\boldsymbol{r}_2)\mathcal{H}^{(2)}(\boldsymbol{r}_2)d\Gamma_2. \tag{66}$$

Варьирование условия аддитивности (65) для энергии $\mathcal{E}_{\kappa}^{(12)}$ и условия квазиаддитивности (28) для совокупной энтропии $S_{\kappa}^{(12)}=S_{\{\kappa\}}^{(1)}\mathcal{I}_{\{\kappa\}}^{(2)}+S_{\{\kappa\}}^{(2)}\mathcal{I}_{\{\kappa\}}^{(1)}$ замкнутой равновесной системы с постоянными значениями энергии $\mathcal{E}_{\kappa}^{(12)}=const$ и энтропии $S_{\kappa}^{(12)}=const$ приводит к равенствам:

$$\delta\mathcal{E}_{\kappa}^{(12)}=\delta\mathcal{E}_{\kappa}^{(1)}+\delta\mathcal{E}_{\kappa}^{(2)}=0, \tag{67}$$

$$\delta S_{\kappa}^{(12)}=\mathcal{I}_{\{\kappa\}}^{(2)}\delta S_{\{\kappa\}}^{(1)}+S_{\{\kappa\}}^{(1)}\delta\mathcal{I}_{\{\kappa\}}^{(2)}+\mathcal{I}_{\{\kappa\}}^{(1)}\delta S_{\{\kappa\}}^{(2)}+S_{\{\kappa\}}^{(2)}\delta\mathcal{I}_{\{\kappa\}}^{(1)}=0. \tag{68}$$

В итоге, учитывая формулу (50) для экстремального значения κ-энтропии, записанную с учетом (57) в виде

$$S_{\{\kappa\}}^{(r)}=k_{\mathcal{B}}\left(\mathcal{I}_{\{\kappa\}}^{(r)}+\gamma_r\right)+\left[\delta S_{\{\kappa\}}^{(r)}/\delta\mathcal{E}_{\{\kappa\}}^{(r)}\right]\mathcal{E}_{\{\kappa\}}^{(r)}, \quad (r=1,2), \tag{69}$$

получим условие

$$\left[\delta S_{\{\kappa\}}^{(1)}/\delta\mathcal{E}_{\{\kappa\}}^{(1)}\right]/\mathcal{I}_{\{\kappa\}}^{(1)}=\left[\delta S_{\{\kappa\}}^{(2)}/\delta\mathcal{E}_{\{\kappa\}}^{(2)}\right]/\mathcal{I}_{\{\kappa\}}^{(2)}, \quad \text{где} \quad \delta S_{\{\kappa\}}^{(r)}/\delta\mathcal{E}_{\{\kappa\}}^{(r)}=\beta^{(r)}, \quad (r=1,2). \tag{70}$$

Равенство (70) удовлетворяется тождественно только при равенстве так называемых физических температур

$$T_{ph}^{(r)}=\frac{\mathcal{I}_{\{\kappa\}}^{(r)}}{\delta S_{\{\kappa\}}^{(r)}/\delta\mathcal{E}_{\{\kappa\}}^{(r)}}, \text{ где} \qquad \delta S_{\{\kappa\}}/\delta\mathcal{E}_{\{\kappa\}}=\beta\equiv 1/T \tag{71}$$

для двух независимых κ-систем при их тепловом контакте.

Отношение эквивалентности (71) является обобщением нулевого закона термодинамики на неэкстенсивные κ-системы, описываемые статистикой Каниадакиса. Оно показывает, что в отличие от классического случая физическая температура T_{ph} не является обратной величиной β^{-1} множителя Лагранжа, но определяется соотношением $T_{ph}(p) \equiv \mathcal{I}_{\{\kappa\}} / \beta$. Если $\kappa = 0$, то $T_{ph} \equiv T$ и законы композиции всех средних и микроскопических величин становятся аддитивными. Подчеркнём важный факт, что температуры $T = 1/\beta$ и $T_{ph} \equiv \mathcal{I}^{eq}_{\{\kappa\}} / \beta$ не зависят от выбора нуля энергий, и поэтому они допускают физическую интерпретацию.

Физическое давление. По аналогии с физической температурой T_{ph}, можно ввести и физическое давление P_{qh}. Для этой цели рассмотрим механическое равновесие двух независимых κ-систем, представляющих собой совокупную замкнутую систему с постоянными значениями энтропии $S^{(12)}_{\{\kappa\}} = const$ и объёма $V^{(12)} = V^{(1)} + V^{(2)} = const$. В этом случае необходимо находить экстремум энтропии $S^{(12)}_{\{\kappa\}}$ с учетом фиксации общего объёма $V^{(12)}$; в результате этой процедуры получим, что

$$\left[\delta S^{(1)}_{\{\kappa\}} / \delta V^{(1)}\right] / \mathcal{I}^{(1)}_{\{\kappa\}} = \left[\delta S^{(2)}_{\{\kappa\}} / \delta V^{(2)}\right] / \mathcal{I}^{(2)}_{\{\kappa\}} = P_{ph} / T_{ph}, \tag{72}$$

где P_{ph} – так называемое физическое давление, которое определяется соотношением

$$P_{ph} := T_{ph}\left[\delta S^{eq}_{\{\kappa\}} / \delta V\right] / \mathcal{I}^{eq}_{\{\kappa\}}. \tag{73}$$

Таким образом, предполагая аддитивный закон энергетического и механического взаимодействия между двумя микроскопическими системами, управляемыми одной и той же неаддитивной κ-энтропией, можно определить физические макроскопические переменные (температуру и давление), отвечающие нулевому принципу термодинамики в контексте неэкстенсивной статистической механики Каниадакиса.

В связи с введением в рассмотрение температуры T_{ph} сделаем следующее общее замечание. В большинстве сложных неэкстенсивных систем важную роль играют длинномасштабные пространственно-временные корреляции в фазовом или геометри-ческом пространстве. Это означает, в частности, что существенное значение имеет та часть внутренней энергии системы, которая связана с силовым взаимодействием отдалённых друг от друга её частей, а именно потенциальная энергия. В классической статистической механике внутренняя энергия определяется, как правило, суммой кинетических энергий всех молекул совокупной системы. В такой системе «тепловой баланс» достигается в основном за счёт локального теплообмена между близко расположенными (в частности, в пограничных областях) её составляющими, т.е. «тепло» связано с передачей кинетической энергии отдельными частицами системы. Поскольку физическая температура T_{ph} отвечает за «глобальный тепловой баланс» между различными частями системы, то её энергетический баланс будет сильно отличаться от локального теплового баланса. Локальный тепловой

баланс описывается абсолютной температурой $\beta = 1/T$, измеряемой термометром. Однако подобное измерение физической температуры T_{ph} нереально, что связано с наличием функционального коэффициента, $\mathcal{I}^{eq}_{\{\kappa\}}$ зависящего, согласно (27), от параметра деформации κ системы, $\mathcal{I}^{eq}_{\{\kappa\}} = \int p^{eq}\sqrt{1+\kappa^2 \ln^2_{\{\kappa\}}(p^{eq})}d\Gamma$.

Важно иметь в виду, что такое переопределение температуры в κ-статистике расходится с основным принципом классической термодинамики, в которой абсолютная температура T является интенсивным параметром, а не функционалом $T_{ph}(p)$. Этот факт требует модификации термостатических соотношений (57) и (58), включая переопределение энтропии Клаузиуса, с учетом использования физической температуры.

5. МАКРОСКОПИЧЕСКАЯ ТЕРМОСТАТИКА

В качестве основных предпосылок, взятых за исходный пункт построения модифицированной макроскопической термодинамики Тсаллиса, работе [38], были выбраны первый закон термодинамики и структура преобразования Лежандра. Используем этот подход и при разработке макроскопической термодинамики в рамках статистики Каниадакиса на основе энтропии Клаузиуса.

Рассмотрим с этой целью структуру преобразования Лежандра. Уравнение (55) $\beta = dS_{\{\kappa\}}/d\mathcal{E}_{\{\kappa\}}$ указывает на то, что параметры β и $\mathcal{E}_{\{\kappa\}}$ образуют пару переменных Лежандра. Это привело к следующему определению свободной энергии Гельмгольца (изохорно-изотермического потенциала) (см.(57)):

$$\mathcal{F}_{\{\kappa\}}(T) := \mathcal{E}_{\{\kappa\}} - TS_{\{\kappa\}} = \mathcal{E}_{\{\kappa\}} - k_{\mathcal{B}}T\ln_{\{\kappa\}}\mathcal{Z}_{\{\kappa\}}. \tag{74}$$

Однако, это определение является неудовлетворительным с точки зрения макроскопической термодинамики. Свободная энергия должна зависеть от физической температуры T_{ph}, а не от переменной $T := 1/\beta$.

Физическая свободная энергия Гельмгольца. По аналогии с подходом, предложенном в работе [39] при модификации первого закона термодинамики в статистике Тсаллиса, переопределим макроскопическую κ-свободную энергию (74) следующим образом:

$$\tilde{\mathcal{F}}_{\{\kappa\}}(T_{ph}) := \mathcal{E}_{\{\kappa\}} - k_{\mathcal{B}}T_{ph}\ln\mathcal{Z}_{\{\kappa\}}, \tag{75}$$

что отличается от соответствующего выражения в традиционной термодинамике. Используя соотношения (51), (52) и (71), можно убедиться, что переопределённая таким образом свободная энергия $\tilde{\mathcal{F}}_{\{\kappa\}}$ является функцией T_{ph}. Дифференцируя функцию $\tilde{\mathcal{F}}_{\{\kappa\}}$, в результате получим

$$d\tilde{\mathcal{F}}_{\{\kappa\}} = d\,\mathcal{E}_{\{\kappa\}} - k_{\mathcal{B}}\ln\mathcal{Z}_{\{\kappa\}}dT_{ph} - T_{ph}w_{\{\kappa\}}dS_{\{\kappa\}}. \tag{76}$$

При написании (76) использовано выражение

$$d\ln \mathcal{Z}_{\{\kappa\}}=\left[k_{\mathcal{B}}u_{\{\kappa\}}(\mathcal{Z}_{\{\kappa\}})\right]^{-1} dS_{\{\kappa\}} = k_{\mathcal{B}}^{-1} w_{\{\kappa\}} dS_{\{\kappa\}}, \tag{77}$$

полученное с учетом соотношений (9), (21) и (52). Здесь введено обозначение $w_{\{\kappa\}} := 1/\sqrt{1+\left(\kappa S^{eq}_{\{\kappa\}}/k_{\mathcal{B}}\right)^2}$ для весовой функции.

Энтропия Клаузиуса. Если теперь использовать первый закон термодинамики

$$d'Q_{\{\kappa\}} = d\mathcal{E}_{\{\kappa\}} + P_{ph}dV, \tag{78}$$

где $Q_{\{\kappa\}}$ – количество теплоты, подводимое к термодинамической κ-системе (или отводимое от нее), то (76) можно переписать в виде

$$d\tilde{\mathcal{F}}_{\{\kappa\}} = d'Q_{\{\kappa\}} - P_{ph}dV - k_{\mathcal{B}}\ln \mathcal{Z}_{\{\kappa\}} dT_{ph} - T_{ph} w_{\{\kappa\}} dS_{\{\kappa\}} \tag{79}$$

Отсюда следует, переопределение термодинамической энтропии Клаузиуса $\tilde{S}_{\{\kappa\}}$ для неаддитивных систем:

$$d\tilde{S}_{\{\kappa\}} = d'Q_{\{\kappa\}} / T_{ph}. \tag{80}$$

где

$$d\tilde{S}_{\{\kappa\}} = w_{\{\kappa\}} dS_{\{\kappa\}} = \frac{dS_{\{\kappa\}}}{\sqrt{1+\left(\kappa S^{eq}_{\{\kappa\}}/k_{\mathcal{B}}\right)^2}}.$$

Введем теперь следующие характеристические функции: обобщённую энтальпию $H_{\{\kappa\}} = \mathcal{E}_{\{\kappa\}} + P_{ph}V$ и обобщённый термодинамический потенциал $G_{\{\kappa\}} = \mathcal{F}_{\{\kappa\}} + P_{ph}V$. Напомним, что все характеристические функции обладают следующим свойством: если известна характеристическая функция, выраженная через соответствующие переменные (свои для каждой функции), то с ее помощью можно вычислить любую термодинамическую величину. В этом нетрудно убедиться из уравнений

$$d\mathcal{E}_{\{\kappa\}} = T_{ph} w_{\{\kappa\}} d\tilde{S}_{\{\kappa\}} - P_{ph}dV, \tag{81}$$

$$dH_{\{\kappa\}} = T_{ph} w_{\{\kappa\}} d\tilde{S}_{\{\kappa\}} + VdP_{ph}, \tag{82}$$

$$d\tilde{\mathcal{F}}_{\{\kappa\}} = -k_{\mathcal{B}}\ln \mathcal{Z}_{\{\kappa\}} dT_{ph} - P_{ph}dV, \tag{83}$$

$$dG_{\{\kappa\}} = -k_{\mathcal{B}}\ln \mathcal{Z}_{\{\kappa\}} dT_{ph} + P_{ph}dV, \tag{84}$$

из которых вытекают следующие обобщённые термодинамические соотношения:

$$\left(\frac{\partial \mathcal{E}_{\{\kappa\}}}{\partial V}\right)_{\tilde{S}_{\{\kappa\}}} = \left(\frac{\partial \tilde{\mathcal{F}}_{\{\kappa\}}}{\partial V}\right)_{T_{ph}} = -P_{ph}, \quad \left(\frac{\partial \mathcal{E}_{\{\kappa\}}}{\partial \tilde{S}_{\{\kappa\}}}\right)_{V} = \left(\frac{\partial H_{\{\kappa\}}}{\partial \tilde{S}_{\{\kappa\}}}\right)_{P_{ph}} = T_{ph}, \tag{85}$$

$$\left(\frac{\partial H_{\{\kappa\}}}{\partial P_{ph}}\right)_{\tilde{S}_{\{\kappa\}}} = \left(\frac{\partial G_{\{\kappa\}}}{\partial P_{ph}}\right)_{T_{ph}} = V, \quad \left(\frac{\partial \tilde{\mathcal{F}}_{\{\kappa\}}}{\partial T_{ph}}\right)_{V} = \left(\frac{\partial \mathcal{G}_{\{\kappa\}}}{\partial T_{ph}}\right)_{P_{ph}} = k_{\mathcal{B}} \ln \mathcal{Z}_{\{\kappa\}}. \tag{86}$$

Уравнение для теплоёмкостей. Как известно, в классической термодинамике теплоёмкость вещества в наиболее общем виде определяется соотношением: $C_\gamma := T\left(\partial S / \partial T\right)_\gamma$. Здесь C_γ – теплоёмкость процесса, в котором сохраняется постоянным параметр γ – любая обобщенная координата. Наиболее распространёнными являются изобарная теплоёмкость и изохорная теплоёмкость, которые в рассматриваемом случае определим соотношениями:

$$C_p := T_{ph}\left(\partial \tilde{S}_{\{\kappa\}} / \partial T_{ph}\right)_{P_{ph}}, \quad C_V := T_{ph}\left(\partial \tilde{S}_{\{\kappa\}} / \partial T_{ph}\right)_{V}. \tag{87}$$

Так как в соответствии с формулой $\left(\partial y / \partial x\right)_z = \left(\partial y / \partial u\right)_z \left(\partial u / \partial x\right)_z$ (справедливой для случая двух переменных, когда $y = y(x,z)$ и $u = u(x,z)$) имеем

$$\left(\frac{\partial \tilde{S}_{\{\kappa\}}}{\partial T_{ph}}\right)_{P_{ph}} = \left(\frac{\partial \tilde{S}_{\{\kappa\}}}{\partial H_{\{\kappa\}}}\right)_{P_{ph}} \left(\frac{\partial H_{\{\kappa\}}}{\partial T_{ph}}\right)_{P_{ph}} \quad \text{и} \quad \left(\frac{\partial \tilde{S}_{\{\kappa\}}}{\partial T_{ph}}\right)_{V} = \left(\frac{\partial \tilde{S}_{\{\kappa\}}}{\partial \mathcal{E}_{\{\kappa\}}}\right)_{V} \left(\frac{\partial \mathcal{E}_{\{\kappa\}}}{\partial T_{ph}}\right)_{V}, \tag{88}$$

а из (85) и (86) следует, что $\left(\partial \tilde{S}_{\{\kappa\}} / \partial H_{\{\kappa\}}\right)_{P_{ph}} = 1/T_{ph}$, $\left(\partial \tilde{S}_{\{\kappa\}} / \partial \mathcal{E}_{\{\kappa\}}\right)_{V} = 1/T_{ph}$. Следовательно выражения (87) можно переписать в виде

$$C_p = \left(\partial H_{\{\kappa\}} / \partial T_{ph}\right)_{P_{ph}}, \quad C_V = \left(\partial \mathcal{E}_{\{\kappa\}} / \partial T_{ph}\right)_{V}. \tag{89}$$

Получим теперь связующее соотношение между теплоёмкостями C_p и C_V. С использованием равенства

$$\left(\partial z / \partial m\right)_n = \left(\partial z / \partial x\right)_y \left(\partial x / \partial m\right)_n + \left(\partial z / \partial y\right)_x \left(\partial y / \partial m\right)_n, \tag{90}$$

являющегося следствием выражения для полного дифференциала функции $z = z(x,y)$, легко получить (при $m = x$) соотношение

$$\left(\partial \tilde{S}_{\{\kappa\}} / \partial T_{ph}\right)_{P_{ph}} = \left(\partial \tilde{S}_{\{\kappa\}} / \partial T_{ph}\right)_{V} + \left(\partial \tilde{S}_{\{\kappa\}} / \partial V\right)_{T_{ph}} \left(\partial V / \partial T_{ph}\right)_{P_{ph}}. \tag{91}$$

Отсюда, при учете уравнения Максвелла $(\partial S_{\{\kappa\}}/\partial V)_{T_{ph}} = (\partial P_{ph}/T_{ph})_V$, следует

$$C_p - C_V = T_{ph}\left(\partial P_{ph}/\partial T_{ph}\right)_V \left(\partial V/\partial T_{ph}\right)_{P_{ph}}. \tag{92}$$

Это соотношение может быть записано и в другом виде, если использовать так называемую «связку трёх производных» $\left(\partial z/\partial x\right)_y \left(\partial x/\partial y\right)_z \left(\partial y/\partial z\right)_x = -1$ (следствие соотношения (90) при $m = x,\ n = z$), из которой следует

$$\left(\partial P_{ph}/\partial T_{ph}\right)_V = -\left(\partial V/\partial T_{ph}\right)_{P_{ph}} \left(\partial P_{ph}/\partial V\right)_{T_{ph}}. \tag{93}$$

С учётом (93) связь между введенными теплоёмкостями приобретает почти обычный вид:

$$C_p - C_V = -T_{ph}\left(\partial V/\partial T_{ph}\right)^2_{P_{ph}} / \left(\partial V/\partial P_{ph}\right)_{T_{ph}}. \tag{94}$$

Таким образом, используя обобщенную энтропию Клаузиуса (80) можно получить основной набор макроскопических термодинамических соотношений (85), (86) и (94) для термодинамики Каниадакиса. Стандартная форма полученных соотношений свидетельствует об их инвариантности относительно неаддитивной модификации их классических аналогов.

6. ДИВЕРГЕНЦИЯ БРЭГМАНА. ОБОБЩЕННАЯ *H*-ТЕОРЕМА

Покажем теперь, что в равновесном состоянии энтропия Каниадакиса достигает своего максимального значения. Рассмотрим с этой целью так называемую дивергенция Брэгмана [40,41]

$$D_{\{\kappa\}}\left[p : p_0\right] := S_{\{\kappa\}}(p_0(\boldsymbol{r})) - S_{\{\kappa\}}(p(\boldsymbol{r})) + \lambda k_{\mathcal{B}} \int \left(p_0(\boldsymbol{r}) - p(\boldsymbol{r})\right) \ln_{\{\kappa\}}\left(\frac{p_0(\boldsymbol{r})}{\alpha}\right) d\Gamma \geq 0, \tag{95}$$

которая относится к наиболее существенным статистическим характеристикам динамической системы [42]. Являясь функционалом, она определяет меру статистической упорядоченности в микросостояниях системы с распределением $p(\boldsymbol{r})$ относительно состояния с распределением $p_0(\boldsymbol{r})$. Выражение (95) представляет собой функционал для двух нормированных распределений $\int p(\boldsymbol{r}) d\Gamma = = \int p_0(\boldsymbol{r}) d\Gamma = 1$.

Различные свойства дивергенции Брэгмана в полном объеме приведены в работе [41]. Здесь же мы отметим, что величина $D_{\{\kappa\}}(p : p_0)$ является вещественным, положительным, выпуклым (в первом аргументе) функционалом. Легко видеть, что при $\kappa \to 0$ эта величина переходит в известную информацию различия Кульбака−Лейблера (см. [10, 43])

$$D_{\{\kappa\}\to 0}(p : p_0) \Rightarrow K(p : p_0) := k_{\mathcal{B}} \iint p(\boldsymbol{r}) \ln\left(\frac{p(\boldsymbol{r})}{p_0(\boldsymbol{r})}\right) d\Gamma. \tag{96}$$

Кроме этого, поскольку при $p = p_0$ имеет место равенство $D_{\{\kappa\}}(p:p) = 0$, то дивергенция Брэгмана является функцией Ляпунова[i].

Принцип максимума энтропии равновесного распределения. Пусть распределение $p_0(\boldsymbol{r})$ является равновесным, для которого справедливо представление (47): $p_0(\boldsymbol{r}) = \alpha \exp_{\{\kappa\}}\{-\mathcal{X}(\boldsymbol{r})/\lambda\}$, а распределение $p(\boldsymbol{r},t)$ соответствует произвольному состоянию системы. Кроме этого, будем полагать, что для обоих представлений справедливо так называемое, условие Гиббса [10]:

$$\mathcal{E}_{\{\kappa\}} = \mathcal{E}^{eq}_{\{\kappa\}}, \quad \int p_0(\boldsymbol{r})\mathcal{X}(\boldsymbol{r})d\Gamma = \int p(\boldsymbol{r},t)\mathcal{X}(\boldsymbol{r})d\Gamma, \quad \text{где} \quad \mathcal{X}(\boldsymbol{r}) := \gamma + \frac{\beta}{k_{\mathcal{B}}}\mathcal{H}(\boldsymbol{r}). \tag{97}$$

Покажем, что в этом случае справедливо следующее равенство

$$\int \big(p_0(\boldsymbol{r}) - p(\boldsymbol{r})\big)\ln_{\{\kappa\}}\big(p_0(\boldsymbol{r})\big)d\Gamma = \int \big(p(\boldsymbol{r}) - p_0(\boldsymbol{r})\big)u_{\{\kappa\}}\big(p_0(\boldsymbol{r})\big)d\Gamma. \tag{98}$$

Действительно, поскольку в силу (21) и (47), мы имеем

$$u_{\{\kappa\}}(p_0) = -\ln_{\{\kappa\}}(p_0) - \lambda \ln_{\{\kappa\}}(\alpha / p_0) = -\ln_{\{\kappa\}}(p_0) - \mathcal{X}, \tag{99}$$

то отсюда, при учете (97), следует соотношение (98).

Из определения дивергенции Брэгмана (95), учете тождества (97) и справедливости принятых выше предположений, следует соотношение

$$S_{\{\kappa\}}(p(\boldsymbol{r})) = S_{\{\kappa\}}(p_0(\boldsymbol{r})) - D_{\{\kappa\}}\big[p:p_0\big] + k_{\mathcal{B}}\int \big(p_0(\boldsymbol{r}) - p(\boldsymbol{r})\big)\mathcal{X}d\Gamma = S^{eq}_{\{\kappa\}}(p_0(\boldsymbol{r})) - D_{\{\kappa\}}\big[p:p_0\big] \tag{100}$$

где информация различия, представленная в виде отрицательного вклада в энтропию, называется негэнтропией. Понятие негэнтропии, т.е. изменения энтропии с обратным знаком, было предложено Э. Шредингером [43]. В общем случае выполняется негэнтропийный принцип Бриллюэна [44]

$$D_{\{\kappa\}}\big[p:p_0\big] + S_{\{\kappa\}}(p(\boldsymbol{r})) - S^{eq}_{\{\kappa\}}(p_0(\boldsymbol{r})) \geq 0, \tag{101}$$

где знак неравенства соответствует необратимым процессам в замкнутой системе. Из неравенства (100) следует, что энтропия равновесного состояния больше, чем энтропия произвольного состояния, $S_{\{\kappa\}}(p(\boldsymbol{r})) \leq S^{eq}_{\{\kappa\}}$.

Теорема Гиббса и *H*-теорема. Как известно, открытая система представляет собой часть большой замкнутой системы и находится с внешним окружением в неравновесном контакте. В окружении отсутствуют неравновесные явления и можно считать, что она находится в равновесном состоянии с распределением $p_0(\boldsymbol{r})$. Сравнивая значения энтропий при условии Гиббса (97), получим из (100) теорему Гиббса [10] в виде неравенства

[i] Напомним, что функцией Ляпунова называется знакоопределённая функция, которая обращается в нуль в точке равновесия системы. Состояние равновесия является аттрактором, когда производная по времени от функции Ляпунова имеет знак, противоположный знаку самой функции.

$$D_{\{\kappa\}}\left[p:p_0\right]=-\left[S_{\{\kappa\}}(p(\boldsymbol{r}),t)+S_{\{\kappa\}}(p_0(\boldsymbol{r}))\right]\geq 0\,. \tag{102}$$

Таким образом, увеличение энтропии системы до ее максимального значения в равновесии происходит совместно с потерей информации различия, то есть имеет место совместное увеличение статистической разупорядоченности и уменьшение статистической упорядоченности микросостояний неэкстенсивной системы.

Поскольку согласно свойству выпуклости [41]) дивергенции Брэгмана $D_{\{\kappa\}}\left[p:p_0\right]$ является знакоопределенной функцией Ляпунова, то для того, чтобы состояние полного равновесия $p_0(\boldsymbol{r})$ было устойчивым, необходимо выполнение следующего неравенства

$$\frac{d}{dt}D_{\{\kappa\}}\left[p:p_0\right]=-\frac{d}{dt}\left[S_{\{\kappa\}}(p(\boldsymbol{r}),t)-S^{eq}_{\{\kappa\}}(p_0(\boldsymbol{r}))\right]\leq 0\,. \tag{103}$$

Таким образом, при стремлении κ-системы к равновесному состоянию во временной эволюции информация различия уменьшается. Из (102) следует H-теорема для открытых неравновесных κ-систем (неравенство для энтропии Каниадакиса)

$$dS_{\{\kappa\}}(p(\boldsymbol{r}),t)/dt>0\,, \tag{104}$$

которое справедливо при приближении к состоянию полного статистического равновесия. Эта теорема утверждает, что κ-энтропия непрерывно растет при приближении системы к равновесию, где энтропия становится максимальной и достигает конечного значения. Таким образом, происходит хаотизация макроскопической системы Каниадакиса при спонтанных переходах.

Заметим, что тот факт, что к-энтропия квазиаддитивна, и энтропия совокупной системы больше, чем сумма энтропий отдельных подсистем, указывает на то, что совокупная система термодинамически более стабильна [7].

7. ЗАКЛЮЧЕНИЕ

В последнее десятилетие все больший интерес привлекает к-статистика Каниадакиса, поскольку она обладает многими важными свойствами (принцип максимальной энтропии, термодинамическая устойчивость, устойчивость Леша, непрерывность и т.п.). Это позволяет к-статистике быть одной из самых удачных обобщений статистической механики Больцмана−Гиббса, особенно в контексте специальной теории относительности и полученных распределений, наблюдающихся в различных физических, природных и искусственных системах.

В представленной работе дается логическая схема построения модифицированной термодинамики неэкстенсивных систем, основанная на κ-энтропии. Найдено универсальное распределение степенного закона на основе максимизации энтропии Каниадакиса при заданном ограничении усредненного значения внутренней энергии системы. Полученные при этом термодинамические равенства для обобщенного канонического ансамбля Гиббса вполне аналогичны обычным равенствам статистической термодинамики для замкнутых и открытых систем и совпадают с ними при приближении параметра деформации κ к нулю. Показано, что энтропия Каниадакиса подчиняется псевдоаддитивному закону для двух статистически независимых подсистем. Доказанная положительность κ-энтропии сово-

купной системы указывает на существование силы притяжения между отдельными подсистемами, и это взаимодействие явно происходит от взаимодействия между их составными частями. Получено обобщение нулевого закона термодинамики для двух независимых неэкстенсивных систем при их тепловом контакте, вводящее в рассмотрение так называемую физическую температуру T_{ph} отличающуюся от инверсии множителя Лагранжа β. Этот факт потребовал переопределения ряда термодинамических соотношений, получаемых естественным путем в рамках статистики Каниадакиса. В качестве основных предпосылок, взятых за исходный пункт нахождения модифицированных термодинамических соотношений в работе выбраны первый закон термодинамики и структура преобразования Лежандра. Путем применения обобщенной энтропии Клаузиуса к аномальной κ-системе, был получен наиболее приемлемый набор макроскопических термодинамических соотношений для неэкстенсивной κ-системы. Наконец, на основе, так называемой дивергенции Брэгмана были сформулированы и доказаны *H*-теорема и теорема Гиббса, описывающие хаотизацию макроскопической системы Каниадакиса при спонтанных переходах.

Развитый в работе подход позволяет моделировать, в частности, сложные космологические и космогонические среды (от галактик и астрофизических дисков до космической пыли), отличительной чертой которых является наличие динамических структур с нецелой топологической размерностью (фракталов), дальнодействующего силового взаимодействия, а также эридитальности и немарковости эволюционных процессов.

REFERENCES

[1] C. Tsallis, “Possible Generalization of Boltzmann-Gibbs-Statistics”, *J. Stat. Phys*., **52**(1-2), 479-487 (1988).

[2] Nonextensive statistical mechanics and thermodynamics: *Full bibliography/ http://tsallis.cat.cbpf.br/biblio.htm*. (accessed 12 February 2020).

[3] A. Renyi, *Probability Theory*. Amsterdam: North-Holland Publ. Co. (1970).

[4] B.D. Sharma, D.P. Mittal, “New Nonadditive Measures of Relative Information”, *J. Comb. Inform. and Syst. Sci*., **2**, 122-133 (1977).

[5] I.J. Taneja, “New Developments in Generalized Information Measures”. Chapter in: Advances in Imaging and Electron Physics, ed. P.W. Hawkes. *London: Academic Press*, **91**, 37-135 (1995).

[6] S. Abe, “A note on the q-deformation-theoretic aspect of the generalized entropies in nonextensive physics”, *Physics Letters A*, **224**, 326-330 (1997).

[7] P.T. Landsberg, V. Vedral, “Distributions and channel capacities in generalized statistical mechanics”, *Phys. Lett. A*. **247**, 211-216 (1998).

[8] T.D. Frank, A.R Plastino, ”Generalized thermostatics based on the Sharma-Mittal entropy and escort mean value”, *Eur. Phys. J. B*, **30**, 543-549 (2002).

[9] R.G. Zaripov, *Samoorganizatsiya i neobratimost' v neekstensivnykh sistemakh*, Kazan': Izd-vo Fen, (2002).

[10] R.G. Zaripov, *Printsipy neekstensivnoy statisticheskoy mekhaniki i geometriya mer besporyadka i poryadka*. Kazan': Izd-vo Kazan. Gos. tekhn. un-ta. (2010).

[11] A.V. Kolesnichenko, ”K razrabotke neadditivnoy termodinamiki kvantovykh sistem na osnove statistiki Tsallisa”, *Mathematica Montisnigri*, **45**, 26-51 (2019).

[12] A.V. Kolesnichenko, ”Modelirovaniye lineynogo otklika kvantovoy neekstensivnoy sistemy na dinamicheskoye vneshneye vozmushcheniye”, *Matemat. Model.*, **31** (12), 97-118 (2019).

[13] A.V. Kolesnichenko, “Dvukhparametricheskiy entropiynyy funktsional Sharma-Mittala kak osnova semeystva obobshchyenykh termodinamik neekstensivnykh system”, *Mathematica Montisnigri*, **42**, 74-101 (2018).

[14] A.V. Kolesnichenko, “K obosnovaniyu v ramkakh neekstensivnoy statistiki Tsallisa sootnosheniy vzaimnosti Onzagera dlya kineticheskikh koeffitsiyentov”, *Mathematica Montisnigri*, **44**, 41-59 (2019).

[15] A.V. Kolesnichenko, *Statisticheskaya mekhanika i termodinamika Tsallisa neadditivnykh system: Vvedenie v teoriyu i prilozheniya*. Moskow: LENAND. (Sinergetika ot proshlogo k budushchemu. № 87), (2019).

[16] 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).

[17] A.V. Kolesnichenko, “Kriteriy termicheskoy ustoychivosti i zakon raspredeleniy chastits dlya samogravitiruyushchikh astrofizicheskikh system v ramkakh statistiki Tsallisa”, *Mathematica Montisnigri*, **37**, 45-75 (2016).

[18] A.V. Kolesnichenko, “Power distributions for self-gravitating astrophysical systems based on nonextensive Tsallis kinetics”, *Solar System Research*, **51**(2), 127-144 (2017).

[19] 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).

[20] 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).

[21] G. Kaniadakis, “Non-linear kinetics underlying generalized statistics”, *Physica A*, **296**, 405-425 (2001)

[22] G. Kaniadakis, A.M. Scarfone, “A new one-parameter deformation of the exponential function”, Physica A, **305**, 69-75 (2002).

[23] G. Kaniadakis,”Statistical mechanics in the context of special relativity II”, *Phys. Rev. E*, **72**, 036108 (2005).

[24] E.M.C. Abreu, J. Neto Ananias., E.M. Barboza, R.C. Nunes, “Holographic considerations on non-gaussian statistics and gravothermal catastrophe”, *Physica A*, **441**, 141-150 (2016).

[25] J.C. Carvalho, J.D. do Nascimento Jr., R. Silva, J.R. De Medeiros, ”Non-Gaussian Statistics and Stellar Rotational Velocities of Main-Sequence Field Stars”, *Astrophys. Journ. Lett*., **696**, L48-L51 (2009).

[26] A.M. Teweldeberhan, H.G. Miller, R. Tegen, “κ-Deformed statistics and the formation of a quark-gluon plasma”, *Int. J. Mod. Phys. E*, **12**, 669-673 (2003).

[27] A. Rossani, A.M. Scarfone, ”Generalized kinetic equations for a system of interacting atoms and photons: theory and simulations”, *J. Physics A: Mathematical and Theoretical*, **37** (18), 4955-4975 (2004).

[28] G. Kaniadakis, “Statistical origin of quantum mechanics”, *Physica A*, **307**, 172-184 (2002).

[29] E.T. Jaynes, “Information theory and statistical mechanics”, *Statistical Physics. Brandeis Ltctures*, **3**, 160 (1963).

[30] G. Kaniadakis, “Theoretical Foundations and Mathematical Formalism of the Power-Law Tailed Statistical Distributions*”, Entropy,* **15**, 3983-4010 (2013).

[31] D.P. Zubarev, *Neravnovesnaya statisticheskaya mekhanika,* M.: Nauka, 1971.

[32] A.M. Scarfone, “On the κ -Deformed Cyclic Functions and the Generalized Fourier Series in the Framework of the κ -Algebra”, Entropy, **17**, 2812-2833 (2015).

[33] A.M. Scarfone, “ κ -Deformed Fourier Transform”, *Physica A: Statistical Mechanics and its Applications*, **480**, 63-78 (2017).

[34] A.M. Scarfone, T.Wada, “Thermodynamic equilibrium and its stability for microcanonical systems described by the Sharma-Taneja-Mittal entropy”, *Physical Review E*, **72**(2), 026123 (2005).
[35] A.M. Scarfone, T. Wada, “Canonical partition function for anomalous systems described by the κ-entropy”, *Prog. Theor. Phys. Suppl.*, **162**, 45 -52 (2006).
[36] C. Tsallis. Introduction to Nonextensive Statistical Mechanics. Approaching a Complex World. New York: Springer, 2009. 382 p.
[37] A.M. Scarfone, T. Wada, “Legendre structure of κ-thermostatistics revisited in the framework of information geometry”, *J. Phys. A*, **47**, 275002 (17 pp) (2014).
[38] S Abe, S. Martinez, F. Pennini, A. Plastino, “Nonextensive thermodynamic relations”, *Physics Letters A*, **281**(2-3), 126-130 (2001).
[39] S. Abe, ”Heat and generalized Clausius entropy of nonextensive systems”, *Eprint arXiv:cond-mat/0012115*, **3,** 1-14 (200).
[40] L. M.Bregman, ”The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming”, *USSR computational mathematics and mathematical physics*, **7**(3), 200-217 (1967).
[41] A. Cichocki, S. Amari, “Families of Alpha- Beta- and Gamma- Divergences: Flexible and Robust Measures of Similarities”, *Entropy*, **12**, 1532-1568 (2010).
[42] A.M. Scarfone, “A Maximal Entropy Distribution Derivation of the Sharma-Taneja-Mittal Entropic Form”, *Open Systems & Information Dynamics*, **25**(1), 1850002-1−1850002-11 (2018).
[43] E. Shredinger, *Chto takoye zhizn' s tochki zreniya fiziki?* M.: Inost. Liter. (1947).
[44] S. Kul'bak, *Teoriya informatsii i statistika*, M.: Nauka. (1967).

Received May 15, 2020

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

MODERN SCIENTIFIC JOURNAL: OVERLAY, CROWDSOURCING, ALTMETRICS

T.A. POLILOVA*

Keldysh Institute of Applied Mathematics of RAS
*Corresponding author. E-mail: polilova@keldysh.ru

DOI: 10.20948/mathmontis-2020-48-11

Summary. The scientific Internet space is actively developing due to the activities of competing participants: publishers of traditional journals with paid access for readers and publishers of Open access journals. Paid journals, using the opportunities of their highly profitable publishing business, offer their users a rich environment of access to the materials of articles. Open access journals are looking for ways to reduce the cost of publishing. The new publication overlay scheme significantly reduces the publisher's expenses. The overlay journal reviews the received article. After the article is accepted for publishing, the article is assigned a DOI and the article's metadata is published on the journal's website along with a link to the full text published in the Open access repository.

The scientific online journal, in contrast to the traditional printed journal, exists in an open information environment, relying on new media and means of communication. The interface with this environment should fully cover all stages of the publishing process — from submission to publication of the article — and further support for changes made by the author based on the results of discussion of the article in the scientific community. One of the key functions of the publishing house is to conduct reviews to ensure the required level of quality of published articles. However, further information technologies come into play, expanding the scope of traditional peer review. In an open environment, any scientist or expert can express their opinion on the merits of an article published on the site within the framework of the problems that they deal with professionally. In the modern world, the reader is not satisfied with the role of a passive consumer of scientific information. Internet technologies and the ideas of Open science have given rise to the ideology of crowdsourcing — all members of the professional scientific community become active participants of the creative process. The opinion of a wide range of specialists should be available to the general reader and participate in the formation of ratings of scientific resources. Articles on the journal's website are accompanied by webometrics and altmetrics indicators that characterize the interest in the article from the scientific Internet community.

1 INTRODUCTION

The growth curve of the number of scientific publications in the world over the past two decades has been going up sharply. The number of scientific journals is also growing. Today, it is safe to say that all scientific journals are published on the Internet. If a scientific journal is not available on the Internet, it can be assumed that the journal does not exist [1]. The scientific Internet space of Western countries is actively developing as a result of the activities of competing participants: publishers of traditional journals with paid access for readers and publishers of Open access journals. In the West, in the 60-80-ies of the last century, several

2010 Mathematics Subject Classification: 00-02, 00A99
Key words and Phrases: Scientific publication, Peer review, Crowdsourcing, Overlay journal, Altmetrics.

large publishing houses appeared that publish scientific journals. Publishing scientific journals has become a stable, highly profitable business.

At the same time, the Open access movement began to take shape: in 2002, European countries adopted the Budapest Declaration on free access to research results that were carried out with governmental funding. The concept of Open access implies that any scientist, representative of the educational community or business should have unrestricted access to the world's scientific knowledge and cultural heritage. Open access to scientific information on the Internet is becoming an attractive philosophy for society.

In Russia, the situation with scientific journals is developing according to a special scenario [2]. Editorial and publishing preparation of issues of scientific journals is usually performed by scientific institutes, universities, and other structures with budget funding that have publishing divisions. Editorial boards of journals consist of employees of institutes who perform editorial work, combining it with their main responsibilities. Until the early 90's, russian printed scientific journals were distributed mainly by subscription, while the price of journals was very low — barely covering printing costs. The subscription was used by large libraries, libraries of scientific organizations and universities, as well as scientists.

In the early 2000s, the printings of scientific journals decreased to a minimum — no more than 2-3 hundred copies. Journals began to move smoothly to the Internet, minimizing their expenses. Scientific Online journals with a paid subscription and an embargo period of 2-3 years have appeared. A pleasant event was the order of the Court of Accounts of the Russian Federation in 2018, obliging publishers to make all academic journals freely available.

It is known that since the beginning of the 2000s, Russia has been undergoing permanent reforms affecting the structures of the Academy of Sciences and higher education. The main measure of a researcher's success has become the number of published papers. A significant increase of the publication activity of scientists again attracted attention to scientific journals. New journals began to appear massively. According to the largest russian aggregator of scientific periodicals eLibrary.ru there are 17432 russian scientific journals in it's database, 6213 of them with full texts are in open access. Thus, in Russia, Open access covers about 35% of scientific journals.

Why have commercial foreign journals ceased to satisfy the scientific community? First of all, for economic reasons: the cost of subscription has become a very significant burden, eating up a huge part of the budget of libraries and scientific institutions. However, major publishers have learned to lobby for their interests, drawing funds from state budgets and scientific foundations. In Russia, lobbyists managed to get paid subscriptions to packages of Western journals for some scientific institutes and universities. The Russian Foundation for basic research conducts annual competitions for access to Western journals.

The model of paid access to scientific journals is implemented by well-known Western publishers with a long history: Springer, Elsevier, Wiley, and Informa. It is claimed that the publication of one peer-reviewed article in the journal Nature costs 40 thousand dollars [3]. Let's compare this amount with the cost of publishing one moderated preprint in arXiv — 10 dollars [3]. Let's ask whether the publisher of the journal Nature is trying to ensure a comfortable existence by obtaining superprofits?

Open access online journals provide their readers with free access to the journal's materials. Who is responsible for the costs associated with editorial and publishing preparation, metadata processing, and so on? One of the business models of Open access journals assumes that expenses are reimbursed by contributions taken from the authors of

articles. In this case, the author will be asked to pay several thousand us dollars for publishing the article. In the scientific community, there is a growing number of supporters of Fair Open access ideas, calling for a more decisive reduction in the fee for publishing an article [4]. At the same time, there are other examples: the author does not pay any fees for publishing an article (preprint) in the Open archive arXiv. For the initial moderation of articles in arXiv, volunteers employees are involved. The costs (very modest) for maintaining the activities of arXiv are borne by Cornell University (USA).

2 OPEN ACCESS PUBLICATIONS

The author has a choice in which journal to publish the article: in the journal of limited (for the reader) access or in the journal of Open access. If an article is submitted to a restricted access journal, the paper will be free of charge. But the readership may be significantly reduced, since not every scientist will be able to pay for access to a closed article. When an open access journal is choosing, the author may have to pay for the costs of publishing the article himself or look for a sponsor.

Note that in addition to the two alternatives (open access and paid access), there are also hybrid models. The journal policy can be flexible. The journal with paid access for the reader opens free access to some publications, if the authors of these articles pay a certain fee. Or an Open access journal with fee for authors may be free of fee for certain categories of authors. In this case, the financial costs are borne by research support funds or publishers themselves.

If the author has decided to publish his article in a journal that is paid for readers, he often retains the opportunity to locate the article as a preprint in an Open repository on the website of his organization, or in a thematic Open archive of preprints, or on his personal page. Often, the author is interested in quick publication. The publishing process of a traditional journal can take several months. During this time, the author may lose priority in their field, and the material of the article may lose relevance. Many restricted access journals allow preprints to appear on the Internet, as Springer published [5]. But the publisher puts forward a condition: if the author's version of the article is published as a preprint, then after accepting the article, the editorial Board suggests that the author indicate in the preprint that the article was accepted and published in the specified journal.

If the article was published as a preprint in an Open repository, it is likely that information about this preprint will reach the mass reader after some time. In ensuring broad access to the article posted on a public website, a greater role plays by search engines such as Google, Yandex. The search service uses its robots (crawlers) to extract metaattributes of articles and place them in its indexes. If a search query for an article appears in search services, it is highly likely that the desired article will appear in the list of responses found. It is more difficult to index an article in Google Scholar. Not every site can be indexed in this service. To do this, the article (or website) must meet certain design requirements [6].

Can an author refuse to re-publish their preprint as a journal article? Unlikely. To obtain the status of a full-fledged scientific work, the article must be published in a peer-reviewed journal. But then there is an ethical problem — the repeated publication of virtually the same text. Some programs for searching for matching texts such as "Antiplagiat" easily detect this kind of duplication and the author will then have to explain that he did not actually violate ethical norms [7, 8]. And now we should pay attention to a new type of scientific journal —

an overlay journal that implements an unconventional scheme for organizing interaction between the author and the editorial Board [9].

Overlay journal adheres to three principles: open access, publish for free, and read without restrictions. To implement these principles, we need to reduce the cost of publishing an overlay journal by significantly reducing the cost of publishing articles in the public domain.

The scheme of interaction between the author and the publisher of an overlay journal is as follows. The author publishes his article in an open repository. The article passes moderation, organized by the repository holders. After that, the author sends the article published in the repository to the overlay journal of the corresponding topic. The journal reviews the article. If a positive review is received and the article is accepted for publication, the article is assigned a DOI. The final version of the reviewed article is re-published in the repository as a new version together with a link to the journal. The overlay journal publishes only the article metadata and a link to the final text of the article on the repository site. Thus, an overlay journal is a set of metadata and links to the full texts of peer-reviewed articles.

One example of support for publishing an overlay journal is the Episciences platform [10]. The interface with the publishing platform fully covers all stages of the publishing process-from submission to publication of the article and further support of the article. Currently, the platform hosts half a dozen open access journals that adhere to the overlay scheme.

The scope of preprint servers is constantly expanding. We can assume that in the near future, overlay journals will occupy a worthy niche and find their authors and readers in various research communities.

3 THE PEER REVIEW PROCESS

Scientific journals with high academic standards are very demanding about the quality of published articles. All articles are subject to mandatory review. Such journals openly publish their publishing policies and ethical standards. Journals, in particular, make sure that the author of the article and the reviewer do not have a "conflict of interests". The author should expect an objective review of his article without the influence of group interests or an antagonism of competing scientific schools.

However, the peer review process in journals is often malfunctioning. Not every reviewer can give an objective assessment of the article. Let's remember, for comparison, how the analysis of dissertations goes. Procedures for the defence of the dissertation suggest the presence of two or three opponents. The review of the organization designated by the dissertation Council as the lead organization is also considered. In addition, the dissertation defense takes into account the feedback of scientists who sent their assessments and comments on the dissertation. The decision on the dissertation is made collectively by all members of the Council. For obvious reasons, reviewers of journals will not be able to provide this level of a review.

It should also be mentioned that one of the time-consuming stages of reviewing is often associated with the detection of plagiarism and "self-plagiarism" in the article. There are no questions about plagiarism: it is not only an obvious ethical violation, but also a violation of copyright law. But that can’t be said about "self-plagiarism": the author's repeated use of his previously written text does not contradict the current legislation. On the contrary, the practice of some editors to combat such reuse of text often violates the rights of the author. The

author's inclusion of text fragments from his previously published works in a new article is often dictated by the need to present the context of the research. Authors can use previously published texts in order to ensure self-sufficiency of the article. It is more convenient for readers to have a self-contained text before their eyes, even if it contains fragments of a previously published article by the author, than to switch to another article, losing the context of the current narrative. However, some ethical codes consider such "self-plagiarism" to be an incorrect publishing practice. Journal editors try to avoid possible conflicts and may reject an article if they find intersections with previously published articles. Thus, the journal does not publish an article that is actually a bona fide scientific work [11].

Another example. If the editorial Board charges the author money for publishing the article, the commercial interests of the journal often obscure the objectivity of the review. It is also known that often journals, having achieved high ratings, begin to trade their popularity. Such "predatory" journals promise authors quick access to prestigious Western databases and force authors to pay thousands of dollars for this opportunity. In such cases, we are no longer talking about objective review.

There are cases when publishers launch advertising market projects under the guise of scientific journals. This story happened to the well-known firm Elsevier [12]. Between 2000 and 2005, Elsevier's Australian division published a series of collections of articles intended for family doctors that were actually advertisements for the products of a pharmaceutical company. When this unpleasant story for Elsevier came to light, Michael Hansen, general director of the health sciences division, issued a statement. He admitted that this was an unacceptable practice, and the publisher regrets that such a case took place [13].

The examples can be continued. In 2012, a book was published by British doctor and scientist Ben Goldacr "The whole truth about medicines: the global conspiracy of pharmaceutical companies", dedicated to the description of contractual schemes between representatives of the pharmaceutical industry and doctors. Doctors often fulfilled the role of advertisers by publishing articles about non-existent achievements. There were cases when there were facts of pressure on researchers from pharmaceutical companies to release untested drugs to the market. In 2013 this book was published in the United States, and in 2015 it was translated into Russian [14].

We can mention another reason for doubting the overall objectivity of the review. Scientific knowledge and technologies are developing rapidly. New directions are emerging, including at the intersection of several disciplines. Often in the scientific environment there is a conflict of generations: young researchers often use modern methods that are not mastered by specialists of the older generation. If a reviewer of a classical journal has no experience in a new field and is not familiar with its specifics, he is unlikely to be able to objectively evaluate a pioneer article. The history of science provides many examples when new revolutionary ideas could not break through the wall of misunderstanding of traditionalists for a long time.

Thus, classical review in a traditional journal has no less serious risks than not too deep moderation of the article, which is used, in particular, in arXiv.org. We should not forget that there are other equally reliable ways to verify the quality of the article. As a rule, articles that are interesting and useful for specialists have good traffic indicators and high indicators of alternative metrics. Here we can also talk about high rates of bibliographic citation, but with some clauses.

4 THE CITATION USING BIBLIOGRAPHIC LINKS

Bibliography in the article began to play a special role after the appearance of bibliographic databases that aggregate data from scientific journals, and after the creation of computer methods for processing bibliographic references from the texts of articles. It is now possible to identify automatically links between articles and journals on the base of the analysis of lists of cited literature. Bibliometrics has developed rapidly as a set of methods for evaluating the publication activity of scientists and methods for building journal ratings. The period of interest in the bibliometrics is not over yet. The pillars of the bibliometrics are the leading Western databases Web of Science and Scopus, and many large national electronic repositories. The largest Russian Scientific electronic library, eLibrary, contributes to the bibliometrics movement.

Some countries, when evaluate the effectiveness of national research, rely on citation indicators for scientific articles. In other countries, such as the United Kingdom, the assessment of scientific effectiveness is mainly based on expert assessment. A recent study by the UK Research Excellence Framework (REF) found notable discrepancies between citation-based metrics and peer review results. It should be noted that earlier similar studies showed a higher degree of agreement between metrics and expert assessments, which was the reason for introducing metric indicators. Now there is considerable uncertainty about the feasibility of relying on citation metrics [15].

If you are planning, for example, to publish an article in the leading journal Nature, you will have to prepare the bibliography of 30-40 items. A short bibliographic list may be interpreted by the editorial Board as the author's lack of the necessary outlook in the field of research under consideration. But the point here is not only and not so much in demonstrating immersion in the subject area. The entire industry of evaluating the quality of articles and the prestige of journals is based on the bibliographic lists. The author's h-index is calculated based on the number of citations of the article in other articles. The journal's impact factor is calculated based on the citation count of articles in the journal. And then the ratings of authors and journals are built on the basis of these indicators. But when a certain technology is superimposed on real publishing practice, there are often unnatural relationships between the participants in the process. Special behavior of authors and editors of journals is formed. The authors can agree on a mutual citation. A similar collusion may occur among a group of journals in order to raise the impact factor to each other. Intermediaries immediately appear with offers to increase the h-index for a fee: you can find dozens of offers of this type on the Internet.

Note that citation indicators based on the analysis of bibliographic lists are not accurate and comprehensive. There is also the problem of lost links. This problem occurs when authors deviate from the prescribed formats of bibliographic references, make significant errors in the spelling of authors ' surnames and names, or make inaccuracies in the title of the article or other parameters of the link. Usually, bibliographic databases have tools that can be used to find lost citations and edit inaccurate bibliographic references. The paper [16] describes the experience of correcting bibliographic references pointing to Russian journals indexed in Web of Science. After correcting the links, the impact factor indicators of a group of Russian journals in Web of Science increased by 4-37%. One more convincing example can be given: after clarifying the links to the articles of the journal "Keldysh Institute Preprints" in the eLibrary.ru in 2017, the number of citations of the journal increased three times.

The disadvantages of indicators based on the analysis of bibliographic references should include their binding to a specific bibliographic database. If your article is referenced from articles and journals that are not related to the subject areas of the bibliographic database, such citations will not be taken into account. As a result, the author has several significantly different indicators of the h-index from different bibliographic databases. The problem of combining citation counters for the same article from different bibliographic databases does not yet have a comprehensive solution. Approaches to solving this problem are presented in [17].

In addition, the h-index and journal impact factor indicators are very inertial, they are formed over a long time and are usually calculated no more than once a year. To calculate a two-year indicator, the database must have issues of the journal for the previous 3 years, and for a five-year indicator you need issues of journals for the previous 5 years.

Citation indicators are formed more dynamically based on the analysis of bibliographic lists of articles in the Google Scholar system. This system uses data from a wide variety of open online sources. The data collection process is automated: programs have been developed that view web pages and extract information from metadata descriptions, as well as directly from article texts. In Google Scholar, the author's h-index is usually higher than similar indicators of such Western bibliographic databases as Web of Science, Scopus, or the Russian electronic library eLibrary.ru.

Another disadvantage of the citation index is that it ignores the context and the reason for citation. In most cases, the citation indicates a high popularity of the article. However, the author can also refer to the article for reasons of disagreement with certain provisions or with the General conclusions of the article. Referring to an article with incorrect content, in the author's opinion, the author further in his article expresses his own alternative opinion. But a traditional bibliographic link does not convey the context of the citation, thus the author of a bad article will increase his citation index and will claim the status of an authoritative scientist.

There is also the problem of forgotten authors who were the first to get significant results. This problem can occur when an author has published their pioneering results in a modest journal without high ratings. The article can be quoted and explained by another author who is published in a more prestigious journal. Further, readers who are interested in the published result will prefer to cite a publication in a prestigious journal, since they thereby increase the authority of their own article by using the authority of the cited journal. Such chains of citations lead to the loss of information about the author who was the first to receive an interesting scientific result. In continuation of this topic, we can give an example of preprints. It is known that after the publication of a preprint, the author usually publishes an article in the journal based on the materials of the preprint. Further, we often see that the number of citations of a preprint is much less than the number of citations of a journal article. This is despite the fact that the preprint is the primary source of the obtained scientific result.

5 ALTMETRICS AND CROWDSOURCING

Developing Internet technologies have significantly changed the approaches to assessing the significance of articles and the influence of ideas presented in articles on the development of science. There are tools that allow you to count the number of requests to a scientific article

and the number of downloads of the article, and show the geography of requests. Most of holders of scientific Internet resources have counters of webometric information on their sites. Web Analytics tools continue to evolve. Attempts are made to identify a set of quantitative indicators that give a clear and unambiguous assessment of the level of motivation of the user to get acquainted with the resource materials [18].

Do not think that the high quality of the article will ensure its success in the web environment. The author must make sure that the article is easy to notice.

One of the most common ways to find an article is to make a request in a universal search service (for example, in Google). If the search was successful, this indicates a good visibility of the article on the Internet. However, if the author of the article actively uses social networks and participates in forums, dedicated to discussing scientific problems, his visibility on the Internet increases further. There is the separate direction [19] related to the assessment of the article's popularity in such systems that are still unconventional for scientists as social networks, forums, and specialized platforms for scientific discussions. This direction is called altimetry. The Altmetrics Manifesto was published in 2010 [20].

The results of scientific research are multidimensional. They can include achievements in a specific scientific direction, in an interdisciplinary methodology, in the development of research technologies. Finally, scientific results can have a social effect, bringing new ideas to broad social groups — from researchers to politicians. The evaluation of such a scientific result should also be multidimensional; it is hardly possible to successfully apply any one metric or one evaluation model [21].

A large amount of the material related to the development of scientific ideas and understanding of the results obtained does not fall into traditional journal articles. Such materials remain in the records of open discussions, on the personal pages of scientists, on the pages of Institute websites, in educational materials for students, in comments to articles, in publications in the media [22]. Using existing services such as Mendeley, CiteULike, or Zotero, scientists can organize their personal library of materials that do not have the format of the scientific article, and thus make these materials available to Internet users [20].

In some cases, a scientist may write a short note that other scientists might later cite. However, such a "nanopublication" traditional journal, most likely, will not be able to publish. As a result, readers will not be able to find a note using search tools focused on finding journal articles. Other search queries are required here. Altmetrics should establish mechanisms accounting for this kind of "nanopublications".

Scientists widely exchange data sets, software codes, experimental techniques, algorithms, etc. ("raw science"). Authors who have received such data do not collect traditional authority ratings — their h-index does not depend on the results associated with obtaining and publishing data sets on websites or in archives. Other indicators are required that take into account this contribution of the scientist. This direction is also included in the area of interest of altmetrics.

One of the characteristics of the article is the review. Until recently, the materials of reviews remained on the shelves of publishers. Currently, the idea of placing reviews together with the article in open access is often implemented, which also fits into the philosophy of altmetrics. In addition, the modern review ceases to be a one-time text accompanying the article. The structure of review materials becomes more complex. The review process develops in a dialogue between the reviewer and the author, and the content of this dialogue is of interest to the readers of the article. There are also versions of the article with regular

changes made after the discussion. After the article is published, the author can get a new review from interested experts, or conduct peer-review in another overlay journal. Reviewing, generally speaking, is not limited in time. Each review gives the author a reason to continue developing his published material. Thus, the article becomes the "alive publication" [23].

Internet technologies and the ideology of Open science make it possible to implement crowdsourcing-style peer review, when a representative community of experts, not just reviewers appointed by the editorial Board, participates in evaluating the quality of an Internet resource. This approach to peer-reviewing is implemented, for example, in the F1000research project [24], where any of several thousand experts of this project can evaluate the article and its additional materials.

F1000research is an Open science publishing platform for rapid publication of scientific articles in the fields of physical and biological sciences, engineering, medicine, social sciences, and humanities. How is the review process organized on this platform? Original articles are accepted for publication, regardless of the intended level of interest or novelty. All articles are published in open access. Authors are invited to attach detailed descriptions of methods, posters, and slides to the text of the article. The author also has the option to provide a link to the original data underlying the study to ensure reproducibility of the results.

The article submitted to F1000Research first passes a quick initial check for compliance with the general editorial practice and is placed on the site with the status "waiting for expert evaluation". Then an open review is conducted, with the authors and reviewers collaborating to make the article as complete as possible. The names of reviewers and the status they assign to the article after reviewing are published together with the article. In the future, any other expert of the f1000research publishing house also has the right to review the article on its own initiative, further clarifying its status. As soon as an article gets two "Approved" status, or two "Approved with reservations" status and one "Approved" status, it will be indexed in various bibliographic databases (in PubMed, PubMed Central, MEDLINE, etc.). If the article is indexed, all versions along with the review reports are sent for storage.

Do experts have incentives to take on the work of reviewing articles? Yes, the F1000Research project has created such a mechanism. The name of the expert and his reviews are open to the entire community, thus actively and conscientiously working expert increases his rating. In addition, there is a noticeable discount for experts who participate in reviewing articles when paying their personal contribution for future publications [25]. By stimulating the activity of experts and authors, the F1000Research project consolidates the scientific community and allows each scientist to participate in the formation of a collective scientific product. Articles in F1000Research can be updated and supplemented at any time after publication, and each version can be independently quoted with its own DOI. The editorial Board suggests the following format for a link to the article [26]:

Author name(s). Article title [version number; details of peer review status]. F1000Research Year, Volume: Publication number (article doi)

All components of the cited bibliographic reference are clear to the reader. The "Status" attribute will require additional explanation. The status indicates the number of checks that are "approved", "approved with reservations", or "not approved".

In addition, regardless of the article, the review itself becomes the object of citation. The review is published under the CC BY 4.0 license, and each review is assigned a DOI. The platform offers the following format for a link to a review [26]:

Reviewer name(s). Peer review report for: Article title [version number; details of peer review status]. F1000Research Year, Volume: Publication number (review doi)

6 CONCLUSION

In our opinion, the evaluation of a scientific article should be multi-dimensional. Classical peer review has risks no less serious than shallow moderation, followed by crowdsourcing, which is active throughout the life of the article. Crowdsourcing in a scientific publication can be defined in simple words: the community of scientists is a powerful resource, and connecting this resource to the review of scientific articles makes it possible to get a better scientific product. On the site of the article, in addition to the review, the reader would be interested to see altmetrics indicators obtained from information from social networks, thematic blogs and forums for professional communication.

The assessment of the scientific significance of an article has been based only on bibliometric indicators based on the analysis of bibliographic references to it for several decades. In our opinion, obtaining indicators based on the analysis of bibliographic references in quoting articles is too long in time, sensitive to errors in the recording of references and unrepresentative, since it is usually limited to one specific bibliographic base.

A promising direction in scientific publishing practice is the overlay journal, which reviews articles from open archives of preprints. After the article is published, the author can get a new review from interested experts, or conduct a review in another overlay journal. Post-review, generally speaking, is not limited in time. Reviews and feedback from colleagues inspire the author to continue developing his article. Thus, the article becomes the "alive publication".

The struggle imposed on the scientific community for high rates of bibliographic citation and for increasing the number of publications in highly rated journals obscures significantly more important tasks: the development of Open access infrastructure, the creation of communication tools for participants in the publishing process, and the enrichment of the means of presenting scientific materials on open publishing platforms.

REFERENCES

[1] M. M. Gorbunov-Posadov, T. A. Polilova, “Tools to Support Scientific Online Publishing”, *Programming and Computer Software*, **45 (**3), 116–120 (2019). https://link.springer.com/article/10.1134%2FS0361768819030046

[2] S. Beliaeva, “Tsena otkrytosti: Vo chto oboidetsia perekhod k Open Access?”, *Poisk*. (2019). https://www.poisknews.ru/skript/czena-otkrytosti-vo-chto-obojdetsya-perehod-k-open-access/

[3] J.R. Adler, T.M. Chan, J.B. Blain, B. Thoma, Atkinson, “OpenAccess: Free online, open-access crowdsource-reviewed publishing is the future; traditional peer-reviewed journals are on the way out”, *Canadian Journal of Emergency Medicine*, **21** (1), 11– 14 (2019). https://doi.org/10.1017/cem.2018.481

[4] Fair Open Access Alliance. https://www.fairopenaccess.org/ (Accessed July 22, 2020)

[5] Springer, Self-archiving policy. https://www.springer.com/gp/open-access/publication-policies/self-archiving-policy (Accessed July 22, 2020)

[6] Google Scholar, Inclusion Guidelines for Webmasters. https://scholar.google.com/intl/en-

US/scholar/inclusion.html#overview (Accessed July 22, 2020)
[7] T.A. Polilova. “Ethical norms and legal framework of scientific publication”. *Mathematica Montisnigri*, **XLV**, 129-136 (2019) http://www.montis.pmf.ac.me/vol45/11.pdf doi: 10.20948/mathmontis-2019-45-11
[8] T.A. Polilova, “Nauchnaia publikatsiia v Rossii: intellektualnye prava”, *Preprinty IPM im. M.V. Keldysha*. 56, 1-24 (2019). http://keldysh.ru/papers/2019/prep2019_56.pdf doi:10.20948/prepr-2019-56
[9] E. Herman, J. Akeroyd, G. Bequet, D. Nicholas, A. Watkinson. “The changed – and changing – landscape of serials publishing: Review of the literature on emerging models”, *Learned Publishing.* (2020). https://doi.org/10.1002/leap.1288 https://onlinelibrary.wiley.com/doi/full/10.1002/leap.1288
[10] Episciences.org, Overlay Journal Platform. www.episciences.org/?lang=en (Accessed July 22, 2020)
[11] T. Polilova, A. Ermakov, “Dissernet and self-plagiarism”, *CEUR Workshop Proceedings*, **2543**, 285-294 (2020). https://www.scopus.com/record/display.uri?eid=2-s2.0-85078459740&origin=resultslist&sort=plf-f&src=s&st1=Dissernet+and+self-plagiarism&st2=&sid=d8727e9f55ec2b7fc177de73ce64e6d1&sot=b&sdt=b&sl=44&s=TITLE-ABS-KEY%28Dissernet+and+self-plagiarism%29&relpos=0&citeCnt=0&searchTerm= (Accessed July 22, 2020)
[12] Wikipedia, Elsevier. https://ru.wikipedia.org/wiki/Elsevier (Accessed July 22, 2020)
[13] Tom Reller. Statement From Michael Hansen, CEO Of Elsevier's Health Sciences Division, Regarding Australia Based Sponsored Journal Practices Between 2000 And 2005. https://www.elsevier.com/about/press-releases/clinical-solutions/statement-from-michael-hansen,-ceo-of-elseviers-health-sciences-division,-regarding-australia-based-sponsored-journal-practices-between-2000-and-2005 (Accessed July 22, 2020)
[14] “Vsia pravda o lekarstvakh, Google i bestsellerah, Galina Iuzefovich — pro knigi, kotorye vse obieiasniaiut”. https://meduza.io/feature/2015/07/04/vsya-pravda-o-lekarstvah-google-i-bestsellerah (Accessed July 22, 2020)
[15] V.A. Traag, L. Waltman, “Systematic analysis of agreement between metrics and peer review in the UK REF”, *Palgrave Communications*, **5**, Article number: 29 (2019). https://doi.org/10.1057/s41599-019-0233-x
[16] D.E. Chebukov, “Poisk poteriannykh tsitirovanii v Web of Science. Ispravlenie oshibok v spiskakh literatury Web of Science”, *Nauchnyi servis v seti Internet*, 461-467 (2017). http://keldysh.ru/abrau/2017/77.pdf doi:10.20948/abrau-2017-77
[17] YongGao QiangWu LinnaZhu, “Merging the citations received by arXiv-deposited e-prints and their corresponding published journal articles: Problems and perspectives”, *Information Processing & Management*, **57** (5), (2020). https://doi.org/10.1016/j.ipm.2020.102267
[18] Iu.G. Reviakin, “Web-analitika dlia nauchnykh publikatsii”, *Preprinty IPM im. M.V. Keldysha*, 50, 1-42 (2020). http://keldysh.ru/papers/2020/prep2020_50.pdf doi:10.20948/prepr-2020-50
[19] M.N. Saushkin, D.E. Chebukov, “Altmetriki na saite nauchnogo zhurnala”, *Nauchnyi servis v seti Internet*, 593-599 (2019). http://keldysh.ru/abrau/2019/theses/40.pdf doi:10.20948/abrau-2019-40
[20] J. Priem, D. Taraborelli, P. Groth, C. Neylon, “Altmetrics: A manifesto”, (2010). http://altmetrics.org/manifesto (Accessed July 22, 2020)
[21] M. V. Vakhrushev, “Altmetriki, vebometriki i informetriki kak vzaimodopolniaiushchie napravleniia v sovremennoi bibliometrii”, *Nauchnye i tekhnicheskie biblioteki*, 8, 67-76 (2019). https://ntb.gpntb.ru/jour/article/viewFile/470/453
[22] A. Grossmann, “Publishing in transition – do we still need scientific journals?”, *ScienceOpen Research*, (2015). https://www.scienceopen.com/document_file/e1dd3665-6406-4a32-befc-

e00d84a72cd1/ScienceOpen/3077_XE696973259861784096.pdf doi: 10.14293/S2199-1006.1.SOR-SOCSCI.ACKE0Y.v1
[23] M.M. Gorbunov-Posadov, “Zhivaia publikatsiia”, (Moscow: KIAM). https://keldysh.ru/gorbunov/live.htm (Accessed July 22, 2020)
[24] F1000Research. Open for Science. https://f1000research.com/ (Accessed July 22, 2020)
[25] F1000Research. Open for Science: Referee Incentives. http://f1000research.com/referee-incentives (Accessed July 22, 2020)
[26] F1000Research. Open for Science: How it Works. https://f1000research.com/about (Accessed July 22, 2020)

Received June 10, 2020

MATHEMATICA MONTISNIGRI
Vol XLVIII (2020)

СОВРЕМЕННЫЙ НАУЧНЫЙ ЖУРНАЛ: ОВЕРЛЕЙ, КРАУДСОРСИНГ, АЛЬТМЕТРИКИ

Т.А. ПОЛИЛОВА*

Институт прикладной математики им. М.В. Келдыша РАН. Москва, Россия

*Ответственный автор. E-mail: polilova@keldysh.ru

DOI: 10.20948/mathmontis-2020-48-11

Ключевые слова: научная публикация, рецензирование, краудсорсинг, оверлейный журнал, альтметрики.

Аннотация. Научное интернет-пространство активно развивается в результате деятельности конкурирующих участников: издателей традиционных журналов платного для читателя доступа и издателей журналов Открытого доступа. Платные журналы, используя возможности своего высокодоходного издательского бизнеса, предлагают своим пользователям насыщенную среду доступа к материалам статей. Журналы Открытого доступа изыскивают возможности удешевления затрат на издание. Новая оверлейная схема издания позволяет заметно снизить расходы издателя. Оверлейный журнал выполняет рецензирование поступившей статьи. После принятия статьи к изданию статье присваивается DOI и на сайте журнала размещаются метаданные статьи вместе с ссылкой на полный текст, размещенный в репозитории Открытого доступа.

Онлайновый научный журнал, в отличие от своего прародителя традиционного печатного журнала, существует в открытой информационной среде, опираясь на новые медиа и средства коммуникации. Интерфейс с этой средой должен полностью охватывать все стадии издательского процесса — от представления до опубликования статьи — и далее сопровождение вносимых автором изменений по результатам обсуждения статьи в научном сообществе. Одна из ключевых функций издательства — проведение рецензирования для обеспечения требуемого уровня качества публикуемых статей. Однако далее в игру вступают информационные технологии, расширяющие сферу применения традиционного рецензирования. В открытой среде любой ученый или эксперт может высказать свое мнение о достоинствах размещенной на сайте статьи в рамках тех проблем, которыми он занимается профессионально. В современном мире читатель не удовлетворяется ролью пассивного потребителя научной информации. Технологии интернета и идеи Открытой науки породили идеологию краудсорсинга — все члены профессионального научного сообщества становятся активными участниками творческого процесса. Мнение широких кругов специалистов должно быть доступно для массового читателя и участвовать в формировании рейтингов научных ресурсов. Статьи на сайте журнала сопровождаются вебометрическими и альтметрическими показателями, которые характеризуют интерес к статье со стороны научного интернет-сообщества.

2010 Mathematics Subject Classification: 00-02, 00A99
Key words and Phrases: Scientific publication, Peer review, Crowdsourcing, Overlay journal, Altmetrics.

1 ВВЕДЕНИЕ

Кривая роста числа научных публикаций в мире в последние два десятилетия резко идет вверх. Растет и количество научных журналов. Сегодня можно с уверенностью заявить, что все научные журналы размещаются в интернете. Если научный журнал не представлен в интернете, можно считать, что он не существует [1]. Научное интернет-пространство западных стран активно развивается в результате деятельности конкурирующих участников: издателей традиционных журналов платного для читателей доступа и издателей журналов Открытого доступа. На Западе еще в 60-80-х годах прошлого столетия появились несколько крупных издательств, занимающихся изданием научных журналов. Выпуск научных журналов превратился в стабильный высокодоходный бизнес.

В то же время стало формироваться движение Открытого доступа: в 2002 году европейские страны приняли Будапештскую декларацию о свободном доступе к результатам исследований, выполненных при государственном финансировании. Концепция Открытого доступа предполагает, что любой ученый, представитель образовательного сообщества или бизнеса должен иметь беспрепятственный доступ к мировому научному знанию и культурному наследию. Открытый доступ к научной информации в интернете становится привлекательной для общества философией.

В России ситуация с научными журналами развивается по особому сценарию [2]. Редакционно-издательскую подготовку выпусков научных журналов выполняют, как правило, научные институты, вузы, иные структуры с бюджетным финансированием, имеющие издательские подразделения. Редколлегии журналов состоят из сотрудников институтов, выполняющих редакционную работу, совмещая ее со своими основными обязанностями. До начала 90-х годов российские печатные научные журналы распространялись преимущественно по подписке, при этом цена журналов была весьма низкой — едва покрывала полиграфические расходы. Подпиской пользовались крупные библиотеки, библиотеки научных организаций и вузов, а также ученые.

В начале 2000-х годов тиражи научных журналов в бумажном исполнении снизились до минимума — не превышали 2-3 сотен экземпляров. Журналы стали плавно перемещаться в интернет, минимизируя свои затраты. Появились онлайновые научные журналы с платной подпиской и периодом эмбарго в 2-3 года. Приятным событием стало распоряжение Счетной палаты РФ в 2018 г., обязывающее издателей выкладывать в свободный доступ все академические журналы.

Известно, что с начала 2000-х годов в России шли перманентные реформы, затрагивающие структуры Академии наук и высшей школы. Главным мерилом успешной работы научного сотрудника стало количество опубликованных им работ. Значительное повышение публикационной активности ученых вновь привлекло внимание к научным журналам. Стали массово появляться новые журналы. По данным крупнейшего российского агрегатора научной периодики eLibrary.ru на текущий момент в его базе насчитывается 17432 российских научных журналов, из них 6213 — с полными текстами в открытом доступе. Таким образом, в России Открытый доступ охватывает около 35% научных журналов.

Почему коммерческие, платные для читателя, западные журналы перестали удовлетворять научное сообщество? Прежде всего, по экономическим причинам: стоимость подписки стала весьма ощутимым бременем, съедая огромную часть

бюджета библиотек и научных учреждений. Однако крупные издатели научились лоббировать свои интересы, вытягивая средства из бюджетов государств и научных фондов. В России лоббистам удалось добиться платной подписки на пакеты западных журналов для части научных институтов и вузов. Ежегодные конкурсы на получение доступа к западным журналам проводит Российский фонд фундаментальных исследований.

Модель платного доступа к научным журналам реализуют известные западные издательства с многолетней историей: Springer, Elsevier, Wiley, Informa. Утверждается, что публикация одной рецензируемой статьи в журнале Nature обходится в 40 тысяч долларов [3]. Сравним эту сумму с себестоимостью публикации одного модерируемого препринта в arXiv — 10 долларов [3]. Зададим вопрос, не пытается ли издатель журнала Nature обеспечить себе безбедное существование за счет получения сверхприбыли?

Онлайновые журналы Открытого доступа предоставляют своим читателям бесплатный доступ к материалам журнала. Кто же несет расходы, связанные с редакционно-издательской подготовкой, обработкой метаданных и пр.? Одна из бизнес-моделей журналов Открытого доступа предполагает, что расходы возмещаются за счет взносов, берущихся с авторов статей. В этом случае за публикацию статьи автору будет предложено заплатить несколько тысяч долларов США. В научном сообществе растет число сторонников идей Справедливого Открытого доступа, призывающих более решительно снизить плату за публикацию статьи [4]. В то же время есть и другие примеры: за размещение статьи (препринта) в Открытом архиве препринтов arXiv автор не платит каких-либо взносов. К первичной модерации статей в arXiv привлекаются сотрудники-энтузиасты. Расходы (весьма скромные) на поддержание деятельности arXiv несет Корнеллский университет (США).

2 ПУБЛИКАЦИЯ В ОТКРЫТОМ ДОСТУПЕ

Перед автором стоит выбор, в каком журнале опубликовать статью: в журнале ограниченного (для читателя) доступа или в журнале Открытого доступа. В случае передачи статьи в журнал ограниченного доступа оплату с автора не потребуют. Но читательская аудитория может заметно сократиться, поскольку не каждый ученый сможет оплатить доступ к закрытой статье. При выборе журнала открытого доступа автору, возможно, придется самому оплачивать затраты на опубликование статьи или искать спонсора.

Отметим, что помимо двух альтернатив (открытый доступ — платный доступ) существуют также гибридные модели. Политика журнала может быть подвижной. Журнал с платным для читателя доступом открывает бесплатный доступ к некоторым публикациям, если авторы этих статей внесут определенную плату. Или платный для авторов журнал Открытого доступа может не взимать плату с некоторых категорий авторов. Финансовые издержки в этом случае берут на себя фонды поддержки научных исследований или сами издатели.

Если автор решил опубликовать свою статью в платном для читателей журнале, у него чаще всего сохраняется возможность предварительно разместить статью в виде

препринта в Открытом репозитории[1] на сайте своей организации, в каком-либо тематическом Открытом архиве препринтов или на своей персональной странице. Часто автор заинтересован именно в оперативной публикации. Издательский процесс традиционного журнала может занимать несколько месяцев. За это время автор может лишиться приоритета в своей области, а материал статьи может потерять актуальность. Многие журналы ограниченного доступа допускают возможность появления в интернете препринтов, например, журналы издательства Springer [5]. Но издательство выдвигает условие: если авторская версия статьи публикуется в виде препринта, то после принятия статьи редакция журнала предлагает автору указать в препринте, что статья была принята и опубликована в указанном журнале.

В том случае, когда статья была опубликована в виде препринта в Открытом репозитории, скорее всего, информация об этом препринте через некоторое время дойдет до массового читателя. В обеспечении широкого доступа к статье, размещенной на открытом сайте, большую роль играют поисковые сервисы, такие как Google, Яндекс. С помощью своих роботов поисковый сервис извлекает метаданные[2] статей и размещает ее в своих индексах. Если в поисковых сервисах возникает запрос на поиск статьи, то с большой вероятностью нужная статья окажется в списке найденных ответов. Сложнее обстоит дело с индексирование статьи в Google Scholar. Далеко не каждый сайт может быть проиндексирован в этом сервисе. Для этого нужно, чтобы статья (или сайт) отвечали определенным требованиям к оформлению [6].

Может ли автор отказаться от повторного опубликования своего препринта в виде журнальной статьи? Вряд ли. Для получения статуса полноценного научного труда статья должна быть опубликована в рецензируемом журнале. Но тогда возникает проблема этического свойства — повторная публикация фактически одного и того же текста. Некоторые программы поиска совпадающих текстов типа «Антиплагиат» без труда выявляют подобного рода совпадения, и автору потом придется объясняться, что он на самом деле не нарушал этические нормы [7, 8]. И вот теперь следует обратить внимание на новый тип научного журнала — оверлейный[3] журнал, реализующий нетрадиционную схему организации взаимодействия автора и редакции журнала [9].

Оверлейный журнал придерживается трех принципов: открытый доступ, бесплатно публиковать, читать без ограничений. Реализовать провозглашенные принципы помогает уменьшение себестоимости выпуска оверлейного журнала за счет существенного снижения стоимости размещения статей в открытом доступе.

Схема взаимодействия автора и издателя оверлейного журнала состоит в следующем. Автор публикует свою статью в каком-либо открытом репозитории.

[1] Репозиторий — хранилище данных. В репозитории на сайте научного учреждения или вуза хранятся, обычно в открытом доступе, препринты, отчеты, статьи, монографии и пр.

[2] Метаданные — данные, относящиеся к дополнительной информации о содержимом или объекте. Метаданные статьи включают название, ФИО авторов, организации авторов, аннотацию статьи, ключевые слова и т.д. Метаданные передаются в библиографические базы для реализации на их основе поисковых запросов.

[3] Оверлейный журнал — тип научного журнала открытого доступа, который реализуется как надстройка «поверх» статей, уже размещенных в открытом доступе. Статьи рецензируются, окончательная версия принятой к публикации статьи вновь размещается на прежнем месте. Журнал размещает только метаданные статьи (название, авторы, аннотация и др.) вместе со ссылкой на полный текст. Источником статей для оверлейного журнала часто служат серверы препринтов.

Статья проходит модерацию, организованную держателями репозитория. После этого автор направляет в оверлейный журнал соответствующей тематики размещенную в репозитории статью. Журнал проводит рецензирование статьи. В случае получения положительной рецензии и принятия статьи к публикации, статье присваивается DOI. Окончательная версия прошедшей рецензирование статьи повторно размещается в репозитории в виде новой версии вместе с ссылкой на журнал. Оверлейный журнал публикует только метаданные статьи и ссылку на окончательный текст статьи на сайте репозитория. Таким образом, оверлейный журнал представляет собой набор метаданных и ссылок на полные тексты статей, прошедших рецензирование.

Одним из примеров поддержки издания оверлейного журнала является платформа Episciences [10]. Интерфейс с издательской платформой полностью охватывает все стадии издательского процесса — от представления до опубликования статьи и дальнейшее сопровождение статьи. В настоящее время на платформе размещено уже полтора десятка журналов открытого доступа, придерживающихся оверлейной схемы.

Сфера деятельности серверов препринтов постоянно расширяется. Можно предположить, что в ближайшем будущем оверлейные журналы займут достойную нишу и найдут своих авторов и читателей в различных исследовательских сообществах.

3 РЕЦЕНЗИРОВАНИЕ

Научные журналы с высокими академическими стандартами весьма требовательно относятся к качеству публикуемых статей. Все статьи проходят обязательное рецензирование. Такие журналы открыто публикуют свою издательскую политику и этические нормы. Журналы, в частности, заботятся, чтобы у автора статьи и рецензента не было «конфликта интересов». Автор должен рассчитывать на объективное рассмотрение своей статьи без влияния групповых интересов или антагонизма конкурирующих научных школ.

Но все же рецензирование в журналах нередко дает сбои. Далеко не каждый рецензент может дать объективную оценку статье. Давайте вспомним, для сравнения, как проходит анализ диссертационных работ. Процедуры защиты предполагают наличие двух или трех оппонентов. Рассматривается также отзыв организации, назначенной диссертационным советом в качестве ведущей организации. Кроме того, при защите диссертации учитываются отзывы ученых, приславших свои оценки и замечания по диссертации. Решение по диссертации принимается коллегиально всеми членами совета. По понятным причинам такой уровень рассмотрения работы рецензенты журналов обеспечить не смогут.

Следует также упомянуть, что один из трудоемких этапов рецензирования часто связан с обнаружением в статье плагиата и «самоплагиата». В отношении плагиата вопросов не возникает: плагиат является не только очевидным этическим нарушением, но и нарушением законодательства об авторском праве. Чего нельзя сказать о «самоплагиате»: повторное использование автором своего ранее написанного текста не противоречит действующему законодательству. Напротив, практикуемая некоторыми изданиями борьба с таким повторным использованием текстов нередко нарушает права автора. Включение автором фрагментов текста из своих ранее опубликованных работ часто диктуется необходимостью представить контекст проведения исследования.

Авторы могут использовать ранее опубликованные тексты с целью обеспечить самодостаточность текста статьи. Для читателя более удобно иметь перед глазами самодостаточный текст, пусть даже и содержащий фрагменты ранее изданной статьи автора, чем делать переход на другую статью, теряя контекст текущего повествования. Тем не менее, некоторые этические кодексы считают подобный «самоплагиат» некорректной издательской практикой. Редакторы журналов стараются избегать возможных конфликтов и могут отклонить статью, если встретят пересечения с ранее опубликованными статьями. Тем самым журнал не публикует статью, которая на самом деле является добросовестной научной работой [11].

Другой пример. Если редакция журнала берет с автора деньги за опубликование статьи, то часто коммерческие интересы журнала заслоняют объективность рецензирования. Известно также, что нередко журналы, добившись высоких рейтингов, начинают торговать своей популярностью. Такие «хищнические» журналы обещают авторам быстрое попадание в престижные западные базы и вынуждают авторов платить тысячи долларов за эту возможность. Об объективном рецензировании в таких случаях речь уже не идет.

Известны случаи, когда под видом научных журналов издательства запускают рекламные рыночные проекты. Такая история случилась с известной фирмой Elsevier [12]. В период с 2000 по 2005 год австралийское отделение Elsevier выпустило серию сборников статей, предназначенных для семейных врачей, которые фактически были рекламой продукции одной фармакологической компании. Когда вскрылась эта неприятная для Elsevier история, генеральный директор отдела наук о здоровье Michael Hansen опубликовал заявление. Он признался, что это была недопустимая практика, издательство сожалеет, что такой случай имел место [13].

Примеры можно продолжить. В 2012 г. была опубликована книга британского врача и учёного Бена Голдакра «Вся правда о лекарствах: мировой заговор фармкомпаний», посвященная описанию договорных схем между представителями фармацевтической промышленности и медиками. Медики часто выполняли роль рекламщиков, публикуя статьи о несуществующих достижениях. Приводились случаи, когда всплывали факты давления на исследователей со стороны фармацевтических компаний с целью выпустить на рынок непроверенные лекарственные препараты. В 2013 г. эта книга была издана в США, а в 2015 г. была переведена на русский язык [14].

Можно упомянуть еще один повод для сомнений во всеобъемлющей объективности рецензирования. Научные знания и технологии развиваются стремительно. Возникают новые направления, в том числе на стыке нескольких дисциплин. Нередко в научной среде возникает конфликт поколений: молодые исследователи часто используют современные методы, которыми не владеют специалисты старшего поколения. Если рецензент классического журнала не имеет опыта в новой области и не знаком с ее спецификой, он вряд ли сможет объективно оценить пионерскую статью. История науки знает множество примеров, когда новые революционные идеи долго не могли пробиться через стену непонимания традиционалистов.

Таким образом, классическое рецензирование в традиционном журнале имеет не менее серьезные риски, чем не слишком глубокая модерация статьи, применяемая, в частности, в arXiv.org. Не следует забывать о том, что существуют и другие не менее надежные способы убедиться в качестве статьи. Как правило, интересные и полезные для специалистов статьи имеют хорошие показатели посещаемости и высокие

показатели альтернативных метрик. Тут можно говорить и о высоких показателях библиографической цитируемости, но с некоторыми оговорками.

4 ЦИТИРОВАНИЕ С ПОМОЩЬЮ БИБЛИОГРАФИЧЕСКИХ ССЫЛОК

Библиография в статье стала играть особую роль после появления библиографических баз, агрегирующих данные научных журналов, и после создания методов компьютерной обработки библиографических ссылок из текстов статей. Появилась возможность автоматически выявлять связи между статьями, между журналами на основе анализа списков цитируемой литературы. Бурное развитие получила библиометрия как совокупность методов оценки публикационной активности ученых, методов построения рейтингов журналов. Период увлечения библиометрией еще не закончился. Столпами библиометрии стали ведущие западные базы Web of Science и Scopus, многие крупные национальные электронные хранилища. Свой вклад в библиометрическое движение вносит крупнейшая российская Научная электронная библиотека eLibrary.

Некоторые страны, проводя оценку эффективности национальных исследований, опираются на показатели цитирования научных статей. В других странах, например в Великобритании, при оценке результативности научной деятельности используют в основном экспертную оценку. Проведенное недавно исследование UK Research Excellence Framework (REF) выявило заметные расхождения между метриками, основанными на цитировании, и результатами экспертных оценок. Следует отметить, что более ранние аналогичные исследования демонстрировали более высокую степень согласия метрик и экспертных оценок, что и стало поводом вводить метрические показатели. Теперь же появилась значительная неопределенность в целесообразности опоры на метрические показатели цитирования [15].

Если вы планируете, например, опубликовать статью в ведущем журнале Nature, то вам придется подготовить библиографический список из 30-40 позиций. Короткий библиографический список редакция может воспринять как отсутствие у автора необходимого кругозора в рассматриваемой области исследования. Но дело здесь не только и даже не столько в демонстрации погруженности в предметную область. На библиографических списках базируется целая индустрия оценки качества статей и престижности журналов. На базе подсчета числа цитирований статьи в других статьях вычисляется индекс Хирша автора. На базе подсчета цитирований статей журнала подсчитывается импакт-фактор журнала. И далее строятся рейтинги авторов и журналов на основе этих показателей. Но, когда некая технология накладывается на реальную издательскую практику, часто возникают противоестественные отношения между участниками процесса. Формируется особое поведение авторов и редакций журналов. Авторы могут договариваться о взаимном цитировании. Аналогичный сговор может возникнуть у группы журналов с целью поднять импакт-фактор друг другу. Тут же возникают посредники с предложениями повысить индекс Хирша за определенную плату: в интернете можно найти десятки предложений такого типа.

Стоит отметить, что показатели цитируемости на основе анализа библиографических списков не являются точными и всеобъемлющими. Существует проблема потерянных ссылок. Эта проблема возникает тогда, когда авторы отклоняются от предписанных

форматов библиографических ссылок, делают существенные ошибки в написании фамилий и имен авторов, допускают неточности в названии статьи или других параметрах ссылки. Обычно в библиографических базах имеются инструменты, с помощью которых можно найти потерянные цитирования, отредактировать неточные библиографические ссылки. В работе [16] описан опыт исправления библиографических ссылок, указывающих на российские журналы, индексируемые в Web of Science. После коррекции ссылок показатели импакт-фактора группы российских журналов в Web of Science увеличились на 4-37%. Можно привести еще один убедительный пример: после уточнения ссылок на статьи издания «Препринты ИПМ им. М.В. Келдыша» в eLibrary.ru в 2017 г. число цитирований издания выросло в три раза.

К недостаткам показателей, основанных на анализе библиографических ссылок, следует отнести их привязку к конкретной библиографической базе. Если на вашу статью ссылаются из статей и журналов, не относящихся по тематике к направлениям библиографической базы, то такие цитирования не будут учитываться. В итоге автор имеет сразу несколько существенно разнящихся показателей индекса Хирша от разных библиографических баз. Проблема объединения счетчиков цитирования одной и той же статьи из разных библиографических баз пока не имеет всеобъемлющего решения. Подходы к решению этой проблемы представлены в работе [17].

Кроме того, показатели индекса Хирша и импакт-фактора журналов весьма инерционны, они формируются длительное время и рассчитываются обычно не чаще одного раза в год. Для расчета двухлетнего показателя требуется наличие в базе выпусков журнала за 3 предыдущих года, а для пятилетнего показателя нужны выпуски журналов за 5 предыдущих лет.

Более динамично показатели цитируемости формируются на основе анализа библиографических списков статей в системе Google Scholar. Эта система использует данные самых разнообразных открытых онлайновых источников. Процесс сбора данных автоматизирован: разработаны программы, которые просматривают веб-страницы и извлекают информацию из описаний метаданных, а также непосредственно из текстов статей. В Google Scholar индекс Хирша автора обычно выше аналогичных показателей таких библиографических баз как Web of Science, Scopus или российской электронной библиотеки eLibrary.ru.

Еще один недостаток показателя цитируемости — игнорирование контекста и причины цитирования. В большинстве случаев цитирование говорит о высокой популярности статьи. Однако автор может сослаться на статью и по причине несогласия с отдельными положениями или с общими выводами статьи. Сославшись на статью с некорректным, по мнению автора, содержанием, автор далее в своей статье высказывает свое собственное альтернативное мнение. Но традиционная библиографическая ссылка не передает контекст цитирования, тем самым автор плохой статьи будет повышать свой показатель цитируемости и претендовать на статус авторитетного ученого.

Существует также проблема забытых авторов, которые первыми получили весомые результаты. Эта проблема может возникнуть, когда автор опубликовал свои пионерские результаты в скромном журнале без высоких рейтингов. Статью может процитировать и раскрыть суть опубликованных результатов другой автор, публикующийся в более престижном журнале. Далее читатели, заинтересовавшиеся опубликованным

результатом, предпочтут процитировать публикацию в престижном журнале, поскольку они тем самым с помощью авторитета цитируемого журнала повышают авторитет своей собственной статьи. Такие цепочки цитирований приводят к потере информации об авторе, первым получившим интересный научный результат. В продолжение этой темы можно привести пример препринтов. Известно, что после публикации препринта автор обычно публикует по материалам препринта статью в журнале. Далее мы нередко видим, что число цитирований препринта намного меньше числа цитирований журнальной статьи. И это несмотря на то, что именно препринт является первоисточником полученного научного результата.

5 АЛЬТМЕТРИКИ И КРАУДСОРСИНГ

Развивающиеся интернет-технологии заметно изменили подходы к оценке значимости статей, влияния изложенных в статьях идей на развитие науки. Появились инструменты, позволяющие подсчитывать количество обращений к научной статье и число скачиваний статьи, показывать географию обращений. Большинство держателей научных интернет-ресурсов заводят на своих сайтах счетчики вебометрической информации. Инструменты веб-аналитики продолжают развиваться. Предпринимаются попытки выделить набор количественных показателей, дающих понятную и однозначную оценку уровня мотивации посетителя к знакомству с материалами ресурса [18].

Не следует думать, что высокое качество статьи заведомо обеспечит ей успех в веб-среде. Автор обязан позаботиться о том, чтобы статью легко было заметить.

Один из самых распространенных способов найти статью — сделать запрос в универсальном поисковом сервисе (например, в Google). Если поиск был успешным, то это свидетельствует о неплохой видимости статьи в интернете. Однако, если автор статьи активно пользуется социальными сетями и участвует в форумах, посвященных обсуждению научных проблем, его видимость в интернете дополнительно возрастает. Появилось отдельное направление [19], связанное с оценкой популярности статьи в таких пока еще нетрадиционных для ученых системах как социальные сети, форумы, специализированные платформы для научных дискуссий. Это направление получило название — альтметрики[4]. Манифест альтметрик был опубликован в 2010 г. [20].

Результаты научных исследований являются многомерными. Они могут включить в себя достижения в конкретном научном направлении, в междисциплинарной методологии, в развитии технологий проведения исследований. Наконец, научные результаты могут иметь социальный эффект, привносящий новые идеи в широкие социальные слои — от исследователей до политических деятелей. Оценка такого научного результата должна быть также многомерной, здесь едва ли можно с успехом применить какую-либо одну метрику или одну модель оценки [21].

Большой объем материала, связанного с развитием научных идей и осмыслением полученных результатов, не попадает в традиционные журнальные статьи. Такие

[4] Альтметрики в научных публикациях — это альтернатива и дополнение к традиционным показателям библиометрии. Альтметрики отражают интерес интернет-аудитории к онлайновой статье. Альтметрики включают количество закладок, сделанных читателями в Mendeley, CiteULike и других системах, количество просмотров и скачиваний статьи, упоминаний в социальных сетях и т.д.

материалы остаются в записях открытых дискуссий, на персональных страницах ученых, на страницах институтских сайтов, в учебных материалах для студентов, в комментариях к статьям, в публикациях в средствах массовой информации [22]. С помощью существующих сервисов, таких как Mendeley, CiteULike или Zotero ученый может организовать свою личную библиотеку материалов, не имеющих формат научной статьи, и тем самым сделать эти материалы доступными для пользователей интернета [20].

В некоторых случаях ученый может написать небольшую заметку, которую в дальнейшем могли бы цитировать другие ученые. Однако подобную «нанопубликацию» традиционный журнал, скорее всего, не сможет опубликовать. Как следствие, читатели не смогут отыскать заметку с помощью поисковых средств, ориентированных на поиск журнальных статей. Здесь требуются другие поисковые запросы. Альметрия должна создавать механизмы учета такого рода «нанопубликаций».

Ученые широко обмениваются наборами данных, программными кодами, методиками проведения экспериментов, алгоритмами и т.д. («сырой наукой», raw science). Авторы, получившие такие данные, не набирают традиционных оценок авторитетности — их индекс Хирша не зависит от результатов, связанных с получением и размещением на сайтах или в архивах наборов данных. Требуются другие индикаторы, которые бы учитывали такой вклад ученого. Это направление также входит в зону интереса альтметрии.

Одной из характеристик статьи является рецензия. До недавнего времени материалы рецензий оставались на полках издательств. В настоящее время нередко реализуется идея размещения рецензий вместе со статьей в открытом доступе, что также укладывается в философию альтметрии. Кроме того, современная рецензия перестает быть одномоментным текстом, сопровождающим статью. Структура рецензионных материалов усложняется. Процесс рецензирования развивается в диалоге рецензента с автором, и содержание этого диалога представляет интерес для читателей статьи. Также возникают версии статьи с очередными изменениями, сделанными по итогам дискуссии. Уже после публикации статьи автор может получить новую рецензию от заинтересованных экспертов, или провести рецензирование в другом оверлейном журнале. Рецензирование, вообще говоря, не ограничено во времени. Каждое рецензирование дает повод для автора продолжать развивать свой опубликованный материал. Тем самым статья превращается в «живую публикацию» [23].

Технологии интернета и идеология Открытой науки позволяют реализовать рецензирование в стиле краудсорсинга[5], когда в оценке качества интернет-ресурса участвует представительное сообщество экспертов, а не только назначенные редакцией рецензенты. Такой подход к рецензированию реализован, например, в проекте F1000research [24], где оценку статье и ее дополнительным материалам может дать любой из нескольких тысяч экспертов данного проекта.

[5] Краудсорсинг (от анг. crowd — толпа) — привлечение широкого круга лиц для выполнения работы на добровольных началах. В издательской деятельности краудсорсинг стал обозначать новую схему рецензирования и обсуждения научной статьи, когда в оценке качества статьи участвует представительное сообщество экспертов, а не только назначенный редакцией рецензент. Обычно, например в проекте F1000Research, по результатам обсуждения с экспертами статья в любое время может быть обновлена и дополнена.

F1000research — издательская платформа Открытой науки для быстрого опубликования научных статей в области физических и биологических наук, инженерии, медицины, социальных и гуманитарных наук. Как организован процесс рецензирования на этой платформе? К публикации принимаются оригинальные статьи вне зависимости от предполагаемого уровня интереса или новизны. Все статьи публикуются в открытом доступе. Авторам предлагается дополнительно приложить к тексту статьи подробные описания методов, постеры, слайды. Также у автора есть возможность дать ссылку на исходные данные, лежащие в основе исследования, чтобы обеспечить воспроизводимость результатов.

Представленная в F1000Research статья сначала проходит быструю первоначальную проверку на соответствие общей редакционной практике и размещается на сайте со статусом «Ожидание экспертной оценки». Далее проводится открытое рецензирование, при этом авторы и рецензенты сотрудничают, чтобы сделать статью как можно более полной. Имена рецензентов и статус, который они присваивают статье после рецензирования, публикуются вместе со статьей. В дальнейшем любой другой эксперт издательства F1000Research также вправе по собственной инициативе прорецензировать статью, дополнительно уточняя ее статус. Как только статья получает два статуса «Одобрено», или два статуса «Одобрено с оговорками» и один статус «Одобрено», она будет проиндексирована в различных библиографических базах данных (в PubMed, PubMed Central, MEDLINE и др.). Если статья проиндексирована, все версии вместе с отчетами о рецензировании отправляются на хранение.

Имеются ли у экспертов стимулы брать на себя работу по рецензированию статей? Да, в проекте F1000Research такой механизм создан. Имя эксперта и его рецензии открыты для всего сообщества, тем самым активно и добросовестно работающий эксперт повышает свой рейтинг. Кроме того, для экспертов, участвующих в рецензировании статей, предусмотрена заметная скидка при оплате его личного взноса за будущие публикации [25]. Путем стимулирования активности экспертов и авторов проект F1000Research консолидирует научное сообщество и дает возможность каждому ученому участвовать в формировании коллективного научного продукта.

Статьи в F1000Research могут быть обновлены и дополнены в любое время после публикации, и каждая версия может быть независимо цитируемой со своим собственным DOI. Редакция предлагает следующий формат ссылки на статью [26]:

Author name(s). Article title [version number; details of peer review status]. F1000Research Year, Volume: Publication number (article doi)

Все компоненты приведенной библиографической ссылки понятны для читателя. Дополнительных разъяснений потребует атрибут «Статус». Статус характеризует количество проверок, которые «одобрены», «одобрены с оговорками» или «не одобрены». Кроме того, независимо от статьи, рецензия сама становится объектом цитирования. Рецензия публикуется под лицензией CC BY 4.0, каждой рецензии присваивается DOI. Платформа предлагает следующий формат ссылки на рецензию [26]:

Reviewer name(s). Peer review report for: Article title [version number; details of peer review status]. F1000Research Year, Volume: Publication number (review doi)

5 ЗАКЛЮЧЕНИЕ

По нашему мнению, оценка научной статьи должна быть многоплановой. Классическое рецензирование имеет риски не менее серьезные, чем неглубокое модерирование, за которым следует краудсорсинг, действующий в течение всего времени существования статьи. Краудсорсинг в научном издании можно определить простыми словами: сообщество ученых является мощным ресурсом, и подключение этого ресурса к рецензированию научных статей дает возможность получить более качественный научный продукт. На сайте статьи, помимо рецензии, читателю интересно было бы увидеть альтметрические показатели, полученные на основе информации из социальных сетей, тематических блогов и форумов для профессионального общения.

Оценка научной значимости статьи уже несколько десятилетий ориентируется только лишь на библиометрические показатели, основанные на анализе библиографических ссылок на статью. По нашему мнению, получение показателей на основе анализа библиографических ссылок в цитирующих статьях оказывается слишком долгим по времени, чувствительным к ошибкам записи ссылок и нерепрезентативным, поскольку обычно ограничено одной конкретной библиографической базой.

Перспективным направлением в научной издательской практике является оверлейный журнал, который рецензирует статьи из открытых архивов препринтов. После публикации статьи автор может получить новую рецензию от заинтересованных экспертов, или провести рецензирование в другом оверлейном журнале. Пост-рецензирование, вообще говоря, не ограничено во времени. Рецензии и отклики коллег вдохновляют автора продолжать развивать свою статью. Тем самым статья превращается в «живую публикацию».

Навязанная научному сообществу борьба за высокие показатели библиографического цитирования и за наращивание числа публикаций в высокорейтинговых журналах заслоняет существенно более значимые задачи: развитие инфраструктуры Открытого доступа, создание инструментов коммуникации для участников издательского процесса, обогащение средств презентации научных материалов на открытых издательских платформах.

REFERENCES

[1] M. M. Gorbunov-Posadov, T. A. Polilova, “Tools to Support Scientific Online Publishing”, *Programming and Computer Software*, **45 (**3), 116–120 (2019). https://link.springer.com/article/10.1134%2FS0361768819030046

[2] S. Beliaeva, “Tsena otkrytosti: Vo chto oboidetsia perekhod k Open Access?”, *Poisk*. (2019). https://www.poisknews.ru/skript/czena-otkrytosti-vo-chto-obojdetsya-perehod-k-open-access/

[3] J.R. Adler, T.M. Chan, J.B. Blain, B. Thoma, Atkinson, “OpenAccess: Free online, open-access crowdsource-reviewed publishing is the future; traditional peer-reviewed journals are on the way out”, *Canadian Journal of Emergency Medicine*, **21** (1), 11– 14 (2019). https://doi.org/10.1017/cem.2018.481

[4] Fair Open Access Alliance. https://www.fairopenaccess.org/ (Accessed July 22, 2020)

[5] Springer, Self-archiving policy. https://www.springer.com/gp/open-access/publication-policies/self-archiving-policy (Accessed July 22, 2020)

[6] Google Scholar, Inclusion Guidelines for Webmasters. https://scholar.google.com/intl/en-US/scholar/inclusion.html#overview (Accessed July 22, 2020)
[7] T.A. Polilova. "Ethical norms and legal framework of scientific publication". *Mathematica Montisnigri*, **XLV**, 129-136 (2019) http://www.montis.pmf.ac.me/vol45/11.pdf doi: 10.20948/mathmontis-2019-45-11
[8] T.A. Polilova, "Nauchnaia publikatsiia v Rossii: intellektualnye prava", *Preprinty IPM im. M.V. Keldysha*. 56, 1-24 (2019). http://keldysh.ru/papers/2019/prep2019_56.pdf doi:10.20948/prepr-2019-56
[9] E. Herman, J. Akeroyd, G. Bequet, D. Nicholas, A. Watkinson. "The changed – and changing – landscape of serials publishing: Review of the literature on emerging models", *Learned Publishing*. (2020). https://doi.org/10.1002/leap.1288 https://onlinelibrary.wiley.com/doi/full/10.1002/leap.1288
[10] Episciences.org, Overlay Journal Platform. www.episciences.org/?lang=en (Accessed July 22, 2020)
[11] T. Polilova, A. Ermakov, "Dissernet and self-plagiarism", *CEUR Workshop Proceedings*, **2543**, 285-294 (2020). https://www.scopus.com/record/display.uri?eid=2-s2.0-85078459740&origin=resultslist&sort=plf-f&src=s&st1=Dissernet+and+self-plagiarism&st2=&sid=d8727e9f55ec2b7fc177de73ce64e6d1&sot=b&sdt=b&sl=44&s=TITLE-ABS-KEY%28Dissernet+and+self-plagiarism%29&relpos=0&citeCnt=0&searchTerm= (Accessed July 22, 2020)
[12] Wikipedia, Elsevier. https://ru.wikipedia.org/wiki/Elsevier (Accessed July 22, 2020)
[13] Tom Reller. Statement From Michael Hansen, CEO Of Elsevier's Health Sciences Division, Regarding Australia Based Sponsored Journal Practices Between 2000 And 2005. https://www.elsevier.com/about/press-releases/clinical-solutions/statement-from-michael-hansen,-ceo-of-elseviers-health-sciences-division,-regarding-australia-based-sponsored-journal-practices-between-2000-and-2005 (Accessed July 22, 2020)
[14] "Vsia pravda o lekarstvakh, Google i bestsellerah, Galina Iuzefovich — pro knigi, kotorye vse obieiasniaiut". https://meduza.io/feature/2015/07/04/vsya-pravda-o-lekarstvah-google-i-bestsellerah (Accessed July 22, 2020)
[15] V.A. Traag, L. Waltman, "Systematic analysis of agreement between metrics and peer review in the UK REF", *Palgrave Communications*, **5**, Article number: 29 (2019). https://doi.org/10.1057/s41599-019-0233-x
[16] D.E. Chebukov, "Poisk poteriannykh tsitirovanii v Web of Science. Ispravlenie oshibok v spiskakh literatury Web of Science", *Nauchnyi servis v seti Internet*, 461-467 (2017). http://keldysh.ru/abrau/2017/77.pdf doi:10.20948/abrau-2017-77
[17] YongGao QiangWu LinnaZhu, "Merging the citations received by arXiv-deposited e-prints and their corresponding published journal articles: Problems and perspectives", *Information Processing & Management*, **57** (5), (2020). https://doi.org/10.1016/j.ipm.2020.102267
[18] Iu.G. Reviakin, "Web-analitika dlia nauchnykh publikatsii", *Preprinty IPM im. M.V. Keldysha*, 50, 1-42 (2020). http://keldysh.ru/papers/2020/prep2020_50.pdf doi:10.20948/prepr-2020-50
[19] M.N. Saushkin, D.E. Chebukov, "Altmetriki na saite nauchnogo zhurnala", *Nauchnyi servis v seti Internet*, 593-599 (2019). http://keldysh.ru/abrau/2019/theses/40.pdf doi:10.20948/abrau-2019-40
[20] J. Priem, D. Taraborelli, P. Groth, C. Neylon, "Altmetrics: A manifesto", (2010). http://altmetrics.org/manifesto (Accessed July 22, 2020)
[21] M. V. Vakhrushev, "Altmetriki, vebometriki i informetriki kak vzaimodopolniaiushchie napravleniia v sovremennoi bibliometrii", *Nauchnye i tekhnicheskie biblioteki*, 8, 67-76 (2019). https://ntb.gpntb.ru/jour/article/viewFile/470/453
[22] A. Grossmann, "Publishing in transition – do we still need scientific journals?", *ScienceOpen*

Research, (2015). https://www.scienceopen.com/document_file/e1dd3665-6406-4a32-befc-e00d84a72cd1/ScienceOpen/3077_XE696973259861784096.pdf doi: 10.14293/S2199-1006.1.SOR-SOCSCI.ACKE0Y.v1

[23] M.M. Gorbunov-Posadov, "Zhivaia publikatsiia", (Moscow: KIAM). https://keldysh.ru/gorbunov/live.htm (Accessed July 22, 2020)

[24] F1000Research. Open for Science. https://f1000research.com/ (Accessed July 22, 2020)

[25] F1000Research. Open for Science: Referee Incentives. http://f1000research.com/referee-incentives (Accessed July 22, 2020)

[26] F1000Research. Open for Science: How it Works. https://f1000research.com/about (Accessed July 22, 2020)

Received June 10, 2020