

CONTENTS

Mathematics

- N. Mihaljevic. The existence and the uniqueness of solutions of integral equations for the capacity q and function of late α 5
- S. Berberyana. On some boundary properties and uniqueness theorems of subharmonic functions..... 16

Mathematical physics

- A.V. Kolesnichenko. To development of the non-additive thermodynamics of the quantum systems on basis statistics of Tsallis..... 26

Mathematical modeling

- K. K. Maevskii. Thermodynamic parameters of mixtures with silicon nitride under shock-wave loading..... 52
- S.V. Ershov, D.D. Zhdanov, A.G. Voloboy. Estimation of noise in calculation of scattering medium luminance by MCRT..... 60
- M. S. Apfelbaum and A. N. Doludenko. Hydrodynamic characteristics of weakly conductive liquid media in the non-uniform electric field 74
- O.N. Koroleva, M.M. Demin, V.I. Mazhukin, A.V. Mazhukin. Modeling of thermal conductivity of Si in the range from the normal to near-critical conditions..... 85

Computer science applications

- A.K. Alekseev, A.E. Bondarev, A.E. Kuvshinnikov. Comparative analysis of the accuracy of OPENFOAM solvers for the oblique shock wave problem..... 95
- V. Frolov, D. Pavlov, M. Trofimov, P. Kazbeev, V. Galaktionov. Open rendering benchmark..... 106

Modeling in economics

- L.Obradović. A note on the perpetual American straddle 123

Academic life

- T.A. Polilova. Ethical norms and legal framework of scientific publication..... 129
- G.K. Borovin, A.V. Grushevskii, A.G. Tuchin, D.A. Tuchin. Russian exploration of Venus: past and prospects..... 137

**СУЩЕСТВОВАНИЕ И УНИКАЛЬНОСТЬ РЕШЕНИЯ
ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ ДЛЯ ПОТЕНЦИАЛА q И ФУНКЦИИ
ЗАПАЗДЫВАНИЯ α**

НИКОЛА МИХАЛЕВИЧ*

Университет Черногории, Морской факультет, Котор, Черногория

*Ответственный автор. E-mail: nikolamih@ac.me

DOI: 10.20948/mathmontis-2019-45-1

Ключевые слова: Оператор, потенциал, функция запаздывания, сжатие

Аннотация. В статье рассматриваются операторы $A_{1\alpha}$ и $A_{2\alpha}$, обеспечивающие сжатие интегральных уравнений для потенциала q и функции запаздывания α . В статье доказывается, что для каждого из этих операторов существует шар, для которого этот оператор является оператором сжатия. Также в статье доказывается, что каждый из операторов отображает шар на себя.

**THE EXISTENCE AND THE UNIQUENESS OF SOLUTIONS OF
INTEGRAL EQUATIONS FOR THE CAPACITY q AND FUNCTION OF
LATE α**

NIKOLA MIHALJEVIC*

Maritime Faculty, University of Montenegro, Kotor, Montenegro

*Corresponding author. E-mail: nikolamih@ac.me

DOI: 10.20948/mathmontis-2019-45-1

Summary. The operators $A_{1\alpha}$ and $A_{2\alpha}$, which provide compression of the integral equations for the potential q and the delay function α , are considered in this paper. In this article proves that for each of these operators there is a sphere for which this operator is a compression operator. Also in this article, it is proved that each of the operators maps the sphere onto itself.

2010 Mathematics Subject Classification: 34B24, 35B05.

Key words and Phrases: Operator, Potential, Function of delay, Compression.

1 ВВЕДЕНИЕ

Мы оцениваем следующие суммы

$$\tilde{S}_{1,l} = \sum_{m=1}^{\infty} \frac{1}{m^{l-1}} \prod_{i=1}^{l-1} \sin m(\alpha(\gamma_1^{-1}(\bar{t}_i)) - \gamma_1^{-1}(\bar{t}_{i+1})) \cos m(\bar{x} - \gamma_1^{-1}(\bar{t}_1) + \alpha(\gamma_1^{-1}(\bar{t}_l))),$$

$$\bar{x} \in [0, \pi], \quad (l = 2, 3, \dots)$$

Для $l = 2$ имеем

$$\tilde{S}_{1,2} = \sum_{m=1}^{\infty} \frac{\sin m(\alpha(\gamma_1^{-1}(\bar{t}_1)) - \gamma_1^{-1}(\bar{t}_2)) \cos m(\bar{x} - \gamma_1^{-1}(\bar{t}_1) + \alpha(\gamma_1^{-1}(\bar{t}_2)))}{m}$$

$$(0 \leq \bar{x} \leq \pi, \quad 0 \leq \bar{t}_1 \leq \pi, \quad 0 \leq \bar{t}_2 \leq \alpha(\bar{t}_1))$$

В результате преобразования правой части получаем

$$\begin{aligned} & \sin m(\alpha(\gamma_1^{-1}(\bar{t}_1)) - \gamma_1^{-1}(\bar{t}_2)) \cos m(\bar{x} - \gamma_1^{-1}(\bar{t}_1) + \alpha(\gamma_1^{-1}(\bar{t}_2))) = \\ & = \frac{1}{2} [\sin m(\alpha(\gamma_1^{-1}(\bar{t}_1)) - \gamma_1^{-1}(\bar{t}_2) + \bar{x} - \gamma_1^{-1}(\bar{t}_1) + \alpha(\gamma_1^{-1}(\bar{t}_2))) + \\ & + \sin m(\alpha(\gamma_1^{-1}(\bar{t}_1)) - \gamma_1^{-1}(\bar{t}_2) - \bar{x} + \gamma_1^{-1}(\bar{t}_1) - \alpha(\gamma_1^{-1}(\bar{t}_2)))]. \end{aligned}$$

Поэтому $\tilde{S}_{1,2}$ разделяется на сумму двух рядов вида $\sum_{n=1}^{\infty} \frac{\sin nx}{n}$. Основываясь на развитии

Фурье $\frac{\pi - x}{2} = \sum_{n=1}^{\infty} \frac{\sin nx}{n}$, ($0 < x < 2\pi$), мы заключаем что $|\tilde{S}_{1,2}| \leq \frac{\pi}{2}$.

Для $l \geq 3$ имеем

$$|\tilde{S}_{1,l}| \leq \sum_{m=1}^{\infty} \frac{1}{m^{l-1}} \leq \sum_{m=1}^{\infty} \frac{1}{m^2} = \frac{\pi^2}{6}.$$

Поэтому оценка действительна

$$|\tilde{S}_{1,l}| \leq \frac{\pi^2}{6}, \quad (l = 2, 3, \dots).$$

Аналогичным образом можно показать, что оценка действительна

$$|\tilde{S}_{2,l}| \leq \frac{\pi^2}{6}, \quad (l = 2, 3, \dots),$$

где

$$\tilde{S}_{2,l} = \sum_{m=1}^{\infty} \frac{1}{m^{l-1}} \prod_{i=1}^{l-1} \sin m(\alpha(\gamma_2^{-1}(\bar{t}_i)) - \gamma_2^{-1}(\bar{t}_{i+1})) \cos m(\bar{x} - \gamma_2^{-1}(\bar{t}_1) - \alpha(\gamma_2^{-1}(\bar{t}_l))),$$

$$\bar{x} \in [0, \pi], \quad (l = 2, 3, \dots).$$

2 СЖИМАЕМОСТЬ

Рассматривающиеся далее операторы $A_{1\alpha}$ и $A_{2\alpha}$, обеспечивающие сжатие интегральных уравнений для потенциала q и функции запаздывания α были введены в работе [1].

Лема 1. Для произвольных функций $f_1(t)$ и $f_2(t)$ имеем

$$\prod_{i=1}^l f_1(t_i) - \prod_{i=1}^l f_2(t_i) = \sum_{i=1}^l (f_1(t_i) - f_2(t_i)) \prod_{j=i+1}^l f_1(t_j) \prod_{k=1}^{i-1} f_2(t_k) \quad (1)$$

(если в произведении Π , нижний индекс больше верхнего, мы берем $\Pi = 1$).

Доказательство. См [2].

Теорема 1. Оператор $A_{1\alpha}$ является оператором сжатия на шаре

$$\tilde{K}_{1\alpha} = \left\{ \tilde{q}_1 : \|\tilde{q}_1\| \leq \max \left\{ 1, \frac{1}{\alpha(\pi)} \right\} \ln \left(1 + \frac{6}{\pi^2 + 3} \right) \right\}.$$

Доказательство. Пусть $\tilde{q}_1^{(1)}, \tilde{q}_1^{(2)} \in \tilde{K}_{1\alpha}$. Тогда

$$\|A_{1\alpha}\tilde{q}_1^{(1)} - A_{1\alpha}\tilde{q}_1^{(2)}\| = \int_0^\pi |A_{1\alpha}\tilde{q}_1^{(1)}(\bar{x}) - A_{1\alpha}\tilde{q}_1^{(2)}(\bar{x})| d\bar{x}.$$

Далее

$$\begin{aligned} & \left| A_{1\alpha}\tilde{q}_1^{(1)}(\bar{x}) - A_{1\alpha}\tilde{q}_1^{(2)}(\bar{x}) \right| = \\ & = \left| \frac{1}{2\pi} \sum_{l=2}^{\infty} \int_{\bar{D}_l^{(1)}} [\tilde{Q}_1^{(1)}(\bar{T}_l) - \tilde{Q}_1^{(2)}(\bar{T}_l)] \prod_{i=1}^{l-1} (\alpha(\gamma_1^{-1}(\bar{t}_i)) - \gamma_1^{-1}(\bar{t}_{i+1})) d\bar{T}_l - \right. \\ & \left. - \frac{1}{\pi} \sum_{m=1}^{\infty} \sum_{l=3}^{\infty} \frac{1}{m^{l-1}} \int_{\bar{D}_l^{(1)}} [\tilde{Q}_1^{(1)}(\bar{T}_l) - \tilde{Q}_1^{(2)}(\bar{T}_l)] \bar{P}_1(\bar{T}_l, m) \cos m(\bar{x} - \gamma_1^{-1}(\bar{t}_1) + \alpha(\gamma_1^{-1}(\bar{t}_2))) d\bar{T}_l \right| \leq \\ & \leq \frac{1}{2\pi} \sum_{l=2}^{\infty} \left| \int_{\bar{D}_l^{(1)}} [\tilde{Q}_1^{(1)}(\bar{T}_l) - \tilde{Q}_1^{(2)}(\bar{T}_l)] \prod_{i=1}^{l-1} (\alpha(\gamma_1^{-1}(\bar{t}_i)) - \gamma_1^{-1}(\bar{t}_{i+1})) d\bar{T}_l \right| + \\ & + \frac{1}{\pi} \sum_{m=1}^{\infty} \sum_{l=2}^{\infty} \frac{1}{m^{l-1}} \left| \int_{\bar{D}_l^{(1)}} [\tilde{Q}_1^{(1)}(\bar{T}_l) - \tilde{Q}_1^{(2)}(\bar{T}_l)] \bar{P}_1(\bar{T}_l, m) \cos m(\bar{x} - \gamma_1^{-1}(\bar{t}_1) + \alpha(\gamma_1^{-1}(\bar{t}_2))) d\bar{T}_l \right| = \\ & = \Sigma_1 + \Sigma_2. \end{aligned}$$

Мы оцениваем обе суммы. Так как

$$\alpha(\gamma_1^{-1}(\bar{t}_i)) - \gamma_1^{-1}(\bar{t}_{i+1}) < \alpha(\gamma_1^{-1}(\bar{t}_i)) < \alpha(\pi),$$

то для первой суммы Σ_1 имеем

$$\begin{aligned}\Sigma_1 &\leq \frac{1}{2\pi} \sum_{l=2}^{\infty} \int_{\bar{D}_l^{(1)}} \left| \left[\tilde{Q}_1^{(1)}(\bar{T}_l) - \tilde{Q}_1^{(2)}(\bar{T}_l) \right] \prod_{i=1}^{l-1} \left(\alpha(\gamma_1^{-1}(\bar{t}_i)) - \gamma_1^{-1}(\bar{t}_{i+1}) \right) \right| d\bar{T}_l \leq \\ &\leq \frac{1}{2\pi} \sum_{l=2}^{\infty} \alpha(\pi)^{l-1} \int_{\bar{D}_l^{(1)}} \left| \tilde{Q}_1^{(1)}(\bar{T}_l) - \tilde{Q}_1^{(2)}(\bar{T}_l) \right| d\bar{T}_l = \\ &= \frac{1}{2\pi} \sum_{l=2}^{\infty} \alpha(\pi)^{l-1} \int_{\bar{D}_l^{(1)}} \left| \prod_{i=1}^l \tilde{q}_1^{(1)}(\bar{t}_i) - \prod_{i=1}^l \tilde{q}_1^{(2)}(\bar{t}_i) \right| d\bar{T}_l.\end{aligned}$$

Для суммы Σ_2 имеем

$$\begin{aligned}\Sigma_2 &= \sum_{l=2}^{\infty} \left| \int_{\bar{D}_l^{(1)}} \left[\tilde{Q}_1^{(1)}(\bar{T}_l) - \tilde{Q}_1^{(2)}(\bar{T}_l) \right] \sum_{m=1}^{\infty} \frac{1}{l-1} \bar{P}_1(\bar{T}_l, m) \cos m \left(\bar{x} - \gamma_1^{-1}(\bar{t}_1) + \alpha(\gamma_1^{-1}(\bar{t}_2)) \right) \right| d\bar{T}_l \leq \\ &\leq \frac{\pi}{6} \sum_{l=2}^{\infty} \int_{\bar{D}_l^{(1)}} \left| \tilde{Q}_1^{(1)}(\bar{T}_l) - \tilde{Q}_1^{(2)}(\bar{T}_l) \right| d\bar{T}_l = \frac{\pi}{6} \sum_{l=2}^{\infty} \int_{\bar{D}_l^{(1)}} \left| \prod_{i=1}^l \tilde{q}_1^{(1)}(\bar{t}_i) - \prod_{i=1}^l \tilde{q}_1^{(2)}(\bar{t}_i) \right| d\bar{T}_l.\end{aligned}$$

Так, как $\alpha(\gamma_1^{-1}(\bar{t}_i)) < \gamma_1^{-1}(\bar{t}_i)$ это $\gamma_1(\alpha(\gamma_1^{-1}(\bar{t}_i))) < \bar{t}_i$, то отсюда и из формулы (1) имеем

$$\begin{aligned}&\int_{\bar{D}_l^{(1)}} \left| \prod_{i=1}^l \tilde{q}_1^{(1)}(\bar{t}_i) - \prod_{i=1}^l \tilde{q}_1^{(2)}(\bar{t}_i) \right| d\bar{T}_l = \\ &= \int_0^{\gamma_1(\pi)} \int_0^{\gamma_1(\alpha(\gamma_1^{-1}(\bar{t}_1)))} \dots \int_0^{\gamma_1(\alpha(\gamma_1^{-1}(\bar{t}_{l-1})))} \left| \sum_{i=1}^l \left(\tilde{q}_1^{(1)}(\bar{t}_i) - \tilde{q}_1^{(2)}(\bar{t}_i) \right) \prod_{j=i+1}^l \tilde{q}_1^{(1)}(\bar{t}_j) \prod_{k=1}^{i-1} \tilde{q}_1^{(2)}(\bar{t}_k) \right| d\bar{t}_1 \dots d\bar{t}_l \leq \\ &\leq \sum_{i=1}^l \int_0^{\gamma_1(\pi)} \int_0^{\gamma_1(\alpha(\gamma_1^{-1}(\bar{t}_i)))} \dots \int_0^{\gamma_1(\alpha(\gamma_1^{-1}(\bar{t}_{i-1})))} \left| \tilde{q}_1^{(1)}(\bar{t}_i) - \tilde{q}_1^{(2)}(\bar{t}_i) \right| \prod_{j=i+1}^l \left| \tilde{q}_1^{(1)}(\bar{t}_j) \right| \prod_{k=1}^{i-1} \left| \tilde{q}_1^{(2)}(\bar{t}_k) \right| d\bar{t}_1 \dots d\bar{t}_l \leq \\ &\leq \sum_{i=1}^l \int_0^{\pi} \int_0^{\bar{t}_i} \dots \int_0^{\bar{t}_{i-1}} \left| \tilde{q}_1^{(1)}(\bar{t}_i) - \tilde{q}_1^{(2)}(\bar{t}_i) \right| \prod_{j=i+1}^l \left| \tilde{q}_1^{(1)}(\bar{t}_j) \right| \prod_{k=1}^{i-1} \left| \tilde{q}_1^{(2)}(\bar{t}_k) \right| d\bar{t}_1 \dots d\bar{t}_l = \\ &= \int_0^{\pi} \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-1}} \left| \tilde{q}_1^{(1)}(\bar{t}_1) - \tilde{q}_1^{(2)}(\bar{t}_1) \right| \prod_{j=2}^l \left| \tilde{q}_1^{(1)}(\bar{t}_j) \right| d\bar{t}_1 \dots d\bar{t}_l + \\ &+ \sum_{i=2}^{l-1} \int_0^{\pi} \int_0^{\bar{t}_i} \dots \int_0^{\bar{t}_{i-1}} \left| \tilde{q}_1^{(1)}(\bar{t}_i) - \tilde{q}_1^{(2)}(\bar{t}_i) \right| \prod_{j=i+1}^l \left| \tilde{q}_1^{(1)}(\bar{t}_j) \right| \prod_{k=1}^{i-1} \left| \tilde{q}_1^{(2)}(\bar{t}_k) \right| d\bar{t}_1 \dots d\bar{t}_l + \\ &+ \int_0^{\pi} \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-1}} \prod_{j=1}^{l-1} \left| \tilde{q}_1^{(2)}(\bar{t}_j) \right| \left| \tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l) \right| d\bar{t}_1 \dots d\bar{t}_l.\end{aligned}$$

Для первого слагаемого имеем

$$\begin{aligned} & \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_1) - \tilde{q}_1^{(2)}(\bar{t}_1)| \prod_{j=2}^l |\tilde{q}_1^{(1)}(\bar{t}_j)| d\bar{t}_l \dots d\bar{t}_1 = \\ & = \int_0^\pi |\tilde{q}_1^{(1)}(\bar{t}_1) - \tilde{q}_1^{(2)}(\bar{t}_1)| \left(\int_0^{\bar{t}_1} |\tilde{q}_1^{(1)}(\bar{t}_2)| f(\bar{t}_2) d\bar{t}_2 \right) d\bar{t}_1 \end{aligned}$$

где

$$f(\bar{t}_2) = \int_0^{\bar{t}_2} |\tilde{q}_1^{(1)}(\bar{t}_3)| \left(\int_0^{\bar{t}_3} |\tilde{q}_1^{(1)}(\bar{t}_4)| \left(\dots \int_0^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_l)| d\bar{t}_l \dots \right) d\bar{t}_4 \right) d\bar{t}_3.$$

Изменив порядок интеграции, а затем роль переменных \bar{t}_1 и \bar{t}_2 далее получаем что

$$\begin{aligned} & \int_0^\pi |\tilde{q}_1^{(1)}(\bar{t}_1) - \tilde{q}_1^{(2)}(\bar{t}_1)| \left(\int_0^{\bar{t}_1} |\tilde{q}_1^{(1)}(\bar{t}_2)| f(\bar{t}_2) d\bar{t}_2 \right) d\bar{t}_1 = \\ & = \int_0^\pi |\tilde{q}_1^{(1)}(\bar{t}_1)| f(\bar{t}_1) \left(\int_{\bar{t}_1}^\pi |\tilde{q}_1^{(1)}(\bar{t}_2) - \tilde{q}_1^{(2)}(\bar{t}_2)| d\bar{t}_2 \right) d\bar{t}_1. \end{aligned}$$

Однако

$$\begin{aligned} f(\bar{t}_1) &= \int_0^{\bar{t}_1} |\tilde{q}_1^{(1)}(\bar{t}_3)| \left(\int_0^{\bar{t}_3} |\tilde{q}_1^{(1)}(\bar{t}_4)| \left(\dots \int_0^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_l)| d\bar{t}_l \dots \right) d\bar{t}_4 \right) d\bar{t}_3 = \\ & = \int_0^{\bar{t}_1} |\tilde{q}_1^{(1)}(\bar{t}_2)| \left(\int_0^{\bar{t}_2} |\tilde{q}_1^{(1)}(\bar{t}_3)| \left(\dots \int_0^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_{l-1})| d\bar{t}_{l-1} \dots \right) d\bar{t}_3 \right) d\bar{t}_2 \end{aligned}$$

и

$$\int_{\bar{t}_1}^\pi |\tilde{q}_1^{(1)}(\bar{t}_2) - \tilde{q}_1^{(2)}(\bar{t}_2)| d\bar{t}_2 = \int_{\bar{t}_1}^\pi |\tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l)| d\bar{t}_l$$

это

$$\begin{aligned} & \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_1) - \tilde{q}_1^{(2)}(\bar{t}_1)| \prod_{j=2}^l |\tilde{q}_1^{(1)}(\bar{t}_j)| d\bar{t}_l \dots d\bar{t}_1 = \\ & = \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-2}} |\tilde{q}_1^{(1)}(\bar{t}_1)| \left(\int_{\bar{t}_1}^\pi |\tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l)| d\bar{t}_l \right) d\bar{t}_{l-1} \dots d\bar{t}_1. \end{aligned}$$

Таким образом, мы можем доказать, что

$$\begin{aligned}
 & \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_i) - \tilde{q}_1^{(2)}(\bar{t}_i)| \prod_{j=i+1}^l |\tilde{q}_1^{(1)}(\bar{t}_j)| \prod_{k=1}^{i-1} |\tilde{q}_1^{(2)}(\bar{t}_k)| d\bar{t}_l \dots d\bar{t}_1 = \\
 & = \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-2}} \prod_{j=i}^{l-1} |\tilde{q}_1^{(1)}(\bar{t}_j)| \prod_{k=1}^{i-1} |\tilde{q}_1^{(2)}(\bar{t}_k)| \left(\int_{\bar{t}_i}^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l)| d\bar{t}_l \right) d\bar{t}_{l-1} \dots d\bar{t}_1
 \end{aligned}$$

для $i = 2, 3, \dots, l-1$.

Мы определяем функцию

$$\tilde{q}_1(\bar{t}) = \max \left\{ |\tilde{q}_1^{(1)}(\bar{t})|, |\tilde{q}_1^{(2)}(\bar{t})| \right\}, \quad \bar{t} \in [0, \pi].$$

Тогда

$$\begin{aligned}
 & \int_{\bar{D}_1^{(1)}} \left| \prod_{i=1}^l \tilde{q}_1^{(1)}(\bar{t}_i) - \prod_{i=1}^l \tilde{q}_1^{(2)}(\bar{t}_i) \right| d\bar{T}_l \leq \\
 & \leq \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-2}} \prod_{j=1}^{l-1} \tilde{q}_1(\bar{t}_j) \left(\int_{\bar{t}_1}^\pi |\tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l)| d\bar{t}_l \right) d\bar{t}_{l-1} \dots d\bar{t}_1 + \\
 & + \sum_{i=2}^{l-1} \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-1}} \prod_{j=1}^{l-1} \tilde{q}_1(\bar{t}_j) \left(\int_{\bar{t}_i}^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l)| d\bar{t}_l \right) d\bar{t}_{l-1} \dots d\bar{t}_1 + \\
 & + \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-2}} \prod_{j=1}^{l-1} \tilde{q}_1(\bar{t}_j) \left(\int_0^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l)| d\bar{t}_l \right) d\bar{t}_{l-1} \dots d\bar{t}_1 = \\
 & = \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-2}} \prod_{j=1}^{l-1} \tilde{q}_1(\bar{t}_j) \left(\int_{\bar{t}_1}^\pi |\tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l)| d\bar{t}_l + \right. \\
 & \quad \left. + \sum_{i=2}^{l-1} \int_{\bar{t}_i}^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l)| d\bar{t}_l + \right. \\
 & \quad \left. + \int_0^{\bar{t}_{l-1}} |\tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l)| d\bar{t}_l \right) d\bar{t}_{l-1} \dots d\bar{t}_1 = \\
 & = \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-2}} \prod_{j=1}^{l-1} \tilde{q}_1(\bar{t}_j) d\bar{t}_{l-1} \dots d\bar{t}_1 \cdot \int_0^\pi |\tilde{q}_1^{(1)}(\bar{t}_l) - \tilde{q}_1^{(2)}(\bar{t}_l)| d\bar{t}_l = \\
 & = \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-2}} \prod_{j=1}^{l-1} \tilde{q}_1(\bar{t}_j) d\bar{t}_{l-1} \dots d\bar{t}_1 \cdot \|\tilde{q}_1^{(1)} - \tilde{q}_1^{(2)}\|.
 \end{aligned}$$

Следующий абзац известен из интегрального учета функций нескольких переменных. Если функция $f(x_1, x_2, \dots, x_n)$ соответствует условию

$$f(x_1, x_2, \dots, x_n) = f(x_{i_1}, x_{i_2}, \dots, x_{i_n}),$$

где (i_1, i_2, \dots, i_n) произвольная перестановка множества $\{1, 2, \dots, n\}$. Тогда

$$n! \int_0^\pi \int_0^{x_1} \dots \int_0^{x_{n-1}} f(x_1, x_2, \dots, x_n) dx_n \dots dx_2 dx_1 = \int_0^\pi \int_0^\pi \dots \int_0^\pi f(x_1, x_2, \dots, x_n) dx_n \dots dx_2 dx_1.$$

Функция $\tilde{Q}_{1, l-1} = \prod_{j=1}^{l-1} \tilde{q}_1(\bar{t}_j)$, очевидно, отвечает этому условию, поэтому

$$\begin{aligned} \int_0^\pi \int_0^{\bar{t}_1} \dots \int_0^{\bar{t}_{l-2}} \prod_{j=1}^{l-1} \tilde{q}_1(\bar{t}_j) d\bar{t}_{l-1} \dots d\bar{t}_1 &= \frac{1}{(l-1)!} \int_0^\pi \int_0^\pi \dots \int_0^\pi \prod_{j=1}^{l-1} \tilde{q}_1(\bar{t}_j) d\bar{t}_{l-1} \dots d\bar{t}_1 = \\ &= \frac{1}{(l-1)!} \prod_{j=1}^{l-1} \int_0^\pi \tilde{q}_1(\bar{t}_j) d\bar{t}_j \leq \frac{1}{(l-1)!} \prod_{j=1}^{l-1} \|\tilde{q}_1\| = \frac{\|\tilde{q}_1\|^{l-1}}{(l-1)!}. \end{aligned}$$

Отсюда следует

$$\begin{aligned} \sum_{l=2}^\infty \alpha(\pi)^{l-1} \int_{\bar{D}^{(l)}} \left| \prod_{i=1}^l \tilde{q}_1^{(1)}(\bar{t}_i) - \prod_{i=1}^l \tilde{q}_1^{(2)}(\bar{t}_i) \right| d\bar{T}_l &\leq \\ \leq \|\tilde{q}_1^{(1)} - \tilde{q}_1^{(2)}\| \cdot \sum_{l=2}^\infty \frac{(\|\tilde{q}_1\| \alpha(\pi))^{l-1}}{(l-1)!} &= (e^{\|\tilde{q}_1\| \alpha(\pi)} - 1) \|\tilde{q}_1^{(1)} - \tilde{q}_1^{(2)}\| \end{aligned}$$

и

$$\begin{aligned} \sum_{l=2}^\infty \int_{\bar{D}^{(l)}} \left| \prod_{i=1}^l \tilde{q}_1^{(1)}(\bar{t}_i) - \prod_{i=1}^l \tilde{q}_1^{(2)}(\bar{t}_i) \right| d\bar{T}_l &\leq \\ \leq \|\tilde{q}_1^{(1)} - \tilde{q}_1^{(2)}\| \cdot \sum_{l=2}^\infty \frac{\|\tilde{q}_1\|^{l-1}}{(l-1)!} &= (e^{\|\tilde{q}_1\|} - 1) \|\tilde{q}_1^{(1)} - \tilde{q}_1^{(2)}\| \end{aligned}$$

следовательно

$$\Sigma_1 \leq \frac{1}{2\pi} (e^{\|\tilde{q}_1\| \alpha(\pi)} - 1) \|\tilde{q}_1^{(1)} - \tilde{q}_1^{(2)}\| \quad \text{и} \quad \Sigma_1 \leq \frac{\pi}{6} (e^{\|\tilde{q}_1\|} - 1) \|\tilde{q}_1^{(1)} - \tilde{q}_1^{(2)}\|.$$

Далее

$$\left| A_{1\alpha} \tilde{q}_1^{(1)}(x) - A_{1\alpha} \tilde{q}_1^{(2)}(x) \right| \leq \left[\frac{1}{2\pi} (e^{\|\tilde{q}_1\| \alpha(\pi)} - 1) + \frac{\pi}{6} (e^{\|\tilde{q}_1\|} - 1) \right] \|\tilde{q}_1^{(1)} - \tilde{q}_1^{(2)}\|,$$

следовательно

$$\|A_{1\alpha} \tilde{q}_1^{(1)} - A_{1\alpha} \tilde{q}_1^{(2)}\| \leq \left[\frac{1}{2\pi} (e^{\|\tilde{q}_1\| \alpha(\pi)} - 1) + \frac{\pi}{6} (e^{\|\tilde{q}_1\|} - 1) \right] \|\tilde{q}_1^{(1)} - \tilde{q}_1^{(2)}\| \int_0^\pi dx =$$

$$\left[\frac{1}{\pi} \left(e^{\|\tilde{q}_1\|^{\alpha(\pi)}} - 1 \right) + \frac{\pi^2}{6} \left(e^{\|\tilde{q}_1\|} - 1 \right) \right] \|\tilde{q}_1^{(1)} - \tilde{q}_1^{(2)}\|.$$

Оператор $A_{1\alpha}$ является оператором сжатия если

$$\frac{1}{\pi} \left(e^{\|\tilde{q}_1\|^{\alpha(\pi)}} - 1 \right) + \frac{\pi^2}{6} \left(e^{\|\tilde{q}_1\|} - 1 \right) < 1.$$

Здесь мы можем выделить два случая. Первый, когда $\alpha(\pi) \leq 1$ и другой, когда $\alpha(\pi) > 1$. В первом случае мы получаем

$$\|\tilde{q}_1\| < \ln \left(1 + \frac{6}{\pi^2 + 3} \right),$$

а во втором

$$\|\tilde{q}_1\| < \frac{1}{\alpha(\pi)} \ln \left(1 + \frac{6}{\pi^2 + 3} \right).$$

Таким образом, мы имеем

$$\|\tilde{q}_1\| \leq \max \left\{ 1, \frac{1}{\alpha(\pi)} \right\} \ln \left(1 + \frac{6}{\pi^2 + 3} \right) = \rho.$$

Теорема доказана. □

Теорема 2. Оператор $A_{2\alpha}$ является оператором сжатия на шаре

$$\tilde{K}_{2\alpha} = \left\{ \tilde{q}_2 : \|\tilde{q}_2\| \leq \max \left\{ 1, \frac{1}{\alpha(\pi)} \right\} \ln \left(1 + \frac{6}{\pi^2 + 3} \right) \right\}.$$

Доказательство. Идентично доказательству теоремы 1, поэтому мы не будем его повторять. □

3 ОТОБРАЖЕНИЕ “НА”

Теорема 3. Оператор $A_{1\alpha}$ отображает шар $\tilde{K}_{1\alpha}$ на себя.

Доказательство. Из выражения для оператора $A_{1\alpha}$ мы заключаем, что

$$|A_{1\alpha} \tilde{q}_1(x)| \leq \frac{1}{2\pi\alpha(\pi)} \left(e^{\alpha(\pi)\|\tilde{q}_1\|} - 1 - \alpha(\pi)\|\tilde{q}_1\| \right) + \frac{\pi}{6} \left(e^{\|\tilde{q}_1\|} - 1 - \|\tilde{q}_1\| \right),$$

соответственно

$$\|A_{1\alpha} \tilde{q}_1(x)\| \leq \frac{1}{2\alpha(\pi)} \left(e^{\alpha(\pi)\|\tilde{q}_1\|} - 1 - \alpha(\pi)\|\tilde{q}_1\| \right) + \frac{\pi^2}{6} \left(e^{\|\tilde{q}_1\|} - 1 - \|\tilde{q}_1\| \right).$$

Предположим, что $\|A_{1\alpha}\tilde{q}_1(x)\| \leq \|\tilde{q}_1\|$. Таким образом, получаем для неравенства

$$\frac{1}{2\alpha(\pi)} \left(e^{\alpha(\pi)\|\tilde{q}_1\|} - 1 - \alpha(\pi)\|\tilde{q}_1\| \right) + \frac{\pi^2}{6} \left(e^{\|\tilde{q}_1\|} - 1 - \|\tilde{q}_1\| \right) \leq \|\tilde{q}_1\|. \quad (2)$$

В дальнейшем мы будем рассматривать случаи $\alpha(\pi) \in (0,1]$ и $\alpha(\pi) \in (1,\pi)$.

Если $\alpha(\pi) \leq 1$, то верхнее неравенство эквивалентно неравенству

$$e^{\|\tilde{q}_1\|} - 1 \leq \left(1 + \frac{6}{3 + \pi^2\alpha(\pi)} \right) \|\tilde{q}_1\|.$$

Вводим функции φ и ψ следующим образом: $\varphi(t) = e^t - 1$ и $\psi(t) = \left(1 + \frac{6}{3 + \pi^2\alpha(\pi)} \right) t$.

Так как $\varphi'(0) = 1$, $\psi'(0) = 1 + \frac{6}{3 + \pi^2\alpha(\pi)}$ и $\varphi(0) = \psi(0) = 0$, понятно, что существует положительное решение ρ_0 уравнения $\varphi(t) = \psi(t)$. Поэтому для всех $t \in (0, \rho_0]$ имеем $\varphi(t) < \psi(t)$. Мы заключаем, что каждый шар \tilde{K}_r , при $r \in (0, \rho_0]$, оператор $A_{1\alpha}$ отображает на себя. Мы докажем что и шар $\tilde{K}_{1\alpha}$ один из них. Поэтому должно быть доказано, что

$$\ln \left(1 + \frac{6}{\pi^2 + 3} \right) < \rho_0.$$

Этот факт равносильен неравенству

$$\frac{6}{\pi^2 + 3} < \left(1 + \frac{6}{3 + \pi^2\alpha(\pi)} \right) \cdot \ln \left(1 + \frac{6}{\pi^2 + 3} \right). \quad (3)$$

Для $\alpha(\pi) = 1$ неравенство (3) будет иметь вид

$$\frac{6}{\pi^2 + 3} < \left(1 + \frac{6}{\pi^2 + 3} \right) \cdot \ln \left(1 + \frac{6}{\pi^2 + 3} \right)$$

Это неравенство является верным, что элементарно проверяется. Таким образом, доказано, что (3) действительно и для всех $\alpha(\pi) \in (0,1]$.

Пусть $\alpha(\pi) \in (1,\pi)$. Аналогично предыдущей процедуре, мы заключаем, что оператор $A_{1\alpha}$ будет отображать на себя шар радиуса $\frac{1}{\alpha(\pi)} \ln \left(1 + \frac{6}{\pi^2 + 3} \right)$, если

$$\frac{6}{\pi^2 + 3} < \left(1 + \frac{6}{\pi^2 + 3} \right) \frac{1}{\alpha(\pi)} \ln \left(1 + \frac{6}{\pi^2 + 3} \right). \quad (4)$$

Последнее условие на самом деле представляет ограничение на значение функции задержки в точке π . □

Теорема 4. Оператор $A_{2\alpha}$ отображают шар $\tilde{K}_{2\alpha}$ на себя.

Доказательство. Аналогично доказательству теоремы 3.

□

4 ЗАКЛЮЧЕНИЕ

В статье [1] мы сформировали интегральные уравнения $\tilde{q}_1(\bar{x}) = \tilde{f}_1(\bar{x}) - A_{1\alpha}\tilde{q}_1(\bar{x})$ и $\tilde{q}_2(\bar{x}) = \tilde{f}_2(\bar{x}) - A_{2\alpha}\tilde{q}_2(\bar{x})$, которые зависят от потенциала q и функции задержки α . В данной статье мы доказали, используя [6-17], что операторы $A_{1\alpha}$ и $A_{2\alpha}$ являются операторами сжатия ($A_{1\alpha}$ на шаре $\tilde{K}_{1\alpha}$, $A_{2\alpha}$ на шаре $\tilde{K}_{2\alpha}$) и “на” отображениях. Это означает, что существует единственное решение вышеупомянутых интегральных уравнений, т.е. существуют уникальные функции q и α , которые имеют абсолютно непрерывные производные на отрезке $[0, \pi]$, с условием $0 < \alpha'(x) < 1$, так что оператор $D_{ij}^{(2)}$ (см [3] и [4]) имеет собственные значения для точно заданных значений λ_{nij} (см [5]).

REFERENCES

- [1] Nikola Mihaljevic, “Formation of integral equations for the potential q and functions of delay α ”, *Math. Montisnigri*, **40**, 14-23 (2017)
- [2] R.Lazović, *Konstrukcija operatora tipa Shturma-Leeuvila sa kashњeњem*, Doktorska disertacija, Beograd (1998)
- [3] N.Mихалъевич, M.Pikula, “Karakteristichna funkcija operatora tipa Shturm-Liuvila sa promenljivim kashњeњem”, *Zbornik Fakulteta za pomorstvo u Kotoru*, **20**, 403-410 (2003).
- [4] Nikola Mihaljevic, “Asymptotics of the eigenvalues of the Sturm-Liouville problem with variable delay” *Math. Montisnigri*, **28**, 5-16 (2013).
- [5] Nikola Mihaljevic, “Representation characteristic function of Sturm-Liouville type over the zeros”, *Math. Montisnigri*, **31**, 25-37 (2014).
- [6] S.B. Norkin, *Differencial'ny'e uravneniya vtorogo porjadka s zapazdy`vayushhim argumentom*, Nauka, Moskva, (1965).
- [7] I.M. Gel'fand, B.M. Levitan, “Ob opredelenii differencial'nogo uravneniya po ego spektral'noj funkcii”, *Izv. AN SSSR, Ser. Mat.*, **15**, 309-360 (1951).
- [8] L.E. Elsgolez, S.B. Norkin, *Vvedenie v teoriyu differencial'ny'x uravnenij s otklonyayushhimsya argumentom*, Nauka, Moskva, (1971).
- [9] R.Lazović, M.Pikula, *Regularized trace of the operator applied to solving inverse problems*, Radovi matematički, (2002).
- [10] M. Pikula, “Ob opredelenii differencial'nogo uravneniya s peremenny`m zapazdy`vanjem”, *Math. Montisnigri*, **6**, 71-91 (1996).
- [11] M. Pikula, “Opredelenie differencial'nogo operatora Shturma-Liuvillya s zapazdy`vayushhim argumentom po dvum spektram”, *Matematicheskij vestnik*, **43**, 159-171 (1991).
- [12] N. Mihaljevich, “A reconstruction of the operator by using given spectral characteristic”, *Math. Montisnigri*, **20-21**, 15-34 (2007-2008).
- [13] N. Mihaljevich, M. Pikula, “The inverse Sturm-Liouville problem with changeable

- delay*”, *Math. Montisnigri*, **16**, 41-68 (2003).
- [14] V.A. Sadovnichij, *Teoriya operatorov*, Moskva, MGU, (1979)
- [15] N. Levinson, “The inverse Sturm-Liouville problem”, *Math. Tidsskr.*, **13**, 25-30 (1949).
- [16] B.V. Shabat, *Vvedenie v kompleksny`j analiz*, Nauka, Moskva, (1985).
- [17] L.V. Ahlfors, *Complex analysis*, Mc Graw-Hill Book Company, (1953).

Received May 16, 2019

О НЕКОТОРЫХ ГРАНИЧНЫХ СВОЙСТВАХ И ТЕОРЕМАХ ЕДИНСТВЕННОСТИ СУБГАРМОНИЧЕСКИХ ФУНКЦИЙ

С. Л. БЕРБЕРЯН

Российско-Армянский (Славянский) университет

Ереван, Армения

E-mail: samvel357@mail.ru

DOI: 10.20948/mathmontis-2019-45-2

Ключевые слова: Субгармонические, логарифмически-субгармонические функции, неевклидово расстояние, угловой предел, подчиненные функции

Аннотация. В статье рассматривается при определенных условиях вопрос ограниченности предельных множеств субгармонических функций в произвольных точках единичной окружности и теоремы единственности для логарифмически-субгармонических функций. Подобного типа задачи ранее рассматривались в основном для мероморфных и голоморфных функций.

ON SOME BOUNDARY PROPERTIES AND UNIQUENESS THEOREMS OF SUBHARMONIC FUNCTIONS

S. BERBERYAN

Russian-Armenian (Slavonic) University

Yerevan, Armenia

E-mail: samvel357@mail.ru

DOI: 10.20948/mathmontis-2019-45-2

Abstract. In the article we study boundedness of cluster sets of subharmonic functions in specific conditions in arbitrary points of a unit circle and theorem of uniqueness for logarithmically subharmonic functions. Similar tasks have already been studied before mainly for meromorphic and holomorphic functions.

2010 Mathematics Subject Classification: 31A05, 31A20, 30C80, 30D40, 30D45

Key words and Phrases: subharmonic, logarithmically subharmonic functions, non-euclidean distance, angular limit, subordinate functions

1 ВВЕДЕНИЕ

В настоящей работе изучаются при определенных условиях на последовательности точек $\{z_n\}$, содержащихся в единичном круге и стремящихся к точкам единичной окружности, ограниченность некоторых предельных множеств, существование угловых пределов и граничные теоремы единственности для некоторых классов субгармонических функций. Подобного типа задачи в основном ранее рассматривались известными математиками для мероморфных и голоморфных функций. Сошлемся только на часть этих работ (см. [1-10]). При изучении субгармонических функций возникают новые трудности и, как показывают примеры, без дополнительных условий утверждение известных теорем для мероморфных и голоморфных функций даже в случае нормальных субгармонических функций не имеют места. Изучаемые точки на единичной окружности, будут охарактеризованы как с помощью линейной меры, так и с помощью понятия категорий. В дальнейшем будем придерживаться общепринятых обозначений (см., например, [4]). Обозначим через D и Γ соответственно единичный круг $|z| < 1$ и единичную окружность $|z| = 1$. Обозначим через $L(\xi, \varphi)$ гиперцикл, проходящий через точки $\xi = e^{i\theta}$, $-\xi$ и который образует угол φ с диаметром Λ^θ , соединяющим точки ξ и $-\xi$. Пусть $H(\xi, \varphi_1, \varphi_2)$ – область, ограниченная двумя гиперциклами $L(\xi, \varphi_1)$ и $L(\xi, \varphi_2)$.

Рассмотрим действительнзначную функцию $f(z)$, определенную в D . Для произвольного подмножества S круга D , для которого точка $\xi \in \Gamma$ является предельной точкой, обозначим через $C(f, \xi, S)$ предельное множество функции $f(z)$ в точке ξ относительно множества S , т.е. $C(f, \xi, S) = \overline{\bigcap f(S \cap U(\xi))}$, где пересечение берётся по всем окрестностям $U(\xi)$ точки ξ , а черта означает замыкание множества относительно двухточечной компактификации \overline{R} множества $R = (-\infty, +\infty)$ в виде отрезка посредством добавления к точкам множества R символов $-\infty$ и $+\infty$. Точку $\xi \in \Gamma$ отнесём к множеству $F(f)$, если $C(f, \xi, \Delta(\xi))$ состоит из единственного значения α . В этом случае говорят, что функция $f(z)$ имеет в точке $\xi \in \Gamma$ угловой предел α . Множество $F(f)$ называется множеством точек Фату для функции $f(z)$.

Точку $\xi \in \Gamma$ относят к множеству $K(f)$ для функции $f(z)$, если $C(f, \xi, \Delta(\xi, \varphi_1, \varphi_2)) = C(f, \xi, \Delta(\xi, \varphi'_1, \varphi'_2))$ для любых углов $\Delta(\xi, \varphi_1, \varphi_2)$, $\Delta(\xi, \varphi'_1, \varphi'_2)$, где $\varphi_1, \varphi_2, \varphi'_1, \varphi'_2 \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Точка $\xi \in \Gamma$ называется точкой Мейера функции $f(z)$ и относится к множеству $M(f)$, если: $C(f, \xi, h(\xi, \varphi)) = C(f, \xi, D) \neq \overline{R}$ для любой хорды

$h(\xi, \varphi)$. Скажем, что точка $\xi \in \Gamma$ является точкой Плеснера для субгармонических функций $f(z)$, если для любых углов $\Delta(\xi, \varphi_1, \varphi_2)$ и $\Delta(\xi, \varphi'_1, \varphi'_2)$ предельные множества $C(f, \xi, \Delta(\xi, \varphi_1, \varphi_2))$, $C(f, \xi, \Delta(\xi, \varphi'_1, \varphi'_2))$ совпадают и $C(f, \xi, \Delta(\xi, \varphi_1, \varphi_2)) = [-\infty, \infty]$ для любых углов $\Delta(\xi, \varphi_1, \varphi_2)$. Множество точек Плеснера обозначим через $I(f)$. Неотрицательная субгармоническая функция $f(z)$ называется логарифмически-субгармонической, если $\ln(f(z))$ – субгармоническая функция. Интерпретируя круг D , как модель плоскости в геометрии Лобачевского, обозначим через $\sigma(z_1, z_2)$ неевклидовое расстояние между точками z_1, z_2 из круга D :

$$\sigma(z_1, z_2) = \frac{1}{2} \ln \frac{1+u}{1-u}, \text{ где } u = \left| \frac{z_1 - z_2}{1 - z_1 \bar{z}_2} \right|.$$

Понятие нормальной функции, рассмотренное для мероморфных функций и состоящее в свойстве порождать нормальное семейство на группе Γ всех конформных автоморфизмов области определения, было затем перенесено на гармонические и субгармонические функции (см., например, [11]). В случае единичного круга D группа Γ состоит из элементов $\Gamma = \{S(z); S(z) = e^{i\alpha}(z+a) \cdot (1+\bar{a}z)^{-1}, a - \text{произвольная точка в } D, \alpha - \text{произвольное действительное число}\}$. Придерживаясь обозначений из работы [4] скажем, что действительная функция $f(z) \in \mathfrak{R}$, если на группе Γ всех конформных автоморфизмов единичного круга D порождаемое ею семейство функций $\Phi: \{f(S(z)); S(z) \in T\}$ нормально в D в смысле Монтеля, т. е. из любой последовательности $\{f(S_n(z))\}$, семейства Φ , где $S_n(z) \in T$ можно извлечь подпоследовательность $\{f(S_{n_k}(z))\}$, равномерно сходящуюся на любом компакте K в D или равномерно расходящуюся к $-\infty$ или к $+\infty$ на K .

Говорят, что функция $u(z)$ подчинена субгармонической функции $f(z)$, если имеет место соотношение

$$u(z) = f(\omega(z)) \tag{1}$$

где $\omega(z)$ есть аналитическая в D функция, причем $\omega(0) = 0, |\omega(z)| < 1$. Известно (см. [3]), что функция $u(z)$ также будет субгармонической в D . Множество $N \subset \Gamma$ называют метрически плотным на некоторой дуге $\gamma \subset \Gamma$, если линейная мера $mes(\gamma' \cap N)$ положительна для каждой дуги $\gamma' \subset \gamma$. Выражение почти в каждой точке $\xi \in E$ означает, что всюду на E , кроме некоторого множества, линейная мера которого равна нулю.

Множество E называется множеством I категории, если оно является объединением счетного семейства нигде не плотных множеств. Множество, не являющееся множеством I категории, является множеством II категории. Множество $E \subset \Gamma$ называется остаточным множеством, если его дополнение до Γ I категории. Для дальнейшего сформулируем известную теорему единственности для логарифмически-субгармонических функций (см.[12]).

Теорема А. Пусть логарифмически-субгармоническая в D функция $f(z)$ имеет угловые пределы, равные нулю на некотором множестве $E \subset \Gamma, mes E > 0$. Тогда $f(z) \equiv 0$.

2 ОГРАНИЧЕННОСТЬ ПРЕДЕЛЬНЫХ МНОЖЕСТВ СУБГАРМОНИЧЕСКИХ ФУНКЦИЙ

Рассмотрим следующую теорему.

Теорема 1. Пусть субгармоническая в D функция $f(z)$ ограничена $|f(z_n)| \leq K$, где $K > 0, n = 1, 2, \dots$ на некоторой последовательности $\{z_n\}, z_n \in D, n = 1, 2, \dots, \lim_{n \rightarrow \infty} |z_n| = 1$,

$$\overline{\lim}_{n \rightarrow \infty} \sigma(z_n, z_{n+1}) < M < +\infty \quad (2)$$

и каждая точка некоторого множества $E \subset \Gamma, mes E > 0$, является предельной для последовательности $\{z_n\}$. Если в некоторой точке ξ множества E имеется такая гиперциклическая область $H(\xi, -\varphi, \varphi)$, для которой существует предел $\lim_{z \rightarrow \xi} f(z) = \alpha$ и неевклидово расстояние от гиперциклов $L(\xi, -\varphi)$ и $L(\xi, \varphi)$ до диаметра Λ^θ , соединяющего точки ξ и $-\xi$, больше или равно $M/2$, то $|\alpha| \leq K$.

Доказательство. Не ограничивая общности можно считать, что множество E целиком лежит на некоторой дуге $\Gamma_1 \subset \Gamma$ и что концевые точки дуги Γ_1 являются предельными точками множества E . Идея доказательства впервые использована в работе [8]. Предположим противное, т.е. что $\alpha > K$. Рассмотрим вначале случай, когда $f(z_n) \leq K$. Тогда имеем, что начиная с некоторого номера N все точки последовательности $\{z_n\}, n = 1, 2, \dots, z_n \in D$, должны лежать вне области $H(\xi, -\varphi, \varphi)$. С другой стороны, из условий теоремы 1 следует, что точка ξ является предельной точкой для последовательности

$\{z_n\}$ и, значит, для бесконечного числа значений n точки z_n и z_{n+1} лежат на противоположных гиперциклах $L(\xi, -\varphi)$ и $L(\xi, \varphi)$ либо вне замкнутой области $\overline{H(\xi, -\varphi, \varphi)}$. Следовательно, для бесконечного числа значений n $\sigma(z_n, z_{n+1}) \geq M$, что невозможно в силу соотношения (2). Полученное противоречие доказывает утверждение теоремы 1 в случае, когда $f(z_n) \leq K$. Аналогично рассматривается случай $f(z_n) \geq -K$. При этом предполагается, что $\alpha < -K$ и снова приходим к противоречию.

Рассмотрим несколько следствий из теоремы 1.

Следствие 1. Пусть субгармоническая в D функция $f(z)$ ограничена $|f(z_n)| \leq K$, $n = 1, 2, \dots$ на некоторой последовательности точек $\{z_n\}$, $n = 1, 2, \dots$, $\lim_{n \rightarrow \infty} |z_n| = 1$, множество предельных точек которой есть множество E , $mes E > 0$ и $\overline{\lim}_{n \rightarrow \infty} \sigma(z_n, z_{n+1}) < M < +\infty$. Если в некоторой точке $\xi \in E$ существует угловой предел α_ξ , то $|\alpha_\xi| \leq K$.

Замечание 1. Отметим, что даже у субгармонических функций $f(z) \in \mathfrak{N}$ из существования углового предела α в некоторой гиперциклической области $H(\xi, -\varphi, \varphi)$ в точке $\xi \in \Gamma$ не следует существование углового предела α_ξ в этой точке. Соответствующий пример приведен в работе [11].

Следствие 2. Пусть $f(z)$ – субгармоническая функция класса \mathfrak{N} и удовлетворяет следующим условиям:

- 1) $|f(z_n)| \leq K$, $n = 1, 2, \dots$ на некоторой последовательности $\{z_n\}$, $z_n \in D$, $n = 1, 2, \dots$, $\lim_{n \rightarrow \infty} |z_n| = 1$;
- 2) множество предельных точек последовательности $\{z_n\}$ есть множество E , $mes E > 0$;
- 3) $\overline{\lim}_{n \rightarrow \infty} \sigma(z_n, z_{n+1}) < M < +\infty$;
- 4) В каждой точке $\xi \in E$ существует последовательность $\{t_n^\xi\} \in \Delta(\xi, \varphi_1^\xi, \varphi_2^\xi)$, такая, что $\overline{\lim}_{n \rightarrow \infty} \sigma(t_n^\xi, t_{n+1}^\xi) < L < +\infty$ и $\lim_{n \rightarrow \infty} f(t_n^\xi) = \alpha_\xi$.

Тогда почти в каждой точке $\xi \in E$ функция $f(z)$ имеет угловые пределы α_ξ , для которых $|\alpha_\xi| \leq K$.

Доказательство. Обозначим через $F = E \cap K(f)$ множество точек $\xi \in E$ на Γ , в которых существуют соответствующие пределы α_ξ вдоль последовательностей $\{t_n^\xi\}$.

Из условий следствия 2 следует, что $mes F = mes E > 0$. В силу известного утверждения из работы [15] в каждой точке $\xi \in F$ функция $f(z)$ имеет угловой предел α_ξ . Отсюда утверждение следствия 2 следует из утверждения следствия 1.

3 ГРАНИЧНЫЕ ТЕОРЕМЫ ЕДИНСТВЕННОСТИ ДЛЯ ЛОГАРИФМИЧЕСКИ-СУБГАРМОНИЧЕСКИХ ФУНКЦИЙ

В качестве применения теоремы 1 рассмотрим одну граничную теорему единственности для логарифмически-субгармонических функций.

Теорема 2. Пусть логарифмически-субгармоническая в D функция $f(z)$ удовлетворяет условию $\lim_{n \rightarrow \infty} f(z_n) = 0$ по некоторой последовательности точек $\{z_n\}$, $z_n \in D$, $n = 1, 2, \dots$ множество предельных точек есть множество $E \subset \Gamma$, $mes E > 0$, $\overline{\lim}_{n \rightarrow \infty} \sigma(z_n, z_{n+1}) < M < +\infty$ и в каждой точке $\xi \in E$ существуют угловые пределы α_ξ . Тогда $f(z) \equiv 0$.

Доказательство. Действительно, рассматривая в теореме 1 в качестве K любое сколь угодно малое положительное число $\varepsilon > 0$, согласно утверждению теоремы 1 получим, что $0 \leq \alpha_\xi \leq \varepsilon$, а, значит, в каждой точке $\xi \in E$ $\alpha_\xi = 0$. Согласно теореме А о единственности для логарифмически-субгармонических функций $f(z) \equiv 0$, что и требовалось доказать.

Следствие 3. Пусть $f(z)$ – логарифмически-субгармоническая функция и удовлетворяются все условия теоремы 2. Если функция $u(z)$ подчинена в D функции $f(z)$, то $u(z) \equiv 0$.

Доказательство. Действительно из соотношения (1) следует, что $\ln(u(z))$ подчинена субгармонической функции $\ln(f(\omega(z)))$ и, значит, согласно известному утверждению (см.[3])

будет субгармонической функцией. Следовательно, $u(z)$ – логарифмически-субгармоническая функция. Из соотношения (1) и из того, что $\omega(0) = 0$ и $|\omega(z)| < 1$ следует, что с помощью аналитической функции $\omega(z)$ произвольная замкнутая область $\overline{G} \subset D$, содержащая точку $z = 0$, отображается в замкнутую область $\overline{G}_1 \subset |\omega| < 1$, содержащую точку $\omega = 0$. В силу утверждения теоремы 2 $f(z) \equiv 0$ и, значит, из соотношения (1) $u(z) \equiv 0$ в любой замкнутой области $\overline{G} \subset D$. Отсюда следует утверждение следствия 3.

Рассмотрим следующую теорему о существовании угловых пределов у субгармонических функций класса \mathfrak{N} .

Теорема 3. Пусть $f(z)$ – субгармоническая функция класса \mathfrak{N} и E – некоторое множество II категории на некоторой дуге $\gamma \subset \Gamma$. Допустим, что в каждой точке $\xi \in E$ существует некоторая последовательность $\{z_n^\xi\} \in \Delta(\xi, \varphi_1, \varphi_2), \lim_{n \rightarrow \infty} z_n = \xi$, для которой $\overline{\lim}_{n \rightarrow \infty} \sigma(z_n, z_{n+1}) < M < +\infty$ и выполнены следующие условия:

- 1) предельное множество $C(f, \xi, \{z_n^\xi\})$ ограничено сверху;
- 2) существует такое метрически плотное на \mathcal{Y} множество N , что для любой точки $\xi \in N$, существует такая последовательность $\{q_n^\xi\}$, для которой $\{q_n^\xi\} \rightarrow \xi, \overline{\lim}_{n \rightarrow \infty} \sigma(q_n^\xi, q_{n+1}^\xi) < L < +\infty, \{q_n^\xi\} \in \Delta(\xi, \varphi'_1, \varphi'_2)$, и $\beta \in C(f, \xi, \{q_n^\xi\})$.

Тогда существует множество $E_2, \text{mes } E_2 > 0$, в каждой точке которого функция $f(z)$ имеет угловые граничные пределы, равные β .

Доказательство. Принимая во внимание, что для произвольной функции $f(z)$, определенной в D , множество $K(f)$ II категории и ее дополнение до Γ I категории [3] получим, что множество $F = K(f) \cap E$ является множеством II категории. Для субгармонических функций $f(z) \in \mathfrak{N}$ известно (см. [15]), что если $\xi \in K(f)$, то $C(f, \xi, \Delta(\xi)) = C(f, \xi, h(\xi, \varphi))$ для любых углов $\Delta(\xi)$ и хорд $h(\xi, \varphi)$.

Из условия 1) согласно известному утверждению (см. [18]) в каждой точке $\xi \in E$ предельные множества $C(f, \xi, \Delta(\xi))$ для любого угла $\Delta(\xi)$ ограничены сверху и, следовательно, эти точки не являются точками Плеснера.

Исходя из известного разложения для субгармонических функций класса \mathfrak{N} (см. [2]) все точки множества F , кроме, быть может, некоторого множества $E_1 \subset E$ I категории, будут принадлежать множеству $M(f)$. Так как любая субгармоническая функция класса \mathfrak{N} непрерывна, то предельное множество $C(f, \xi, D)$ замкнутое и связное множество. В силу теоремы Коллингвуда о максимальности (см. [1]) и определению точек Мейера

$$C(f, \xi, h(\xi, \varphi)) = C(f, \xi, D) \quad (3)$$

в каждой точке $\xi \in \Gamma \setminus E_1$. Так как предельное множество $C(f, \xi, h(\xi, \varphi))$ (см. [18]) ограничено сверху, то в силу соотношения (3) предельное множество $C(f, \xi, D)$ также ограничено сверху в любой точке $\xi \in F \setminus E_1$. Поэтому в каждой такой точке ξ существует некоторая окрестность $U(\xi)$, в которой функция $f(z)$ ограничена сверху. Обозначим через γ^1 пересечение γ с указанной окрестностью на Γ .

С помощью функции $z = \varphi(\omega)$ конформно отображим единичный круг $D_1 : |\omega| < 1$ на указанную окрестность $U(\xi)$. Тогда функция $\psi(\omega) = f(\varphi(\omega))$ субгармоническая функция, ограниченная сверху в круге D_1 , и поэтому почти всюду на $\Gamma_1 : |\omega| = 1$ имеет радиальные пределы. Так как при конформных отображениях почти всюду некасательные пути переходят в некасательные пути (см. [7]), то почти в каждой точке $\xi \in \text{mes}(\gamma^1 \cap K(f))$ функция $f(z)$ имеет предел β_ξ по некоторому некасательному пути L_ξ . В силу известного утверждения о том, что если $\xi \in \text{mes}(\gamma^1 \cap K(f))$, то (см. [15]) функция $f(z)$ имеет угловой предел β_ξ , получим существование угловых пределов почти всюду на дуге γ^1 .

Но с другой стороны, согласно условию 2) на множестве $E_2 = N \cap \gamma^1, \text{mes } E_2 > 0$, угловые пределы функции $f(z)$ должны быть равны β , что и требовалось доказать.

Из утверждения теоремы 3 следует еще одна теорема единственности для логарифмически-субгармонических функций класса \mathfrak{N} .

Теорема 4. Пусть $f(z)$ – логарифмически-субгармоническая функция класса \mathfrak{N} и E – некоторое множество II категории на дуге Γ . Допустим, что в каждой точке $\xi \in E$

существует некоторая последовательность $\{z_n^\xi\} \in \Delta(\xi, \varphi_1, \varphi_2)$, $\lim_{n \rightarrow \infty} z_n = \xi$, для которой $\overline{\lim}_{n \rightarrow \infty} \sigma(z_n, z_{n+1}) < M < +\infty$ и выполнены следующие условия:

- 1) предельное множество $C(f, \xi, z_n^\xi)$ ограничено сверху;
- 2) существует такое метрически плотное на γ^1 множество N , что для любой точки $\xi \in N$, $0 \in C(f, \xi, \{q_n^\xi\})$, где последовательность $\{q_n^\xi\} \in \Delta(\xi, \varphi'_1, \varphi'_2)$, $\{q_n^\xi\} \rightarrow \xi$, и $\overline{\lim}_{n \rightarrow \infty} \sigma(q_n^\xi, q_{n+1}^\xi) < L < +\infty$.

Тогда $f(z) \equiv 0$.

Доказательство. Действительно из утверждения теоремы 3 следует, что угловые пределы функции $f(z)$ на некотором множестве E , $mes E > 0$, равны 0. Согласно теореме А о единственности логарифмически-субгармонических функций получим, что $f(z) \equiv 0$.

Замечание 2. Отметим, что теоремы, подобные теореме 4, рассматривались для мероморфных функций в работе [10].

Следствие 4. Пусть $f(z)$ – логарифмически-субгармоническая функция класса \mathfrak{N} и удовлетворяются все условия теоремы 4. Если функция $u(z)$ подчинена в D функции $f(z)$, то $u(z) \equiv 0$.

Для доказательства следствия 4 проводятся те же рассуждения, что и при доказательстве следствия 3.

4 ЗАКЛЮЧЕНИЕ

Таким образом, в настоящей работе получены теоремы об ограниченности предельных множеств, о существовании угловых пределов и о единственности некоторых классов субгармонических функций, определенных в единичном круге. Эти теоремы можно рассматривать как субгармонические аналоги известных теорем, полученных ранее для мероморфных и голоморфных функций, определенных в единичном круге.

Благодарности: Работа выполнена при финансовой поддержке государственного комитета по науке МОН РА в рамках научного проекта № 18Т-1А019 и в рамках финансовой поддержки проекта развития Российско-Армянского Университета.

REFERENCES

- [1] E.F. Collingwood, A.J. Lohwater, *The theory of cluster sets*, Cambridge University Press (1966).
- [2] S.L. Berberyan, V.I. Gavrilov, “Predelnie mnozhestva neprerivnich i garmonicheskikh funktsii po nekasatel'nim granichnim putyam”, *Math. Montisnigri*, **1**, 17-25 (1993).
- [3] I.I. Privalov, *Subgarmonicheskie funktsii*, M. L., NKTP SSSR, (1937).
- [4] V.I. Gavrilov, “Normalnie funktsii i pochti periodicheskie funktsii”, *DAN SSSR*, **240** (4), 768-770 (1978).
- [5] M.M. Djrbashyan, V.S. Zakaryan, *Klassi i granichnie svoistva funktsii meromorfnykh v krugakh*, Moskva, Nauka, (1993).
- [6] F. Bagemihl, “Some boundary properties of normal functions bounded on nontangential arcs”, *Arch. Math.*, **14** (6), 399-406 (1963).
- [7] I.I. Privalov, *Granichnie svoistva analiticheskikh funktsii*, M. L., GITTL, (1950).
- [8] F. Bagemihl, W. Seidel, “Sequential and continuous limits of meromorphic functions”, *Ann. Acad. Sci. Fenn. Ser. AI*, 280, 1-17 (1960).
- [9] V.I. Gavrilov, “O golomorfnykh funktsiyakh ograniachennikh na posledovatelnostyakh tochek”, *Sibirskii matematicheski zhurnal*, **6** (6), 1227-1233 (1965).
- [10] V.I. Gavrilov, “O nekotorykh teoremax edinstvennosti dlya meromorfnykh funktsii”, *Trudi seminarov imeni I.G. Petrovskogo, MGU*, 1, 57-62 (1975).
- [11] J. Meek, “Of Fatous points of normal subharmonic functions”, *Mathematica Japonica*, **22** (3), 309-314 (1977).
- [12] S. Lozinski, “O subgarmonicheskikh funktsiyakh I ikh prilozheniyakh k teorii poverkhnostei”, *Izv. AN. SSSR, Seriya matem.*, **8**, 175-194 (1944).
- [13] M. Arsove, “The Lusin-Privalov theorem for subharmonic functions”, *Proc. London Math. Soc.*, **14** (54), 260-270 (1964).
- [14] Z. Pavicevic and I. Susic, “Boundary behavior of subharmonic functions”, *Matematiqki vesnik*, 50, 83-87 (1998).
- [15] S.L. Berberyan, “O granichnykh osobennostyakh normalnykh subgarmonicheskikh funktsii”, *Math. Montisnigri*, **18-19**, 5-14 (2005-2006).
- [16] S.L. Berberyan, “O teoreme Plesnera dlya garmonicheskikh funktsii”, *Math. Montisnigri*, **24**, 150-154 (2012).
- [17] S.L. Berberyan, “O granichnykh teoremax edinstvennosti dlya logarifmicheskikh subgarmonicheskikh funktsii”, *Izvestiya vuzov, Matematika*, 9, 3-9 (2016).
- [18] S.L. Berberyan, “Ob uglovnykh granichnykh znacheniyakh normalnykh neprerivnykh funktsii”, *Izvestiya vuzov, Matematika*, 3, 22-28 (1986).
- [19] S.L. Berberyan, “Ob ograniachennosti normalnykh garmonicheskikh funktsii”, *Vestnik MGU, Matematika, Mekhanika*, 2, 57-61 (2013).
- [20] S.L. Berberyan “O nekotorykh teoremax edinstvennosti dlya subgarmonicheskikh funktsii”, *National Academy of Sciences of Armenia, reports*, **110** (2), 128-136 (2010).

Received May 20, 2019

К РАЗРАБОТКЕ НЕАДДИТИВНОЙ ТЕРМОДИНАМИКИ КВАНТОВЫХ СИСТЕМ НА ОСНОВЕ СТАТИСТИКИ ТСАЛЛИСА

А.В. КОЛЕСНИЧЕНКО

Институт прикладной математики им. М.В. Келдыша РАН. Москва, Россия

Ответственный автор. E-mail: kolesn@keldysh.ru

DOI: 10.20948/mathmontis-2019-45-3

Ключевые слова: Квантовая неэкстенсивная статистика, энтропия Тсаллиса, степенной закон распределения матрицы плотности.

Аннотация. В настоящее время теории неэкстенсивных сложных систем существенно развиваются в ускоренном ритме, 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, а степенными распределениями. Среди множества неэкстенсивных систем особое значение имеют малые квантовые системы, основанные на неаддитивной параметрической энтропии Тсаллиса, связанной с матрицей плотности, описывающей системы, квантовые состояния которых известны не полностью. При изучении подобных систем возникают многочисленные новые математические проблемы, требующие своего решения. В их числе, одной из важных, является проблема разработки неаддитивной термодинамики квантовых систем в рамках статистики Тсаллиса.

В данной работе при описании квантово-механической неэкстенсивной системы вслед за фон Нейманом предполагается правильность основных двух начал термодинамики. Выполненный анализ основывается на степенном равновесном распределении матрицы плотности, полученном из условия абсолютного экстремума квантовой энтропии Тсаллиса при заданности средней энергии и среднего числа частиц для ансамбля квантовых систем, а также на осреднении наблюдаемых величин по эскортному распределению. В результате получено обобщение на квантовый случай 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 термодинамические соотношения в квантовой статистике Тсаллиса, которые отличны от выведенных ранее традиционным для неэкстенсивной статистики способом. На основе свойства выпуклости различающей информации Ратье–Каннапана, обобщённой на квантовый случай, обсуждается второй закон термодинамики. Изучены спонтанные переходы между стационарными состояниями сложной квантово-механической системы и доказана H -теорема Больцмана.

Развитый подход предполагает использование неэкстенсивной квантовой термодинамики в различных контекстах, касающихся, в частности, моделирования квантовых тепловых эффектов в наноразмерных устройствах и других квантовых технологиях.

2010 Mathematics Subject Classification: 85A35, 91B50, 82C40.

Key words and Phrases: Quantum Nonextensive Statistics, Entropy of Tsallis, Power Law Distribution of Density Matrix.

TO DEVELOPMENT OF THE NON-ADDITIVE THERMODYNAMICS OF THE QUANTUM SYSTEMS ON BASIS STATISTICS OF TSALLIS

A.V. KOLESNICHENKO

Keldysh Institute of Applied Mathematics, Russian Academy of Science

Corresponding author. E-mail: kolesn@keldysh.ru

DOI: 10.20948/mathmontis-2019-45-3

Summary. At present, theories of nonextensive complex systems develop substantially in an accelerated rhythm, at which new ideas emerge, which allow a deeper understanding of their nature, possibilities 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, but by power distributions. Among the set of nonextensive systems, small quantum systems based on the nonadditive parametric entropy of Tsallis associated with the density matrix, describing systems whose quantum states are not fully known, are of particular importance. When studying such systems, there are numerous new mathematical problems that need to be solved. Among them, one of the most important is the problem of developing nonadditive thermodynamics of quantum systems based on Tsallis statistics.

In this paper, when describing a quantum-mechanical nonextensive system, following von Neumann, the correctness of both basic principles of thermodynamics is assumed. The analysis of the quantum system is based on the power-law equilibrium distribution of the density matrix obtained from the absolute extremum condition of the Tsallis quantum entropy with a given average energy and average number of particles for an ensemble of quantum systems, as well as averaging the observed values for the escort distribution. As a result, a generalization to the quantum case of the zero law of thermodynamics is obtained for two independent subsystems with their thermal contact, and a so-called physical temperature is introduced that is different from the inversion of the Lagrange multiplier β . Using the generalized first law of thermodynamics and the Legendre transformation on the basis of the introduced Clausius entropy, modified thermodynamic relations were found in the Tsallis quantum statistics that are different from the previously derived method for a non-extensive statistical mechanics. The second law of thermodynamics is discussed on the basis of the convexity property of the Ratier – Kannapan information, which is generalized to the quantum case. Spontaneous transitions between stationary states of a complex quantum-mechanical system are studied and the Boltzmann H -theorem is proved..

The developed approach assumes the use of nonextensive quantum thermodynamics in various contexts, concerning, in particular, the simulation of quantum thermal effects in nanodevices and other quantum technologies.

1 ВВЕДЕНИЕ

Как теперь стало понятно, статистическая механика Больцмана–Гиббса и стандартная термодинамика не являются вполне универсальными теориями, поскольку они имеют ограниченные области применимости. Это связано, в частности, с тем, что в основе статистики Больцмана–Гиббса лежит постулат о полном перемешивании потока «фазовых точек» в фазовом пространстве (гипотеза молекулярного хаоса). А это означает, в свою очередь, что фазовое пространство не содержит запрещённых состояний и обладает обычными свойствами непрерывности, гладкости, евклидовости. При этом гипотеза перемешивания, дополненная предположением о бесконечном числе степеней свободы, приводит к экспоненциальному распределению вероятности состояний системы (из которого следует, в частности, свойство аддитивности экстенсивных термодинамических переменных), или, в случае кинетической теории газов, к максвелловскому распределению скоростей.

Вместе с тем существует широкий класс сложных систем, элементы которых взаимодействуют глобально, чему предшествует снижение симметрии системы, связанное с формированием коллективных мод интенсивных переменных. В физике и в других естественных науках, использующих методы статистической механики, известны многочисленные примеры подобных систем, поведение и свойства которых являются аномальными с точки зрения классической статистики. Существует множество систем, в которых имеются нелокальные корреляции, сильные взаимозависимости между отдельными (всеми) элементами системы. Сложная пространственно-временная структура подобных систем приводит к нарушению принципа аддитивности для таких важнейших термодинамических величин, как энтропия или внутренняя энергия.

Довольно широкий класс подобных систем (хотя далеко не всех) адекватно описывается неэкстенсивной (неаддитивной) статистической механикой, основанной на параметрических энтропиях Тсаллиса [1] и Реньи [2], которые, однако, сохраняют гносеологическую структуру (логическую схему построения) классической статистики [3-6]. Важным преимуществом неэкстенсивных статистик по сравнению с классической статистикой Больцмана–Гиббса является асимптотический степенной закон распределения вероятностей (проявляющийся при максимизации соответствующих параметрических энтропий), который не зависит от экспоненциального поведения, обусловленного распределением Гиббса.

Следует заметить, что неэкстенсивные статистики (например, статистики Тсаллиса и Реньи) представляет собой всё же обобщение, а не альтернативу классической статистике Больцмана–Гиббса, поскольку они распространяют область применимости стандартной статистической теории на неэкстенсивные системы только путём расширения математической формы их энтропийного функционала [7-10].

В настоящее время теории неэкстенсивных сложных систем существенно развиваются в ускоренном ритме, при котором появляются новые идеи, позволяющие глубже понять их природу, возможности и ограничения (см. Библиографию, представленную на сайте <http://tsallis.-cat.cbpf.br/biblio.htm>, которая постоянно обновляется). Каждая теория имеет широкий спектр важных приложений, связанных с физикой статистических систем, вероятностные свойства которых описываются не гиббсовыми (не гауссовыми), а степенными распределениями. В частности,

неэкстенсивная статистика Тсаллиса успешно применяется ко многим сложным системам, начиная от нелинейных диффузионных уравнений], обобщённых кинетических уравнений [11], систем Фоккера-Планка, H -теоремы Больцмана [6], удельной теплоемкости гармонического осциллятора, до изучения космических систем с дальним силовым взаимодействием, межзвездной турбулентности и теории фракталов [9], эволюции астрофизических дисков [10, 12, 13], городской транспортной системы [14], биофизики, экономики, нейрофизики и многое другое.

Среди множества сложных систем особую важность имеют малые квантовые системы, основанные на неаддитивной параметрической энтропии Тсаллиса $S_q(\hat{\rho})$, связанной с матрицей плотности $\hat{\rho}$, описывающей различные квантовые состояния. При изучении подобных систем возникают многочисленные новые математические проблемы, требующие своего решения. В их числе, одной из важнейших, является проблема построения термодинамики квантово-механических ансамблей в рамках неэкстенсивной статистики Тсаллиса.

В данной работе для описания квантовой физической системы мы воспользуемся формализмом матрицы (оператора) плотности $\hat{\rho}$, с помощью которого наиболее удобно описывать системы, квантовые состояния которых известны не полностью [15]. Кроме этого, следуя фон Нейману (1964), будем использовать обычный формализм феноменологической термодинамики, при этом роль квантовой механики сведётся лишь к тому, что наше рассмотрение будет относиться к квантово-механическим объектам – правильность же обоих основных начал термодинамики предполагается [15]. С учётом этого будет показано, как можно получить равновесную статистическую термодинамику неэкстенсивных систем и определить её свойства на основе двух функционалов – квантовой параметрической энтропии Тсаллиса $S_q(\hat{\rho})$ и обобщённой квантовой относительной энтропии (квантовой различающей информации Ратье–Каннаппана $K_q(\hat{\rho}:\hat{\sigma})$). Это исследование будет базироваться на степенном равновесном распределении матрицы плотности, полученном из условия абсолютного экстремума энтропии $S_q(\hat{\rho})$ при заданности средней энергии и среднего числа частиц для ансамбля квантовых систем, а также на осреднении его случайных динамических параметров (наблюдаемых) по эскортному распределению. Будет получено обобщение на квантовый случай нулевого закона термодинамики для двух независимых подсистем при их тепловом контакте и введена так называемая физическая температура, отличная от инверсии множителя Лагранжа β . При её использовании, с привлечением обобщённого первого закона термодинамики и преобразования Лежандра, будут найдены модифицированные термодинамические соотношения, отличные от выводимых «обычным» для статистики способом. И, наконец, на основе функционала для квантовой относительной энтропии, обобщённого на неэкстенсивные системы, будет обсуждаться второй закон термодинамики, будут исследованы спонтанные переходы между произвольными состояниями сложной квантовой системы и доказана H -теорема Больцмана в рамках неэкстенсивной квантовой статистики.

2. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ И СТАТИСТИЧЕСКИЕ СВОЙСТВА КВАНТОВОЙ ЭНТРОПИИ ДЛЯ НЕЭКСТЕНСИВНЫХ СИСТЕМ

В основу изучения различных статистических квантовых ансамблей неэкстенсивных систем можно положить экстремальные свойства квантовой информационной энтропии (введённой впервые в работе [35]) и использовать их для нахождения различных матриц плотности, заменяющих функции распределения вероятностей в классической статистике. Отметим, кстати, что при обобщении ансамблей Гиббса на случай квантовой статистики, фон Нейман [34] исходил именно из экстремальных свойств введённой им энтропии квантового состояния $S(\hat{\rho}) \equiv -\text{Tr}(\hat{\rho} \ln \hat{\rho})$, где $\hat{\rho}$ – матрица (операторⁱ⁾) плотности микросистемы, при помощи формализма которой, согласно теореме Глисона [16], описывается любая квантово-механическая система.

$$\text{Tr} \hat{\rho}(x, x') = 1. \quad (1)$$

В квантовой неэкстенсивной статистике при вероятностной нормировке матрицы плотности $\hat{\rho}(x, x') = \sum_r w_r \psi_r(x) \psi_r^*(x')$ (в матричном x -представлении, описывающей смешанные квантовые состояния [15]), квантовая информационная энтропия Тсаллиса $S_q(\hat{\rho})$ задаётся следующим обобщённым функционалом от оператора плотности [1]:

$$S_q(\hat{\rho}) \equiv \frac{1}{q-1} \text{Tr}(\hat{\rho} - \hat{\rho}^q). \quad (2)$$

Здесь энтропийный индекс q (параметр деформации) представляет собой вещественное число (принадлежащее области $q \in \mathbb{R}$), которое характеризует неэкстенсивную особенность (неаддитивность) квантовой системы. Заметим, что шпуровая структура определения энтропии (2) важна тем, что делает энтропию функционально независимой от унитарных преобразований в пространстве состояний, т.е. эта формула справедлива при любом представлении оператора $\hat{\rho}$, а не только при его матричном x -представлении [17].

Можно показать, что параметрическая квантовая энтропия (2) может быть представлена также в следующих эквивалентных формах:

$$S_q(\hat{\rho}) = -\text{Tr}(\hat{\rho}^q \ln_q \hat{\rho}) = -\text{Tr}(\hat{\rho} \text{Ln}_q \hat{\rho}) \equiv -\langle \text{Ln}_q \hat{\rho} \rangle_1, \quad (2^*)$$

$$\ln_q \hat{A} \equiv \frac{\hat{A}^{1-q} - 1}{1-q}, \quad \text{Ln}_q \hat{A} \equiv \frac{\hat{A}^{q-1} - 1}{q-1} = \hat{A}^{q-1} \ln_q \hat{A} \quad (3)$$

– так называемые деформированные логарифмы [18], обладающие, как легко убедиться, следующим свойством: при $q \rightarrow 1$, $\ln_q \hat{A} \rightarrow \ln \hat{A}$, $\text{Ln}_q \hat{A} \rightarrow \ln \hat{A}$. При его

ⁱ⁾ Далее операторы будем обозначать буквой со «шляпкой» над ней.

использовании квантовая энтропия Тсаллиса $S_q(\hat{\rho}) = -\text{Tr}(\hat{\rho}^q \ln_q \hat{\rho})$ переходит в классическую квантовую энтропию фон Неймана $S_1(\hat{\rho}) \equiv -\text{Tr}(\hat{\rho} \ln \hat{\rho})$, являющуюся, в свою очередь, квантовым обобщением энтропии Гиббса в классической статистической механике.

В традиционной квантовой статистике любой случайной динамической переменной A ставится в соответствие эрмитовый оператор [15] \hat{A} так, что среднее значение этой переменной в состоянии микросистемы, описываемом матрицей плотности $\hat{\rho}$, вычисляется по формуле: $\langle A \rangle_1 = \text{Tr}(\hat{\rho} \hat{A})$. В неэкстенсивной квантовой статистике Тсаллиса для вычисления среднего значения $\langle \hat{A} \rangle_q$ динамической переменной \hat{A} и её флуктуации $\Delta_q \hat{A}$ могут быть использованы различные формулировки [3]. Далее мы воспользуемся следующим их определением:

$$\langle \hat{A} \rangle_q \equiv \text{Tr}(\hat{P}_q \hat{A}) = \frac{\text{Tr}(\hat{\rho}^q \hat{A})}{\text{Tr} \hat{\rho}^q}, \quad \Delta_q \hat{A} \equiv \hat{A} - \langle \hat{A} \rangle_q, \quad \text{Tr}(\hat{\rho}^q \Delta_q \hat{A}) = 0, \quad (4)$$

где

$$\hat{P}_q(x) \equiv \hat{\rho}^q / \text{Tr}(\hat{\rho}^q) \quad (5)$$

– так называемое нормированное эскортное распределение [3], для которого $\text{Tr}(\hat{P}_q) = 1$.

2.1. Неаддитивность квантовой энтропии Тсаллиса для независимых систем

Энтропия (2) имеет много полезных свойств. Покажем сначала, что для независимых квантовых физических систем она неаддитивна. Действительно, пусть состояние системы, состоящей из двух подсистем, описывается совместным мультипликативным статистическим оператором $\hat{\rho}^{(1,2)} \equiv \hat{\rho}^{(1)} \otimes \hat{\rho}^{(2)}$, где $\hat{\rho}^{(1)}$ и $\hat{\rho}^{(2)}$ – операторы плотности отдельных подсистем (здесь и далее символом \otimes обозначено матричное произведение). Тогда энтропии отдельных подсистем и общая энтропия системы задаются следующими выражениями:

$$S_q^{(1)} \equiv S_q(\hat{\rho}^{(1)}) = \frac{1}{q-1} \left[1 - \text{Tr}(\hat{\rho}^{(1)})^q \right], \quad S_q^{(2)} = S_q(\hat{\rho}^{(2)}) = \frac{1}{q-1} \left[1 - \text{Tr}(\hat{\rho}^{(2)})^q \right],$$

$$S_q^{(1,2)} = S_q[\hat{\rho}^{(1)} \otimes \hat{\rho}^{(2)}] = \frac{1}{q-1} \left[1 - \text{Tr}(\hat{\rho}^{(1,2)})^q \right] = \frac{1}{q-1} \left[1 - \text{Tr}(\hat{\rho}^{(1)} \otimes \hat{\rho}^{(2)})^q \right] \quad (*)$$

при условии нормировки $\text{Tr}(\hat{\rho}^{(1,2)}) = \text{Tr}(\hat{\rho}^{(1)}) = \text{Tr}(\hat{\rho}^{(2)}) = 1$. Используя формулу (*), получим, с учётом известного соотношения] $\text{Tr}(\hat{A} \otimes \hat{B}) = \text{Tr}(\hat{A}) \text{Tr}(\hat{B})$ [15, следующее выражение

$$\begin{aligned}
(q-1)S_q^{(1)}S_q^{(2)} &= \frac{1}{q-1} \left[1 - \text{Tr}(\hat{\rho}^{(1)})^q \right] \times \left[1 - \text{Tr}(\hat{\rho}^{(2)})^q \right] = \\
&= \frac{1 - \text{Tr}(\hat{\rho}^{(1)})^q}{q-1} + \frac{1 - \text{Tr}(\hat{\rho}^{(2)})^q}{q-1} - \frac{1 - \text{Tr}(\hat{\rho}^{(1)})^q \text{Tr}(\hat{\rho}^{(2)})^q}{q-1} = S_q^{(1)} + S_q^{(1)} - S_q^{(1,2)}.
\end{aligned}$$

Отсюда следует свойство неаддитивности энтропии двух независимых систем в квантовой статистике Тсаллиса

$$S_q(\hat{\rho}^{(1)} \otimes \hat{\rho}^{(2)}) = S_q(\hat{\rho}^{(1)}) + S_q(\hat{\rho}^{(2)}) + (1-q)S_q(\hat{\rho}^{(1)})S_q(\hat{\rho}^{(2)}). \quad (6)$$

Таким образом, неаддитивная квантовая энтропия $S_q(\hat{\rho}^{(1)} \otimes \hat{\rho}^{(2)})$ является субэкстенсивным (суперэкстенсивным) функционалом при $q > 1$ ($q < 1$) и экстенсивным функционалом только в пределе слабой связи двух подсистем, когда $q \rightarrow 1$.

Вместе с тем, эскортное осреднение приводит к свойству аддитивности для осреднённой энергии совокупной квантовой системы

$$E_q^{(1,2)} = E_q^{(1)} + E_q^{(2)}, \quad \text{где} \quad E_q \equiv \langle \hat{H} \rangle_q = \text{Tr}(\hat{P}_q \hat{H}).$$

Здесь $\hat{H}(\mathbf{x}, \mathbf{x}') = \left\{ -\frac{\hbar^2}{2m} \sum_{j=1}^n \frac{\partial^2}{\partial \mathbf{x}_j^2} + \frac{1}{2} \sum_{j \neq k}^N \Phi(|\mathbf{x}_j - \mathbf{x}_k|) \right\} \prod_{i=1}^n \delta(\mathbf{x}_i - \mathbf{x}'_i)$ – оператор энергии

системы из N одинаковых бесспиновых частиц массы m , взаимодействующих между собой с потенциалом $\Phi(|\mathbf{x}|)$.

2.2. Экстремальность большого канонического распределения

Прежде всего, отметим, что различные статистические ансамбли квантовых систем (подобно традиционным) эквивалентны в термодинамическом отношении, что связано, в частности, с малостью флуктуаций энергии, числа частиц и объёма [17]. Далее мы воспользуемся наиболее удобным для наших целей большим каноническим ансамблем квантовых систем, описывающим контакт с термостатом и резервуаром частиц и определяемым заданием средней энергии и среднего числа частиц.

Рассматривая ансамбль систем с постоянным объёмом (находящийся в тепловом и материальном контакте с окружением), определим матрицу равновесной плотности $\hat{\rho}(\mathbf{x})$ (т.е. статистический оператор равновесного распределения) из условия абсолютного экстремума квантовой информационной энтропии Тсаллиса (2), найденного при заданности осреднённых операторов плотности энергии $\hat{H}(\mathbf{x})$ и полного числа частиц $\hat{N}(\mathbf{x})$:

$$\langle \hat{H} \rangle_q \equiv E_q = \text{Tr}(\hat{P}_q \hat{H}) = \text{const}, \quad \langle \hat{N} \rangle_q \equiv N_q = \text{Tr}(\hat{P}_q \hat{N}) = \text{const}, \quad (7)$$

а также при сохранении нормировки (1).

Согласно вариационному принципу Джейнса [20], равновесная матрица плотности $\hat{\rho}$, «экстремизирующая» энтропию S_q при указанных ограничениях, определяется из условия равенства нулю первой вариации по $\hat{\rho}$ следующего обобщённого Лагранжиана

$$L(\hat{\rho}) \equiv -\text{Tr}(\hat{\rho} \text{Ln}_q \hat{\rho}) - \beta \frac{\text{Tr}(\hat{\rho}^q \hat{H})}{c_q} + \beta \mu \frac{\text{Tr}(\hat{\rho}^q \hat{N})}{c_q} - \lambda \text{Tr} \hat{\rho}. \quad (8)$$

Здесь

$$c_q \equiv \text{Tr}(\hat{\rho}^q) \quad (9)$$

– так называемый коэффициент Тсаллиса; β , $(\beta\mu)$ и λ – определяемые из уравнений (1) и (7) лагранжевы множители, которые связаны с ограничением на осреднённые операторы плотности энергии и полного числа частиц квантовой системы в неаддитивной статистике Тсаллиса; при этом величина параметра μ имеет смысл химического потенциала квантовых частиц.

Определяя абсолютный экстремум функционала (8) из условия $\delta L(\hat{\rho}) / \delta \hat{\rho} = 0$, находим для неэкстенсивных квантовых систем следующее представление большого канонического распределения оператора плотности $\hat{\rho}(\mathbf{x}, \beta_q)$:

$$\begin{aligned} \hat{\rho}(\mathbf{x}, \beta_q) &= \frac{1}{\tilde{Q}_q(\beta)} \left\{ 1 - (1-q)\beta_q \left[(\hat{H}(\mathbf{x}) - \tilde{E}_q) - \mu(\hat{N}(\mathbf{x}) - \tilde{N}_q) \right] \right\}^{1/(1-q)} = \\ &= \tilde{Q}_q^{-1} \exp_q \left\{ -\beta_q \left[(\hat{H}(\mathbf{x}) - \tilde{E}_q) - \mu(\hat{N}(\mathbf{x}) - \tilde{N}_q) \right] \right\}, \end{aligned} \quad (10)$$

где

$$\tilde{Q}_q(\beta_q) = \text{Tr} \left\{ \exp_q \left[-\beta_q \left((\hat{H} - \tilde{E}_q) - \mu(\hat{N} - \tilde{N}_q) \right) \right] \right\} \quad (11)$$

– статистическая сумма состояний для большого квантового ансамбля, определяемая из условия нормировки (1); параметр $\beta_q \equiv \beta / \tilde{c}_q$ называется обратной *физической температурой* равновесной квантовой системы, $T_{pf} \equiv 1 / \beta_q$;

$$\tilde{c}_q \equiv \text{Tr}(\hat{\rho}^q) = \text{Tr} \left\{ \exp_q \left[-\beta_q \left((\hat{H} - \tilde{E}_q) - \mu(\hat{N} - \tilde{N}_q) \right) \right] \right\}^q / (\tilde{Q}_q)^q \quad (12)$$

– значение коэффициента Тсаллиса в равновесном случае; знак тильды « $\tilde{}$ » здесь и далее

над осреднёнными динамическими переменными \hat{A} означает, что осреднение проведено с помощью равновесного распределения (10).

2.3. Экспонента Тсаллиса и деформированный логарифм

В формуле (10) использована так называемая деформированная экспонента Тсаллиса [3]

$$\exp_q \hat{A} = \begin{cases} \left[1 + (1-q)\hat{A} \right]^{1/(1-q)}, & \text{если } \text{Spec} \left[1 + (1-q)\hat{A} \right] \geq 0; \\ 0, & \text{в других случаях,} \end{cases} \quad (13)$$

где неравенство $\text{Spec} \left[1 + (1-q)\hat{A} \right] \geq 0$ означает, что существует естественное «отключение», когда спектр оператора в квадратных скобках имеет отрицательные значения, связанные с действительностью следа.

Легко проверить, что в пределе $q \rightarrow 1$ функция (13) принимает стандартный вид:

$$\exp_1 \hat{A} \equiv \lim_{q \rightarrow 1+0} \exp_q \hat{A} = \lim_{q \rightarrow 1-0} \exp_q \hat{A} = \exp \hat{A}, \quad (\forall x). \quad (14)$$

Используя определения (3) и (13), можно убедиться, что имеют место следующие соотношения для деформированной экспоненты [3]:

$$\exp_q (\ln_q \hat{A}) = \ln_q (\exp_q \hat{A}) = \hat{A}, \quad (\forall x; \forall q), \quad (15)$$

$$(\exp_q \hat{A})(\exp_q \hat{B}) = \exp_q \left[\hat{A} + \hat{B} + (1-q)\hat{A}\hat{B} \right], \quad (16)$$

$$d(\exp_q \hat{A}) / d\hat{A} = (\exp_q \hat{A})^q. \quad (17)$$

Соответственно для деформированного логарифма $\ln_q \hat{A}$ имеем:

$$\ln_q (\hat{A}\hat{B}) = \ln_q \hat{A} + \ln_q \hat{B} + (1-q)(\ln_q \hat{A})(\ln_q \hat{B}), \quad (18)$$

$$\ln_q \hat{A}^{-1} = -\hat{A}^{q-1} \ln_q \hat{A}, \quad \ln_q (\hat{B}\hat{A}^{-1}) = \hat{A}^{q-1}(\ln_q \hat{B} - \ln_q \hat{A}), \quad (19)$$

$$d \ln_q \hat{A} / d\hat{A} = 1 / \hat{A}^q. \quad (20)$$

Эти формулы будут использованы далее.

2.4. Некоторые свойства равновесного распределения

Из выражения (10) следует соотношение

$$\left(\hat{\rho} \tilde{Q}_q \right)^{1-q} = 1 - (1-q)\beta_q \left[\Delta_q \hat{H} - \mu \Delta_q \hat{N} \right]. \quad (21)$$

Если умножить (21) на $\hat{\rho}^q$ и затем взять шпур, то получим равенство

$$(\tilde{Q}_q)^{1-q} \text{Tr}(\hat{\rho}) = \text{Tr} \left\{ \hat{\rho}^q \left[1 - (1-q)\beta_q \left(\Delta_q \hat{H} - \mu \Delta_q \hat{N} \right) \right] \right\} = \text{Tr}(\hat{\rho}^q), \quad (22)$$

из которого, с использованием (1) и (7), можно найти важное соотношение для равновесного значения коэффициента Тсаллиса

$$\tilde{c}_q \equiv \text{Tr}(\hat{\rho}^q) = (\tilde{Q}_q)^{1-q} = 1 + (1-q)\tilde{S}_q. \quad (23)$$

Используя (23) и вытекающее из формулы (10) выражение

$$\tilde{c}_q \equiv \text{Tr}(\hat{\rho}^q) = (\tilde{Q}_q)^{-q} \text{Tr} \left\{ \exp_q \left[-\beta_q \left(\Delta_q \hat{H} - \mu \Delta_q \hat{N} \right) \right] \right\}^q, \quad (24)$$

получим ещё одно представление обобщённой статистической суммы

$$\tilde{Q}_q(\beta, \mu) = \text{Tr} \left\{ \exp_q \left[-\beta_q \left(\Delta_q \hat{H} - \mu \Delta_q \hat{N} \right) \right] \right\}^q. \quad (25)$$

Далее везде квантово-механическую флуктуацию $\Delta_q \hat{A}$ наблюдаемой величины \hat{A} будем задавать соотношением

$$\Delta_q \hat{A} \equiv \hat{A} - \tilde{A} = \hat{A} - \text{Tr}(\hat{P}_q \hat{A}), \quad (26)$$

где «равновесное» эскортное распределение \hat{P}_q определяется, как легко можно убедиться, формулой

$$\hat{P}_q(\mathbf{x}, \beta_q, \mu) = \left\{ \exp_q \left[-\beta_q \left(\Delta_q \hat{H} - \mu \Delta_q \hat{N} \right) \right] \right\}^q / \tilde{Q}_q(\beta_q, \mu). \quad (27)$$

Здесь

$$\tilde{Q}_q(\beta_q, \mu) \equiv \text{Tr} \left\{ \exp_q \left[-\beta_q \left(\Delta_q \hat{H} - \mu \Delta_q \hat{N} \right) \right] \right\} = \text{Tr} \left\{ \exp_q \left[-\beta_q \left(\Delta_q \hat{H} - \mu \Delta_q \hat{N} \right) \right] \right\}^q.$$

Наконец, при использовании распределения (27) и формулы (25) можно получить следующую форму записи для среднего значения $\tilde{A}_q \equiv \text{Tr}(\hat{P}_q \hat{A})$ любой наблюдаемой \hat{A} в равновесной квантовой системе

$$\begin{aligned} \tilde{A}_q &= \tilde{Q}_q^{-1} \text{Tr} \left\{ \left[\exp_q \left(-\beta_q \Delta_q \hat{H} + (\mu \beta_q) \Delta_q \hat{N} \right) \right]^q \hat{A} \right\} = \\ &= \text{Tr} \left\{ \hat{A} \left[\exp_q \left(-\beta_q \Delta_q \hat{H} + (\mu \beta_q) \Delta_q \hat{N} \right) \right]^q \right\} / \text{Tr} \left[\exp_q \left(-\beta_q \Delta_q \hat{H} + (\mu \beta_q) \Delta_q \hat{N} \right) \right]^q. \end{aligned} \quad (28)$$

2.5. Экстремальность большого канонического распределения

Заметим, что экстремальные свойства любых квантовых неэкстенсивных ансамблей можно получить из следующего неравенства:

$$\text{Tr}(\hat{\sigma} \text{Ln}_q \hat{\sigma}) - \text{Tr} \left[\hat{\sigma} \left(\frac{\hat{\sigma}}{\hat{\rho}} \right)^{q-1} \text{Ln}_q \hat{\rho} \right] \begin{cases} \geq 0, & \text{если } q > 0; \\ \leq 0, & \text{если } q < 0, \end{cases} \quad (29)$$

где $\hat{\rho}$ и $\hat{\sigma}$ – произвольные статистические операторы.

Действительно, поскольку для числа $r > 0$ имеем (см. теорему № 42 в монографии [19])

$$\begin{aligned} \text{Ln}_q r &= \frac{r^{q-1} - 1}{q-1} \geq 1 - \frac{1}{r}, & \text{если } q > 0, \\ &= 1 - \frac{1}{r}, & \text{если } q = 0, \\ &\leq 1 - \frac{1}{r}, & \text{если } q < 0, \end{aligned} \quad (30)$$

то, подставляя в (30) $r \equiv \hat{\sigma} \hat{\rho}^{-1}$ (где $\hat{\rho}$ и $\hat{\sigma}$ – положительно-определённые операторы) и усредняя полученное выражение по распределению $\hat{\sigma}$, получим, в случае когда $q > 0$, неравенство

$$\text{Tr} \left\{ \hat{\sigma} \text{Ln}_q \left(\frac{\hat{\sigma}}{\hat{\rho}} \right) \right\} \geq \text{Tr} \left\{ \hat{\sigma} \left(1 - \frac{\hat{\rho}}{\hat{\sigma}} \right) \right\} = 1 - 1 = 0, \quad (31)$$

Поскольку оба оператора $\hat{\rho}$ и $\hat{\sigma}$ нормированы на единицу и операторы под знаком шпура перестановочны. Воспользовавшись формулой (19), будем иметь

$$\text{Ln}_q \left(\frac{\hat{\sigma}}{\hat{\rho}} \right) = \left(\frac{\hat{\sigma}}{\hat{\rho}} \right)^{q-1} \text{Ln}_q \left(\frac{\hat{\sigma}}{\hat{\rho}} \right) = \hat{\sigma}^{q-1} (\text{Ln}_q \hat{\sigma} - \text{Ln}_q \hat{\rho}) = \text{Ln}_q \hat{\sigma} - \left(\frac{\hat{\sigma}}{\hat{\rho}} \right)^{q-1} \text{Ln}_q \hat{\rho}. \quad (32)$$

Подставляя теперь это соотношение в неравенство (31), получим «верхнее» неравенство (29). Используя аналогичный подход, легко убедиться в том, что при $q < 0$ имеет место «нижнее» неравенство (29).

Докажем теперь, что в случае квантовой неэкстенсивной статистики экстремум функционала (8), т.е. большое каноническое распределение $\hat{\rho}$, соответствует максимуму (минимуму) квантовой энтропии Тсаллиса $S_q(\hat{\rho}) = -\text{Tr}(\hat{\rho} \text{Ln}_q \hat{\rho})$ соответственно при $q > 0$ ($q < 0$) среди всех вероятностных распределений с одинаковыми средней энергией и средним числом частиц.

Пусть $\hat{\rho}$ – большое каноническое распределение, а $\hat{\sigma}$ – нормированный статистический оператор, соответствующий той же средней энергии и среднему числу

частиц, как и (10), а в остальном – произвольный. При подстановке (10) в неравенство (29), получим

$$\tilde{S}_q(\hat{\sigma}) \equiv -\text{Tr}(\hat{\sigma} \text{Ln}_q \hat{\sigma}) \leq -\text{Tr} \left[\hat{\sigma} \left(\frac{\hat{\sigma}}{\hat{\rho}} \right)^{q-1} \text{Ln}_q \hat{\rho} \right] \leq \tilde{S}_q(\hat{\rho}). \quad (33)$$

Действительно, поскольку $\text{Ln}_q \hat{A} = \hat{A}^{q-1} \text{Ln}_q \hat{A}$, то, с учётом (15) и (19), будем иметь

$$\begin{aligned} \text{Ln}_q \hat{\rho} &= \text{Ln}_q \left\{ \tilde{Q}_q \exp_q \left[-\beta_q \left((\Delta_q \hat{H} - \mu \Delta_q \hat{N}) \right) \right] \right\} = \\ &= \left\{ \exp_q \left[-\beta_q \left((\Delta_q \hat{H} - \mu \Delta_q \hat{N}) \right) \right] \right\}^{q-1} \left[-\beta_q \left((\Delta_q \hat{H} - \mu \Delta_q \hat{N}) \right) - \tilde{S}_q(\hat{\rho}) \right], \end{aligned} \quad (34)$$

откуда следует

$$-\text{Tr} \left\{ \hat{\sigma} \left(\frac{\hat{\sigma}}{\hat{\rho}} \right)^{q-1} \text{Ln}_q \hat{\rho} \right\} = -\text{Tr} \left\{ \hat{\sigma}^q \left[-\beta_q \left((\Delta_q \hat{H} - \mu \Delta_q \hat{N}) \right) - \tilde{S}_q(\hat{\rho}) \right] \right\} \tilde{Q}_q^{q-1} = \tilde{S}_q(\hat{\rho}). \quad (35)$$

Таким образом, большое каноническое распределение (10) соответствует максимуму квантовой информационной энтропии Тсаллиса при $q > 0$ среди всех вероятностных распределений с одинаковыми средней энергией и среднем числом частиц. Используя аналогичный подход, можно легко убедиться в том, что при подстановке (10) в неравенство (29) получим, что распределение (10) соответствует минимуму энтропии Тсаллиса при $q < 0$.

2.6. Равновесные термодинамические соотношения

Подставляя распределение (10) в (2) и используя (23), получим следующее выражение для равновесной энтропии Тсаллиса для большого канонического ансамбля квантовых систем:

$$\tilde{S}_q = -\text{Tr}(\hat{\rho} \text{Ln}_q \hat{\rho}) = -\frac{1 - \tilde{c}_q}{1 - q} = \frac{(\tilde{Q}_q)^{1-q} - 1}{1 - q} = \ln_q \tilde{Q}_q. \quad (36)$$

Определим теперь деформированный большой термодинамический потенциал в квантовой статистике Тсаллиса следующим соотношением:

$$\tilde{\Omega}_q \equiv \tilde{E}_q - \frac{1}{\beta} \tilde{S}_q - \mu \hat{N}_q = \tilde{E}_q - \frac{1}{\beta} \frac{(\tilde{Q}_q)^{1-q} - 1}{1 - q} - \mu \hat{N}_q. \quad (37)$$

При дифференцировании энтропии \tilde{S}_q по \tilde{E}_q в результате будем иметь

$$\partial \tilde{S}_q / \partial \tilde{E}_q = \partial \ln_q \tilde{Q}_q / \partial \tilde{E}_q = \beta. \quad (38)$$

Действительно, используя формулы (17) и (20), а также соотношения (11) и (25), получим

$$\begin{aligned} \frac{\partial \tilde{S}_q}{\partial \tilde{E}_q} &= \frac{\partial \ln_q \tilde{Q}_q}{\partial \tilde{Q}_q} \frac{\partial \tilde{Q}_q}{\partial \tilde{E}_q} = \tilde{Q}_q^{-q} \frac{\partial \tilde{Q}_q}{\partial \tilde{E}_q} = \tilde{Q}_q^{-q} \text{Tr} \frac{\partial \exp_q \{..\}}{\partial \{..\}} \frac{\partial \{..\}}{\partial \tilde{E}_q} = \\ &= \tilde{Q}_q^{-q} \text{Tr} \left\{ \exp_q \{..\} \right\}^q \frac{\partial \{..\}}{\partial \tilde{E}_q} = \tilde{Q}_q^{1-q} \frac{\beta}{\tilde{c}_q} = \beta. \end{aligned} \quad (39)$$

Здесь введено обозначение $\{..\} \equiv -\beta_q (\Delta_q \hat{H} - \mu \Delta_q \hat{N})$. Аналогично получим

$$\partial \tilde{S}_q / \partial \hat{N}_q = -\beta \mu_\alpha, \quad (\alpha = 1, \dots, M). \quad (40)$$

В результате будем иметь следующие соотношения равновесной термодинамики квантовых неэкстенсивных систем:

$$\tilde{S}_q = \ln_q \tilde{Q}_q, \quad \tilde{\Omega}_q = \tilde{E}_q - \beta^{-1} \tilde{S}_q - \mu \hat{N}_q, \quad (41)$$

$$\partial \tilde{S}_q / \partial \tilde{E}_q = \partial (\ln_q \tilde{Q}_q) / \partial \tilde{E}_q = \beta, \quad \partial \tilde{S}_q / \partial \hat{N}_q = -\beta \mu, \quad (42)$$

$$\frac{\partial}{\partial \beta} (\beta \tilde{\Omega}_q) = \tilde{E}_q, \quad \frac{\partial}{\partial (\beta \mu)} (\beta \tilde{\Omega}_q) = -\hat{N}_q. \quad (43)$$

По поводу соотношений (41)-(43) следует иметь в виду следующее. Величина \tilde{Q}_q характеризуется микроскопическими значениями энергии $\hat{H}(\mathbf{x})$ и числа частиц \hat{N}_q , вычисленными относительно средней энергии \tilde{E}_q и среднего числа частиц \tilde{N}_q соответственно. Если ввести новую статистическую сумму Q_q^* , которая определяется микроскопическими величинами $\hat{H}(\mathbf{x})$ и \hat{N}_q относительно нулевой точки, согласно формуле

$$(Q_q^*)^{1-q} \equiv (\tilde{Q}_q)^{1-q} - \beta(1-q) (\tilde{E}_q - \mu \tilde{N}_q) \quad (44)$$

и переопределить квантовый большой термодинамический потенциал формулой

$$\tilde{\Omega}_q^* \equiv -\beta^{-1} \ln_q Q_q^*, \quad (45)$$

то соотношения (41)-(43) равновесной квантовой термодинамики принимают почти классическую форму

$$\tilde{S}_q = \beta (\tilde{E}_q - \mu \tilde{N}_q - \tilde{\Omega}_q^*), \quad (46)$$

$$\frac{\partial}{\partial \beta} (\ln_q Q_q^*) = -\frac{\partial}{\partial \beta} (\beta \tilde{\Omega}_q^*) = -\tilde{E}_q, \quad (47)$$

$$\frac{\partial}{\partial (\beta \mu)} (\ln_q Q_q^*) = -\frac{\partial}{\partial (\beta \mu)} (\beta \tilde{\Omega}_q^*) = \tilde{N}_q, \quad (48)$$

$$\partial \tilde{S}_q / \partial \tilde{E}_q = \beta, \quad \partial \tilde{S}_q / \partial \tilde{N}_q = -\beta \mu. \quad (49)$$

Следует заметить, что введённые выше термодинамические соотношения обусловлены стандартной процедурой нахождения равновесного распределения матрицы плотности на основе принципа Джейнса экстремума квантовой параметрической энтропии Тсаллиса и потому применимы для весьма широкого круга квантовых аномальных явлений, описываемых неаддитивной статистической механикой.

3. МОДИФИЦИРОВАННАЯ КВАНТОВАЯ ТЕРМОДИНАМИКА НЕЭКСТЕНСИВНЫХ СИСТЕМ

Вместе с тем, принимая во внимание тот факт, что структура основных соотношений статистической термодинамики Гиббса существенно зависит от аддитивности классической энтропии, крайне важно выяснить, каким образом, в случае использования физической температуры T_{ph} могут быть модифицированы аналогичные соотношения (40)-(43) в неаддитивной статистике Тсаллиса.

3.1. Термодинамическое равновесие двух независимых систем

Рассмотрим для этого термодинамическое равновесие двух независимых неэкстенсивных квантовых систем с энтропиями Тсаллиса $S_q^{(1)} \equiv S_q(\hat{\rho}^{(1)})$ и $S_q^{(2)} \equiv S_q(\hat{\rho}^{(2)})$, представляющих собой общую замкнутую систему с постоянными значениями совокупной энтропии $S_q^{(1,2)} \equiv S_q(\hat{\rho}^{(1,2)}) = const$ и совокупной энергии системы $E_q^{(1,2)} = E_q^{(1)} + E_q^{(2)} = const$.

Согласно свойству (6) неаддитивности q -энтропии Тсаллиса, совокупную энтропию квантовой системы можно записать в следующем виде

$$S_q^{(1,2)} = S_q^{(1)} [1 + \varepsilon S_q^{(2)}] + S_q^{(2)} [1 + \varepsilon S_q^{(1)}] - \varepsilon S_q^{(1)} S_q^{(2)}. \quad (50)$$

Здесь и далее $\varepsilon \equiv (1 - q)$. Варьирование $\delta S_q^{(1,2)}$ и $\delta E_q^{(1,2)}$ для полной замкнутой системы с постоянными значениями энтропии $S_q^{(1,2)}$ и энергии $E_q^{(1,2)}$ приводит к равенству $\delta S_q^{(1,2)} = 0 = \delta S_q^{(1)} [1 + \varepsilon S_q^{(2)}] + \delta S_q^{(2)} [1 + \varepsilon S_q^{(1)}]$ для энтропии и равенству $\delta E_q^{(1,2)} = 0 = \delta E_q^{(1)} + \delta E_q^{(2)}$ для средней энергии. Объединяя их, в итоге получим уравнение

$$\frac{\delta S_q^{(1)} / \delta E_q^{(1)}}{1 + \varepsilon S_q^{(1)}} = \frac{\delta S_q^{(2)} / \delta E_q^{(2)}}{1 + \varepsilon S_q^{(2)}}, \quad (51)$$

или, с учётом (37) и (49),

$$\frac{\beta}{1 + (1-q)S_q^{(1)}} = \frac{\beta}{1 + (1-q)S_q^{(2)}} = \frac{\beta}{c_q} \equiv \beta_q. \quad (52)$$

Отношение эквивалентности (52) определяет условие теплового равновесия двух квантовых q -систем и является обобщением нулевого закона термодинамики на неаддитивные квантово-механические системы. Оно показывает, что в отличие от классического квантового случая ($q \rightarrow 1$) физическая температура T_{ph} не является обратной величиной множителя Лагранжа, β^{-1} , но

$$T_{ph} \equiv \frac{1}{\beta_q} = \frac{c_q}{\beta} = \left(1 + \frac{1-q}{k_B} S_q \right) T = c_q T. \quad (53)$$

Очевидно, что квантовая физическая температура T_{ph} , отличная от инверсии множителя Лагранжа β , отвечает за «глобальный» энергетический (тепловой) баланс между различными частями неаддитивной квантовой системы, который сильно отличается от локального теплового баланса. Локальный баланс можно охарактеризовать общей абсолютной температурой $T = 1/\beta$, измеряемой термометром, но любое измерение физической температуры T_{ph} связано с необходимостью вычисления коэффициента Тсаллиса \tilde{c}_q , зависящего от параметра неаддитивности q .

Таким образом, отличие физической температуры от инверсии множителя Лагранжа β приводит, вообще говоря, к необходимости видоизменения полученных выше термодинамических соотношений (40)-(43) для неаддитивных квантовых систем. В работе [21] в качестве основных предпосылок, взятых за исходный пункт построения модифицированной макроскопической термодинамики Тсаллиса с физической температурой, предлагается использование первого закона термодинамики и преобразования Лежандра.

3.2. Модифицированные термодинамические равенства для квантово - механического ансамбля

Введём по аналогии с физической температурой квантовой системы T_{ph} квантовое физическое давление P_q путём рассмотрения механического равновесия двух независимых q -систем. В этом случае энтропия совокупной квантовой системы должна максимизироваться с фиксацией общего объёма $V^{(1,2)} = V^{(1)} + V^{(2)} = const$. В результате будем иметь

$$\frac{\delta S_q^{(1)} / \delta V^{(1)}}{1 + \varepsilon S_q^{(1)}} = \frac{\delta S_q^{(2)} / \delta V^{(2)}}{1 + \varepsilon S_q^{(2)}} = \frac{P_{ph}}{T_{ph}}, \quad (54)$$

где P_{ph} – квантовое физическое давление, которое определяется соотношением

$$P_{ph} \equiv \frac{T_{ph}}{1 + (1-q)S_q} \frac{\delta S_q}{\delta V} = \frac{T_{ph}}{c_q} \frac{\delta S_q}{\delta V}. \quad (55)$$

Ясно, что определённые таким образом квантовые физическая температура и физическое давление обязательно должны привести к модификации определения термодинамической энтропии Клаузиуса [23].

Уравнение (38) ($\beta = \partial S_q / \partial E_q$) указывает на то, что параметры β и $E_q = E_q(S_q, P_{ph}, N)$ образуют пару переменных Лежандра. Это приводит к следующему определению большого термодинамического потенциала:

$$\Omega'_q(T, V, \mu) \equiv E_q - TS_q - \mu N_q = E_q - T \ln_q \left[c_q^{1/(1-q)} \right] - \mu N_q$$

(в этом разделе знак «тильды» будем опускать). Это выражение, однако, является неудовлетворительным, поскольку оно не написано с точки зрения физической температуры: термодинамический потенциал должен быть функцией T_{ph} , а не зависеть от абсолютной температуры T .

По аналогии с работой [22], которой мы далее воспользуемся, переопределим большой термодинамический потенциал (38) следующим образом:

$$\Omega_q(T_{ph}, V, \mu_\alpha) = E_q - T_{ph} \ln \left[c_q^{1/(1-q)} \right] - \mu N_q. \quad (56)$$

При учёте соотношений (23), (43) и (53) можно убедиться в том, что величина Ω_q на самом деле является функцией T_{ph} . Варьируя функцию Ω_q , в результате получим

$$\delta \Omega_q = \delta E_q - \frac{\ln c_q}{(1-q)} \delta T_{ph} - \frac{T_{ph}}{c_q} \delta S_q - \mu \delta N_q - N_q \delta \mu. \quad (57)$$

Если теперь использовать первый закон термодинамики в формулировке Каратеодори [23]

$$\delta' Q_q = \delta E_q + \delta' W_q = \delta E_q + P_{ph} \delta V - \mu \delta N_q, \quad (58)$$

где $\delta' Q_q$ и $\delta' W_q$ – малые изменения количества теплоты (так называемой некомпенсированной теплоты), подводимой к q -системе или отводимой от неё, и работы, которые определяются выражениями [24-26]

$$\delta'Q_q \equiv \frac{\text{Tr}[\delta\hat{\rho}^q(\hat{H} - E_q)]}{\text{Tr}\hat{\rho}^q}, \quad \delta'W_q \equiv -\langle\delta\hat{H}\rangle_q = -\frac{\text{Tr}(\hat{\rho}^q\delta\hat{H})}{\text{Tr}\hat{\rho}^q}, \quad (59)$$

то выражение (57) можно переписать в виде

$$\delta\Omega_q = \delta'Q_q - P_{ph}d\delta - \frac{\ln c_q}{(1-q)}\delta T_{ph} - \frac{T_{ph}}{c_q}\delta S_q - N_q\delta\mu. \quad (60)$$

Отсюда следует, что определение термодинамической энтропии Клаузиуса модифицируется для неаддитивных квантовых систем следующим образом:

$$\delta S_q = c_q\delta'Q_q / T_{ph}. \quad (61)$$

Введём теперь в рассмотрение следующие модифицированные характеристические функции квантовой системы: обобщённую энтальпию $H_q(S_q, P_{ph}, N_q) = E_q + P_{ph}V$, свободную энергию Гельмгольца $F_q(T, V, N_q) = E_q - T_{ph}\left[\ln c_q^{1/(1-q)}\right]$ и свободную энергию Гиббса $G_q(T, P_{ph}, N_q) = F_q + P_{ph}V$. Заметим, что все характеристические функции обладают следующим свойством: если известна характеристическая функция, выраженная через соответствующие (свои для каждой характеристической функции) переменные, то из неё можно вычислить любую термодинамическую величину.

В этом нетрудно убедиться. В частности, из уравнений

$$\delta E_q = \frac{T_{ph}}{c_q}\delta S_q - P_{ph}\delta V + \mu\delta N_q, \quad (62)$$

$$\delta H_q = \frac{T_{ph}}{c_q}\delta S_q + V\delta P_{ph} + \mu\delta N_q, \quad (63)$$

$$\delta F_q = -\left[\frac{\ln c_q}{(1-q)}\right]\delta T_{ph} - P_{ph}\delta V + \mu\delta N_q, \quad (64)$$

$$\delta G_q = -\left[\frac{\ln c_q}{1-q}\right]\delta T_{ph} + V\delta P_{ph} + \mu\delta N_q, \quad (65)$$

$$\delta\Omega_q = -\left[\frac{\ln c_q}{(1-q)}\right]\delta T_{ph} - P_{ph}\delta V - N_q\delta\mu \quad (66)$$

следуют обобщённые термодинамические соотношения

$$\left(\frac{\partial E_q}{\partial V}\right)_{S_q, N_q} = \left(\frac{\partial F_q}{\partial V}\right)_{T_{ph}, N_q} = -P_{ph}, \quad (67)$$

$$\left(\frac{\partial E_q}{\partial S_q}\right)_{V, N_q} = \left(\frac{\partial H_q}{\partial S_q}\right)_{P_{ph}, N_q} = T_{ph} / c_q, \quad (68)$$

$$\left(\frac{\partial H_q}{\partial P_{ph}}\right)_{S_q, N_q} = \left(\frac{\partial G_q}{\partial P_{ph}}\right)_{T_{ph}, N_q} = V, \quad (69)$$

$$\left(\frac{\partial F_q}{\partial T_{ph}}\right)_{V, N_q} = \left(\frac{\partial G_q}{\partial T_{ph}}\right)_{P_{ph}, N_q} = -\ln c_q / (1-q). \quad (70)$$

Как известно, в термодинамике теплоёмкость вещества в наиболее общем виде определяется следующим образом: $c_z = T \left(\frac{\partial S}{\partial T} \right)_z$. Здесь c_z – теплоёмкость в таком процессе, в котором сохраняется постоянным параметр z , где z – любые обобщённые координаты. Наиболее распространёнными являются изобарная теплоёмкость и изохорная теплоёмкость:

$$c_{qp} = T_{ph} c_q^{-1} \left(\frac{\partial S_q}{\partial T_{ph}} \right)_{P_{ph}}, \quad c_{qV} = T_{ph} c_q^{-1} \left(\frac{\partial S_q}{\partial T_{ph}} \right)_V. \quad (71)$$

Так как в соответствии с формулой $\left(\frac{\partial y}{\partial x} \right)_z = \left(\frac{\partial y}{\partial u} \right)_z \left(\frac{\partial u}{\partial x} \right)_z$ (справедливой для случая двух переменных, когда $y = y(x, z)$ и $u = u(x, z)$) имеем

$$\left(\frac{\partial S_q}{\partial T_{ph}} \right)_{P_{ph}} = \left(\frac{\partial S_q}{\partial H_q} \right)_{P_{ph}} \left(\frac{\partial H_q}{\partial T_{ph}} \right)_{P_{ph}} \quad \text{и} \quad \left(\frac{\partial S_q}{\partial T_{ph}} \right)_V = \left(\frac{\partial S_q}{\partial E_q} \right)_V \left(\frac{\partial E_q}{\partial T_{ph}} \right)_V, \quad (72)$$

а из (68) и (70) следует, что $\left(\frac{\partial S_q}{\partial H_q} \right)_{P_{ph}} = c_q / T_{ph}$, $\left(\frac{\partial S_q}{\partial E_q} \right)_V = c_q / T_{ph}$, то соотношения (71) можно записать в виде

$$c_{qp} = \left(\frac{\partial H_q}{\partial T_{ph}} \right)_{P_{ph}}, \quad c_{qV} = \left(\frac{\partial E_q}{\partial T_{ph}} \right)_V. \quad (73)$$

Наконец, уравнение, устанавливающее связь между теплоёмкостями c_p и c_V , может быть получено следующим образом. В соответствии с соотношением

$$\left(\frac{\partial z}{\partial t} \right)_n = \left(\frac{\partial z}{\partial x} \right)_y \left(\frac{\partial x}{\partial t} \right)_n + \left(\frac{\partial z}{\partial y} \right)_x \left(\frac{\partial y}{\partial t} \right)_n, \quad (74)$$

являющимся следствием выражения для полного дифференциала функции $z = z(x, y)$, можно записать (полагая $t = x$)

$$\left(\frac{\partial S_q}{\partial T_{ph}} \right)_{P_{ph}} = \left(\frac{\partial S_q}{\partial T_{ph}} \right)_V + \left(\frac{\partial S_q}{\partial V} \right)_{T_{ph}} \left(\frac{\partial V}{\partial T_{ph}} \right)_{P_{ph}}. \quad (75)$$

Отсюда, используя уравнение Максвелла $\left(\frac{\partial S_q}{\partial V} \right)_{T_{ph}} = \left(\frac{\partial P_{ph}}{\partial T_{ph}} \right)_V$ [23], получим

$$c_p - c_V = c_q^{-2} T_{ph} \left(\frac{\partial P_{ph}}{\partial T_{ph}} \right)_V \left(\frac{\partial V}{\partial T_{ph}} \right)_{P_{ph}}. \quad (76)$$

Это выражение можно представить в другом более удобном виде, если использовать связку трёх производных $(\partial z / \partial x)_y (\partial x / \partial y)_z (\partial y / \partial z)_x = -1$ (следствие соотношения (74) при $m = x$, $n = z$), из которой следует

$$\left(\partial P_{ph} / \partial T_{ph}\right)_V = -\left(\partial V / \partial T_{ph}\right)_{P_{ph}} \left(\partial P_{ph} / \partial V\right)_{T_{ph}}. \quad (77)$$

С учётом (77) связь между теплоёмкостями приобретает почти классический вид:

$$c_p - c_V = -c_q^{-2} T_{ph} \left(\partial V / \partial T_{ph}\right)_{P_{ph}}^2 / \left(\partial V / \partial P_{ph}\right)_{T_{ph}}. \quad (78)$$

Таким образом, стандартная форма для термодинамических соотношений (67)-(70) и теплоёмкостей (73) позволяет заключить, что они остаются инвариантными относительно неаддитивной модификации их классических аналогов. Подчеркнём тот важный факт, что температуры $T = 1/\beta$ и $T_{ph} = 1/\beta_q$ не зависят от выбора нуля энергий, и поэтому они допускают физическую интерпретацию. Заметим, что в дополнение к структуре Лежандра различные другие важные теоремы и свойства также остаются q -инвариантными [3].

4. КВАНТОВАЯ ОТНОСИТЕЛЬНАЯ ЭНТРОПИЯ В СТАТИСТИКЕ ТСАЛЛИСА. ОБОБЩЕННАЯ H -ТЕОРЕМА БОЛЬЦМАНА

Наряду с квантовой параметрической энтропией $S_q(\hat{\rho}) = (q-1)^{-1} \text{Tr}(\hat{\rho} - \hat{\rho}^q)$ обобщённая квантовая относительная энтропия (или квантовая информация различия Ратье–Каннаппана [25, 27, 28])

$$\begin{aligned} K_q(\hat{\rho} : \hat{\sigma}) &\equiv \left[1 - \text{Tr}(\hat{\rho}^q \hat{\sigma}^{1-q})\right] / (1-q) = \\ &= \text{Tr} \left[\hat{\rho} \ln_q(\hat{\sigma} \hat{\rho}^{-1}) \right] = \text{Tr} \left[\hat{\rho}^q (\ln_q \hat{\rho} - \ln_q \hat{\sigma}) \right] = -\text{Tr}(\hat{\rho}^q \ln_q \hat{\sigma}) - S_q(\hat{\rho}) \end{aligned} \quad (79)$$

также относится к наиболее существенным статистическим характеристикам квантово-механической q -системы. Являясь функционалом, она характеризует переход системы от состояния с матрицей плотности $\hat{\rho}$ в состояние с матрицей $\hat{\sigma}$, когда статистические наблюдения ведутся относительно состояния $\hat{\rho}$.

Можно показать, что в пределе слабой связи $q \rightarrow 1$ величина $K_q(\hat{\rho} : \hat{\sigma})$ переходит в традиционную квантовую относительную энтропию матрицы $\hat{\sigma}$ по отношению к матрице $\hat{\rho}$ (или в квантовую различающую информацию Кульбака–Лейблера [6])

$$\lim_{q \rightarrow 1} K_q(\hat{\rho} : \hat{\sigma}) = K_1(\hat{\rho} : \hat{\sigma}) = \text{Tr} \left[\hat{\rho} \left(\ln \frac{\hat{\rho}}{\hat{\sigma}} \right) \right] = -\text{Tr}[\hat{\rho} \ln \hat{\sigma}] - S_1(\hat{\rho}). \quad (80)$$

Действительно, в пределе $q \rightarrow 1$ имеем

$$\begin{aligned} \varepsilon^{-1} \left[1 - \text{Tr}(\hat{\rho}^q \hat{\sigma}^{1-q}) \right] &= \varepsilon^{-1} \text{Tr} \left\{ \hat{\rho} \left[1 - (\hat{\rho} / \hat{\sigma})^{q-1} \right] \right\} = \varepsilon^{-1} \text{Tr} \left\{ \hat{\rho} \left[1 - \exp \ln (\hat{\rho} / \hat{\sigma})^{q-1} \right] \right\} \\ &\square \varepsilon^{-1} \text{Tr} \left\{ \hat{\rho} \left[1 - 1 - \varepsilon \ln (\hat{\rho} / \hat{\sigma}) \right] \right\} = \text{Tr} \left\{ \hat{\rho} \left[\ln \left(\frac{\hat{\rho}}{\hat{\sigma}} \right) \right] \right\} = K_1 (\hat{\rho} : \hat{\sigma}). \end{aligned} \quad (81)$$

4.1. Выпуклость обобщённой квантовой относительной энтропии

Покажем, что энтропия $K_q (\hat{\rho} : \hat{\sigma})$ является вещественным, выпуклым и положительным (отрицательным) функционалом с минимумом (максимумом) при $q > 0$ ($q < 0$). Это можно сделать, используя квантовую относительную энтропию (79), преобразованную к виду

$$K_q (\hat{\rho} : \hat{\sigma}) = \text{Tr} (\hat{\rho} \text{Ln}_q \hat{\rho}) - \text{Tr} \left[\hat{\rho} (\hat{\rho} / \hat{\sigma})^{q-1} \text{Ln}_q \hat{\sigma} \right]$$

и неравенства (29), записанные в виде

$$\begin{aligned} \text{Tr} (\hat{\rho} \text{Ln}_q \hat{\rho}) - \text{Tr} \left[\hat{\rho} (\hat{\rho} / \hat{\sigma})^{q-1} \text{Ln}_q \hat{\sigma} \right] &\geq 0, \quad \text{если } q > 0; \\ &\leq 0, \quad \text{если } q < 0. \end{aligned}$$

В результате получим

$$K_q (\hat{\rho} : \hat{\sigma}) \geq 0, \quad q \geq 0; \quad K_q (\hat{\rho} : \hat{\sigma}) \leq 0, \quad q < 0. \quad (82)$$

Таким образом, обобщённая квантовая относительная энтропия неотрицательна, т.е. функционал $K_q (\hat{\sigma} : \hat{\rho})$ при $q > 0$ удовлетворяет такому же обобщённому неравенству Клейна $K_1 (\hat{\rho} : \hat{\sigma}) \geq 0$ (являющемуся квантовым аналогом неравенства Гиббса для информации различия Кульбака–Лейблера [7], как и квантовая относительная энтропия фон Неймана ($K_1 (\hat{\rho} : \hat{\sigma}) \geq 0$), а потому может использоваться для тех же целей. Однако в рассматриваемом случае имеется свобода выбора параметра деформации q , что позволяет адекватно исследовать неэкстенсивную квантовую систему.

При $\hat{\rho} \equiv \hat{\sigma}$ имеем равенство $K_q (\hat{\rho} : \hat{\rho}) = 0$. Таким образом, квантовая информация различия Ратье–Каннапана является знакоопределённой функцией Ляпунова.

4.2. Неаддитивность квантовой относительной энтропии для независимых систем

Пусть состояние физической квантовой системы описывается нормированными совместными распределениями операторов плотности $\hat{\rho}_{(1,2)} = \hat{\rho}_1 \otimes \hat{\rho}_2$ и $\hat{\sigma}_{(1,2)} = \hat{\sigma}_1 \otimes \hat{\sigma}_2$ при статистической независимости двух физических систем. Квантовые относительные энтропии для неэкстенсивных совокупной и отдельных систем определяются выражениями

$$K_q(\hat{\rho}_{(1,2)} : \hat{\sigma}_{(1,2)}) \equiv \varepsilon^{-1} \left[1 - \text{Tr}(\hat{\rho}_{(1,2)}^q \hat{\sigma}_{(1,2)}^{1-q}) \right], \quad (83)$$

$$K_q(\hat{\rho}_1 : \hat{\sigma}_1) \equiv \varepsilon^{-1} \left[1 - \text{Tr}(\hat{\rho}_1^q \hat{\sigma}_1^{1-q}) \right], \quad K_q(\hat{\rho}_2 : \hat{\sigma}_2) \equiv \varepsilon^{-1} \left[1 - \text{Tr}(\hat{\rho}_2^q \hat{\sigma}_2^{1-q}) \right], \quad (84)$$

где условия нормировки имеют вид

$$\text{Tr} \hat{\rho}_{(1,2)} = \text{Tr} \hat{\rho}_1 = \text{Tr} \hat{\rho}_2, \quad \text{Tr} \hat{\sigma}_{(1,2)} = \text{Tr} \hat{\sigma}_1 = \text{Tr} \hat{\sigma}_2. \quad (85)$$

Отсюда следует, что квантовая информация различия Ратье–Каннаппана обладает следующим свойством псевдоаддитивности для независимых систем

$$\begin{aligned} K_q(\hat{\rho}_{(1,2)} : \hat{\sigma}_{(1,2)}) &= \\ &= K_q(\hat{\rho}_1 : \hat{\sigma}_1) + K_q(\hat{\rho}_2 : \hat{\sigma}_2) + (q-1)K_q(\hat{\rho}_1 : \hat{\sigma}_1)K_q(\hat{\rho}_2 : \hat{\sigma}_2). \end{aligned} \quad (86)$$

При $q=1$ из (86) следует свойство аддитивности для квантовой информации различия Кульбака–Лейблера $K_1(\hat{\rho} : \hat{\sigma}) \equiv \text{Tr}[\hat{\rho} \ln(\hat{\rho} \hat{\sigma}^{-1})]$ в традиционной квантовой механике с мерой фон Неймана.

Пусть равновесная неаддитивная квантовая система находится в термостате с температурой $1/\beta$. Для определения равновесного оператора плотности $\hat{\sigma}$ находим безусловный экстремум лагранжиана

$$L(\hat{\sigma}) \equiv -\text{Tr}(\hat{\sigma} \text{Ln}_q \hat{\sigma}) - \beta c_q^{-1} \text{Tr}(\hat{\sigma}^q \hat{H}) - \lambda \text{Tr} \hat{\sigma}, \quad \text{где } c_q \equiv \text{Tr} \hat{\sigma}^q.$$

В результате получим следующее каноническое распределение для нормированной матрицы плотности в квантовой статистике Тсаллиса:

$$\hat{\sigma}(\mathbf{x}, \tilde{\beta}_q) = \frac{1}{\tilde{Q}_q(\tilde{\beta}_q)} \left\{ 1 - (1-q)\tilde{\beta}_q \left[(\hat{H}(\mathbf{x}) - \tilde{E}_q) \right] \right\}^{1/(1-q)}, \quad (87)$$

где

$$\tilde{E}_q = \text{Tr}[\hat{\sigma}^q \hat{H}(\mathbf{x})] / \text{Tr} \hat{\sigma}^q, \quad \tilde{\beta}_q = \beta / \tilde{c}_q, \quad \tilde{c}_q = \text{Tr} \hat{\sigma}^q, \quad (88)$$

$$\tilde{Q}_q(\tilde{\beta}_q) \equiv \text{Tr} \left\{ \exp_q \left[-\tilde{\beta}_q (\hat{H}(\mathbf{x}) - \tilde{E}_q) \right] \right\} \quad (89)$$

– статистическая сумма.

Используя каноническое распределение (87) для термостата с температурой $1/\beta$ можно найти соответствующие значения энтропии $\tilde{E}_q = \ln_q \tilde{Q}_q$, энергии $\tilde{E}_q = \text{Tr}(\hat{\sigma}^q \hat{H}) / \text{Tr} \hat{\sigma}^q$ и свободной энергии $\tilde{F}_q = \tilde{E}_q - \tilde{\beta}_q^{-1} \tilde{S}_q$ для квантовой системы,

находящейся в равновесном состоянии. Дифференцируя выражение $\tilde{S}_q = \ln_q \tilde{Q}_q$, можно получить полную систему термодинамических равенств для замкнутой системы, обобщающую классические квантовые соотношения на равновесные квантовые q -системы (ср. с формулами (41)-(43)):

$$d\tilde{S}_q = \beta d\tilde{E}_q, \quad d(\beta\tilde{F}_q) = \tilde{E}_q d\beta. \quad (90)$$

Рассмотрим теперь спонтанный переход между произвольным состоянием системы с матрицей плотности $\hat{\rho}$ и его равновесным состоянием с матрицей плотности $\hat{\sigma}$. Подставляя (87) в выражение $\tilde{c}_q K_q(\hat{\sigma} : \hat{\rho})$, в результате получим неравенство

$$\begin{aligned} \tilde{c}_q K_q(\hat{\rho} : \hat{\sigma}) &= \varepsilon^{-1} \left[\tilde{c}_q - \tilde{c}_q \operatorname{Tr}(\hat{\rho}^q \hat{\sigma}^{1-q}) \right] = \\ &= \varepsilon^{-1} \left\{ \tilde{c}_q - \operatorname{Tr} \left[\hat{\rho}^q \left(1 - (1-q)\tilde{\beta}_q (\hat{H}(\mathbf{x}) - \tilde{E}_q) \right) \right] \right\} = -\frac{1-\tilde{c}_q}{(1-q)} + \frac{1-c_q}{(1-q)} + c_q \tilde{\beta}_q (E_q - \tilde{E}_q) = \\ &= \tilde{S}_q - S_q + c_q \tilde{\beta}_q \left[(E_q - \tilde{E}_q) \right] = \tilde{S}_q - S_q + \tilde{\beta}_q \operatorname{Tr} \left[\hat{\rho}^q (\hat{H}(\mathbf{x}) - \tilde{E}_q) \right], \end{aligned} \quad (91)$$

совпадающее при $q=1$ с известным термодинамическим неравенством для информации различия Кульбака–Лейблера для аддитивных квантовых систем [6, 17].

4.3. Неравенство Клаузиуса

Варьируя оператор $\tilde{c}_q K_q(\hat{\rho} : \hat{\sigma})$ относительно матрицы плотности $\hat{\rho}$, т.е. считая, что $\hat{\rho} \rightarrow \hat{\rho} + \delta\hat{\rho}$ и $\operatorname{Tr}(\delta\hat{\rho}) = 0$, получим

$$\tilde{c}_q \delta K_q(\hat{\rho} : \hat{\sigma}) = -\delta S_q(\hat{\rho}) + \tilde{\beta}_q \operatorname{Tr} \left[\delta\hat{\rho}^q (\hat{H}(\mathbf{x}) - \tilde{E}_q) \right]. \quad (92)$$

Здесь использовано преобразование

$$S_q(\hat{\rho}) \equiv \varepsilon^{-1} \operatorname{Tr}(\hat{\rho} - \hat{\rho}^q) \rightarrow \varepsilon^{-1} \operatorname{Tr}(\hat{\rho} - \hat{\rho}^q) + \varepsilon^{-1} \delta \operatorname{Tr}(\hat{\rho} - \hat{\rho}^q) = \tilde{S}_q + \delta S_q(\hat{\rho}).$$

Учитывая теперь выражение (59) для вариации количества теплоты, поступающей в систему $\delta'Q_q \equiv \operatorname{Tr} \left[\delta\hat{\rho}^q (\hat{H} - E_q) \right] / \tilde{c}_q$, перепишем выражение (88) следующим образом

$$\tilde{c}_q \delta K_q(\hat{\rho} : \hat{\sigma}) = \beta \delta'Q_q - \delta S_q(\hat{\rho}). \quad (93)$$

Используя теперь положительное унитарное отображение $\hat{\rho} \rightarrow \hat{\rho} + \delta\hat{\rho} \equiv \Lambda(\hat{\rho})$ (полностью сохраняющее след), представим вариацию $\delta K_q(\hat{\rho} : \hat{\sigma})$ в виде:

$$\delta K_q(\hat{\rho} : \hat{\tilde{\rho}}) = \delta K_q(\Lambda(\hat{\rho}) : \hat{\tilde{\rho}}) - \delta K_q(\hat{\rho} : \hat{\tilde{\rho}}). \quad (94)$$

В работе [25], было показано, что только когда параметр деформации q лежит в интервале $q \in (0, 2)$, то справедливо неравенство

$$\delta K_q(\Lambda(\hat{\rho}) : \hat{\tilde{\rho}}) - \delta K_q(\hat{\rho} : \hat{\tilde{\rho}}) \leq 0. \quad (95)$$

Следовательно, при учёте (87)-(89), можно получить следующее фундаментальное неравенство Клаузиуса

$$\beta \delta' Q_q \leq \delta S_q(\hat{\rho}), \quad q \in (0, 2), \quad (96)$$

связывающее энтропию замкнутой квантовой системы с теплотой и температурой.

Таким образом, второй закон термодинамики (90) в квантовой термодинамике Тсаллиса также справедлив, что согласуется с принципом термодинамики в классической квантовой термодинамике [15]. Однако в отличие от последней, он справедлив только для значений энтропийного индекса, ограниченного интервалом $q \in (0, 2)$. Однако для квантовых систем с $q > 2$ он нарушается. Отметим ещё одно обстоятельство. В классической термодинамике второй закон определяет направление реально осуществляющихся процессов. Следовательно, адиабатические необратимые процессы ($\delta' Q_q = 0$) в квантовой термодинамике Тсаллиса, согласно неравенству (90), могут происходить в направлении роста энтропии только для значений параметра деформации $q \in (0, 2)$.

4.4. H -теорема в квантовой статистике Тсаллиса

Сравним теперь значения энтропий Клаузиуса для двух состояний квантово-механической системы с распределениями $\hat{\rho}$ и $\hat{\tilde{\rho}}$ при условии Гиббса, т.е. при одинаковости средних энергий

$$\text{Tr}(\hat{H}\hat{P}_q) = \text{Tr}(\hat{H}\hat{\tilde{P}}_q). \quad (97)$$

С учётом условия $\tilde{c}_q \equiv (\tilde{Q}_q)^{1-q} > 0$ и свойства выпуклости (82) информации различия Ратье–Каннапана, из (85) получим

$$K_q(\hat{\tilde{\rho}} : \hat{\rho}) = \tilde{S}_q(\hat{\rho}) - S_q(\hat{\rho}^q) \geq 0 \quad q \in (0, 2) \quad (98)$$

Из неравенств (92) следует обобщённая теорема Гиббса: квантовая q -энтропия равновесного состояния максимальна $\tilde{S}_q(\hat{\rho}) \geq S_q(\hat{\rho}^q)$ при $0 < q < 2$.

Из (92) также вытекает, что при увеличении энтропии $S_q(\hat{\rho}^q) \rightarrow \tilde{S}_q(\hat{\rho})$, положительная мера информации различия уменьшается, т.е. имеет место уменьшение

статистической упорядоченности в микросостояниях квантовой неэкстенсивной системы.

Согласно свойству выпуклости (82), квантовая информация различия Ратье–Каннаппана является знакоопределённой функцией Ляпунова. Чтобы состояние полного равновесия было устойчивым необходимо выполнение следующего неравенстваⁱⁱ⁾:

$$\frac{d}{dt} [\tilde{c}_q K_q(\hat{\rho} : \hat{\rho})] = -\frac{d}{dt} [S_q(\hat{\rho}) - \tilde{S}_q(\hat{\rho})] < 0 \quad \text{при } 0 < q < 2. \quad (99)$$

Таким образом, при стремлении системы к равновесному состоянию во временной эволюции информация различия уменьшается. Из (93) следует закон возрастания энтропии Тсаллиса со временем в квантовой неаддитивной статистической механике

$$dS_q(p) / dt > 0 \quad \text{при } 0 < q < 2, \quad (100)$$

который справедлив при приближении к состоянию полного статистического равновесия (H -теорема Больцмана). Таким образом, происходит хаотизация макроскопической системы при спонтанных переходах.

5. ЗАКЛЮЧЕНИЕ

Область неэкстенсивной (неаддитивной) статистической механики в настоящее время переживает фазу интенсивного развития. Обсуждаются различные подходы и идеи, в том числе методы исследования равновесных и неравновесных состояний, базирующиеся на статистических моделях с параметрическими функционалами для энтропий. В случае квантовых систем, описываемых методом статистических операторов (матрицами плотности) комплексов частиц, были ранее получены обобщения ряда этих функционалов для статистик Ферми–Дирака и Бозе–Эйнштейна.

В представленной работе для описания квантово-механической неэкстенсивной системы использован формализм матрицы плотности, описывающий системы, квантовые состояния которых известны не полностью. Кроме этого, широко использован обычный формализм феноменологической термодинамики, причём правильность обоих основных начал термодинамики, вслед за фон Нейманом, подразумевается. С учётом этих предположений получена равновесная статистическая термодинамика квантовых неэкстенсивных систем и определены её термодинамические свойства. Проведённое рассмотрение базируется на двух функционалах – на квантовой параметрической энтропии Тсаллиса и на обобщённой квантовой различающей информации Ратье–Каннаппана. Выполненный в работе анализ квантовой системы основывается на степенном равновесном распределении матрицы плотности, полученном из условия абсолютного экстремума квантовой энтропии Тсаллиса при заданности средней энергии и среднего числа частиц для ансамбля систем, а также на осреднении наблюдаемых по эскортному распределению.

ⁱⁱ⁾ Напомним, что функцией Ляпунова для данной системы называется знако-определённая 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.

В результате получено обобщение на квантовый случай 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 найдены модифицированные термодинамические соотношения на основе введённой энтропии Клаузиуса, которые отличны от выведенных ранее традиционным для статистической механики способом. С использованием свойства выпуклости обобщённой различающей информации Ратье–Каннаппана обсуждается второй закон термодинамики квантовых систем. Изучены также спонтанные переходы между стационарными состояниями сложной квантово-механической системы и доказана H -теорема Больцмана.

Развитый подход предполагает использование неэкстенсивной квантовой термодинамики в различных квантовых технологиях, связанных, в частности, с моделированием тепловых квантовых эффектов в различных наноструктурах.

REFERENCES

- [1] C. Tsallis, “Possible generalization of Boltzmann–Gibbs statistics”, *J. Stat. Phys.*, **52** (1-2), 479-487 (1988).
- [2] A. Rényi, *Probability Theory*, Amsterdam: North-Holland Publ. Co., (1970).
- [3] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics. Approaching a Complex World*, New York: Springer, (2009).
- [4] E.K. Lenzi, R.S. Mendes, “Collisionless Boltzmann equation for systems obeying Tsallis distribution”, *Eur. J. Phys., B*, **21** (3), 401-406 (2001).
- [5] S. Abe, “Heat and entropy in nonextensive thermodynamics: transmutation from Tsallis theory to Rényi-entropy-based theory”, *Physica A: Statistical Mechanics and its Applications*, **300** (3), 417-423 (2001).
- [6] R.G. Zaripov, *Printsipy neekstensivnoy statisticheskoy mekhaniki i geometrii mer besporyadka i poryadka*. Kazan: Izdatelstvo Kazanskogo Gosudarstvennogo tekhnicheskogo universiteta, (2010).
- [7] A.V. Kolesnichenko, *K postroyeniyu neadditivnoy termodinamiki slozhnykh sistem na osnove statistiki Kurado–Tsallisa*, Preprint IPM No 25 (Moscow: KIAM), (2018). doi:10.20948/prepr-25.
- [8] A.V. Kolesnichenko, *K konstruirovaniyu termodinamiki neadditivnykh sred na osnove statistiki Tsallisa–Mendes–Plastino*, Preprint IPM No 23 (Moscow: KIAM), (2018). doi:10.20948/prepr-23.
- [9] A.V. Kolesnichenko, “Dvukhparametricheskii entropiynny funktsional Sharma–Mittala kak osnova semeystva obobshchennykh termodinamik neekstensivnykh sistem”, *Mathematica Montisnigri*, **42**, 74-101 (2018).
- [10] A.V. Kolesnichenko, *Statisticheskay mekhanika i termodinamika Tsallisa neadditivnykh system. Vvedenie v teoriyu i prilozheniia*, Moskva: LENAND. (Sinergetika: ot proshlogo k budushchemu № 87), (2019).
- [11] A.V. Kolesnichenko, B.N. Chetverushkin, “Kinetic derivation of a quasihydrodynamic system of equations on the base of nonextensive statistics”, *Russian Journal of Numerical Analysis and Mathematical Modelling*, **28**, 547-576 (2013).
- [12] A.V. Kolesnichenko, M.Ya. Marov, “Modification of the jeans instability criterion for fractal-structure astrophysical objects in the framework of nonextensive statistics”, *Solar System Research*, **48** (5), 354-365 (2014).
- [13] A.V. Kolesnichenko, “Modifikatsiya v ramkakh statistiki Tsallisa kriteriev gravitatsionnoy neustoychivosti astrofizicheskikh diskov s fraktalnoy strukturoy fazovogo prostranstva”, *Mathematica Montisnigri*, **32**, 93-118 (2015).

- [14] A. V. Kolesnichenko, “On construction of the entropy transport model based on the formalism of nonextensive statistics”, *Mathematical Models and Computer Simulations*, **6** (6), 587-597 (2014).
- [15] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton (NJ): Princeton University Press, (1955).
- [16] A. M. Gleason, “Measures on the closed subspaces of a Hilbert space”, *Mathematics Journal* (Indiana University), **6**, 885-893 (1957).
- [17] D. Zubarev, V. Morozov and G. Röpke, *Statistical Mechanics of Nonequilibrium Processes*, Vol. 1, *Basic Concepts, Kinetic Theory*, Akademie-Verlag, Berlin, (1996).
- [18] C. Tsallis, “Nonextensive Statistics: Theoretical, Experimental and Computational Evidences and Connections”, *Brazilian Journal of Physics*, **29** (1), 1-35 (1999).
- [19] G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities* (2nd ed.). Cambridge: Cambridge University Press, (1952).
- [20] E.T. Jaynes, “Information theory and statistical mechanics”, *In com.: «Statistical Physics». Brandeis Lectures*, **3**, 160 (1963).
- [21] S. Abe, Y. Eds. Okamoto, *Nonextensive Statistical Mechanics and Its Applications, Series Lecture Notes in Physics*, Springer: Verlag, Berlin, New York. (2001).
- [22] S. Abe, “Heat and generalized Clausius entropy of nonextensive systems”, *Eprint arXiv:condmat/0012115*, **3**, 1-14 (2000).
- [23] A. Münster. *Chemische thermodynamic*, Akademie-Verlag Berlin, (1969).
- [24] C. Jarzynski, “Equalities and inequalities: irreversibility and the second law of thermodynamics at the nanoscale”, *Annu. Rev. Cond. Matt. Phys*, **2**, 329-335 (2011).
- [25] S. Abe, A.K. Rajagopal, “Validity of the Second Law in Nonextensive Quantum Thermodynamics”, *Physical Review Letters*, **91** (12), id. 120601 (2003).
- [26] C. Jarzynski, D. Wójcik, “Classical and Quantum Fluctuation Theorems for Heat Exchange”, *Physical Review Letters*, **92** (23), id. 230602 (2004).
- [27] S. Abe, “Quantum q -divergence”, *Physica A: Statistical Mechanics and its Applications*, **344** (3), 359-365 (2004).
- [28] S. Abe, “Nonadditive generalization of the quantum Kullback–Leibler divergence for measuring the degree of purification”, *Physical Review A*, **68** (3), id. 032302 (2003).

Received March 23, 2019

THERMODYNAMIC PARAMETERS OF MIXTURES WITH SILICON NITRIDE UNDER SHOCK-WAVE LOADING

K. K. MAEVSKII

Lavrentyev Institute of Hydrodynamics of the Siberian Branch of the Russian Academy of Sciences
Lavrentyev Avenue 15, 630090 Novosibirsk, Russia
E-mail: konstantinm@hydro.nsc.ru

DOI: 10.20948/mathmontis-2019-45-4

Summary. The results of numerical experiments on modeling thermodynamic parameters such as value of pressure, compression and temperature for shock wave loading of pure silicon nitride Si_3N_4 and mixtures based on it are presented. The model TEC (thermodynamic equilibrium components) and its modification model TEC2 allow to describe thermodynamic parameters of Si_3N_4 with different value of porosity, in a wide range of pressure from 5 to 3000 GPa using model relation and the experimental data of Si_3N_4 . The developed model is used to take into account the Si_3N_4 phase transition under shock-wave action. The results of calculations are compared with the data obtained on the basis of experiments for pure Si_3N_4 value of porosity 1 and 1.037; for mixture of Si_3N_4 and periclase MgO value of porosity 1.12. The value of pressure and compression for shock wave loading of nonporous mixture of aluminum nitride AlN and Si_3N_4 are calculated.

1 INTRODUCTION

In this paper, thermodynamically equilibrium model TEC [1, 2] is adapted to shock wave loading of a mixture of powders, including nitrides, when one or more components of the mixture undergo phase transformation. The researches of the behavior of porous mixtures of powder materials under shock-wave loading are of interest for many problems of modern science and technology [3–8]. Given the large variety of mixtures, it is desirable to use the equation of state (EoS) only components of the mixture. The construction of EoSs have been carried out for many years, but given the complexity and diversity of the materials, work in this direction continues [9–17]. The developed models well describe the behavior of pure substances, but are difficult to use in researching of mixtures. It is necessary to take into account the interaction of components for researching shock-wave loading of mixtures. It is preferable to use a simple equation in this case. The construction of simple EoSs is of particular interest in the modeling of thermodynamic parameters of mixtures [18–25]. In the developed model, the fairly simple EoS is used; it gives good correspondence to the experiment in wide range of pressures, both for solid and porous pure materials.

Interest in the investigation of compressibility of such mixtures is associated as with the possibility of creating materials that have the required properties and the properties of the materials themselves. The synthesis of materials, including multicomponent materials, under in-

2010 Mathematics Subject Classification: 74A15, 74J40, 76L05, 80A10, 82D20.

Key words and phrases: equation of state, silicon nitride, shock wave, high pressure, thermodynamic equilibrium.

tense shock-wave action is an effective way to create new functional and structural materials, in particular different types of ceramics with desired properties on the basis of nitride mixtures. The active researches of the properties and methods of obtaining such materials are carried out [26–32]. Mixtures of different compositions, including Si_3N_4 as a component, are considered in [33–36]. The researchers of Si_3N_4 and mixtures based on it are the purpose of this work.

In phase transition region, thermodynamic characteristics of materials were determined from the assumption that material in this region is a mixture of low-pressure phase and high-pressure phase [37–40]. The results of modeling for this method with the phase transition are given in [41–44] for nitrides and a number of other materials. The beginning of phase transition process is determined by the condition that pressure reaches the critical value obtained on the basis of experiments.

2 CALCULATION MODEL

The numerical simulation of thermodynamic parameters of shock-wave loading of both pure materials and mixtures is based on the assumption that all components of the mixture, including gas in pores, under shock-wave loading are in thermodynamic equilibrium model TEC [1, 2, 45, 46], it takes into account the interaction of components and uses the author's EoS. The thermal EoS for the condensed component with current density ρ and initial density ρ_0 is as follows:

$$P(\rho, T) = \rho_0 c_0^2 / n ((\rho / \rho_0)^n - 1) + \Gamma \rho c_V (T - T_0). \quad (1)$$

Here c_0 is the value of sound speed of under normal conditions, T_0 is initial temperature, c_V is specific heat capacity in this case constant. The value n is determined by the derivative of the volume elasticity modulus under initial conditions.

The function Γ reflects the contribution of the thermal components in the model [1, 2, 45, 46]. The conditions of dynamic compatibility at the wave front are written: the mass-flow conservation conditions for each mixture component and the momentum and energy flux conservation conditions for the mixture as a whole. The obtained equations together with EoS for each component are sufficient to find dependencies of type $P(U)$ or $D(U)$ (P , U , and D —pressure, mass and wave velocities), which can be interpreted as shock adiabat of a multicomponent mixture [1, 2]. The model TEC for high-porous samples well describes experimental data up to pressure value of 300 GPa [46].

The presence of experimental data on the dependence of heat capacity c_P on temperature allows to modify the equations that determine the state of the condensed component, increasing the area of reliable description to 3000 GPa. The cold component of the pressure is described also as well as in model TEC by the equation of Tait. Tait equation and its modified versions are one of the most well known empirical state equations. The Tait equation of state for describing

data on the density of sea water up to 50 MPa was originally published by Peter Guthrie Tait in 1888 [47]. The pressure form of the Tate equation of state is identical to that of the Murnaghan equation of state. Thermal components of pressure and specific energy P_T and E_T are considered in the following form:

$$P(\rho, \tau)_T = c_{V0}\rho T_0(\tau - 1)k \frac{1 + a_1\tau}{1 + a_2\tau}, \quad (2)$$

$$E(\rho, \tau)_T = F(\tau) - c_{V0}T_0k \frac{1 + 2a_2\tau + (a_1 - a_2 + a_1a_2)\tau^2}{(1 + a_2\tau)^2} \ln(\rho/\rho_0), \quad (3)$$

$$F(\tau) = c_{V0}T_0 \left[(\tau - 1) \left(\lambda - \frac{b_1}{b_2} \right) - \frac{1}{b_2} \left(1 - \frac{b_1}{b_2} \right) \ln \left(\frac{1 + b_2\tau}{1 + b_2} \right) \right]. \quad (4)$$

In this case, $\tau = T/T_0$ is the relative temperature, c_{V0} is specific heat capacity under initial conditions. Parameters are obtained according to the experimental data. For determining parameters k , a_1 , and a_2 , we can use function Γ . The values of λ , b_1 , and b_2 are determined from the correspondence to experimental data for $c_P(T)$. The model with this EoS is defined as model TEC2 [48].

3 MODELING RESULTS

The results of modeling the thermodynamic parameters for Si_3N_4 density $\rho_0 = 3.184 \text{ g/cm}^3$ and $\rho_0 = 3.07 \text{ g/cm}^3$ and the data obtained by experiments from [26, 28] are shown in figures 1 and 2. The beginning of the phase transition for silicon nitride is determined at pressure of 30 GPa. The calculations for the silicon nitride was considered transition phase $\beta\text{-Si}_3\text{N}_4$ in the high-density phase of $c\text{-Si}_3\text{N}_4$. The calculated lines and data obtained from the experiment are shown in figure 2 with density shift for $m = 1.037$ (porosity m is ratio of density of solid substance to the density of sample). It was shown earlier that the model describes the decrease in the wave velocity with the growth of mass velocity [49] in the phase transition region, and the temperature dependences of nitrides under shock wave loading [46]. The results allow modeling of thermodynamic parameters of the mixture, which contains nitrides as components, using the model parameters for pure nitrides [41].

Figure 3 illustrates the results obtained for nonporous mixture of AlN and Si_3N_4 with equal volume fractions and an initial density $\rho_0 = 3.335 \text{ g/cm}^3$. For comparison, the data obtained on the basis of experiments for AlN and Si_3N_4 are presented. It is assumed that the phase transition of the components of the mixture begins under the same conditions as for pure substances. The beginning of the phase transition for AlN is determined at pressure of 20 GPa, as well as for Si_3N_4 —at 30 GPa. The beginning of the phase transition is assumed in the pressure range 33–36 GPa for a mixture of Si_3N_4 and potassium bromide KBr in a ratio of 10/90 by weight in [50]. The phase transition from $\beta\text{-Si}_3\text{N}_4$ to $c\text{-Si}_3\text{N}_4$ is discovered above 36 GPa [28].

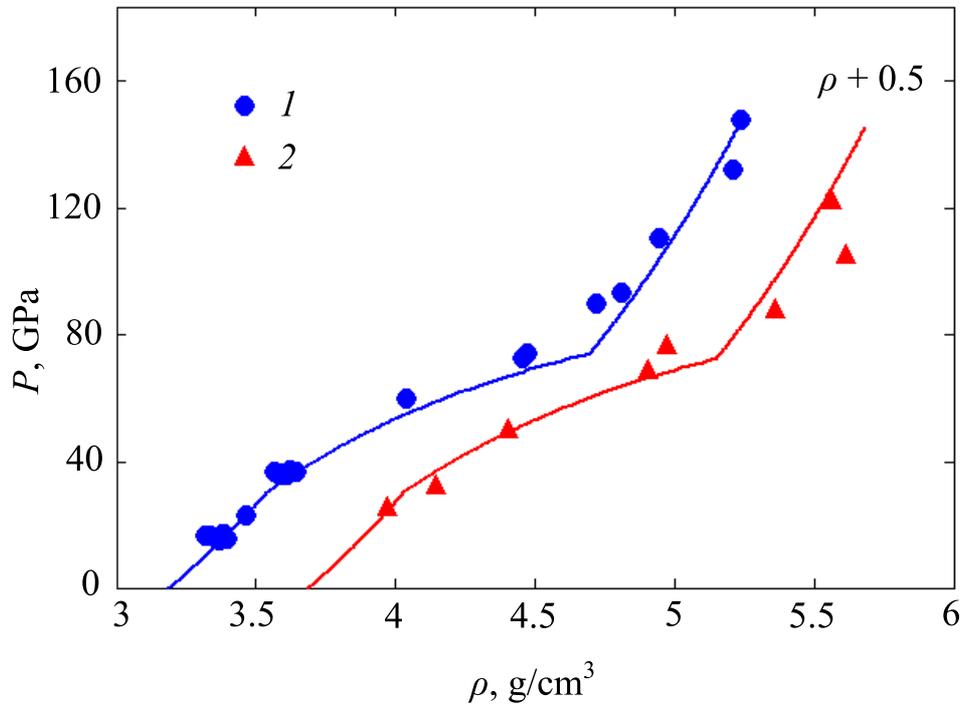


Figure 1: Shock adiabat of Si₃N₄. Calculation by model TEC at $m = 1$ (1) and 1.037 (2). Experimental data: 1—[28]; 2—[26].

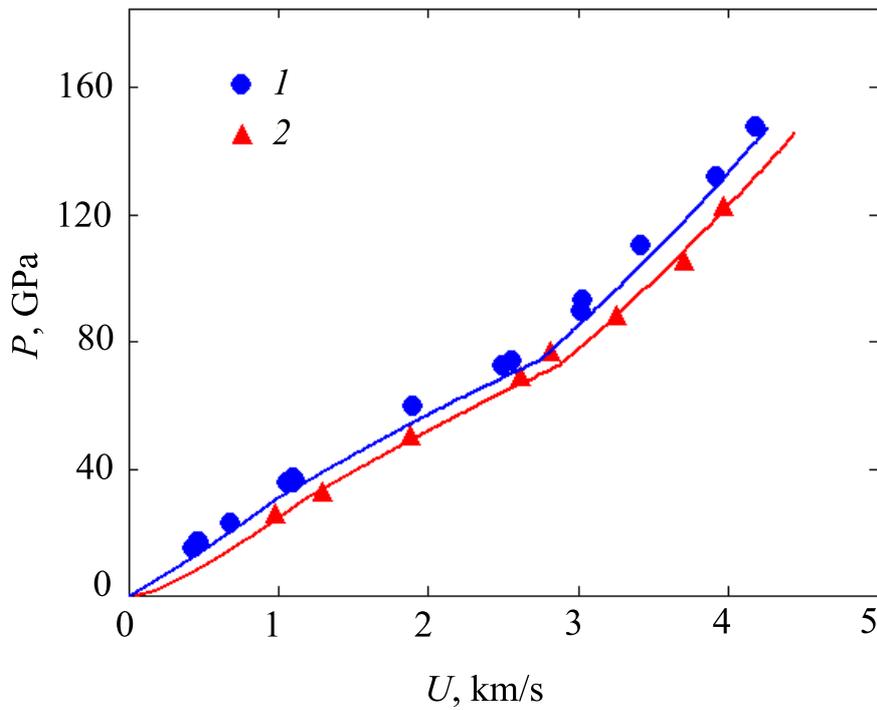


Figure 2: Shock adiabat of Si₃N₄. Symbols are as in figure 1.

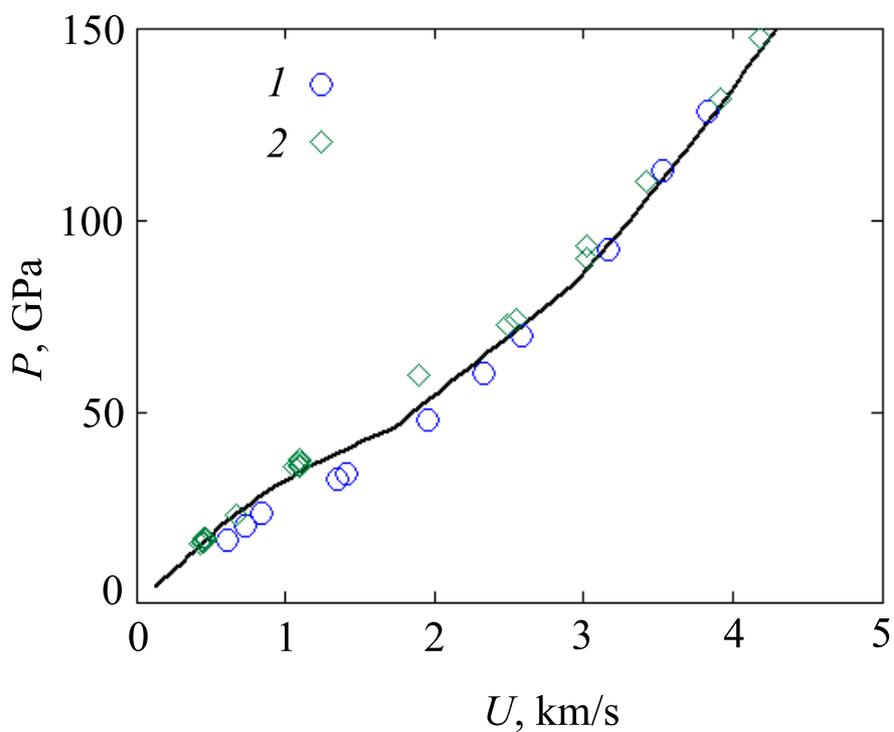


Figure 3: Shock adiabat of AlN and Si₃N₄ mixture. Experimental data: 1—AlN [26]; 2—Si₃N₄ [28].

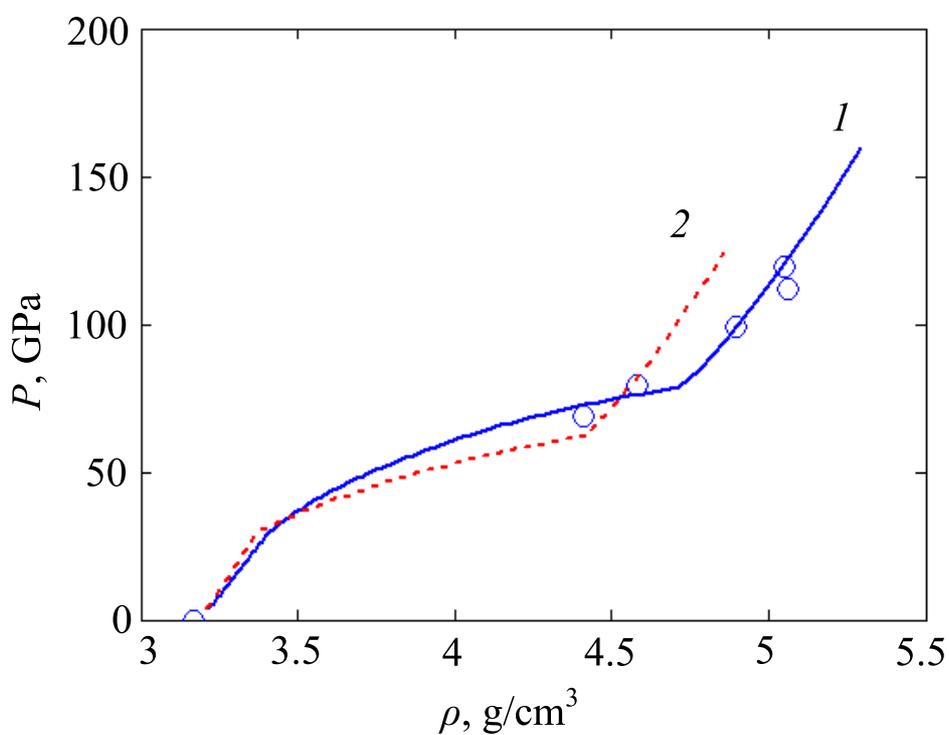


Figure 4: Shock adiabat in pressure–density coordinates for Si₃N₄ and MgO mixture with $m = 1.012$ (1) and 1.25 (2). Experimental data [51].

The parameters of mixtures of Si_3N_4 with oxides were modeled, for which there were data in [51]. Figure 4 shows the calculation of the mixture of Si_3N_4 and periclase MgO . The mixture with mass fractions $\text{Si}_3\text{N}_4(95)\text{MgO}(5)$ and density $\rho_0 = 3.164 \text{ g/cm}^3$ was modeled. The average porosity value determined from [51] $m = 1.012$. For comparison, the calculation for a higher value $m = 1.25$ is shown. The curve 2 for the mixture with $m = 1.25$ is above curve 1 with $m = 1.012$ before and after the phase transition region, which is typical for more porous materials. The phase transition ends earlier for more porous material, so curve 2 is below curve 1 in the phase transition region, taking into account the higher internal energy at the same density values. The modeling of this mixture is of interest due to the fact that MgO has a high-pressure phase transition [52]. The calculations are preferably performed on TEC2 for mixture with MgO .

4 CONCLUSIONS

Thus, the model allows calculating thermodynamic parameters of mixtures with nitrides under shock wave loading. The model parameters selected on basis of experimental data allow reliably describing the thermodynamic characteristics of silicon nitride and mixtures based on it. The assumption about thermodynamic equilibrium makes it possible to take into account the interaction between components, which becomes essential especially in case of porous media. The main objective of the research is achieved. The numerical experiments on modeling thermodynamic parameters for shock wave loading of pure Si_3N_4 with value of porosity 1 and 1.037 and for two mixtures Si_3N_4 with MgO and Si_3N_4 with AlN are performed. This model makes it possible to determine the compositions and ratios of the components of mixtures in order to obtain the specified characteristics under shock-wave loading of solid and porous heterogeneous materials.

Acknowledgments: The paper is based on the proceedings of the XXXIV International Conference on Interaction of Intense Energy Fluxes with Matter, Elbrus, the Kabardino-Balkar Republic of the Russian Federation, March 1 to 6, 2019.

REFERENCES

- [1] S. A. Kinelovskii and K. K. Maevskii, “Model of the behavior of aluminum and aluminum based mixtures under shock-wave loading”, *High Temp.*, **52**, 821–829 (2014).
- [2] S. A. Kinelovskii and K. K. Maevskii, “Modeling shock loading of multicomponent materials including bismuth”, *High Temp.*, **54**, 675–681 (2016).
- [3] B. R. Krueger and T. Vreeland, “A Hugoniot theory for solid and powder mixtures”, *J. Appl. Phys.*, **69**, 710–716 (1991).
- [4] A. A. Golyshev, V. V. Kim, A. N. Emel’yanov, and A. M. Molodets, “Model for calculation shock compression parameters of a platelet gradient mixture”, *J. Appl. Mech. Tech. Phys.*, **56**, 618–625 (2015).
- [5] V. V. Yakushev, S. Yu. Ananov, A. V. Utkin, A. N. Zhukov, and A. Yu. Dolgoborodov, “Shock compressibility of mixtures of micro- and nano-sized nickel and aluminum powders”, *Combust., Explos. Shock Waves*, **54**, 552–557 (2018).

-
- [6] B. Nayak and S. V. G. Menon, “Non-equilibrium theory employing enthalpy-based equation of state for binary solid and porous mixtures”, *Shock Waves*, **28**, 141–151 (2018).
- [7] S. A. Kinelovskii and K. K. Maevskii, “Model of the behavior of the mixture with different properties of the species under high dynamic loads”, *J. Appl. Mech. Tech. Phys.*, **54**, 524–530 (2013).
- [8] K. K. Maevskii and S. A. Kinelovskii, “Numerical simulation of thermodynamic parameters of lithium deuteride and its mixtures under shock wave loading”, *AIP Conf. Proc.*, **2051**, 020181 (2018).
- [9] Ya. B. Zel’dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, New York: Academic Press, (1967).
- [10] A. V. Bushman and V. E. Fortov, “Model equations of state”, *Sov. Phys. Usp.*, **26**, 465–496 (1983).
- [11] A. V. Bushman, V. E. Fortov, G. I. Kanel’, and A. L. Ni, *Intense Dynamic Loading of Condensed Matter*, Washington, DC: Taylor and Francis, (1993).
- [12] V. Fortov, *Thermodynamics and Equations of State for Matter: From Ideal Gas to Quark–Gluon Plasma*, Singapore: World Scientific, (2016).
- [13] I. V. Lomonosov and S. V. Fortova, “Wide-range semiempirical equations of state of matter for numerical simulation on high-energy processes”, *High Temp.*, **55**, 585–610 (2017).
- [14] W. B. Holzapfel, “Equations of state for Cu, Ag, and Au and problems with shock wave reduced isotherms”, *High Pressure Res.*, **30**, 372–394 (2010).
- [15] S. A. Kinelovskii and K. K. Maevskii, “Simple model for calculating shock adiabats of powder mixtures”, *Combust., Explos. Shock Waves*, **47**, 706–714 (2011).
- [16] M. A. Kadatskiy and K. V. Khishchenko, “Theoretical investigation of the shock compressibility of copper in the average-atom approximation”, *Phys. Plasmas*, **25**, 112701 (2018).
- [17] D. V. Minakov, M. A. Paramonov, and P. R. Levashov, “Consistent interpretation of experimental data for expanded liquid tungsten near the liquid–gas coexistence curve”, *Phys. Rev. B*, **97**, 024205 (2018).
- [18] K. V. Khishchenko, “The equation of state for magnesium at high pressures”, *Tech. Phys. Lett.*, **30**, 829–831 (2004).
- [19] K. V. Khishchenko, “Equation of state for tungsten over a wide range of densities and internal energies”, *J. Phys.: Conf. Ser.*, **653**, 012081 (2015).
- [20] R. K. Belkheeva, “Equation of state for a highly porous material”, *High Temp.*, **53**, 348–357 (2015).
- [21] E. I. Kraus and I. I. Shabalin, “Calculation of elastic modules behind strong shock wave”, *J. Phys.: Conf. Ser.*, **774**, 012009 (2015).
- [22] K. V. Khishchenko, “Equation of state of sodium for modeling of shock-wave processes at high pressures”, *Mathematica Montisnigri*, **40**, 140–147 (2017).
- [23] S. D. Gilev, “Few-parameter equation of state of copper”, *Combust., Explos. Shock Waves*, **54**, 482–495 (2018).
- [24] K. V. Khishchenko, “Equation of state for magnesium hydride under conditions of shock loading”, *Mathematica Montisnigri*, **43**, 70–77 (2018).
- [25] K. K. Maevskii, “Thermodynamic parameters of lithium deuteride in pressure range 5–1000 gigapascals”, *Mathematica Montisnigri*, **41**, 123–130 (2018).
- [26] A. A. Bakanova, V. A. Bugaeva, I. P. Dudoladov, and R. F. Trunin, “Udarnaya szhimaemost’ nitridov i karbidov metallov”, *Izv. Akad. Nauk SSSR, Fiz. Zemli*, **6**, 58–63 (1995).
- [27] T. Sekine, H. He, T. Kobayashi, M. Zhang, and F. Xu, “Shock-induced transformation of β -Si₃N₄ to a high-pressure cubic-spinel phase”, *Appl. Phys. Lett.*, **76**, 3706–3708 (2000).
- [28] H. He, T. Sekine, T. Kobayashi, and H. Hirosaki, “Shock-induced phase transition of β -Si₃N₄ to *c*-Si₃N₄”, *Phys. Rev. B*, **62**, 11412–11417 (2000).
- [29] T. Sekine, “Shock synthesis of cubic silicon nitride”, *J. Am. Ceram. Soc.*, **85**, 113–116 (2002).
- [30] V. V. Yakushev, A. V. Utkin, A. N. Zhukov, D. V. Shakhrai, and V. V. Kim, “Determination of *P, T*-conditions developed at high-temperature shock compression of silicon nitride in planar recovery ampoules”, *High Temp.*, **54**, 197–205 (2016).
- [31] N. Nishiyama, R. Ishikawa, H. Ohfuji, H. Marquardt, A. Kurnosov, T. Taniguchi, B. Kim, H. Yoshida, A. Masuno, J. Bednarcik, E. Kulik, Y. Ikuhara, F. Wakai, and T. Irifune, “Transparent polycrystalline cubic silicon nitride”, *Sci. Rep.*, **7**, 44755 (2017).

- [32] A. Yamakawa, “Shock-yielding properties of Si_3N_4 ceramics: Correlation with the microstructure and static mechanical strength”, *J. Ceram. Soc. Jpn.*, **101**, 1358–1362 (1993).
- [33] A. S. Yunoshev, “Shock-wave synthesis of cubic silicon nitride”, *Combust., Explos. Shock Waves*, **40**, 370–373 (2004).
- [34] A. N. Zhukov, S. E. Zakiev, and V. V. Yakushev, “Assessment of the effect of particle size on the rate of temperature alignment in systems used for shock-wave synthesis of diamond, cubic boron nitride, and γ -silicon nitride, based on a simple model”, *High Temp.*, **54**, 62–67 (2016).
- [35] T. Mashimo, M. Uchino, and A. Nakamura, “Yield properties, phase transition, and equation of state of aluminum nitride (AlN) under shock compression up to 150 GPa”, *J. Appl. Phys.*, **86**, 6710–6716 (1999).
- [36] L. A. Gömze and L. N. Gömze, “Rheological principles of development hetero-modulus and hetero-viscous complex materials with extreme dynamic strength”, *IOP Conf. Ser.: Mater. Sci. Eng.*, **175**, 012001 (2017).
- [37] I. V. Lomonosov, V. E. Fortov, A. Frolova, K. Khishchenko, A. Charakhch’yan, and L. Shurshalov, “The simulation of transformation of graphite to diamond under conditions of dynamic compression in a conic target”, *High Temp.*, **41**, 447–458 (2003).
- [38] A. A. Charakhch’yan, V. V. Milyavskii, and K. V. Khishchenko, “The use of models of mixture for analysis of shock-wave experiments with incomplete phase transformation”, *High Temp.*, **47**, 235–242 (2009).
- [39] K. V. Khishchenko, A. A. Charakhch’yan, V. Fortov, A. Frolova, V. Milyavskiy, and L. Shurshalov, “Hydrodynamic simulation of converging shock waves in porous conical samples enclosed within solid targets”, *J. Appl. Phys.*, **110**, 053501 (2011).
- [40] E. Bourasseau, N. Pineau, J. B. Maillat, and V. Dubois, “Calculation of the shock temperature of moderately porous graphites”, *Carbon*, **103**, 464–472 (2016).
- [41] K. K. Maevskii and S. A. Kinelovskii, “Model of behavior of nitrides and their mixtures under high dynamic loads”, *AIP Conf. Proc.*, **1623**, 391–394 (2014).
- [42] K. K. Maevskii and S. A. Kinelovskii, “Thermodynamic parameters for mixtures of quartz under shock wave loading in views of the equilibrium model”, *AIP Conf. Proc.*, **1683**, 020132 (2015).
- [43] K. K. Maevskii, “Numerical investigations of oxides and silicates under shock-wave loading”, *J. Phys.: Conf. Ser.*, **894**, 012057 (2017).
- [44] K. K. Maevskii and S. A. Kinelovskii, “Thermodynamic parameters for mixtures of quartz under shock wave loading in views of the equilibrium model”, *J. Phys.: Conf. Ser.*, **946**, 012113 (2018).
- [45] S. A. Kinelovskii and K. K. Maevskii, “Estimation of the thermodynamic parameters of a shock-wave action on high-porosity heterogeneous materials”, *Tech. Phys.*, **61**, 1244–1249 (2016).
- [46] K. K. Maevskii and S. A. Kinelovskii, “Thermodynamic parameters of mixtures with silicon nitride under shock-wave impact in terms of equilibrium model”, *High Temp.*, **56**, 853–858 (2018).
- [47] Y. H. Li, “Equation of state of water and sea water”, *J. Geophys. Res.*, **72**, 2665–2678 (1967).
- [48] K. K. Maevskii and S. A. Kinelovskii, “Numerical simulation of thermodynamic parameters of high-porosity copper”, *Tech. Phys.*, **64**, 1090–1095 (2019).
- [49] R. F. Trunin, *Issledovaniya e’kstremal’ny’x sostoyanij kondensirovanny’x veshhestv metodom udarny’x voln. Uravneniya Gyugonio*, Sarov: Rossijskij federal’ny’j yaderny’j centr—VNIIEF, (2006).
- [50] V. V. Yakushev, A. N. Zhukov, A. V. Utkin, A. I. Rogacheva, and V. A. Kudakina, “Formation of cubic silicon nitride from the low-pressure phase by high-temperature shock compression”, *Combust., Explos. Shock Waves*, **51**, 603–610 (2015).
- [51] S. P. Marsh (ed.), *LASL Shock Hugoniot Data*, Berkeley: University of California Press, (1980).
- [52] K. Miyaniishi, Y. Tange, N. Ozaki, T. Kimura, T. Sano, Y. Sakawa, T. Tsuchiya, and R. Kodama, “Laser-shock compression of magnesium oxide in the warm-dense-matter regime”, *Phys. Rev. E*, **92**, 023103 (2015).

Received May 15, 2019

ESTIMATION OF NOISE IN CALCULATION OF SCATTERING MEDIUM LUMINANCE BY MCRT

S. V. ERSHOV¹, D. D. ZHDANOV², A. G. VOLOBOY¹

¹Keldysh Institute of Applied Mathematics of RAS,
Miuskaya Sq. 4, Moscow, Russia, 125047
e-mail: voloboy@gin.keldysh.ru, web page: <http://keldysh.ru/>

²ITMO University, 49 Kronverksky Pr., St. Petersburg, 197101, Russia

DOI: 10.20948/mathmontis-2019-45-5

Summary. We investigate new method for calculation of radiance of scattering medium by bi-directional Monte-Carlo ray tracing with photon maps. Usually photons are collected by an integration spheres at the ends of camera ray segments, or a cylinder along that segments. Meanwhile in our method several integration spheres are distributed at random along the first camera ray segment. The rest segments do not collecting photons. The method optimal for a particular scene is the one which produces the least noise, so one need to be able to estimate it. In this paper an analytic calculation of noise in the general bi-directional Monte-Carlo ray tracing is derived and then applied to the proposed method. Then the analytic estimates of noise can be used to find optimal parameters and/or to choose between single integration sphere, multiple integration spheres and integration cylinders.

1 INTRODUCTION

A bi-directional Monte Carlo ray tracing is well-known as a powerful method of calculation of a virtual camera image. The forward part of the method traces the light rays from light sources and creates photon maps which allow calculating luminance and illuminance of scene surfaces. Then the backward part traces the light rays from camera, estimates the luminance in the hit point and then accumulates it along the camera ray path into the pixel.

The method can be implemented via several techniques [1, 2, 3, 4] of which the photon map visualization approach [5, 6, 7, 8] is the mostly usable now. Most of these methods calculate global illumination in the form of photon maps and then visualize them as the luminance of secondary and caustic illumination [5, 7, 9]. There is an approach that implements the reverse calculation scheme, i.e. generates a visibility map as spheres of the illuminance integration in the direction of observation, which are “filled” with the light photons related to the caustic and the secondary illumination [6, 8]. In these methods the camera ray is traced stochastically until it terminates due to some criterion, for example, after the given number of diffuse events is reached. An integration sphere to

2010 Mathematics Subject Classification: 78-04, 65C05, 65C20.

Key words and Phrases: scattering medium, bi-directional ray tracing, Monte-Carlo ray tracing.

collect Forward Monte Carlo ray tracing (FMCRT) rays is set in each point of scattering. After a forward ray hits that sphere we calculate the surface luminance for the view direction equal to the camera ray direction before the sphere center. This luminance is then scaled by the camera ray attenuation accumulated to this point and added to the pixel luminance, [1, 6, 10].

In scattering medium the integration spheres are distributed over the volume, not only on boundaries, see e.g. [11], [12]. Frequently it is advantageous to use integration volumes other than spheres, e.g. cylinders [13]. One can find comparison and analysis in [5, 12, 14, 13]; [15]. Alternatively, the integration volume can be a union of spheres [16].

Usually the bi-directional ray tracing with photon maps is applied progressively, i.e. ray tracing goes iteration by iteration, and the error vanishes as time goes on. In each iteration we trace camera rays and place integration spheres in the ends of the segments. Then during FMCRT phase we estimate illuminance in the centers of integration spheres from the photons that hit them.

It is important to have a reliable analytic estimate of the noise in the stochastic rendering. Firstly this allows optimizing parameters of the method so that the noise remaining after the fixed simulation time was minimal (or do that with the constraint on the RAM used). Secondly there are several other inaccuracies besides the noise in the method, e.g. bias. This latter is unavoidable because of the final size of the integration area used for the illuminance estimation. And it grows rather quickly with area size. Therefore it is advantageous to decrease this size. But in such a case number of used photons is also reduced and thus the noise is increased. Therefore there exists the optimal size which minimizes the total error. And to calculate it one needs to predict the noise.

In [17] the authors operated the simplest case when the integration spheres are in the first camera rays hit so that position of the integration sphere for the given pixel is not random. In [18] the camera ray is allowed to be “glossy” (nearly specular diffuse) scattered before, so position of the integration sphere and view direction for it is random. This introduces additional source of noise which is analyzed. The authors consider progressive rendering when the radius of integration spheres decrease from iteration to iteration, so after large simulation time the results are for a very small radius. They derive the estimates of the bias and variance (noise) for some class of the radius decrease rule and show that there is a range of parameters for which both errors vanish in course of time.

They however assume that there is only one camera ray per pixel in each iteration. This is good when the integration sphere is set prior to the first random scattering of the camera ray as in [17] or when this scattering is “weakly diffuse” (“glossy”) as in [18]. But this strategy is not optimal when the camera ray undergoes a wide diffuse scattering before setting the integration sphere. Indeed, now the centers of integration spheres for adjacent pixels are not close but can spread over the whole scene and even beyond the visible area. The density of integration spheres can then be low and they capture only a small fraction of the forward photons. In other words a substantial part of calculations in the FMCRT phase are lost because these photons miss the sparsely distributed integration spheres. The natural remedy is to increase the number of camera rays through pixel so that the integration spheres to cover a decent fraction of illuminated area. This introduces an additional control parameter: the number of backward rays per pixel. In principle it can vary across the image and change with time. Setting it too low is not advantageous

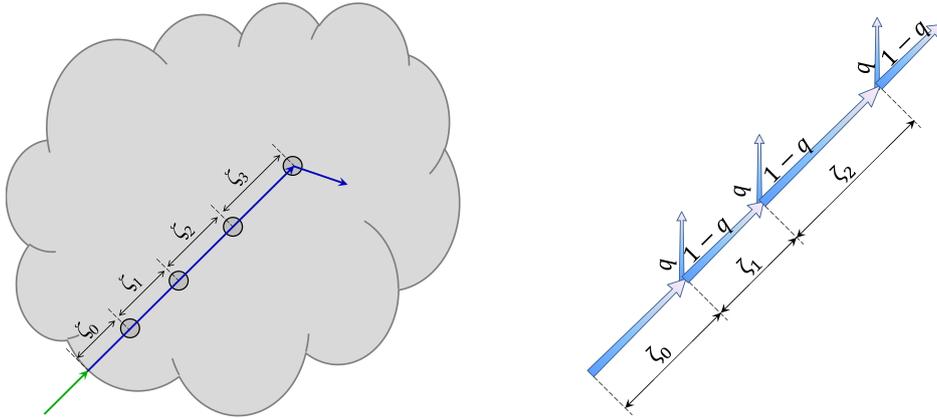


Figure 1: Several integration spheres along the camera ray segment. Left: positioning of the spheres. Right: Camera ray propagation with sub-steps: it strides by ζ_m then with probability q it undergoes extinction otherwise goes the same direction

because a many FMCRT rays whose tracing took time are not used. But using too much camera rays per pixel is also bad because this increases simulation time while the effective number of FMCRT rays used in the given pixel saturates. Therefore there exists an optimal relation between the number of forward and backward rays.

In this paper we derive an estimation of the noise as a function of the number of camera rays per pixel which applies to a general case of bi-directional MCRT with progressive photon maps. Then we apply it to calculation of luminance of turbid medium with multiple integration spheres [16] and derive an analytic dependence of noise as a function of the control parameters of the method: the number of forward and backward rays, the average number of integration spheres per camera ray segment and so on. This analytic estimate can be used to find optimal parameters.

2 MULTIPLE INTEGRATION SPHERES

In [16] we suggested to use integration volume composed of *several* integration spheres distributed randomly over a camera ray segment. The basic idea is that when tracing an camera ray we perform many “sub-steps” of length ζ so that the ray goes straight during several of them and only then an extinction which can be scattering or absorption occurs, see Figure 1.

These sub-steps are independent from each other, the density of ζ being always the same $p_\zeta(\zeta)$.

After the ray propagated the next step length ζ its further destiny is decided at random. With probability q there is an extinction event (scattering or absorption is then decided at random); otherwise the next step is made *retaining the ray direction*. Absorption is processed as a “Russian roulette” killing the ray at random.

In [16] it was obtained that step length must be distributed as

$$p_\zeta(\zeta) = \alpha e^{-\alpha\zeta} \tag{1}$$

$$\alpha \equiv \sigma_{ext}/q \tag{2}$$

The total ray length after n such steps then has the density

$$p_n(s) = \alpha e^{-\alpha s} \frac{(\alpha s)^n}{n!} \quad (3)$$

The contribution of an FMCRT ray segment that crosses the BMCRT ray at distance s to the pixel luminance $C(s)$ is

$$C(s) = q\mathcal{C}(s) \times \sigma_{ext} e^{-\sigma_{ext}s} ds \quad (4)$$

see [16], notice this applies to both *single segment* and *whole forward ray trajectory*. Here

$$\mathcal{C}(s) = \frac{\sigma_{sc} f(-\mathbf{u}, \mathbf{v})}{\sigma_{ext} \pi R^2} \mathcal{F} \quad (5)$$

and \mathcal{F} is the total flux (sum over all light sources), f is the phase function and σ_{sc} is scattering of the medium and σ_{ext} is its extinction.

Here and below we use the term “luminance” as an equivalent of “radiance”

3 GENERAL ESTIMATE OF NOISE IN BI-DIRECTIONAL MCRT

In bi-directional MCRT each forward ray interacts with each camera ray, and the luminance of a pixel is

$$L = \frac{1}{N_F} \sum_{i=1}^{N_F} \frac{1}{N_B} \sum_{j=1}^{N_B} \hat{C}(i, j) \quad (6)$$

where i enumerates forward rays, j enumerates backward rays, $\hat{C}(i, j)$ is the increase of pixel luminance from interaction of the i -th forward with the j -th camera ray, N_F is the number of forward rays and N_B is the number of camera rays ***traced through this pixel***.

The above sum L is a random variable, and its *noise* is $\langle L^2 \rangle - \langle L \rangle^2$ where the average is taken over the ensembles of rays. It can be thus understood as a *repeated* average over the forward and backward ensembles. Let us square (6)

$$\begin{aligned} (N_F N_B)^2 L^2 &= \sum_{i=1}^{N_F} \sum_{j=1}^{N_B} \hat{C}^2(i, j) + \sum_{i=1}^{N_F} \sum_{j \neq j'}^{N_B} \hat{C}(i, j) \hat{C}(i, j') \\ &+ \sum_{i \neq i'}^{N_F} \sum_{j=1}^{N_B} \hat{C}(i, j) \hat{C}(i', j) + \sum_{i \neq i'}^{N_F} \sum_{j \neq j'}^{N_B} \hat{C}(i, j) \hat{C}(i', j') \end{aligned}$$

and average over the forward ray ensemble and over the backward ray ensemble. These averaging are denoted as $\langle \cdot \rangle_F$ and $\langle \cdot \rangle_B$ and their order is arbitrary. This gives

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

where we used the obvious fact that

$$\begin{aligned}\langle \hat{C}(i, j) \hat{C}(i', j') \rangle_F &= \langle \hat{C} \rangle_F(j) \langle \hat{C} \rangle_F(j') & \text{if } i' \neq i \\ \langle \hat{C}(i, j) \hat{C}(i', j') \rangle_B &= \langle \hat{C} \rangle_B(i) \langle \hat{C} \rangle_B(i') & \text{if } j' \neq j\end{aligned}$$

The noise is therefore

$$\langle \langle L^2 \rangle \rangle - \langle \langle L \rangle \rangle^2 = \frac{1}{N_F N_B} \left(\langle \langle \hat{C}^2 \rangle_B \rangle_F - \langle \langle \hat{C} \rangle_B \rangle_F^2 \right) \quad (7)$$

$$+ \frac{1}{N_B} \left(1 - \frac{1}{N_F} \right) \left(\langle \langle \hat{C} \rangle_F^2 \rangle_B - \langle \langle \hat{C} \rangle_B \rangle_F^2 \right) \quad (8)$$

$$+ \frac{1}{N_F} \left(1 - \frac{1}{N_B} \right) \left(\langle \langle \hat{C} \rangle_B^2 \rangle_F - \langle \langle \hat{C} \rangle_B \rangle_F^2 \right) \quad (9)$$

$$\langle \langle L \rangle \rangle = \langle \langle \hat{C} \rangle_B \rangle_F^2 \quad (10)$$

4 ESTIMATION OF NOISE FOR MULTIPLE SPHERES

The noise is as sum of three terms, see (7), (8), (9), (10) where now \hat{C} is the contribution of one forward ray to *all* n integration spheres set by the given camera ray:

$$\hat{C} = \sum_{m=0}^n C(\hat{\zeta}_m), \quad (11)$$

$C(s)$ is contribution from *one* FMCRT ray to the single integration sphere at distance s from the camera ray start,

$$\hat{\zeta}_m \equiv \sum_{i=0}^m \zeta_i \quad (12)$$

is position of the m -th sphere set after m sub-steps. Here and below $\zeta_0, \zeta_1, \zeta_2, \dots$ are successive ray sub-steps between the 1st, 2nd, ..., integration spheres.

Notice the above estimate assume that each of N_F forward rays is checked for interaction with each of N_B camera rays. If the whole process runs progressively, iteration by iteration and for the next iteration *both* forward and backward rays are all new, then this estimates applies to one iteration only. The relative variance obtained in M (identical) iterations will be M times lower (instead of replacing $N_F \mapsto MN_F, N_B \mapsto MN_B$).

4.1 Calculation of the averages $\langle \hat{C} \rangle_B$ and $\langle \hat{C}^2 \rangle_B$

The contribution from one forward ray (11), averaged over the BMCRT ensemble i.e. sphere positions i.e. the $\langle C \rangle_B$ is obviously

$$\begin{aligned}
 \langle \hat{C} \rangle_B &= \sum_{n=0}^{\infty} P(n) \int_0^{\infty} \cdots \int_0^{\infty} \hat{C} p_{\zeta}(\zeta_0) \cdots p_{\zeta}(\zeta_n) d\zeta_0 \cdots d\zeta_n \\
 &= q \sum_{n=0}^{\infty} (1-q)^n \int_0^{\infty} \cdots \int_0^{\infty} \left(\sum_{m=0}^n C(\hat{\zeta}_m) \right) p_{\zeta}(\zeta_0) \cdots p_{\zeta}(\zeta_m) d\zeta_0 \cdots d\zeta_m
 \end{aligned}$$

where $\hat{\zeta}_m$ is given by (12).

The outer sum is over the number of integration spheres n and changing the order of summation we arrive at

$$\langle \hat{C} \rangle_B = \sum_{m=0}^{\infty} (1-q)^m \int_0^{\infty} \cdots \int_0^{\infty} C(\hat{\zeta}_m) p_{\zeta}(\zeta_0) \cdots p_{\zeta}(\zeta_m) d\zeta_0 \cdots d\zeta_m$$

The sum of steps $\hat{\zeta}_m$ has density $\frac{(\alpha \hat{\zeta}_m)^m}{m!} \alpha e^{-\alpha \hat{\zeta}_m}$ (3), and using (2) and (4) we obtain

$$\langle \hat{C} \rangle_B = \int_0^{\infty} \mathcal{E}(t) \sigma_{ext} e^{-\sigma_{ext} t} dt \quad (13)$$

Similarly, the *squared* contribution of one FMCRT ray, averaged over the BMCRT ensemble i.e. the $\langle \hat{C}^2 \rangle_B$ is

$$\begin{aligned}
 \langle \hat{C}^2 \rangle_B &= \sum_{n=0}^{\infty} P(n) \int_0^{\infty} \cdots \int_0^{\infty} \left(\sum_{m=0}^n C(\hat{\zeta}_m) \right)^2 p_{\zeta}(\zeta_0) \cdots p_{\zeta}(\zeta_n) d\zeta_0 \cdots d\zeta_n \\
 &= q \sum_{n=0}^{\infty} (1-q)^n \sum_{m=0}^n \int_0^{\infty} \cdots \int_0^{\infty} C^2(\hat{\zeta}_m) p_{\zeta}(\zeta_0) \cdots p_{\zeta}(\zeta_m) d\zeta_0 \cdots d\zeta_m \\
 &\quad + 2q \sum_{n=0}^{\infty} (1-q)^n \int_0^{\infty} \cdots \int_0^{\infty} \sum_{m=1}^n \sum_{m'=0}^{m-1} C(\hat{\zeta}_m) C(\hat{\zeta}_{m'}) p_{\zeta}(\zeta_0) \cdots p_{\zeta}(\zeta_n) d\zeta_0 \cdots d\zeta_n
 \end{aligned}$$

or, changing the order of summation,

$$\begin{aligned}
 \langle \hat{C}^2 \rangle_B &= \sum_{m=0}^{\infty} (1-q)^m \int_0^{\infty} \cdots \int_0^{\infty} C^2(\hat{\zeta}_m) p_{\zeta}(\zeta_0) \cdots p_{\zeta}(\zeta_m) d\zeta_0 \cdots d\zeta_m \\
 &\quad + 2 \sum_{m=1}^{\infty} \sum_{m'=0}^{m-1} (1-q)^m \int_0^{\infty} \cdots \int_0^{\infty} C(\hat{\zeta}_m) C(\hat{\zeta}_{m'}) p_{\zeta}(\zeta_0) \cdots p_{\zeta}(\zeta_n) d\zeta_0 \cdots d\zeta_n
 \end{aligned}$$

The sum of steps $\hat{\zeta}_{m'}$ has density $\frac{(\alpha \hat{\zeta}_{m'})^{m'}}{(m')!} \alpha e^{-\alpha \hat{\zeta}_{m'}}$ (3). Then, $\zeta_m = \zeta_{m'} + \zeta_{m'+1} + \cdots + \zeta_m$ is a sum of two independent random variables, $\zeta_{m'}$ and $\zeta_{m'+1} + \cdots + \zeta_m$. Since the all sub-steps are equally distributed, the density of the sum of $m - m' - 1$ of them is $p_{m-m'-1}(s) = \alpha e^{-\alpha s} \frac{(\alpha s)^{m-m'-1}}{(m-m'-1)!}$, so

$$\begin{aligned} \langle \hat{C}^2 \rangle_B &= \int_0^\infty C^2(t) \alpha e^{-q\alpha t} dt + 2(1-q) \\ &\quad \int_0^\infty \int_0^\infty C(t)C(t+s) \sum_{m=0}^\infty \sum_{m'=0}^m \frac{(\alpha(1-q)t)^{m'}}{(m')!} \frac{(\alpha(1-q)s)^{m-m'}}{(m-m')!} \alpha^2 e^{-\alpha(s+t)} ds dt \end{aligned}$$

and using the obvious identity

$$\sum_{m=0}^\infty \sum_{m'=0}^m \frac{x^{m'}}{(m')!} \frac{y^{m-m'}}{(m-m')!} = \sum_{m'=0}^\infty \frac{x^{m'}}{(m')!} \sum_{m=m'}^\infty \frac{y^{m-m'}}{(m-m')!} = \sum_{m'=0}^\infty \frac{x^{m'}}{(m')!} \sum_{m=0}^\infty \frac{y^m}{m!} = e^x e^y$$

together with (2) and (4) we arrive at

$$\langle \hat{C}^2 \rangle_B = q \int_0^\infty \mathcal{C}^2(t) \sigma_{ext} e^{-\sigma_{ext} t} dt + 2(1-q) \int_0^\infty \int_0^\infty \mathcal{C}(s) \mathcal{C}(s+t) \sigma_{ext}^2 e^{-\sigma_{ext}(s+t)} ds dt \quad (14)$$

4.2 Combined averages

Averaging (13) and (14) over the FMCRT ensemble, we arrive at

$$\langle \langle \hat{C} \rangle_B \rangle_F = \int_0^\infty \langle \mathcal{C} \rangle(t) \sigma_{ext} e^{-\sigma_{ext} t} dt \quad (15)$$

$$\begin{aligned} \langle \langle \hat{C}^2 \rangle_B \rangle_F &= q \int_0^\infty \langle \mathcal{C}^2 \rangle(t) \sigma_{ext} e^{-\sigma_{ext} t} dt \\ &\quad + 2(1-q) \int_0^\infty \int_0^\infty \langle \mathcal{C}(s) \mathcal{C}(s+t) \rangle \sigma_{ext}^2 e^{-\sigma_{ext}(s+t)} ds dt \end{aligned} \quad (16)$$

where $\langle \mathcal{C} \rangle(t)$ (respectively $\langle \mathcal{C}^2 \rangle(t)$) is contribution (respectively squared contribution) of a single forward ray to the integration sphere located at distance t from the camera ray start averaged over the forward ray ensemble.

Notice that because of linearity of averaging, $\langle \langle \hat{C} \rangle_B \rangle_F = \langle \langle \hat{C} \rangle_F \rangle_B$ and $\langle \langle \hat{C}^2 \rangle_B \rangle_F = \langle \langle \hat{C}^2 \rangle_F \rangle_B$.

For (8) and (9) we also need the averages $\langle \langle \hat{C} \rangle_F^2 \rangle_B$ and $\langle \langle \hat{C} \rangle_B^2 \rangle_F$. The former can be calculated as in the above Section just replacing $C(x)$ with $\langle C \rangle_F(x)$, so (14) transforms into

$$\begin{aligned} \langle \langle \hat{C} \rangle_F^2 \rangle_B &= q \int_0^\infty \langle \mathcal{C} \rangle^2(t) \sigma_{ext} e^{-\sigma_{ext} t} dt \\ &\quad + 2(1-q) \int_0^\infty \int_0^\infty \langle \mathcal{C} \rangle(s) \langle \mathcal{C} \rangle(s+t) \sigma_{ext}^2 e^{-\sigma_{ext}(s+t)} ds dt \end{aligned} \quad (17)$$

The remaining $\langle \langle \hat{C} \rangle_B^2 \rangle_F$ is calculated even simpler. Squaring the $\langle C \rangle_B$ is given by (13) and averaging over the FMCRT ensemble yields

$$\langle \langle \hat{C} \rangle_B^2 \rangle_F = \left\langle \left(\int_0^\infty \mathcal{C}(t) \sigma_{ext} e^{-\sigma_{ext} t} dt \right)^2 \right\rangle = \int_0^\infty \int_0^\infty \langle \mathcal{C}(t) \mathcal{C}(s) \rangle \sigma_{ext}^2 e^{-\sigma_{ext}(t+s)} dt ds \quad (18)$$

4.3 Averages over the forward ray ensemble and correlations

For the above, we need contribution (or squared contribution) of a forward ray to the *given* integration sphere at point t averaged over the forward ray ensemble,

If we assume a forward ray hits the given integration sphere no more than once (which is so if the latter is small), then one ray contribution *if it hits the sphere* is (5) and 0 if it misses the sphere, so

$$\langle \mathcal{C} \rangle(t) = \mathcal{F}H_1(t) \quad (19)$$

$$\langle \mathcal{C}^2 \rangle(t) = \frac{\mathcal{F}^2}{\pi R^2} H_2(t) \quad (20)$$

where

$$H_m(t) \equiv \int_0^\pi \int_0^{2\pi} f^m \frac{F}{\mathcal{F}} \sin \vartheta d\vartheta d\varphi$$

and F is the angular distribution of radiance in the sphere center t .

Besides, we need correlations, i.e. average of the product of contributions of *the same* forward ray to two integration spheres with centers at s and at $s + t$, $\langle \mathcal{C}(s)\mathcal{C}(s+t) \rangle$.

The product $\mathcal{C}(s)\mathcal{C}(s+t)$ is not 0 only if the forward ray hits *both* spheres. In principle it is possible that it hits one sphere, then bounces somewhere, returns and hits the second one, but in case of small spheres it has too small probability (about $O(R^4)$), so we neglect this and consider only the case when the same *one segment* hits both spheres. The probability of that event equals the area of overlap of the projection of the two spheres onto the plane perpendicular to the segment times attenuation of the ray between the spheres. The above area is the of overlap of two circles of radius R whose centers are separated by $l \equiv t \sin \vartheta$ where ϑ is the angle between the view direction and the forward ray. This area is $\left(\arccos \frac{l}{2R} - \frac{l}{2R} \sqrt{1 - \left(\frac{l}{2R}\right)^2} \right) / \pi$ times lower than the whole sphere projection πR^2 , and attenuation of the ray is $e^{-\sigma_{ext} t |\cos \vartheta|}$, so the probability that $\mathcal{C}(s)\mathcal{C}(s+t) \neq 0$ is

$$R^2 \times \left(\arccos \frac{l}{2R} - \frac{l}{2R} \sqrt{1 - \left(\frac{l}{2R}\right)^2} \right) e^{-2\sigma_{ext} t |\cos \vartheta|}$$

and if this happens, $\mathcal{C}(s)\mathcal{C}(s+t)$ is the square of (5).

Recalling that $\langle \mathcal{C}^2 \rangle$ is that same squared (5) times probability of hitting the sphere i.e. πR^2 , we obtain

$$\langle \mathcal{C}(s)\mathcal{C}(s+t) \rangle = \frac{\mathcal{F}^2}{\pi^2 R^2} \int_0^{2\pi} \int_0^\pi A \left(\frac{t \sin \vartheta}{R} \right) e^{-\sigma_{ext} t |\cos \vartheta|} h(\vartheta, \varphi, p) \sin \vartheta d\vartheta d\varphi \quad (21)$$

where φ is the azimuthal angle of rotation about the view direction,

$$A(x) \equiv \begin{cases} \arccos x - x\sqrt{1-x^2}, & |x| \leq 1 \\ 0, & x > 1 \end{cases}$$

and

$$h \equiv f^2 \frac{F}{\mathcal{F}}$$

is taken at the center of sphere which the forward ray with direction (ϑ, φ) hits *first* i.e. at the camera ray point p

$$p = \begin{cases} s, & \cos \vartheta \geq 0 \\ t, & \cos \vartheta \leq 0 \end{cases} \quad (22)$$

4.4 The total noise

Substituting (15), (16), (17), (18), (19) and (20) into (7)–(9) we obtain

$$\begin{aligned} \langle\langle L^2 \rangle\rangle - \langle\langle L \rangle\rangle^2 &= q \frac{\int_0^\infty \langle \mathcal{C}^2 \rangle(t) \sigma_{ext} e^{-\sigma_{ext} t} dt}{N_F N_B} \\ &+ 2(1-q) \frac{\int_0^\infty \int_0^\infty \langle \mathcal{C}(s) \mathcal{C}(s+t) \rangle \sigma_{ext}^2 e^{-\sigma_{ext}(s+t)} ds dt}{N_F N_B} \\ &+ \frac{q}{N_B} \left(1 - \frac{1}{N_F}\right) \int_0^\infty \langle \mathcal{C} \rangle^2(t) \sigma_{ext} e^{-\sigma_{ext} t} dt \\ &+ \frac{2(1-q)}{N_B} \left(1 - \frac{1}{N_F}\right) \int_0^\infty \int_0^\infty \langle \mathcal{C} \rangle(s) \langle \mathcal{C} \rangle(s+t) \sigma_{ext}^2 e^{-\sigma_{ext}(s+t)} ds dt \\ &+ \frac{1}{N_F} \left(1 - \frac{1}{N_B}\right) \int_0^\infty \int_0^\infty \langle \mathcal{C}(s) \mathcal{C}(t) \rangle \sigma_{ext}^2 e^{-\sigma_{ext}(t+s)} dt ds \\ &- \frac{1}{N_F N_B} \left(\int_0^\infty \langle \mathcal{C} \rangle(t) \sigma_{ext} e^{-\sigma_{ext} t} dt \right)^2 \end{aligned}$$

It contains two integrals of correlations which for $R \rightarrow 0$ are estimated as (25) and (26).

Using (19) and (20), we have

$$\begin{aligned} \frac{\langle\langle L^2 \rangle\rangle - \langle\langle L \rangle\rangle^2}{\langle\langle L \rangle\rangle^2} &\approx \frac{1}{R^2 N_F} \left(\frac{q D_1 + \sigma_{ext} R (1-q) D_2}{N_B} + \left(1 - \frac{1}{N_B}\right) \sigma_{ext} R D_3 \right) \\ &+ \frac{1}{N_B} \left(1 - \frac{1}{N_F}\right) (q D_4 + (1-q) D_5) - \frac{1}{N_F N_B} \end{aligned}$$

where D_i are some constants composed from space integrals of

$$\begin{aligned}
 H_m(t) &\equiv \int_0^\pi \int_0^{2\pi} f^m \frac{F}{\mathcal{F}} \sin \vartheta d\vartheta d\varphi \\
 \tilde{H}_m(s) &\equiv \int_0^\pi \int_0^{2\pi} f^m \frac{F}{\mathcal{F}} d\vartheta d\varphi
 \end{aligned}$$

and are thus independent from any ray tracing parameters, i.e. q , R and the number of rays.

One can see that for small R but not very large number of rays the noise is smaller for $q < 1$ then for $q = 1$ i.e. for the standard method.

5 CONCLUSION

In previous work [16] we developed a method of gathering of photons in scattering medium which uses several integration spheres stochastically distributed over a camera ray segment. The resulting disjoint integration volume is intermediate between usual integration sphere (one per segment) and integration cylinder. In our method the control parameter q allows to vary integration volume from a single integration sphere to practically a cylinder which is a limit of the union of a large number of spheres.

We obtained analytical estimates of noise in the bi-directional MCRT and applied them to the particular case of calculation of luminance of turbid medium, suggested in [16]. It happened possible to derive a closed-form analytical expressions for the noise and its dependence on the method parameters. These can be used to find the optimal parameters and/or choose between single and multiple integration spheres or cylinders.

One must realize that the method (or parameter) that provides the lowest noise for the given *number of rays* is not always really the best. One reason is that if integration volumes occupy most of the medium domain, interaction of them with an FMCRT ray is slow and this seriously decelerate ray tracing, thus the number of rays traced in *the same time* of calculation drops and this increases noise. Second, the own luminance of the medium is not the only image component but there is also the luminance of objects “seen through” the medium. Usually both are calculated from with the same rays thus it may happen that although because of better integration volumes the noise of the “own” luminance of medium is still decreased in spite of the lower number of traced rays, the noise in the rest part of image increases because there is no such “compensation” for the rays count.

APPENDIX A. INTEGRALS OF CORRELATIONS

Let us calculate approximation for $R \rightarrow 0$ of

$$\int_0^\infty \int_0^\infty \langle \mathcal{L}(s)\mathcal{L}(s+t) \rangle \sigma_{ext}^2 e^{-\sigma_{ext}(t+s)} dt ds \quad (23)$$

and

$$\begin{aligned}
 & \int_0^\infty \int_0^\infty \langle \mathcal{C}(s)\mathcal{C}(t) \rangle \sigma_{ext}^2 e^{-\sigma_{ext}(s+t)} ds dt \\
 &= 2 \int_0^\infty \int_0^\infty \langle \mathcal{C}(s)\mathcal{C}(s+t) \rangle \sigma_{ext}^2 e^{-\sigma_{ext}(2s+t)} ds dt
 \end{aligned} \tag{24}$$

where $\langle \mathcal{C}(s)\mathcal{C}(s+t) \rangle$ is given by (21) and (22).

Both integrals are much similar and we begin with (23); as (24) differs from it only by the extra factor $e^{-\sigma_{ext}s}$ these calculations will apply to it as well. We thus begin with (23).

Let us subdivide interval of integration in ϑ into 3 parts: the ‘‘main’’ $[\Theta, \pi - \Theta]$ and two short intervals near the poles: $[0, \Theta]$ and $[\pi - \Theta, \pi]$. Then

$$\int_0^\infty \int_0^\infty \langle \mathcal{C}(s)\mathcal{C}(s+t) \rangle \sigma_{ext}^2 e^{-\sigma_{ext}(t+s)} dt ds = \frac{\mathcal{F}^2}{\pi^2 R^2} (I_1 + I_{2a} + I_{2b})$$

where

$$\begin{aligned}
 I_1 &\equiv \int_0^\infty \left(\int_0^\infty \left(\int_0^{2\pi} \int_\Theta^{\pi-\Theta} A\left(\frac{t \sin \vartheta}{R}\right) e^{-\sigma_{ext}t|\cos \vartheta|} h(\vartheta, \varphi, p) \sin \vartheta d\vartheta d\varphi \right) dt \right) ds \\
 I_{2a} &\equiv \int_0^\infty \left(\int_0^\infty \left(\int_0^{2\pi} \int_0^\Theta A\left(\frac{t \sin \vartheta}{R}\right) e^{-\sigma_{ext}t|\cos \vartheta|} h(\vartheta, \varphi, p) \sin \vartheta d\vartheta d\varphi \right) dt \right) ds \\
 I_{2b} &\equiv \int_0^\infty \left(\int_0^\infty \left(\int_0^{2\pi} \int_{\pi-\Theta}^\pi A\left(\frac{t \sin \vartheta}{R}\right) e^{-\sigma_{ext}t|\cos \vartheta|} h(\vartheta, \varphi, p) \sin \vartheta d\vartheta d\varphi \right) dt \right) ds
 \end{aligned}$$

The idea is to take such small Θ that in the ‘‘main’’ part I_1 the integration area in t is very narrow, so $p \approx s$ thus integration over t can be done analytically and after some tedious transformations we obtain a simple approximation to I_1 . For the rest parts near the poles, the range of t is wide, and the point p is s in I_{2a} and $s+t$ in I_{2b} . The range of integration in Θ is small which allows to estimate the integrals from above and it happens that for $R \rightarrow 0$ they are negligible as compared to I_1 , so the sought-for integral of correlations is close to I_1 for which we have a simple approximation. We shall see that the small angle Θ must be chosen so that it as $R \rightarrow 0$ it goes to 0 but *slower* than R so that $\Theta \rightarrow 0$ but $R/\Theta \rightarrow 0$.

Now let us apply the above intentions quantitatively.

As the integrand vanishes for $\frac{t \sin \vartheta}{R} > 1$, in I_1 the $t \leq R/\Theta \rightarrow 0$. Therefore the point $p \approx s$ even for $\cos \vartheta \leq 0$. Changing then the order of integration so that the one in t to be the first, we obtain

$$\begin{aligned}
 I_1 &\approx 2R \int_0^\infty \int_0^{2\pi} \int_\Theta^{\pi-\Theta} \left(\int_0^1 A(y) dy \right) h(\vartheta, \varphi, s) \sigma_{ext}^2 e^{-\sigma_{ext}s} d\vartheta d\varphi ds \\
 &= \frac{4R}{3} \int_0^\infty \int_0^{2\pi} \int_\Theta^{\pi-\Theta} h(\vartheta, \varphi, s) \sigma_{ext}^2 e^{-\sigma_{ext}s} d\vartheta d\varphi ds \\
 &\approx \frac{4R}{3} \int_0^\infty \left(\int_0^{2\pi} \int_0^\pi h(\vartheta, \varphi, s) d\vartheta d\varphi \right) \sigma_{ext}^2 e^{-\sigma_{ext}s} ds
 \end{aligned}$$

Now let us come to the integrals in the near-pole areas. Assuming illumination and phase function are not singular near the view direction, we can take h exactly at the pole direction. Then, we can replace $\sin \vartheta$ with ϑ or $\pi - \vartheta$:

$$I_{2a} \approx 8\pi \int_0^\infty \left(\int_0^\infty \left(\frac{R^2}{t^2} \int_0^{\min(\frac{t\Theta}{2R}, 1)} A(y)ydy \right) e^{-2\sigma_{ext}t} \sigma_{ext}^2 dt \right) e^{-\sigma_{ext}s} h(0, 0, s) ds$$

$$I_{2b} \approx 8\pi \int_0^\infty \left(\int_0^\infty \left(\frac{R^2}{t^2} \int_0^{\min(\frac{t\Theta}{2R}, 1)} A(y)ydy \right) e^{-2\sigma_{ext}t} \sigma_{ext}^2 h(\pi, 0, s+t) dt \right) e^{-\sigma_{ext}s} ds$$

Since

$$\int_0^X A(y)ydy = \frac{\frac{\pi}{2} - (1 - 4X^2) \arccos X - X(1 + 2X^2) \sqrt{1 - X^2}}{8}$$

with very good accuracy is limited by $\pi X^2/4$, we have

$$I_{2a} \leq \frac{\pi^2}{2} \int_0^\infty \left(\int_0^\infty \left(\min\left(\frac{2R}{t}, \Theta\right) \right)^2 e^{-2\sigma_{ext}t} \sigma_{ext}^2 dt \right) e^{-\sigma_{ext}s} h(0, 0, s) ds$$

$$\leq 2\pi^2 \Theta \sigma_{ext} R \int_0^\infty \sigma_{ext} e^{-\sigma_{ext}s} h(0, 0, s) ds$$

where we used the inequality $\int_X^\infty u^{-2} e^{-u} du = X^{-1} (e^{-X} - X \text{Ei}(1, X)) \leq X^{-1}$.

Similarly,

$$I_{2b} \leq \frac{\pi^2}{2} \int_0^\infty \left(\Theta^2 \int_0^{\frac{2R}{\Theta}} e^{-2\sigma_{ext}t} \sigma_{ext}^2 h(\pi, 0, s+t) dt \right. \\ \left. + 4R^2 \int_{\frac{2R}{\Theta}}^\infty t^{-2} e^{-2\sigma_{ext}t} \sigma_{ext}^2 h(\pi, 0, s+t) dt \right) e^{-\sigma_{ext}s} ds$$

The two its terms can be estimated as

$$\int_0^\infty \left(\int_0^{\frac{2R}{\Theta}} e^{-2\sigma_{ext}t} \sigma_{ext}^2 h(\pi, 0, s+t) dt \right) e^{-\sigma_{ext}s} ds$$

$$\leq \left(\int_0^\infty h(\pi, 0, s) \sigma_{ext} e^{-\sigma_{ext}s} ds \right) \left(1 - e^{-\frac{2\sigma_{ext}R}{\Theta}} \right)$$

$$\approx \frac{2\sigma_{ext}R}{\Theta} \left(\int_0^\infty h(\pi, 0, s) \sigma_{ext} e^{-\sigma_{ext}s} ds \right)$$

and

$$\begin{aligned}
 & \int_0^\infty \left(\int_{\frac{2R}{\Theta}}^\infty t^{-2} e^{-2\sigma_{ext}t} \sigma_{ext}^2 h(\pi, 0, s+t) dt \right) e^{-\sigma_{ext}s} ds \\
 & \leq \sigma_{ext}^2 \left(\int_{\frac{2\sigma_{ext}R}{\Theta}}^\infty t^{-2} e^{-t} dt \right) \left(\int_0^\infty h(\pi, 0, s) \sigma_{ext} e^{-\sigma_{ext}s} ds \right) \\
 & \leq \sigma_{ext} \frac{\Theta}{2R} \left(\int_0^\infty h(\pi, 0, s) \sigma_{ext} e^{-\sigma_{ext}s} ds \right)
 \end{aligned}$$

so

$$I_{2b} \leq 2\pi^2 \Theta \sigma_{ext} R \left(\int_0^\infty h(\pi, 0, s) \sigma_{ext} e^{-\sigma_{ext}s} ds \right)$$

Therefore for $R \rightarrow 0$ we can neglect $I_{2a} + I_{2b}$ as compared to I_1 , and obtain

$$\begin{aligned}
 & \int_0^\infty \int_0^\infty \langle \mathcal{C}(s) \mathcal{C}(s+t) \rangle \sigma_{ext}^2 e^{-\sigma_{ext}(t+s)} dt ds \\
 & \approx \frac{4\sigma_{ext} \mathcal{F}^2}{3\pi^2 R} \int_0^\infty \left(\int_0^{2\pi} \int_0^\pi h(\vartheta, \varphi, p) d\vartheta d\varphi \right) \sigma_{ext} e^{-\sigma_{ext}s} ds \quad (25)
 \end{aligned}$$

The second integral, (24) differs only by the extra factor $e^{-\sigma_{ext}s}$. Therefore the calculation done above applies to it as well and give in this case

$$\begin{aligned}
 & \int_0^\infty \int_0^\infty \langle \mathcal{C}(s) \mathcal{C}(t) \rangle \sigma_{ext}^2 e^{-\sigma_{ext}(s+t)} ds dt \\
 & \approx \frac{8\sigma_{ext} \mathcal{F}^2}{3\pi^2 R} \int_0^\infty \left(\int_0^{2\pi} \int_0^\pi h(\vartheta, \varphi, p) d\vartheta d\varphi \right) \sigma_{ext} e^{-2\sigma_{ext}s} ds \quad (26)
 \end{aligned}$$

REFERENCES

- [1] E. P. Lafortune and Y. D. Willems, “Bi-directional path tracing”, *Proceedings of Third International Conference on Computational Graphics and Visualization Techniques (Compugraphics '93)*, Alvor, Portugal, December 1993, 145–153 (1993).
- [2] R. A. Redner, M. E. Lee, and S. P. Uselton, “Smooth b-spline illumination maps for bidirectional ray tracing”, *ACM Trans. Graph.*, **14**(4), 337–362 (1995). (Online). Available: <http://doi.acm.org/10.1145/225294.225296>
- [3] A. G. Voloboi, V. A. Galaktionov, K. Dmitriev, and E. A. Kopylov, “Bidirectional ray tracing for the integration of illumination by the quasi-Monte Carlo method”, *Programming and Computer Software*, **30**(5), 258–265 (2004). (Online). Available: <https://doi.org/10.1023/B:PACS.0000043051.47471.aa>
- [4] V. A. Frolov, A. A. Kharlamov, and A. V. Ignatenko, “Biased solution of integral illumination equation via irradiance caching and path tracing on GPUs”, *Programming and Computer Software*, **37**(5), 252–259 (2011). (Online). Available: <https://doi.org/10.1134/S0361768811050021>

-
- [5] W. Jarosz, M. Zwicker, and H. W. Jensen, “The Beam Radiance Estimate for Volumetric Photon Mapping”, *Computer Graphics Forum*, **27**(2), 557–566 (2008).
- [6] D. D. Zhdanov, A. A. Garbul, I. S. Potemin, A. G. Voloboy, V. A. Galaktionov, S. V. Ershov, and V. G. Sokolov, “Photorealistic volume scattering model in the bidirectional stochastic ray tracing problem”, *Programming and Computer Software*, **41**(5), 295–301 (2015). (Online). Available: <https://doi.org/10.1134/S0361768815050102>
- [7] T. Hachisuka and H. W. Jensen, “Stochastic progressive photon mapping”, *ACM Trans. Graph.*, **28**(5), 141:1–141:8 (2009). (Online). Available: <http://doi.acm.org/10.1145/1618452.1618487>
- [8] V. Havran, R. Herzog, and H.-P. Seidel, “Fast Final Gathering via Reverse Photon Mapping”, *Computer Graphics Forum*, **24**(3), 323–333 (2005).
- [9] H. W. Jensen, “Global illumination using photon maps”, *Proceedings of the Eurographics Workshop on Rendering Techniques '96*. London, UK: Springer-Verlag, 1996, 21–30 (1996). (Online). Available: <http://dl.acm.org/citation.cfm?id=275458.275461>
- [10] S. V. Ershov, D. D. Zhdanov, A. G. Voloboy, and M. I. Sorokin, “Treating diffuse elements as quasi-specular to reduce noise in bi-directional ray tracing”, *Preprint IPM*, no. 122 (2018). (Online). Available: <https://doi.org/110.20948/prepr-2018-122-e>
- [11] E. P. Lafortune and Y. D. Willems, “Rendering participating media with bidirectional path tracing”, in *Rendering Techniques '96*, X. Pueyo and P. Schröder, Eds. Vienna: Springer Vienna, 91–100 (1996).
- [12] W. Jarosz, D. Nowrouzezahrai, I. Sadeghi, and H. W. Jensen, “A comprehensive theory of volumetric radiance estimation using photon points and beams”, *ACM Trans. Graph.*, **30**(1), 5:1–5:19, (2011). (Online). Available: <http://doi.acm.org/10.1145/1899404.1899409>
- [13] J. Křivánek, I. Georgiev, T. Hachisuka, P. Vévoda, M. Šik, D. Nowrouzezahrai, and W. Jarosz, “Unifying points, beams, and paths in volumetric light transport simulation”, *ACM Trans. Graph.*, **33**(4), 1–13 (2014).
- [14] W. Jarosz, D. Nowrouzezahrai, R. Thomas, P.-P. Sloan, and M. Zwicker, “Progressive photon beams”, *ACM Trans. Graph.*, **30**(6), 181:1–181:12 (2011). (Online). Available: <http://doi.acm.org/10.1145/2070781.2024215>
- [15] B. Bitterli and W. Jarosz, “Beyond points and beams: higher-dimensional photon samples for volumetric light transport”, *ACM Trans. Graph.*, **36**(4), 112:1–112:12 (2017). (Online). Available: <https://doi.org/10.1145/3072959.3073698>
- [16] S. V. Ershov, D. D. Zhdanov, and A. G. Voloboy, “Calculation of luminance of scattering medium by mcert using multiple integration spheres”, *Math. Montisnigri*, **44**, 122–134 (2019). (Online). Available: <https://doi.org/10.20948/mathmon-2019-44-10>
- [17] T. Hachisuka, W. Jarosz, and H. W. Jensen, “A progressive error estimation framework for photon density estimation”, *ACM Trans. Graph.*, **29**(6), 144:1–144:12 (2010). (Online). Available: <http://doi.acm.org/10.1145/1882261.1866170>
- [18] C. Knaus and M. Zwicker, “Progressive photon mapping: A probabilistic approach”, *ACM Trans. Graph.*, **30**(3), 25:1–25:13 (2011). (Online). Available: <http://doi.acm.org/10.1145/1966394.1966404>

Received June 10, 2019

HYDRODYNAMIC CHARACTERISTICS OF WEAKLY CONDUCTIVE LIQUID MEDIA IN THE NON-UNIFORM ELECTRIC FIELD

M. S. APFELBAUM AND A. N. DOLUDENKO

Joint Institute for High Temperatures of the Russian Academy of Sciences
Izhorskaya 13 Bldg 2, 125412 Moscow, Russia
e-mail: adoludenko@gmail.com

DOI: 10.20948/mathmontis-2019-45-6

Summary. Theoretical model of the ions formation in a liquid dielectric and flows caused by high electric field is proposed. The three-dimensional system of macroscopic pre-breakdown electro-hydrodynamic equations is written. The influence of electric field on the molecule dissociation rate is taken into account. The system includes the Poisson equation for electric field potential, equation of ion formation and the Navier–Stokes equations with the electric force. Author’s steady analytical electrodynamic solution of these equations for the electric field distribution and potential of spherical high voltage capacitor with liquid transformer oil type dielectric is described. Analytical non-stationary and numerical steady solutions for velocity distributions in liquid dielectric flows are obtained.

1 INTRODUCTION

Deviations from the Ohm’s law for slightly ionized solid media in pre-breakdown uniform electric fields were experimentally discovered by Poole about 100 years ago [1]. In a weakly conductive liquid media the same experimental effect was obtained by M. Wien about 10 years later [2, 3].

Theoretically this exponential effect for considerable number of media was explained by Frenkel for solid dielectrics and by Onsager [3] for liquid weak electrolytes and for weakly conductive liquid dielectrics. The space charge and electro-hydrodynamic (EHD) flows have been observed in these dielectrics at the pre-breakdown conditions [4].

The space charge formation, according to [5], occurs in the pre-breakdown fields until all the EHD characteristics become steady. Herein, steady conduction can be as unipolar (corona-discharge type), as quasi-neutral (plasma or electrolyte type). The last was considered early in [5] and in the present work.

The pre-breakdown current–voltage theoretical and experimental characteristics of considerable media in non-uniform electric fields are described by us in [6]. Purpose of present work is researching of the electro-hydrodynamic flows, caused by these high non-uniform electric fields. These intense flows are observed in transformer oil type liquids [7, 8] with complex molecular structure [9]. The hydrodynamic transfer of high voltage space charge, appeared in considerable liquids, is described in [10]. In the review [8] the surface high voltage electrode effects influence on considerable pre-breakdown electro-hydrodynamic flows is researched. This

2010 Mathematics Subject Classification: 76W05, 76D05, 76-04.

Key words and phrases: electro-hydrodynamics, liquid insulator, weakly conductive liquid media.

influence must be taken into account at plan high voltage electrodes. When applied high voltage field is non-uniform, the volume effects influence on considerable flows is dominated.

The transport properties of charged and neutral components of weakly conducting media under consideration are investigated in [11, 12]. In [11], using the quantum formula Kubo, a new method for calculating the conductivity of dense media has been developed. The calculations were carried out using the Monte-Carlo method as well as the molecular dynamics method. The Wigner–Liouville equation was used. The graphs of dependence of the studied media electrical conductivity on the initial parameters and their plasma non-ideality are presented. In [12], the thermal conductivity of the silicon electron gas is simulated, using the corresponding Fermi–Dirac distributions. Coefficients of thermal conductivity are obtained, and their dependence on temperature is presented graphically.

The transfer processes in the considered irreversible thermodynamics are used in [13]. The approach of the continuous medium and the local thermodynamic equilibrium are used. The transfer equations are derived with recommendations for their use in the analysis of thermal conductivity, electrical conductivity and diffusion. The formation of mixtures of liquids is used when applying a simple mathematical model from the equations of irreversible thermodynamics. The Navier–Stokes and heat inflow equations are solved. The formation of droplets during the evaporation of a liquid is substantiated [14].

In [15], the equations of a low-density plasma magnetosphere are written in a strong magnetic field. The Vlasov equation of kinetic is used. The equation of the magnetohydrodynamic mixture motion is derived. This method is used to study the evolution of the plasma under consideration.

2 THEORETICAL MODEL OF ELECTRO-HYDRODYNAMIC PROCESSES IN THE LIQUID INSULATOR

The following inequalities are valid for bulk charge concentrations n_{\pm} , impurities concentrations n_p and neutrals n_a concentrations in a low-conducting medium:

$$n_{\pm} \ll n_a, \quad n_p \ll n_a. \quad (1)$$

The above relations can be considered as a condition for a weak ionization (dissociation) or low conductivity in the considered medium. The rates of volume ionization (or dissociation) and recombination are supposed to be known as the thermodynamics functions of the above mentioned concentrations, temperature and field intensity $|\mathbf{E}|$ (up to the breakdown values). They can be represented in the following form:

$$W_i = W_i(n_a, n_p, T, |\mathbf{E}|) = W_i(n_a, n_p, T, 0)f(|\mathbf{E}|). \quad (2)$$

Here, T is the absolute temperature; $|\mathbf{E}|$ is the intensity of the electric field. The rate of the ions recombination is

$$\begin{cases} W_r = K_r n_+ n_-, \\ K_r = \frac{(b_+ + b_-) |e| Z}{\varepsilon \varepsilon_0}; \end{cases} \quad (3)$$

W_i denotes the rate of ionization (dissociation); W_r is the rate of ion recombination; Z is the valence; ε is the dielectric constant of liquid insulator; b_+ and b_- are the mobilities of the corresponding ions. The expression for K_r (recombination constant) was obtained by Langevin in 1903 and by Onsager in 1934 for particular case of weak electrolytes. This is the two-particle ion–ion recombination, when the energy excess is absorbed by a medium.

Onsager had shown weak influence of sub-breakdown electrical fields on K_r when direct and reverse processes are ionization and recombination correspondingly. But in the case when these processes are dissociation and reverse dissociation the influence of \mathbf{E} on K_r can not be neglected. Besides weak electrolytes, the Langevin's formula is valid for dense gases with chemical reactions, where ionic conductivity is much greater than electronic one. Corresponding conductivity occurs due to various processes.

Among them, there are the neutral molecules (with ionic chemical bonds) dissociation and the neutral molecules (with covalent chemical bonds) ionization. The latter goes through the electron ionization from some neutral particle with relatively low ionization potential and attaches it to another with relatively high electron affinity.

For the ion diffusion coefficient, we have used more known Einstein–Nernst relation:

$$ZD_{\pm} = \frac{k_B T b_{\pm}}{|e|}, \quad (4)$$

where k_B is the Boltzmann constant, T is supposed to be a constant and equals to 300 K, because the Joule heat is small for the pre-breakdown phenomena.

Function $f(\mathbf{E})$ describes the dependence of the ionization (dissociation) rate on the electric field intensity. The expression, describing it, was obtained by Frenkel for the solid dielectric and generalized by Ostroumov in [4] for the case when $Z > 1$ with ion chemical bonds in molecules. It was used in the present research, and has the form

$$f(\mathbf{E}) = \exp(\beta |\mathbf{E}|^{1/2}), \quad \beta = \frac{|Ze|^{3/2}}{\sqrt{\pi \varepsilon \varepsilon_0 k_B T}}, \quad Z = 2, \quad \varepsilon = 2.5. \quad (5)$$

The value of valence Z for dissociating molecules of transformer oil was taken from [9], and the dielectric constant of transformer oil ε was taken from [4]. This function f of the high field shock ionization rate is not well known for gases.

The equations, describing creation and annihilation of the space charge, high voltage conductivity and the electric field distribution can be written as [5]

$$\begin{cases} \frac{\partial q}{\partial t} + (\mathbf{V}, \nabla q) - \frac{k_B T b}{|Ze|} \Delta q + (\mathbf{E}, \nabla \sigma) = -\frac{q\sigma}{\varepsilon \varepsilon_0}, & \Delta \phi = -\frac{q}{\varepsilon \varepsilon_0}, \\ \frac{\partial \sigma}{\partial t} + (\mathbf{V}, \nabla \sigma) + b^2 (\mathbf{E}, \nabla q) - \frac{k_B T b \Delta \sigma}{|Ze| Z} - \frac{\sigma_0^2}{\varepsilon \varepsilon_0} \exp(\beta |\mathbf{E}|^{1/2}) + \frac{\sigma^2}{\varepsilon \varepsilon_0} = 0, \end{cases} \quad (6)$$

where q is the bulk charge, \mathbf{V} is the velocity vector, σ is the electrical conductivity.

The well-known hydrodynamics equations should be added to equations (6) to construct the closed system of equations. The first of them is the continuity equation for incompressible media

$$\operatorname{div} \mathbf{V} = 0. \quad (7)$$

We used the equation (7) to obtain electrodynamic equations (6). For our purpose it is enough to consider the law of momentum conservation without considering the law of energy conservation. The momentum conservation equation can be written as

$$\begin{cases} \rho \left(\frac{\partial u}{\partial t} + u \left(\frac{\partial u}{\partial x} \right) + v \left(\frac{\partial u}{\partial y} \right) \right) = -\frac{\partial p}{\partial x} + \mu \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + q E_x, \\ \rho \left(\frac{\partial v}{\partial t} + u \left(\frac{\partial v}{\partial x} \right) + v \left(\frac{\partial v}{\partial y} \right) \right) = -\frac{\partial p}{\partial y} + \mu \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) + q E_y. \end{cases} \quad (8)$$

Some properties of oil and the initial conditions for the system of equations (6)–(8) are

$$\begin{aligned} \mu &= 0.00016 \text{ Pa s}, & \rho &= 800 \text{ kg/m}^3, & \sigma_0 &= 10^{-13} \text{ Ohm}^{-1} \text{ cm}^{-1}, \\ P|_{t=0} &= P_0 = 10^5 \text{ Pa}, & q|_{t=0} &= 0, & V|_{t=0} &= 0. \end{aligned} \quad (9)$$

Upon further calculation, the following boundary conditions were adopted: the condition of adhesion on the surface of the electrode-wire and flat electrodes was set:

$$\mathbf{V}|_{\partial \Delta} = 0,$$

where $\partial \Delta$ is the boundary of the electrodes.

At the remaining boundaries of the computational domain, the condition of free flow was set:

$$\left. \frac{\partial \mathbf{V}}{\partial \mathbf{n}} \right|_{\partial \Omega} = 0,$$

where \mathbf{n} is the external unit normal to the boundary $\partial \Omega$ of the calculation domain Ω .

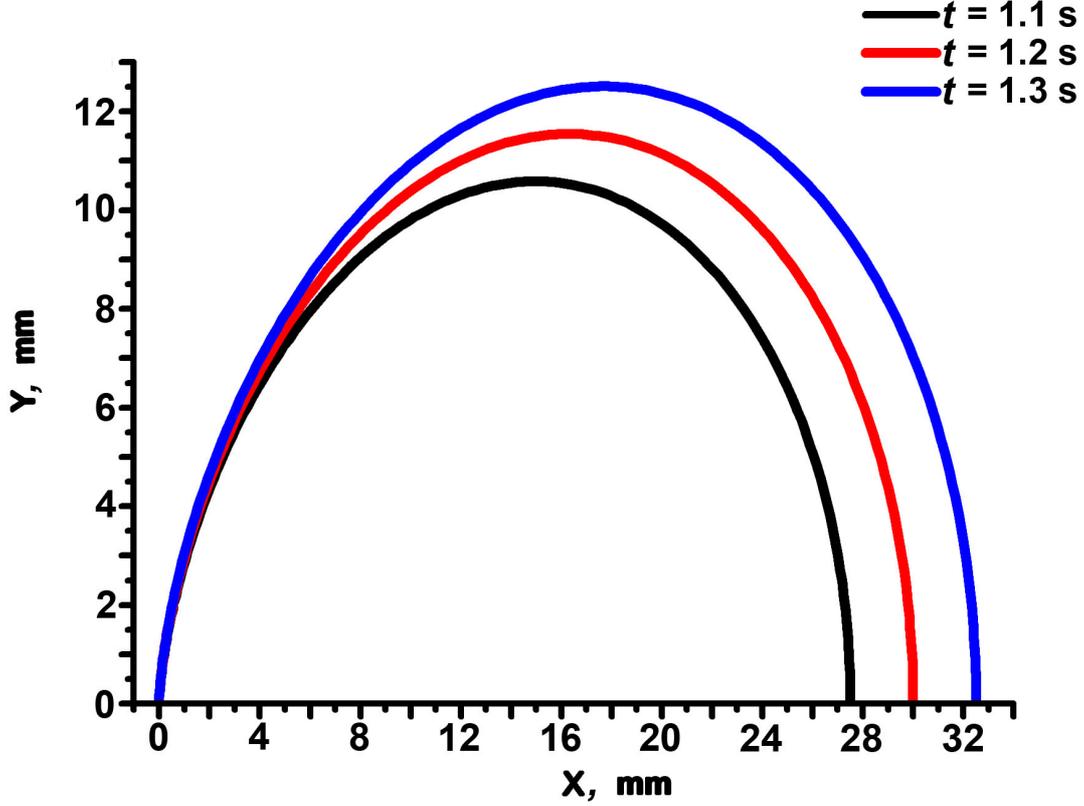


Figure 1: Graphics of the constant flow jets (12) for $U = 4$ kV and debit 0.1 L/s for different time moments.

The quasi-stationary analytical electrodynamic solution of equations (6) for the spherical symmetry electrical potential distribution in quasi-neutral medium can be obtained from charge conservation law (this well-known law can be obtained from equations (6) too). This solution looks like

$$\begin{aligned} \phi(r) &= \left(\frac{I}{4\pi\sigma_0} \right)^{1/2} \left[\frac{8}{\beta} - \sqrt{\frac{\varepsilon\varepsilon_0}{\tau\sigma_0}} \left(|\mathbf{E}|^{1/2} + \frac{8}{\beta} \right) \right] \text{sign}(\phi(r_0)), \\ |\mathbf{E}| \exp\left(\frac{\beta}{2}|\mathbf{E}|^{1/2}\right) &= \frac{I}{4\pi\sigma_0 r^2}, \quad |\phi(r_0)| = U. \end{aligned} \quad (10)$$

This quasi-neutral solution (10) is zero approaching of hydrodynamic space charge transfer differential operator series [10]

$$q = \sum_{i=0}^{\infty} (\tau\nabla)^i \sigma \mathbf{E} \nabla \tau, \quad (11)$$

where τ is the charge relaxation time. Mathematical space of differential operators, obtained in [10], is not the Banach one. The quasi-exponential dependence as dependence [6] for volt-ampere pre-breakdown characteristics calculations of high voltage spherical capacitors can be

obtained from (10). These deviations were explained in [3] for plan capacitors early. The Laplas condition of pre-breakdown electric field, obtained in [7], can be also obtained from (10).

In contrary to the unipolar conduction solutions of our equations (6), the analytical formulae for pre-breakdown volt-ampere characteristics can be obtained analytically only for the case of cylindrical symmetry. This solution was obtained for slightly ionized gases in [16] early. The non-stationary hydrodynamic solution of equations (6)–(8) with the use of (9) and (10) for development of weakly conductive liquid jet flows from high voltage pinpoint electrode is

$$\Psi = \frac{\varepsilon\varepsilon_0 U^2 (t - \tau) \sin^2 \theta}{32\pi\rho r}. \quad (12)$$

The graph of solution (12) is shown in figure 1. According to this graph, the jet flows are more developed, when time increases.

3 NUMERICAL CALCULATIONS

Numerical calculations of the isothermal two-dimensional system of equations (7)–(10) were carried out for the “wire above the plane” configuration of electrodes mentioned above. The need to determine the maximum speed of the fluid is important task due to the cooling problem of low-power non-pressure 10 kV transformers. We believe that the voltage inside such a transformer is proportional to the generated one and can reach value of 500 V and higher. The method applied and solving system of equations was described in detail earlier [17].

Calculations, mentioned in this paper, were carried out for the wire with square cross-section. This was done because the transverse dimension of the wire is much smaller than the linear dimensions of the calculation region. This region has to be not less than the distance from the wire to the flat electrode, and, in its turn, should be quite large to minimize the influence of boundary conditions on the calculation. As a compromise on the speed of calculation and flow details outside the region of wire electrode such a grid was chosen, in which the size of each side of the wire electrode equals to 8 computational cells.

When conducting full-scale experiments and in reality the cross section of wire electrode is a circle. There is a question: how reliable are the simulation results with the square-shape electrode and how much are they differ from the simulation results with a round-shape electrode? To determine this, four calculations were carried out: two on a coarse grid, figure 2(*a*, *b*), with a wire of square and “round” (as far as possible) cross-sections, and two on a finer grid with a similar sectional view, figure 2(*c*, *d*). The calculations were carried out until the flow was completely established.

Coarse grid parameters: 100×100 cells, cell size equals to 0.0005 m or 0.5 mm. Parameters of the finer grid: 500×500 cells, cell size equals to 0.0001 m or 0.1 mm.

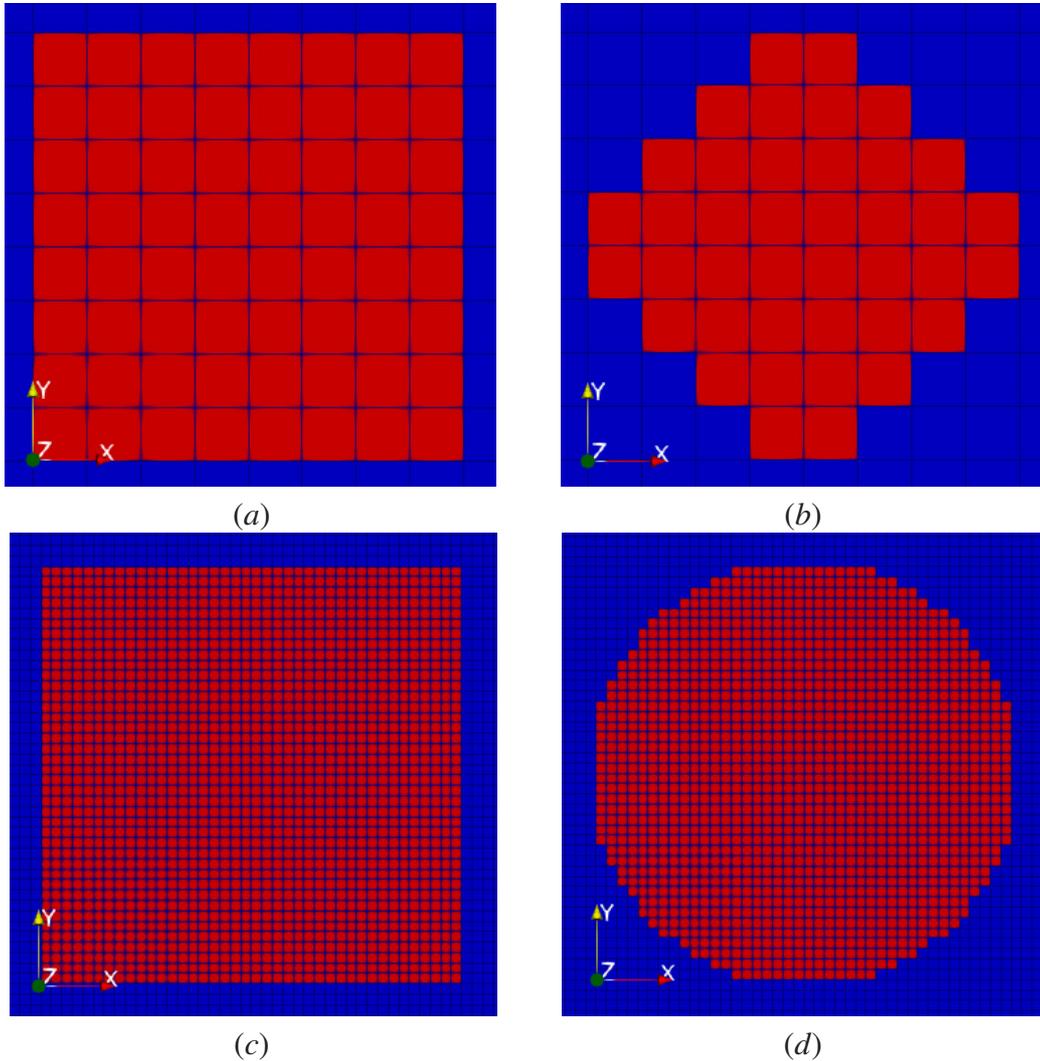


Figure 2: Square (*a, c*) and round (*b, d*) wire cross-sections on a coarse (*a, b*) and more detailed (*c, d*) grids.

Figure 3 demonstrates the axial velocities of fluid motion between the electrode-wire and the electrode-plane without taking into account the effect of the bulk charge. Voltage between the electrodes equals to 500 V. Inter-electrode distance equals to 2 cm. Different configuration of the wire cross-section and various fineness of the calculation grid are presented without taking into account the effect of space charge. One can see in figure 3 that the result has not changed qualitatively, but the shape of the electrode affects the quantitative result quite strongly. The maximum flow speed on the straight line connecting the electrodes increases with increasing fineness of the grid. In addition, when calculating on a fine grid in the region of a square-section electrode, the reverse flow is noticeable at a distance of 18 to 20 mm from the flat electrode, shown by the arrow in figures 3 and 4. Besides, the maximum of fluid velocity is greater if the shape of the wire cross-section is close to rounded.

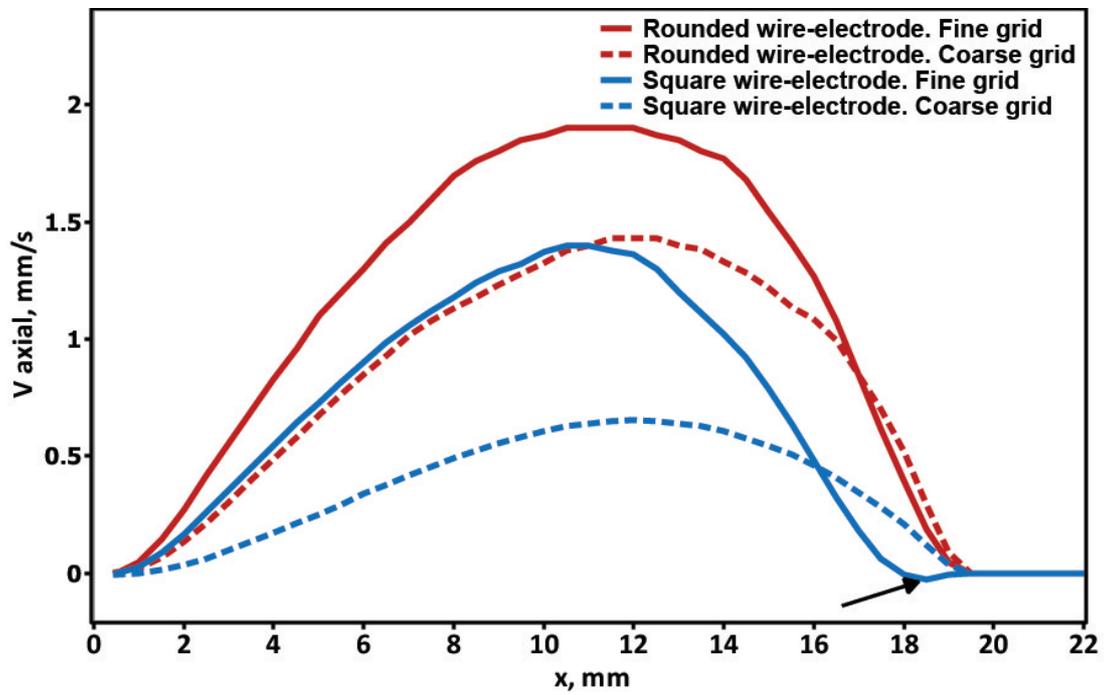


Figure 3: Distribution of the axial velocity between the electrodes in the electrode system “wire above the plane” in the axisymmetric electro-hydrodynamic flow of a weakly conducting liquid.

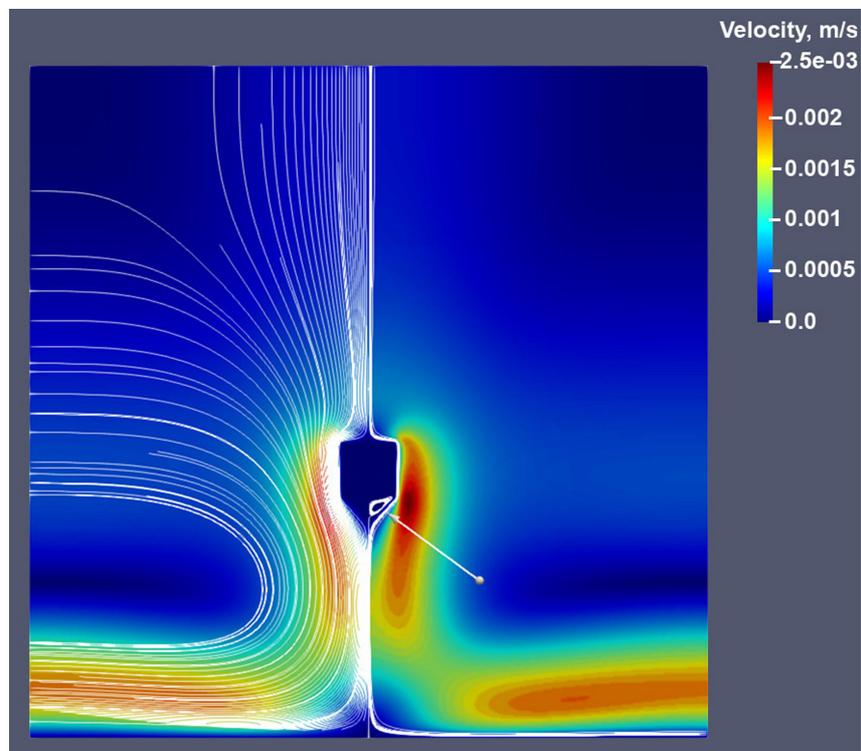


Figure 4: Streamlines and magnitude of the velocity for a weakly conducting liquid in an electric field without taking into account the effect of space charge. The cross section of the wire-electrode is square.

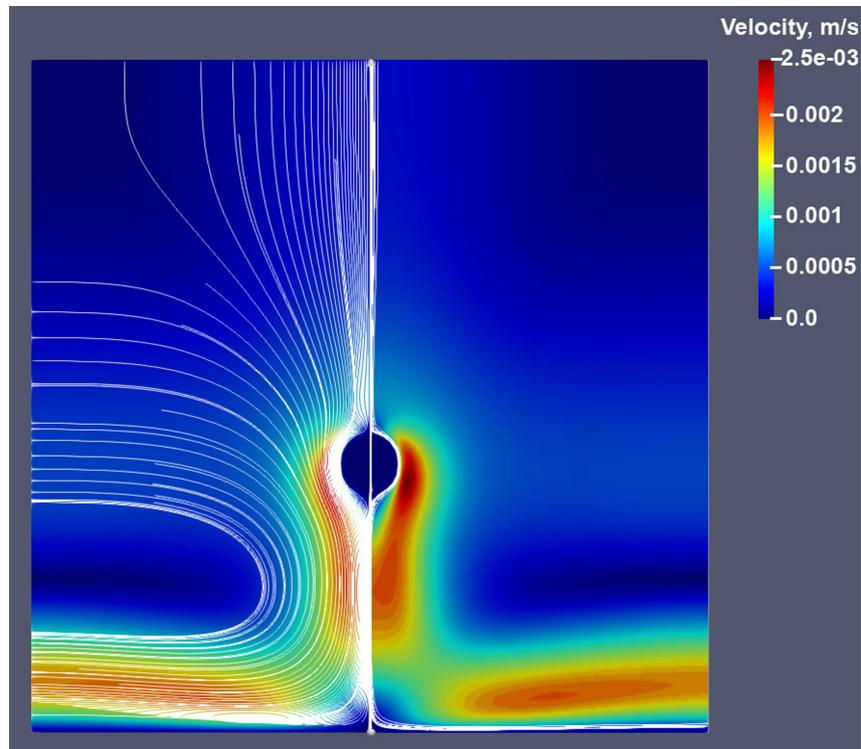


Figure 5: Same as in figure 4 but for the round cross section of the wire-electrode.

Figures 4 and 5 present comparative pictures of the motion of a weakly conducting liquid in the region of electrodes at a voltage of 500 V using a grid of 500×500 cells. It can be seen that in a case of the electrode with a circular cross section, the jets on both sides of the electrode are combined in a single jet at a distance of 10 mm from the flat electrode. In a case of the quadrate electrode, these two streams remain disconnected. It is also evident that the maximum fluid velocities are observed not in the central jet on the connection line of the two electrodes, but on both sides of the electrode-wire. And, as it was already mentioned earlier, there is a noticeable upward flow near the square electrode. This is not observed near the electrode with circular cross-section.

4 CONCLUSIONS

- For the pre-breakdown volt-ampere characteristics of transformer oil and liquid heptane, the squared current–voltage dependencies are obtained, and with decreasing electrode distance they become quasi-exponential. In addition, as the electrode distance decreases, the pre-breakdown current increases; this agrees with the results of the experiments.
- This work shows how fineness of the calculation grid and the cross-section of the wire electrode influence on the weakly conducting liquid flow in the inter-electrode space.

Flow structures and velocity distribution are obtained both for square and round shape of the wire electrode. The maximum axial velocity between the electrodes increases if the wire with the round cross-section is considered. Besides that, it is shown that this maximum axial velocity increases while using finer calculation grid. Further conclusions about what cross-section form should be chosen can be made after conducting relevant experiments.

Acknowledgments: This work was partly supported by the Russian Foundation for Basic Research (project No. 18-08-00136).

The paper is based on the proceedings of the XXXIII International Conference on Equations of State for Matter, which was held in Elbrus and Tegenekli settlements, in the Kabardino-Balkar Republic of the Russian Federation, from March 1 to 6, 2018.

REFERENCES

- [1] H. H. Poole, "On the dielectric constant and electrical conductivity of mica in intense field", *Philos. Mag.*, **2**(187), 112–120 (1916).
- [2] M. Wien, "Über die abweichungen der electrolyte vom ohmschen gesetz", *Phys. Z.*, **29**, 751 (1928).
- [3] L. Onsager, "Deviations from Ohm's law in a weak electrolytes", *J. Chem. Phys.*, **2**(9), 599–615 (1934).
- [4] G. A. Ostroumov, *Vzaimodejstvie e'lektricheskix i gidrodinamicheskix polej: fizicheskie osnovy e'lektrogidrodinamiki*, Moscow: Nauka, (1979).
- [5] M. S. Apfelbaum, "Ob odnoj raschetnoj sxeme e'lektrogidrodinamicheskix techenij", *Sovetskaja E'lektrohimija*, **11**, 1463–1471 (1986).
- [6] M. S. Apfelbaum, V. I. Vladimirov, and V. Ya. Pecherkin, "Stationary pre-breakdown volt-ampere characteristics of a weakly conductive dielectric liquids and slightly ionized gases", *Surface Engineering and Applied Electrochemistry*, **51**(3), 260–268 (2015).
- [7] M. S. Apfelbaum and V. A. Polyansky, "Ob obrazovanii ob'emnogo zaryada v slaboprovodyashix sredax", *Magn. Gidrodin.*, (1), 71–76 (1982).
- [8] A. I. Zhakin, "Ionnaya e'lektroprovodnost' i kompleksobrazovanie v zhidkix die'lektrikax", *Phys. Usp.*, **173**(1), 51–67 (2003).
- [9] V. G. Arakelyan, *Physical and Chemical Bases of Operation of Oil-Filled Electrical Equipment: Reference Data, Analysis, Research, Diagnostics, Monitoring*, Moscow: Tetraprint, (2012).
- [10] M. S. Apfelbaum, "O primenenii rezol'ventny'x operatorov v e'lektrogidrodinamike", *Elektron. Obrab. Mater.*, **23**(2), 59–62 (1987).
- [11] V. S. Filinov and A. S. Larkin, "Electrical conductivity of strongly plasma media", *Mathematica Montisnigri*, **41**, 112–118 (2018).
- [12] O. N. Koroleva and A. V. Mazhukin, "Determination of thermal conductivity and heat capacity of silicon electron gas", *Mathematica Montisnigri*, **40**, 99–109 (2017).
- [13] A. G. Kaptilnyy and A. A. Karabutov, "Transport processes in extended non-equilibrium thermodynamics", *Mathematica Montisnigri*, **39**, 86–103 (2016).
- [14] A. A. Sviridenkov, P. D. Toktaliev, and V. V. Tretyakov, "Numerical simulation of heat and mass transfer, mixture formation in combustion of gas turbine", *Mathematica Montisnigri*, **40**, 127–139 (2017).
- [15] A. V. Kolesnichenko, "To description of motion of rarefied magnetospheric plasma in a strong magnetic field", *Mathematica Montisnigri*, **41**, 131–150 (2018).
- [16] N. N. Tikhodeev, "Differencialnoe uravnenie unipoliarnoi korony i ego integrirovanie v prostesih sluchaiakh", *Zh. Tekh. Fiz.*, **25**, 1449–1457 (1955).

- [17] M. S. Apfelbaum and A. N. Doludenko, “The formation of weakly conductive liquid dielectric flows caused by a pre-breakdown electric field”, *Mathematica Montisnigri*, **40**, 90–98 (2017).

Received November 1, 2018

MODELING OF THERMAL CONDUCTIVITY OF SI IN THE RANGE FROM THE NORMAL TO NEAR-CRITICAL CONDITIONS

O.N. KOROLEVA^{1,2*}, M.M. DEMIN¹, V.I. MAZHUKIN^{1,2}, A.V. MAZHUKIN^{1,2}

¹Keldysh Institute of Applied Mathematics of RAS, Russia, Moscow

²National Research Nuclear University MEPhI, Russia, Moscow

* Corresponding author. E-mail: koroleva.on@mail.ru

DOI: 10.20948/mathmontis-2019-45-7

Summary. In a wide temperature range, including the semiconductor-metal phase transition region and the near-critical region, the results of modeling the silicon phonon thermal conductivity are presented. Since the transfer of thermal energy is carried out by phonons and free charge carriers, it is necessary to take into account both the contribution of phonons and electrons in the total thermal conductivity. In contrast to metals, heat transfer in silicon in the solid state is determined by phonon thermal conductivity. Although the contribution of the electronic component to the total thermal conductivity increases with increasing temperature, the inclusion of phonon thermal conductivity is of particular importance in liquid silicon. At higher temperatures, phonon thermal conductivity plays an important role in the modeling of the mechanisms of interaction of pulsed laser radiation with silicon in the framework of the two-temperature continuum model. Obtaining the temperature dependence of phonon thermal conductivity in such a wide temperature range from experiment is problematic. In this work, phonon thermal conductivity was obtained in the range $300 \leq T \leq 6500$ K from molecular dynamics simulation using the KIHS potential.

1 INTRODUCTION

Over the past decades, laser processing of silicon by short pulses has been developing rapidly. The possibility of introducing of a large amount of energy into a densely localized region through laser irradiation has been used in a large number of applications. Among them are such as the production of nanomaterials [1–4], surface nanostructuring [5, 6], chemical and physical synthesis [6–8]. Therefore, the fundamental mechanisms of the interaction of short-pulse laser radiation with a silicon target cause a long-term steady interest of researchers. A recognized tool for the theoretical study of laser interaction with a target is mathematical modeling. The problem of the adequacy and reliability of simulation results is associated with determining the properties of the target. The information on the temperature dependence of thermal conductivity is a necessary characteristic of a target when modeling laser heating of a substance using continuum models [9]. It is known that in accordance with the classical concepts, in a solid there are two mechanisms of heat transfer: elastic lattice vibrations and free electrons, therefore, thermal conductivity can be represented by lattice and electronic components. However, the contributions of phonon and electronic thermal conductivities at high temperatures are difficult to obtain from experiment. Therefore, in this work, the study of the thermal conductivity of the silicon lattice is carried out in the

2010 Mathematics Subject Classification: 82C26, 68U20, 74A15.

Key words and Phrases: Molecular Dynamics Simulation, Near-Critical Region, Phonon Thermal Conductivity of Silicon.

framework of the atomistic approach, which gives an understanding of the thermal behavior of silicon at the nanoscale.

Atomistic models are a system of differential equations that describes a set of interacting particles (atoms, ions, molecules). When using atomistic models to study various properties of substances, the most important role is played by the choice of the interaction potential between the particles, since the reliability of the results obtained directly depends on it.

The construction of the interatomic interaction potential for silicon has a number of features and is more complicated than for metals. Therefore, in order to determine the applicability of the selected interaction potentials in certain particular conditions, careful test calculations are required. This problem is especially acute in materials with covalent bonds, which include silicon. In [10], based on a comparison of the interaction potentials, it was found that the most suitable for modeling silicon are the Stillinger – Weber potentials (SW) and KIHS [11-13]. In this study, the KIHS potential was used to simulate the phonon thermal conductivity of silicon.

The thermal conductivity of phonons of single-crystal silicon was simulated over a wide range of temperature values using the widespread LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) application package [14]. It implements support for many pair and many-particle short-range potentials, it is possible to write atomic configurations to a text file, and thermostats and barostats are built-in. The computational algorithm is based on the Verlet finite-difference scheme [15]. The velocity and pressure for the ensemble of particles were adjusted using a thermostat and a Berendsen barostat [16].

The determination of phonon thermal conductivity in the framework of classical molecular dynamics is a complex problem. Despite the fact that research on the theoretical determination of the thermal conductivity of a silicon lattice has been going on for several decades, it has not been possible to obtain the temperature dependence in a wide range necessary for mathematical modeling, which confirms the complexity of the problem. The results of modeling the phonon thermal conductivity of silicon, as a rule, contain values at one, two, or three temperatures [17–21], either a temperature dependence, but only for the solid phase [22–28], or a narrow range dependence near the melting point [29].

The most common methods for calculating thermal conductivity are the direct method “heat source - sink” [17, 30, 31] and the Green-Kubo method [17, 30, 32]. The direct method is nonequilibrium, it is based on the application of a temperature gradient on the modeling cell, and therefore is similar to the experimental situation. The Green-Kubo method is equilibrium; it uses current fluctuations to calculate the thermal conductivity using the fluctuation-dissipation theorem [33]. Thermal conductivity, according to the Green-Kubo method, can be expressed through the autocorrelation function of the microscopic heat flux. A detailed analysis of the use of both methods for calculating the phonon thermal conductivity of silicon was carried out in [17]. One of the advantages of the direct method, in comparison with the Green-Kubo method, is the saving of computational resources, which is very important, sometimes determining for choosing a modeling method. For example, as noted in [17], for a direct method, a simulation time of 1 ns is sufficient to obtain a smooth temperature profile, and the value of κ converges to within $\pm 10\%$, the same simulation time using the Green-Kubo method leads to statistical errors up to $\pm 50\%$. Both methods demonstrate the effects of finite size. These effects arise if the mean free path of phonons is comparable to the size of the simulation cell. For a system such as Si, the required size of the computational domain to achieve a completely convergent value of κ may be beyond the

reach of atomistic modeling. Finite size effects impose a restriction on the smallest length of the computational domain. The thermal conductivity of silicon can be obtained by a direct method from modeling systems of different sizes and extrapolating the results to a system of infinite size. The equilibrium Green-Kubo method, in comparison with the direct method, has an advantage for calculating the thermal conductivity of highly anisotropic materials [30].

The Evans method [34, 35] is characterized by the combination of the elements of the equilibrium (Green-Kubo) and nonequilibrium perturbation method developed in [36,37]. Like the direct method, the Evans method reduces the computational time required to obtain the phonon thermal conductivity. Like the Green-Kubo method, this method is well suited only to describe the properties of homogeneous systems.

Along with classical molecular dynamics, the ab-initio approach [19, 22], the method of lattice dynamics and molecular dynamics based on the solution of the Boltzmann equation [38, 39], are also used to determine phonon thermal conductivity.

In this paper, the direct method (DM) was chosen as the most simple and economical from a computational point of view as an approach to determining the temperature dependence of the phonon thermal conductivity of silicon using molecular dynamics (MD).

The aim of this work is to obtain the temperature dependence of phonon thermal conductivity in the range from normal conditions to the critical region using the KIHS potential [11–13] in the framework of the DM of molecular dynamics modeling. The obtained temperature dependence of phonon thermal conductivity is planned to be used in the continuum simulation of laser evaporation of silicon.

2 STATEMENT OF THE PROBLEM. COMPUTATIONAL ALGORITHM.

To determine the thermal conductivity of the phonon subsystem of silicon, a series of calculations was performed based on molecular dynamics models. The phonon thermal conductivity was determined on the basis of the DM.

DM is a method of nonequilibrium molecular dynamics (NEMD), where regions of a heat source and sink are created in a modeling cell to apply a constant heat flux along the direction of interest. The thermal conductivity κ_{lat} is determined by the known heat flux according to the Fourier law [40]

$$W = -\kappa_{lat} \frac{\partial T}{\partial x} \quad (1)$$

where W is the heat flux, x is the coordinate in the direction of flux. The difficulty in applying the direct method to solids lies in the fact that the size of the modeling region should be much larger than the mean free path of phonons in a substance. For a crystal, this is difficult to do, because requires a very large size of the computational domain and, accordingly, a very large number of atoms. For example, 500 million atoms were used in [41]. Therefore, when calculating with a small number of atoms, the thermal conductivity coefficient turns out to be dependent on the length of the region due to phonon scattering at the boundary. Therefore, to limit the size of the simulation cell and make the simulation feasible, you can use the scaling procedure in which thermal conductivity is determined for several lengths of the simulation cell along the x direction. Then, the inverse dependence of the thermal conductivity $1/\kappa_{lat}$ is constructed with respect to the inverse value of the length of the simulation cell, $1/L$, and the thermal conductivity is determined by extrapolating the data $1/L \rightarrow 0$ [17]. Such a procedure

is justified by the expression for thermal conductivity obtained from the kinetic theory [17, 31]. Another disadvantage of the direct method is that the temperature gradient must be large in order to reduce the influence of temperature fluctuations. Also, the average temperature is reached only in the middle of the region, and the rest of the region is at a temperature different from the average value.

The simulation domain in the form of a parallelepiped with sizes of $10 \times 10 \times 20$ unit cells (lattice constant 0.543 nm) with periodic boundary conditions along three axes was considered. As the interaction potential, the KIHS potential is used [13]. The region along the x axis was divided into 20 intervals, and heating is performed in the first interval, and the heat sink is in the 11th. At each time step, a fixed amount of heat dQ_N was pumped into the heating region, and the same amount was taken from the sink region. The heat flux W was calculated as

$$W = dQ/(SNdt)/2, \quad (2)$$

where $dQ = N \times dt \times \delta Q_N$ is the total energy released, where δQ_N is the energy released in 1 time step, N is the number of time steps, dt is the time step size, S is the cross-sectional area of the region. Division by 2 is used due to periodic boundary conditions, i.e. heat distribution goes in 2 directions. Then, the resulting temperature gradient was calculated, and the Fourier law (1) was used to obtain thermal conductivity. The choice of time step depended on temperature, and took the values from 3 fs at 300K to 1 fs at 4000K and higher.

3 MODELING RESULTS

Fig. 1 shows the time-averaged spatial temperature profile used to calculate the thermal conductivity. The average temperature is 1690 K. In a small region (~ 8 nm), a very nonlinear temperature profile is observed in the immediate vicinity of the source. The same strongly nonlinear temperature profile is also observed near the sink. In the intermediate region, the temperature profile is close to a linear dependence. This gap between the heat source and the heat sink is indicated on the graph (Fig. 1) by dashed lines. In this interval, a temperature gradient was measured. The presence of a heat source and heat sink and the use of periodic boundary conditions create a current in two opposite directions. Since the heat flux passes along a clearly defined direction in the lattice, a single simulation can be used to obtain thermal conductivity along only one direction of the crystal lattice. Therefore, the heat flux was determined by a series of calculations. For the crystal, calculations were carried out for 5 different sizes of the region from 20 to 320 unit cells with a constant cross section of 10×10 cells. The maximum number of atoms was 256,000.

The heat flux (2) was determined from the temperature difference between the heating and heat sink areas, for which the instantaneous temperature difference was averaged over the entire calculation time after establishing the stationary distribution. To increase the accuracy of the calculations, the temperature difference was calculated not over the entire interval between the source and the sink, but in its central part with a length of 0.8 of the full length. The calculation time varied from 480 ps for a temperature of 350 K to 3.8 ns at 1550 K and a higher temperature.

To calculate the thermal conductivity from the Fourier law (1), the scaling procedure described above was used. The inverse dependence of the thermal conductivity $1/\kappa_{lat}$ was constructed with respect to the reciprocal of the length of the simulation cell $1/L$, and the

thermal conductivity was determined by extrapolating the data $1/L \rightarrow 0$. For the liquid state, only one calculation was performed for a size of 40 cells, for large sizes the results were approximately the same. Fig. 2 shows the dependences of the reciprocal of the thermal conductivity on the reciprocal of the size of the region for two temperatures 350 K and 1550 K.

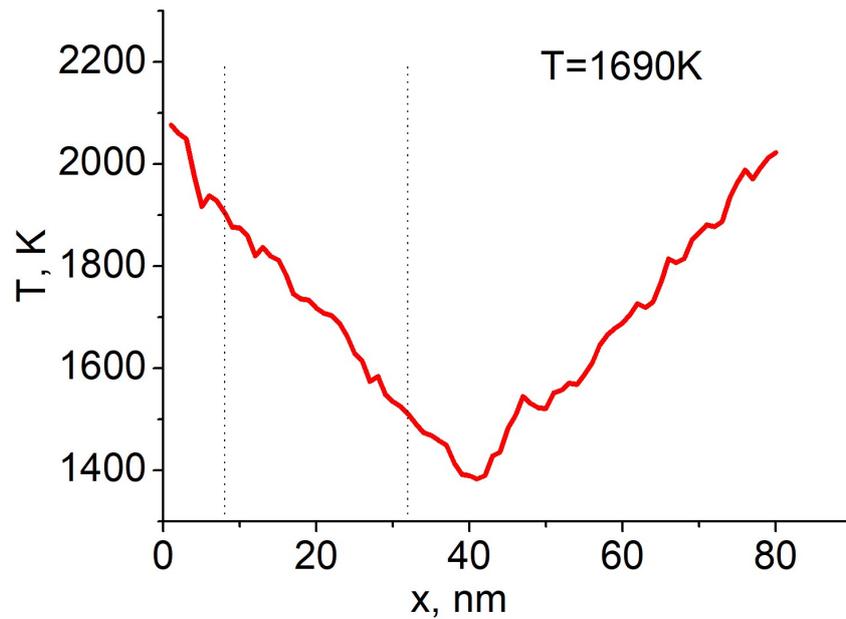


Fig. 1. Spatial profile of the temperature at one of the time instants.

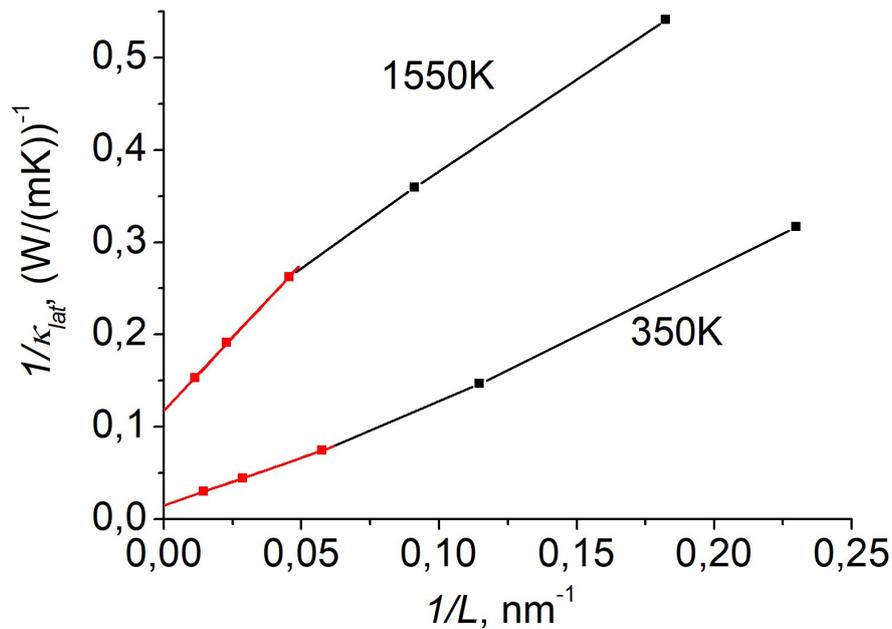


Fig. 2. Dependence of the reciprocal of the thermal conductivity on the reciprocal of the size of the region for two temperatures.

The results of calculations on the graph are shown by red lines and markers. Extrapolation was performed using the least squares method at the three points corresponding to the longest lengths of the region (red lines and markers in Fig. 2).

Fig. 3 shows the obtained temperature dependence of the phonon thermal conductivity with the experimental data [42] and the data from [29] for the Stillinger – Weber potential, as well as a theoretical estimate [44, 45]. The dashed line shows the extrapolation of the phonon thermal conductivity to the critical point. In this work, we used the silicon critical point parameters $T_{crit} = 6750 \pm 250$ K, $\rho_{crit} = 0.24 \pm 0.06$ g/cm³, $P_{crit} = 800 \pm 300$ bar obtained from molecular dynamics modeling with the KIHS potential in [43].

According to theoretical concepts [45], the thermal conductivity of the lattice of crystalline silicon is calculated as follows

$$\kappa_{lat}(T) = \frac{1}{3} c_v \rho \omega_F \bar{l}_F \quad (3)$$

where ω_F is the velocity of sound, or the group velocity of thermal vibrations of phonons; \bar{l}_F is the average mean free path of elastic waves - phonons; c_v is the specific heat of the lattice, ρ is the density. Figure 7 shows the approximation of the theoretical estimate (blue line) obtained in [44]

$$\kappa_{lat}(T) = 15,85 \times T_{lat}^{-1.23} \text{ W/mK}$$

Phonon thermal conductivity is difficult to obtain from experiment, therefore, experimental data are given for the total thermal conductivity of silicon.

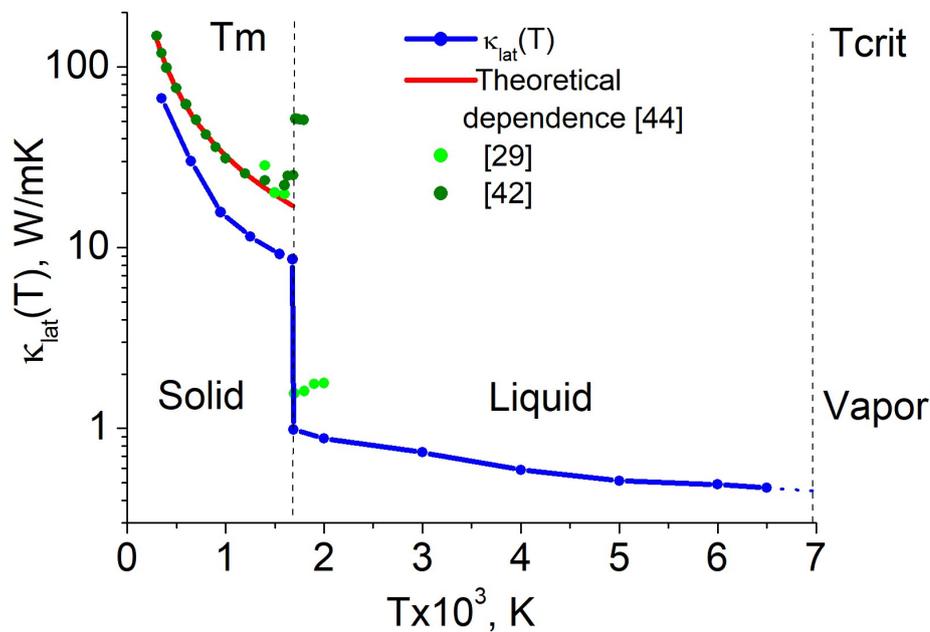


Fig. 3. Temperature dependence of the phonon thermal conductivity of silicon (this work), experimental data for total thermal conductivity [42], and data from [29] for the Stillinger-Weber potential. A theoretical estimate for the solid phase is also given [44].

In crystalline silicon, phonon thermal conductivity is decisive. The difference between the calculation data with the KIHS potential and the experimental values of the total thermal conductivity is approximately 60% at 350K. At the melting temperature, the difference from experiment doubles, which is explained by an increase in the contribution of electronic thermal conductivity to the total thermal conductivity. A comparison of the results of calculations with a theoretical estimate [44, 45] shows a qualitative coincidence.

At the semiconductor-metal phase transition, the phonon thermal conductivity decreases stepwise, which is explained by the destruction of long-range bonds. At the equilibrium melting temperature $T_m = 1687$ K, the jumps in the phonon thermal conductivity of silicon is 88.9%. In the calculation of thermal conductivity with the Stillinger – Weber potential [29], the jump estimate is 92.08%.

After melting, silicon from a semiconductor becomes a metal, because of which the electronic component in the thermal conductivity becomes decisive [46, 47], and the phonon component differs significantly from the experimental data on total thermal conductivity.

The difference between the phonon thermal conductivity in the liquid of the potentials KIHS and SW [29] is 60%. At a temperature of 2000 K, the difference between the results obtained in this work and the data obtained using the ab-initio approach from [19] is 30.8% (0.91 and 0.63 W/(mK), respectively).

At the temperature $T > 5000$ K, phonon thermal conductivity tends to decrease, which does not contradict theoretical concepts. In the near-critical region, the phonon thermal conductivity obtained from the simulation results is approximately 0.47 W/(mK) at $T = 6500$ K, approximately the same value at temperature $T \approx T_{crit}$ (0,45 W/(mK)).

4. CONCLUSION

1. Based on a series of calculations by the direct method using molecular dynamics modeling, the temperature dependence of the phonon thermal conductivity of silicon was obtained in the temperature range $300 < T < 6500$ K. The KIHS potential was used in the simulation.

2. Comparison of the simulation results with the KIHS potential (this work) with the simulation results with the Stillinger – Weber potential [29] showed that for high-temperature processes ($T_m < T < T_{crit}$) in silicon, the best results are achieved when the KIHS potential is used, in contrast to low-temperature processes ($T \leq T_m$), where the classic Stillinger –Weber potential has an advantage.

3. Comparison of the results obtained at a temperature $T = 2000$ K with the results obtained using the ab-initio approach [19] showed a slight discrepancy of 30.8% (0.91 and 0.63 W/(mK), respectively).

4. Comparison of the results of modeling of the phonon thermal conductivity with the experimental values of total thermal conductivity [42] makes sense only in crystalline silicon in the range $300 \text{ K} < T < 1000 \text{ K}$, since at a temperature above $T > 1000$ K the degeneracy of charge carriers increases, the band gap decreases, which leads to a rapid increase in the concentration of carriers and, accordingly, to an increase in the thermal conductivity of electrons [46, 47]. The difference between the phonon thermal conductivity and the total thermal conductivity with increasing temperature will only increase, especially after a phase transition in liquid silicon. After melting at the equilibrium melting temperature $T_m = 1687$ K from the side of the solid phase, the phonon thermal conductivity obtained from molecular

dynamic modeling (MDM) is 2.9 times less than the experimental data and theoretical estimate. From the side of the liquid phase, the phonon thermal conductivity obtained from MDM is ~ 50 times lower than the experimental values of the total thermal conductivity.

5. In the near-critical temperature range of $5000 \text{ K} < T < T_{\text{crit}}$, the phonon thermal conductivity decreases, which does not contradict theoretical concepts, and at $T \approx T_{\text{crit}}$ its value is approximately $\kappa_{\text{lat}} \approx 0.45 \text{ W/(mK)}$.

REFERENCES

1. Urs Zywietz, Andrey B. Evlyukhin, Carsten Reinhardt & Boris N. Chichkov, “Laser printing of silicon nanoparticles with resonant optical electric and magnetic responses”, *Nat. Comm.*, **5**, 3402 (2014).
2. Urs Zywietz, Mikolaj K. Schmidt, Andrey B. Evlyukhin, Carsten Reinhardt, Javier Aizpurua, and Boris N. Chichkov, “Electromagnetic Resonances of Silicon Nanoparticle Dimers in the Visible”, *ACS Photonics*, **2** (7), 913–920 (2015).
3. Boris Chichkov, Urs Zywietz and Lothar Koch. “Laser printing of silicon nanoparticles and living cells”, *Biomedical Optics & Medical Imaging, Proc. SPIE Int. Soc. Opt. Eng.*, (2015).
4. M E Shaheen, J E Gagnon and B J Fryer, “Femtosecond laser ablation behavior of gold, crystalline silicon, and fused silica: a comparative study”, *Las. Phys.*, **24**, 106102(1-8) (2014).
5. Vinod Parmar, Yung C. Shin, “Wideband anti-reflective silicon surface structures fabricated by femtosecond laser texturing”, *Appl. Sur. Sci.*, **459**, 86-91 (2018).
6. Benjamin Franta, Eric Mazur and S. K. Sundaram, “Ultrafast laser processing of silicon for photovoltaics”, *Int. Mater. Rev.*, **63**(4), 227-240 (2018).
7. A. Al-Kattan, Y.V. Ryabchikov, T. Baati, V. Chirvony, J. F. Sánchez-Royo, M. Sentis, D. Braguer, V.Yu. Timoshenko, M. Anne Esteve, A.V. Kabashin, “Ultrapure laser-synthesized Si nanoparticles with variable oxidation states for biomedical applications”, *J. Mater. Chem. B*, **4**, 7852-7858 (2016).
8. R.F. Balderas-Valadez, J.O. Estévez-Espinoza, U. Salazar-Kuri, C. Pacholski, W.L. Mochan, V. Agarwal, “Fabrication of ordered tubular porous silicon structures by colloidal lithography and metal assisted chemical etching: SERS performance of 2D porous silicon structures”, *Appl. Sur. Sci.*, **462**, 783-790 (2018).
9. V. I. Mazhukin, A. V. Mazhukin, M. M. Demin, A. V. Shapranov, “Nanosecond laser ablation of target Al in a gaseous medium: explosive boiling”, *Applied Physics A: Material Science and Processing*, **124** (3), 237(1-10), (2018).
10. V.I. Mazhukin, A.V. Shapranov, A.V. Rudenko, “Comparative analysis of potentials of interatomic interaction for crystalline silicon”, *Math. Montis.*, **30**, 56-75 (2014).
11. F.H. Stillinger, T.A. Weber, “Computer simulation of local order in condensed phases of silicon”, *Phys. Rev. B*, **31**, 5262-5271 (1985).
12. L. Pizzagalli, J. Godet, J. Guenole, S. Brochard, E. Holmstrom, K. Nordlung, T. Albaret, “A new parametrization of the Stillinger-Weber potential for an improved description of defects and plasticity of silicon”, *J. Phys, Condens. Matter*, **25**, 055801(1-12) (2013).
13. T. Kumagai, S. Izumi, S. Hara, S. Sakai, “Development of bond-order potentials that can reproduce the elastic constants and melting point of silicon for classical molecular dynamics simulation”, *Comp. Mater. Sci.*, **39** (2), 457-464 (2007).
14. Lammmps project website: <http://lammmps.sandia.gov/> (accessed July 7, 2019).
15. L. Verlet, “Computer “Experiments” on Classical Fluids. I. Thermodynamically Properties of Lennard-Jones Molecules”, *Phys. Rev.*, **159**, 98-103 (1967).
16. H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, A. DiNola, J.R. Haak, “Molecular dynamics with coupling to an external bath”, *J. Chem. Phys.*, **81**, 3684-3690 (1984).

17. P. K. Schelling, S. R. Phillpot, P. Keblinski, “Comparison of atomic-level simulation methods for computing thermal conductivity”, *Phys. Rev. B*, **65**, 144306 (2002)
18. Wang Zenghui, Li Zhixin, “Lattice dynamics analysis of thermal conductivity in silicon nanoscale film Lattice dynamics analysis of thermal conductivity in silicon nanoscale film”, *Applied Thermal Engineering*, **26**, 2063–2066 (2006)
19. Jun Kang and Lin-Wang Wang, “First-principles Green-Kubo method for thermal conductivity calculations”, *Phys. Rev. B*, **96**, 020302(R) (2017)
20. Yongjin Lee and Gyeong S. Hwang, “Force-matching-based parameterization of the Stillinger-Weber potential for thermal conduction in silicon”, *Phys. Rev. B*, **85**, 125204 (2012)
21. D. P. Sellan, E. S. Landry, J. E. Turney, A. J. H. McGaughey, and C. H. Amon, “Size effects in molecular dynamics thermal conductivity predictions”, *Phys. Rev. B*, **81**, 214305 (2010)
22. Christian Carbogno, Rampi Ramprasad, and Matthias Scheffler, “Ab Initio Green-Kubo Approach for the Thermal Conductivity of Solids”, *Phys. Rev. Lett.*, **118**, 175901 (2017).
23. Sebastian G. Volz, Gang Chen, “Molecular-dynamics simulation of thermal conductivity of silicon crystals”, *Phys. Rev. B*, **61**(4), 2651-2656 (2010)
24. Keivan Esfarjani and Gang Chen, *Heat transport in silicon from first principles calculations*, arXiv: 1107.5288v1, (2011)
25. Shenghong Ju and Xingang Liang, “Thermal conductivity of nanocrystalline silicon by direct molecular dynamics simulation”, *J. Appl. Phys.*, **112**, 064305 (2012)
26. D. A. Broido, M. Malorny, G. Birner, Natalio Mingo, D. A. Stewart, “Intrinsic lattice thermal conductivity of semiconductors from first principles”, *Appl. Phys. Lett.*, **91**, 231922 (2007)
27. Lin Sun, Jayathi Y. Murthy, “Molecular dynamics simulation of phonon transport in EDIP silicon”, *Proceedings of HT2005, San Francisco, California, USA*, 1-6 (2005)
28. P.C. Howell, “Comparison of molecular dynamics methods and interatomic potentials for calculating the thermal conductivity of silicon”, *J. Chem. Phys.*, **137**, 224111 (2012).
29. C. H. Baker, C. Wu, R. N. Salaway, L. V. Zhigilei, P. M. Norris, “Resolving the vibrational and electronic contributions to thermal conductivity of silicon near the solid-liquid transition: molecular dynamics study”, *I. J. Trans. Phenomena*, **13**, 143–150 (2013)
30. L. Hu, W. J. Evans, P. Keblinski, “One-dimensional phonon effects in direct molecular dynamics method for thermal conductivity determination”, *J. Appl. Phys.*, **110**, 113511 (2011)
31. Florian Müller-Plathe, “A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity”, *J. Chem. Phys.*, **106**, 6082 (1997).
32. R. Zwanzig, “Time-correlation functions and transport coefficients in statistical mechanics”, *Annu. Rev. Phys. Chem.*, **16**, 67-102 (1965).
33. R Kubo, “The fluctuation-dissipation theorem”, *Rep. Prog. Phys.*, **29**, 255-284 (1966)
34. Denis J. Evans, “Homogeneous nemd algorithm for thermal conductivity - application of non-canonical linear response theory”, *J. Phys. Lett.*, **91A**(9), 457-460 (1982).
35. Kranthi K. Mandadapu, Reese E. Jones, Panayiotis Papadopoulos, “A homogeneous non-equilibrium molecular dynamics method for calculating thermal conductivity with a three-body potential”, *J. Chem. Phys.*, **130** (20), 204106(1-18) (2009).

36. G. Ciccotti, G. Jacucci and I.R. McDonald, “Thermal response to a weak external field”, *J. Phys. C: Solid State Phys.*, **11** (13), 1509-1513 (1978).
37. G. Ciccotti, G. Jacucci, and I.R. McDonald, “Thought experiments by molecular dynamics”, *J. Stat. Phys.*, **21**, 1–22 (1979).
38. J. E. Turney, E. S. Landry, A. J. H. McGaughey, and C. H. Amon, “Predicting phonon properties and thermal conductivity from anharmonic lattice dynamics calculations and molecular dynamics simulations”, *Phys. Rev. B*, **79**, 064301 (2009).
39. A. J. H. McGaughey and M. Kaviani, “Quantitative validation of the Boltzmann transport equation phonon thermal conductivity model under the single-mode relaxation time approximation”, *Phys. Rev. B*, **69**, 094303 (2004).
40. W. Evans and P. Keblinski, “Thermal conductivity of carbon nanotube cross-bar structures”, *Nanotechnology*, **21**(47), 475704, (2010) doi: 10.1088/0957-4484/21/47/475704
41. Chaofeng Hou, Ji Xu, Wei Ge and Jinghai Li, “Molecular dynamics simulation overcoming the finite size effects of thermal conductivity of bulk silicon and silicon nanowires”, *Modelling Simul. Mater. Sci. Eng.*, **24**, 045005 (1-9) (2016)
42. Eiji Yamasue, Masahiro Susa, Hiroyuki Fukuyama, Kazuhiro Nagata, “Thermal conductivities of silicon and germanium in solid and liquid states measured by non-stationary hot wire method with silica coated probe”, *Journal of Crystal Growth*, **234**, 121–131 (2002).
43. V.I. Mazhukin, A.V. Shapranov, O.N. Koroleva, A.V. Rudenko. Molecular dynamics simulation of critical point parameters for silicon, *Math. Montis.*, **31**, 56-76 (2014).
44. H.M. Van Driel, “Kinetics of high-density plasmas generated in Si by 1.06- and 0.53-mm picosecond laser pulses”, *Phys. Rev. B*, **35**, 8166 – 8176 (1987).
45. W. Neil, N. Ashcroft, D. Mermin, *Solid state physics*, Saunders College, Vol. 1, 1976.
46. O.N. Koroleva, A.V. Mazhukin, “Determination of thermal conductivity and heat capacity of silicon electron gas”, *Math. Montis.*, **40**, 99-109 (2017)
47. O.N. Koroleva, A.V. Mazhukin, V.I. Mazhukin, “Modeling of silicon characteristics in the semiconductor-metal phase transition region”, *Math. Montis.*, **41**, 73-90 (2018).

Received May 25, 2019

COMPARATIVE ANALYSIS OF THE ACCURACY OF OPENFOAM SOLVERS FOR THE OBLIQUE SHOCK WAVE PROBLEM

A. K. ALEKSEEV¹, A.E. BONDAREV^{1*}, A.E. KUVSHINNIKOV¹

¹Keldysh Institute of Applied Mathematics Russian Academy of Sciences

Moscow, Russia

*Corresponding author. E-mail: bond@keldysh.ru

DOI: 10.20948/mathmontis-2019-45-8

Summary. The article is devoted to a comparative assessment of the accuracy for solvers of the OpenFOAM open software package. As a test problem, we consider the classical two-dimensional problem of a supersonic inviscid compressible flow falling on a flat plate at an angle of attack. As a result, an oblique shock wave is formed before the start of the plate. The simulation results for the solvers considered in comparison are compared with the known exact solution. Calculations for all solvers participating in the comparison were carried out with the same setting of the parameters of the incident flow and angle of attack. Special attention was paid to QGDFoam solver, which has controlled dissipative properties. For this solver, within the framework of a general comparison, calculations were carried out with a variation of the parameter that allows controlling dissipative properties. The results of estimates of deviations from the exact solution in various norms for all solvers are given.

1 INTRODUCTION

A comparative assessment of the accuracy and efficiency of numerical methods and algorithms for mathematical modeling of CFD problems has been the subject of special attention of researchers throughout the development of mathematical modeling. Over a long period, a certain set of CFD test problems has developed, and the verification of the efficiency of the developed numerical method was based on testing the method on this set of problems. These aspects are reflected in a fairly large number of reviews, for example, [1, 2].

Currently, the task of comparative assessment of numerical methods accuracy has not lost its relevance. New technical problems and the appearance of new mathematical models entail intensive development of numerical methods. Developed new numerical methods and algorithms are often implemented in the form of solvers integrated into various software packages, both commercial and open. In this process, not always and not all solvers pass a full test on the classical set of test problems. This set can include such well-known problems as: falling an oblique shock wave onto a plate, rarefaction wave, forming a boundary layer on a plate or a smooth curved surface, flow in front of the obstacle, flow in front of a spherically blunt obstacle, flow around a cone, flow behind a ledge, flow in the far wake. Such a set of tests provides testing of a numerical method and its software implementation for mathematical models, describing both inviscid flows and viscous ones.

It should be noted that the analysis of the accuracy of numerical methods in the simulation of discontinuities has been relevant since the main directions and approaches to the simulation of flows with shock waves were formed. Two major directions were formed here - methods for gas-dynamic flows modeling without marking discontinuities and methods where discontinuities were defined as boundaries of a flowfield. A detailed description of these directions can be found in [3,4]. Both directions are characterized by a large number of

2010 Mathematics Subject Classification: 49Q99, 76M27.

Key words and Phrases: comparative estimation of accuracy, OpenFOAM solvers, oblique shock wave

numerical methods implemented in their framework. As examples of approaches without isolating of discontinuities, one can cite papers [5–9].

Methods that use the isolating of discontinuities in the form of the boundaries of the computational domain can be found, for example, in [10–12]. Practical experience of using both approaches has shown that both approaches have both clear advantages and obvious disadvantages. Thus, numerical methods without isolating of discontinuities require to make the computational grid more detailed in the vicinity of discontinuities and, as a result, give a vague picture of the flow structure. In turn, methods that use discontinuity detection encounter problems when modeling rapidly changing structures of shock waves, which leads to the need for a rapid restructuring of the geometry of the computational regions and the introduction of scenario approaches for organizing such a restructuring. Nevertheless, serious attempts are being made to overcome such difficulties in both directions.

For methods without isolating discontinuities, it is possible to note the work [13]. The approach proposed in this paper makes it possible to clearly identify the locations of discontinuities by automating the processing of the computed results. In this approach, gas-dynamic functions are considered as the intensity of the image, and the values of the functions at each point as the elements of the image (pixels). A differential detector is used and a detectable fracture is classified using discrete analogs of gas-dynamic relations performed at the discontinuity. The stated approach does not depend on the specific type of the problem being solved and does not require any a priori information about the flow. As an example of an approach combining both directions, we can mention the method of dynamic adaptation presented in [14–18]. The method is based on the transition to an arbitrary nonstationary coordinate system in which not only grid functions, but also coordinates of grid nodes are unknown. According to [18], this approach allows for calculations using methods without isolating discontinuities with automatic condensing of grid nodes to the solution features, and with explicit selection of moving boundaries and discontinuities when necessary. It should be emphasized that for methods that do not use the selection of boundaries, an analysis of the solution behavior at a discontinuity and an assessment of the accuracy are necessary.

This paper is devoted to comparing the accuracy of solvers of an open software package OpenFOAM [19] at an oblique shock wave and continues the research on the comparative assessment of numerical methods accuracy on classical test problems. At the previous stages of this study, a comparative assessment was made of the accuracy of the OpenFOAM solver group for the task of flow around a cone under the angle of attack. Studies were performed with a variation of the Mach number, the angle of the cone and the angle of attack in wide ranges with the selected step. Thus, a study was performed for a class of problems defined for these determining parameters within the ranges of variation. The results of the comparison with the well-known tabular solution allowed us to construct the dependence of the error on the determining parameters for each solver and to make a comparison for the class of problems in question. The main results are presented in [20–23].

It should be noted that these numerical studies were based on the principles of constructing a generalized computational experiment [24–27]. The construction of such an experiment is based on numerical parametric studies and the solution of optimization analysis problems. Solving such problems implies a multiple solution to the direct problem of numerical modeling of a gas-dynamic process with various input data. The defining parameters of a class of problems, such as the characteristic Mach number, Reynolds number, geometric parameters, etc., vary in certain ranges with a certain partitioning step. As a result, the

resulting solution is a multidimensional volume of data. To analyze this volume, modern methods of data analysis and visual analytics are used.

In this paper, a comparison is made for the solvers of the OpenFOAM (Open Source Field Operation And Manipulation CFD Toolbox) open source software package. This is a free software product created for solving problems of hydro and gas dynamics. Widely used in many areas of science and technology, the OpenFOAM package contains a number of solvers with different computational properties. The OpenFOAM package also allows one to develop new solvers on the platform of package. Four solvers participated in the comparison: two standard solvers - *rhoCentralFoam* and *sonicFoam* and two new solvers - *pisoCentralFoam* [28] and *QGDFoam* [29,30]. The last two solvers were developed by teams of Ivannikov Institute for System Programming and Keldysh Institute of Applied Mathematics of the Russian Academy of Sciences. It should be noted that the *QGDFoam* solver has a controllable parameter that allows one to adjust the dissipative properties of the numerical method, which is extremely important in suppressing unwanted oscillations at shock waves. The research in this paper for *QGDFoam* was performed with a variation of this parameter.

Previously, most comparative estimates of numerical methods accuracy in simulating a shock wave were reduced to a comparison of the width of the shock wave spreading zone along a selected line crossing the discontinuity. In this paper, we use error estimates for the entire flow field in the computational domain in different norms.

2 FORMULATION OF THE TEST PROBLEM

In this paper, the classical two-dimensional inviscid problem of modeling an oblique shock wave is used to compare solvers. The general flow scheme is shown in Fig. 1. A supersonic gas flow with Mach number M at an angle β falls on a flat plate. At the beginning of the plate, an oblique shock wave S occurs. This problem is considered within the framework of the Euler system of equations and has an exact analytical solution.

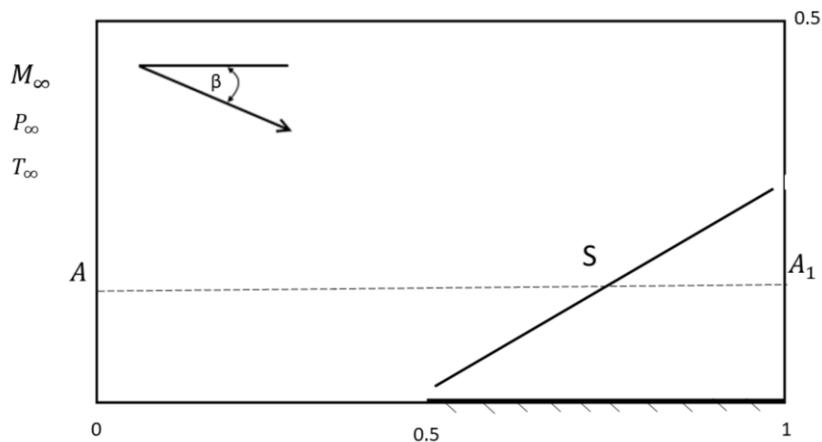


Fig.1. Flow scheme.

At the input boundary, the parameters of the external flow are specified for the Mach number M and a certain value β . On the part of the lower boundary corresponding to a flat plate, a no-flow condition is specified. At the output boundary, we set the derivatives of gas-dynamic functions equal to zero along the normal to the boundary. On the upper boundary for

the velocity components the boundary conditions are set similarly to the conditions for the input boundary. For the remaining gas-dynamic functions of the upper boundary the conditions are set similarly to the conditions for the output boundary.

3 THE PROCEDURE OF COMPARISON

The solution of the problem was performed using 4 solvers of the OpenFOAM software package. These solvers were: *rhoCentralFoam*, *sonicFoam*, *pisoCentralFoam*, *QGDFoam*. We give below their brief characteristics.

Solver *rhoCentralFoam* — is based on a central-upwind scheme, which is a combination of central-differential and upwind schemes [31,32]. The essence of the central-upwind flow schemes consists in a special selection of the control volume containing two types of domains: around the boundary points - the first type; around the center point - the second type. The boundaries of the control volumes of the first type are determined by means of local propagation velocities. The advantage of these schemes is that, using the appropriate technique to reduce the numerical viscosity, it is possible to achieve good solvability for discontinuous solutions — shock waves in gas dynamics, and for solutions in which viscous phenomena play a major role.

Solver *sonicFoam* is based on the PISO algorithm (Pressure Implicit with Splitting of Operator) [33]. The basic idea of the PISO method is that two difference equations are used to calculate the pressure for the correction of the pressure field obtained from discrete analogs of the equations of moments and continuity. This approach is due to the fact that the velocities corrected by the first correction may not satisfy the continuity equation, therefore, a second corrector is introduced which allows us to calculate the velocities and pressures satisfying the linearized equations of momentum and continuity.

Solver *pisoCentralFoam* is a combination of a central-upwind scheme [28] with the PISO algorithm.

Solver *QGDFoam* [29,30] is based on a system of quasi-gas dynamic equations [34–36] developed by a research team led by B.N. Chetverushkin. A quasi-gas dynamic algorithm is built on the basis of a mathematical model that generalizes the Navier-Stokes system of equations and differs from it by additional dissipative terms, having the form of second spatial derivatives with a small parameter in the form of a coefficient [36]. The principal difference of QGD (quasi-gas dynamic and quasi-hydrodynamic) systems from the Navier-Stokes system of equations is the space-time averaging for determining the main gas dynamic quantities. The presence of a controlled parameter with dissipative terms makes it possible to successfully suppress unwanted oscillations at discontinuities. The calculations used the values of this parameter in the range from 0.1 to 0.3.

To organize the comparison, the unification of calculations was performed. There are two ways in the OpenFOAM package to select the approximation variant of differential operators: directly in the solver's code or using the *fvSchemes* and *fvSolution* configuration files. To make the comparison correct, we used the same parameters, where it was possible, acting in the same way as [20-23]. The following parameters were selected in the *fvSchemes* file: *ddtSchemes* – Euler, *gradSchemes* – Gauss linear, *divSchemes* – Gauss linear, *laplacianSchemes* – Gauss linear corrected, *interpolationSchemes* – vanLeer. In the *fvSolution* file: *solver* – *smoothSolver*, *smoother* – *symGaussSeidel*, *tolerance* – $1e-09$, *nCorrectors* – 2, *nNonOrthogonalCorrectors* – 1.

To estimate the deviation of the obtained numerical results from the known exact solution in the entire computational domain, analogs of the L2 norms were used

$$\partial_{L2} = \sqrt{\sum_m |y_m - y_m^{exact}|^2 V_m} / \sqrt{\sum_m |y_m^{exact}|^2 V_m}$$

and L1

$$\partial_{L1} = \sum_m |y_m - y_m^{exact}| V_m / \sum_m |y_m^{exact}| V_m.$$

Here, y_m is the pressure p , the local Mach number Ma , the density ρ in the cell, and V_m is the volume of the cell. All calculations were carried out with setting the following flow parameters: flow angle $\beta = 6^\circ$, Mach number $M_\infty = 2$, pressure $P_\infty = 101325$ Pa, temperature $T_\infty = 300$ K.

4 CALCULATION RESULTS

Calculations for all solvers allowed us to obtain a well-known flow pattern for the simulated oblique shock problem. A typical flow pattern is shown in Fig. 2 as a pressure distribution in the computational domain. The presented pressure distribution was obtained using *rhoCentralFoam* solver. The destruction of the solution was not observed for any of the solvers, which testifies to the high stabilizing properties of all solvers participating in the study.



Fig. 2. Typical pressure distribution.

For all solvers, comparisons were made with the known exact solution [2]. The results are presented in tables 1 and 2 for the norms L1 and L2, respectively. The bold font indicates the minimum values. Further, in the tables for solvers, the abbreviations are used: rCF (*rhoCentralFoam*), pCF (*pisoCentralFoam*), sF (*sonicFoam*), QGDF (*QGDFoam*). The deviations from the exact solution over the entire computational domain were calculated for the local Mach number Ma , pressure p and density ρ .

The upper row of both tables shows the value of the parameter α given for the *QGDFoam* solver, which allows adjusting the additional artificial viscosity.

	rCF	pCF	sF	QGDF, $\alpha=0.1$	QGDF, $\alpha=0.15$	QGDF, $\alpha=0.2$	QGDF, $\alpha=0.3$
Ma	0.000592	0.000768	0.001014	0.000646	0.000668	0.000757	0.001005
p	0.001755	0.001902	0.003182	0.002245	0.002203	0.002406	0.003061
ρ	0.001350	0.001480	0.002211	0.001549	0.001532	0.001677	0.002131

Table 1. Norm L1, $M=2$, $\beta = 6^\circ$

	rCF	pCF	sF	QGDF, $\alpha=0.1$	QGDF, $\alpha=0.15$	QGDF, $\alpha=0.2$	QGDF, $\alpha=0.3$
Ma	0.004231	0.004572	0.005504	0.004086	0.004318	0.004699	0.005500
p	0.013287	0.013744	0.017505	0.014393	0.014734	0.015647	0.017753
ρ	0.009331	0.009633	0.012146	0.009940	0.010222	0.010860	0.012305

Table 2. Norm L2, $M=2$, $\beta = 6^\circ$

The results in the tables show that the smallest deviation from the exact solution for the flow field in almost all cases is provided by the solver *rhoCentralFoam*. It can also be noted that for the *QGDFoam* solver, decreasing the α parameter significantly reduces the error. When evaluated in the L2 norm for the Mach number, the result of the *QGDFoam* solver provides the smallest deviation from the exact solution.

We now turn from general integral estimates to a more careful consideration of the behavior of gas-dynamic functions in the vicinity of the shock wave. Fig. 3,4,5 show the results for all solvers in the form of a density, pressure and a local Mach number distribution along the horizontal line AA_1 , crossing the computational domain at a distance from the lower boundary equal to $y = 0.15$ (Fig 1). The exact solution is indicated by a dotted line. All solvers are indicated by the colors shown in the corresponding table in the figures.

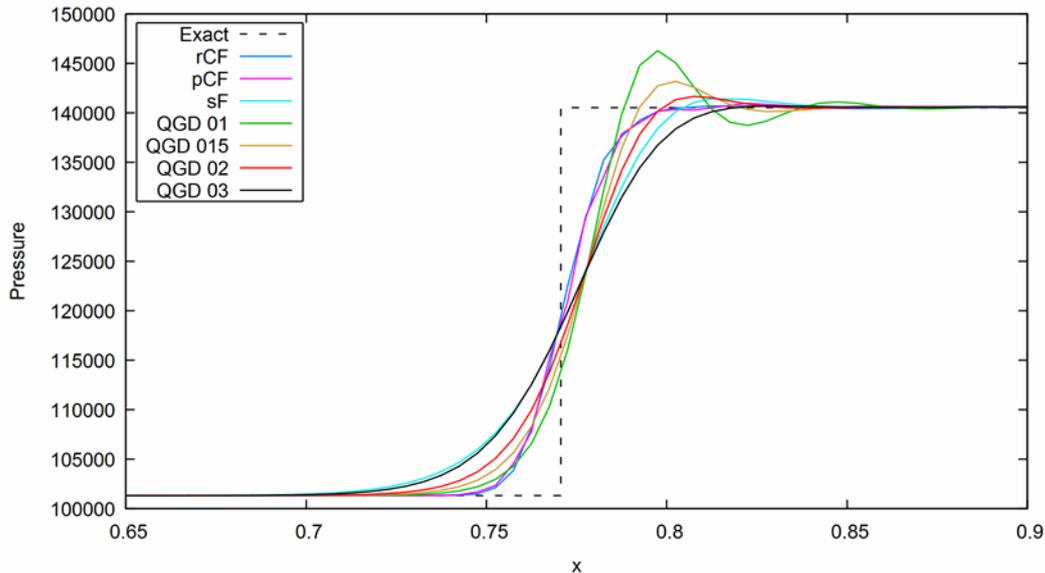


Fig.3. The distribution of pressure in the vicinity of the shock wave.

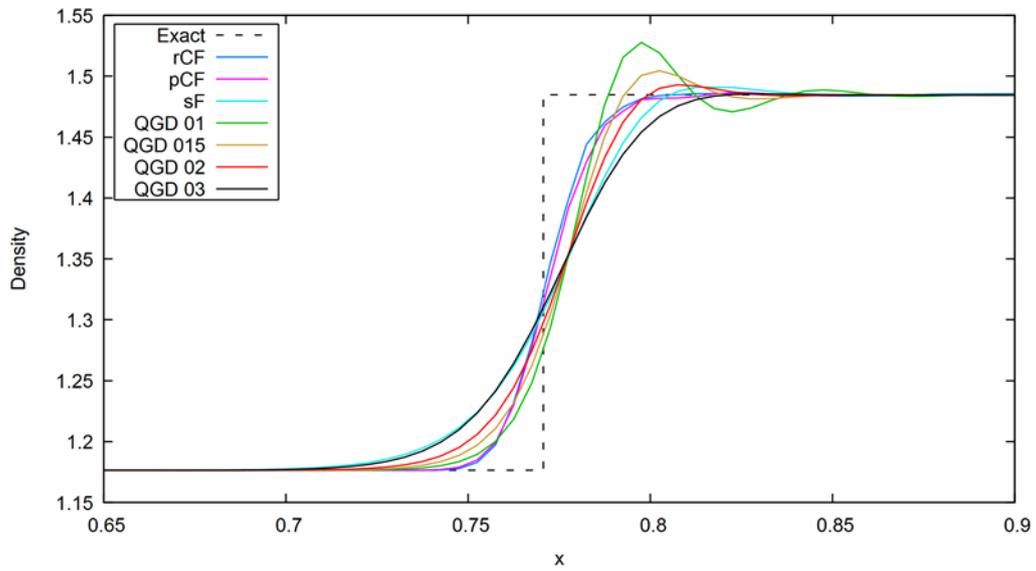


Fig.4. The distribution of density in the vicinity of the shock wave.

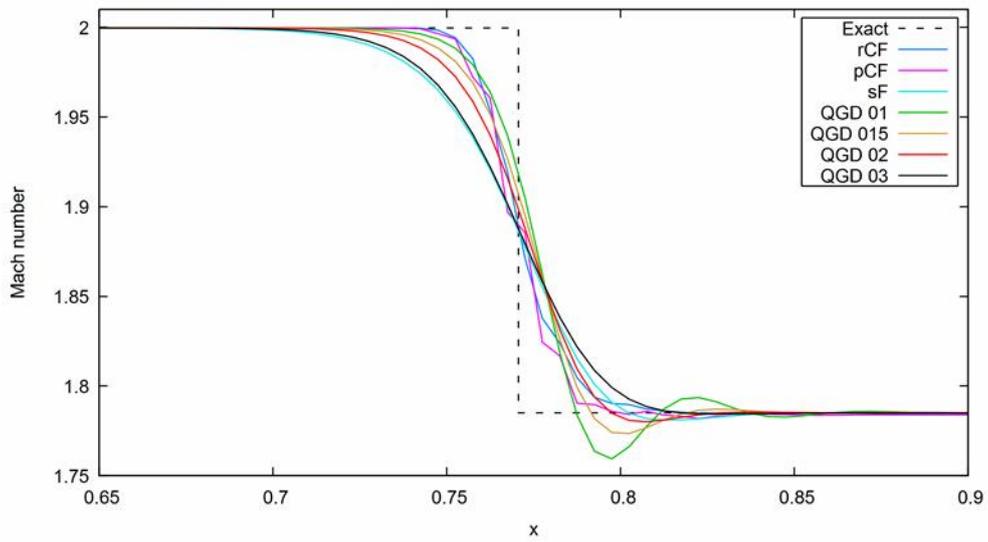


Fig.5. The distribution of the local Mach number in the vicinity of the shock wave.

The pattern of pressure distribution in the lower part of the shock and in the upper part is shown in close-ups in Figures 6 and 7, respectively.

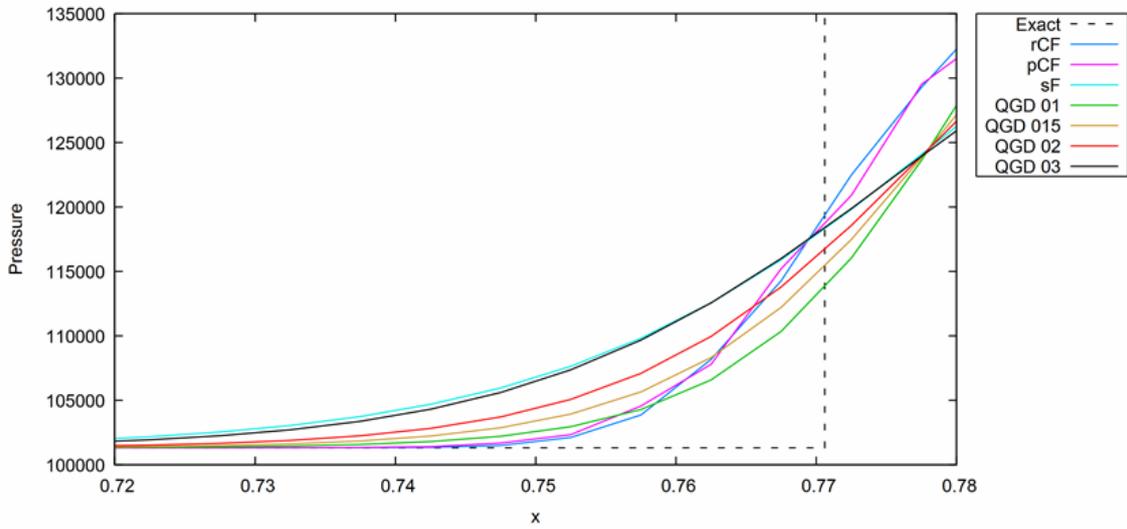


Fig.6. Pressure distribution in the vicinity of the lower part of the shock wave.

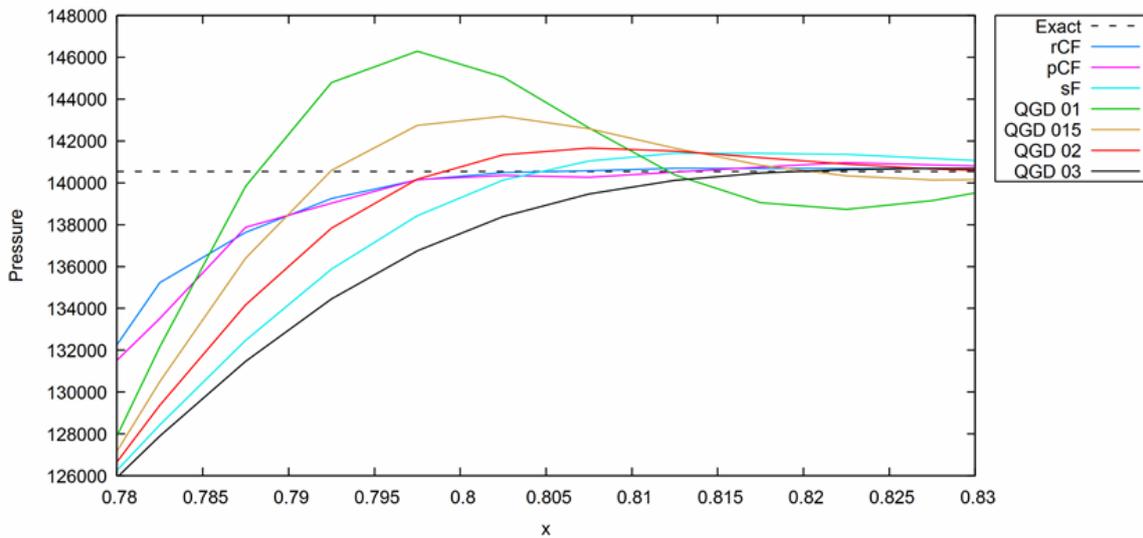


Fig.7. Pressure distribution in the vicinity of the upper part of the shock wave.

The presented figures make it possible to judge the degree of spreading of the shock for all the solvers considered in comparison. The result closest to the exact solution is provided by the *rhoCentralFoam* solver. For solver *QGDFoam*, the effect of variation of the parameter α is very clearly represented. A decrease in the parameter α brings the calculated results closer to the exact solution; however, oscillations that appear are noticeable in the upper part of the shock wave. This confirms the well-known fact that often the general assessment of accuracy in the norm of monotone schemes that provide smooth solutions shows worse results than for less monotonic schemes with oscillations. In the case of the *QGDFoam* solver, the user of the solver has the opportunity to choose either to obtain a smooth solution, or to improve the estimate at the norm.

5 CONCLUSIONS

A comparative evaluation of solution accuracy for the four OpenFOAM solvers has been made. As a test problem, we used the classical two-dimensional oblique shock problem caused by the fall of a supersonic flow of an inviscid compressible gas on a flat plate at an angle. The calculations were performed for fixed values of the parameters of the incident flow and angle of attack. Comparison of results with the exact solution was carried out over the entire field of calculated data using analogs of the norms L1 and L2.

The results obtained showed that in almost all cases the solver *rhoCentralFoam* provides the smallest deviation from the exact solution. For solver *QGDFoam*, reducing the parameter that controls the artificial viscosity can significantly reduce the deviation from the exact solution, but at the same time, oscillations appear in the upper part of the shock wave that do not destroy the solution.

The results of the comparative evaluation can be useful both for users of the OpenFoam software package and for developers of the software content of this package.

In the future, the authors consider the expansion of a comparative estimate based on the construction of a generalized computational experiment by varying the Mach number and angle of attack for the oblique shock wave problem considered. It is also planned to solve an optimization problem for the *QGDFoam* solver in order to find the optimal control of the dissipative properties of the solver on strong discontinuities.

Acknowledgements: This work was supported by grant of RSF № 19-11-00169.

REFERENCES

- [1] V.M. Paskonov, V.I. Polezhaev, L.A. Chudov, *Chislennoe modelirovanie processov teplo- i massoobmena*, M.: Nauka (1993).
- [2] E.N. Bondarev, V.T. Dubasov, Y.A. Ryzhov et al., *Aerigidromeckanika*, M.: Mashinostroenie (1993).
- [3] A.G. Kulikovskiy, N.V. Pogorelov, A.Yu. Semenov, *Matematicheskie voprosy chislennogo resheniya giperbolicheskikh uravneniy*, M.: Phizmatlit (2001)..
- [4] A.A. Samarsky, Yu.P. Popov, *Raznosnye shemy gazovoy dinamiki*, M.: Nauka (1981).
- [5] A. Harten, "High resolution schemes for hyperbolic conservation laws", *Journal of Computational Physics*, **49**, 357-393 (1983).
- [6] S. Osher, "Riemann solvers, the entropy condition, and difference approximation", *SIAM J. Numer. Anal.*, **21** (2), 217-235 (1984).
- [7] A. Harten, "ENO schemes with subcell resolution", *Journal of Computational Physics*, **83**, 148-184 (1989).
- [8] X-D. Liu, S. Osher, T. Chan, "Weighted essentially non-oscillatory schemes", *Journal of Computational Physics*, **115**, 200-212 (1994).
- [9] Ph. Colella, P.R. Woodward, "The piecewise parabolic method (PPM) for gas-dynamical simulations", *Journal of Computational Physics*, **54** (1), 174-201 (1984).
- [10] S.K. Godunov, A.V. Zabrodin, M.Ya. Ivanov et al., *Chislennoe reshenie mnogomernykh zadach gazovoy dinamiki*, M.: Nauka (1976).
- [11] A. V. Zabrodin, G. P. Prokopov, V.A. Cherkashin, "Self-adapted algorithms of gas dynamics", *Lect. Notes Phys.*, **90**, 587-593 (1979).
- [12] A. V. Zabrodin, "O problemah chislennogo modelirovaniya gazodinamicheskikh techeniy co

- slozhnoy strukturoy", *Konstruirovaniye algoritmov b reshenie zadach matematicheskoy fiziki*, M.: IPMatem. AN SSSR (1987).
- [13] S. B. Bazarov, "Digital Image Processing for Visualization of Numerical Results in Fluid Dynamics", *Application of Scientific Visualization in Applied Problems*, M: MGU, 39–42 (2000).
- [14] N. A. Dar'in, V. I. Mazhukin, A. A. Samarskii, "A finite-difference method for solving the equations of gas dynamics using adaptive grids which are dynamically connected with the solution", *U.S.S.R. Comput. Math. Math. Phys.*, **28** (4), 164–174 (1988).
- [15] V.I. Mazhukin, A.A. Samarskii, M.M. Chuiko, "Metod dinamicheskoi` adaptatsii dlia chislennogo resheniia nestatsionarnykh mnogomernykh zadach Stefana", *Doclady RAN*, **368**(3), 307 – 310 (1999).
- [16] V. Mazhukin, M. Chuiko. "Solution of the Multi-interface Stefan Problem by the Method of Dynamic Adaptation", *Computational Methods in Applied Mathematics*, **2**(3), 283–294 (2002). DOI: <https://doi.org/10.2478/cmam-2002-0017>.
- [17] P.V. Breslavskii, V.I Mazhukin. "Dynamic adaptation method in gasdynamic simulations with nonlinear heat conduction", *Comput. Math. Math. Phys.*, **48**(11), 2102-2115 (2008). DOI: 10.1134/S0965542508110158).
- [18] P. V. Breslavskii, V.I Mazhukin. "Dynamically adapted grids for interacting discontinuous solutions", *Comput. Math. Math. Phys.*, **47**(4), 687-706 (2007), DOI: 10.1134/S0965542507040124
- [19] OpenFOAM, website: <http://www.openfoam.org> (accessed July 7, 2019)
- [20] A.E. Bondarev, A.E. Kuvshinnikov, "Comparative study of the accuracy for OpenFOAM solvers", *Proceedings of Ivannikov ISPRAS Open Conference. IEEE Xplore*, 132-136 (2017). doi: 10.1109/ISPRAS.2017.00028
- [21] A. E. Bondarev, A. E. Kuvshinnikov, "Analysis of the Accuracy of OpenFOAM Solvers for the Problem of Supersonic Flow Around a Cone", *Lecture Notes in Computer Science*, **10862**, 221–230 (2018). doi:10.1007/978-3-319-93713-7_18
- [22] A. E. Bondarev, A. E. Kuvshinnikov, "Comparative Estimation of QGDFoam Solver Accuracy for Inviscid Flow Around a Cone", *Proceedings of the 2018 Ivannikov ISPRAS Open Conference, IEEE*, 82-87 (2018). doi: 10.1109/ISPRAS.2018.00019
- [23] A.K. Alekseev, A.E. Bondarev, A.E. Kuvshinnikov, "Verification on the Ensemble of Independent Numerical Solutions", *Lecture Notes in Computer Science*, **11540**, 315–324 (2019). doi: 10.1007/978-3-030-22750-0_25
- [24] A.E. Bondarev, V.A. Galaktionov, L.Z. Shapiro, "Postroenie I vizualniy analiz parallelnykh resheniy dlya optimizacionnykh I parametricheskikh issledovaniy v vychislitelnoy gazovoy dinamike", *Proceedings of 26-th International conference of computer graphics and vision*, 211-215 (2016).
- [25] A.E. Bondarev, "On the Construction of the Generalized Numerical Experiment in Fluid Dynamics", *Mathematica Montisnigri*, **42**, 52-64 (2018).
- [26] A.E. Bondarev, "On visualization problems in a generalized computational experiment", *Scientific Visualization*, **11** (2), 156-162 (2019). doi: 10.26583/sv.11.2.12
- [27] A.E. Bondarev, "On the Estimation of the Accuracy of Numerical Solutions in CFD Problems", *Lecture Notes in Computer Science*, **11540**, 325–333 (2019). doi: 10.1007/978-3-030-22750-0_26
- [28] M. Kraposhin, A. Bovtrikova, 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). doi:10.1016/j.procs.2015.11.007.
- [29] M. V. Kraposhin, D. A. Ryazanov, E. V. Smimova, T. G. Elizarova, M. A. Istomina, "Development of OpenFOAM Solver for Compressible Viscous Flows Simulation Using Quasi-Gas Dynamic Equations", *Ivannikov ISPRAS Open Conference*, 117-123 (2017).

- [30] M. A. Istomina, "About realization of one-dimensional quasi-gas dynamic algorithm in the open program OpenFOAM complex", Preprint IPM No. 1 (Moscow: KIAM), (2018).
- [31] A. Kurganov, E. Tadmor, "New high-resolution central schemes for nonlinear conservation laws and convection-diffusion equations", *J. Comput. Phys.*, **160**, 241–282 (2000), doi:10.1006/jcph.2000.6459.
- [32] C. J. Greenshields, H. G. Wellerr, L. Gasparini, J. M. 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), 1–21 (2010), doi:10.1002/flid.2069.
- [33] R. Issa, "Solution of the implicit discretized fluid flow equations by operator splitting", *J. Comput. Phys.*, **62** (1), 40–65 (1986). doi:10.1016/0021-9991(86)90099-9
- [34] B. N. Chetverushkin, T. N. Elizarova "Kinetic algorithms for calculating gas dynamic flows," *U.S.S.R. Comput. Math. Math. Phys.*, **5** (5), 164–169 (1985).
- [35] B. N. Chetverushkin, *Kinetic schemes and Quasi-Gas Dynamic system of equations*, CIMNE, Barcelona (2008).
- [36] T. G. Elizarova, *Quasi-Gas Dynamic Equations*, Springer (2009).

Received June 18, 2019

ОТКРЫТОЕ СОРЕВНОВАНИЕ РЕНДЕР-СИСТЕМ

В.А. ФРОЛОВ^{1,2}, Д.С. ПАВЛОВ², М.А. ТРОФИМОВ², П.А. КАЗБЕЕВ²,

В.А. ГАЛАКТИОНОВ^{1*}

¹Институт прикладной математики им. М.В. Келдыша РАН. Москва, Россия

²Московский Государственный университет имени М.В. Ломоносова. Москва, Россия

*Ответственный автор. E-mail: vlgal@gin.keldysh.ru

DOI: 10.20948/mathmontis-2019-45-9

Ключевые слова: Фотореалистичный рендеринг, интеграл освещённости, трассировка лучей методом Монте-Карло.

Аннотация. Сравнение алгоритмов и методов — краеугольный камень любого исследования, позволяющий обосновать, что предложенный подход делает шаг вперёд по сравнению с предыдущими исследованиями. В таких областях, как компьютерное зрение или машинное обучение, создано огромное количество открытых наборов данных, на которых исследователи могут тестировать свои алгоритмы и соревноваться друг с другом. В компьютерной графике и фотореалистичном рендеринге в настоящий момент ситуация обстоит по-другому. Открытых наборов сцен, на которых различные исследователи в разных рендер-системах могли бы получать совпадающие (или хотя бы близкие) изображения, не существует. Это приводит к тому, что в научных работах практикуется т. н. “cherry-picking” — педантичный выбор сцен и условий освещения таким образом, чтобы продемонстрировать преимущества разработанного алгоритма. Такой подход сильно понижает практическую значимость исследований — даже если новый метод хорошо работает на определённом классе сцен, это ещё не значит что от него можно добиться хороших результатов в других случаях. По этой причине многие рендер-системы до сих не выходят за рамки базовых алгоритмов, не доверяя результатам исследований.

В данной работе мы начали заполнять этот пробел. Нами был создан специальный набор сцен (т. н. «бенчмарк»), позволяющий оценить производительность интегрирования освещения в различных ситуациях и, таким образом, проявить положительные и отрицательные стороны рендер-систем и алгоритмов, используемых ими в различных условиях (изображения некоторых сцен представлены на рис. 1). Мы воссоздали многие сцены из известных статей по компьютерной графике и добавили некоторые сцены, исходя из своего опыта. Наша цель — сделать как можно более полное покрытие, задействовав как можно меньше сцен и функциональности рендер-системы для того, чтобы такое сравнение было легко воспроизвести в любой существующей системе.

Для валидации нашего подхода было проведено пилотное сравнение среди 4-х популярных продуктов для Autodesk 3ds Max (VRay, Corona, Octane, Hydra Renderer) на скорость интегрирования освещения на различных сценариях. Несмотря на то, что 3 из 4-х систем являются закрытыми коммерческими продуктами, нам удалось получить для всех сцен совпадающие или близкие изображения, что свидетельствует о жизнеспособности предлагаемого нами подхода сравнения.

2010 Mathematics Subject Classification: 68U05, 68M20, 62M05.

Key words and Phrases: Photorealistic rendering, integration of luminance, Monte-Carlo ray tracing.

OPEN RENDERING BENCHMARK

V. FROLOV^{1,2}, D. PAVLOV², M. TROFIMOV², P. KAZBEEV²,
V. GAVAKTIONOV^{1*}

¹Keldysh Institute of Applied Mathematics Russian Academy of Sciences
Moscow, Russia

² Moscow State University (MSU). Moscow, Russia

*Corresponding author. E-mail: vlgal@gin.keldysh.ru

DOI: 10.20948/mathmontis-2019-45-9

Summary. Comparison of algorithms and methods is the cornerstone of any research which allows to prove that the proposed approach takes a step forward with a previous one. In areas such as computer vision or machine learning a huge amount of open data sets have been created, in which researchers can test their algorithms and compete with each other. In computer graphics and photorealistic rendering at the current moment the situation is different. Open sets of scenes in which different researchers using different render systems could get matching (or at least close) images do not exist. This leads to the fact that in scientific papers so-called “Cherry-picking” is practiced. This one is a pedantic selection of scenes and lighting conditions in such a way as to demonstrate the advantages of the developed algorithm. Such approach greatly reduces the practical significance of research - even if the new method works well on a certain class of scenes, it does not mean it will work for other cases. For this reason, many render systems still do not go beyond the basic algorithms, not trusting the results of research.

In this paper, we began to fill this gap. We have created a special set of scenes (the so-called “benchmark”), which allows us to evaluate the performance of lighting integration in various situations and, thus, to show the positive and negative aspects of render systems and algorithms used by them in various conditions. We recreated many scenes from well-known computer graphics papers and added some scenes based on our experience. Our goal is to make the most complete coverage possible, using as few scenes and functionality of the render system as possible so that such a comparison can be easily reproduced in any existing system.

To validate our approach, we conducted a pilot comparison among 4 popular products for 3D Studio Max (VRay, Corona, Octane, Hydra Renderer) for the speed of integrating the lighting on various scenarios. Despite the fact that 3 out of 4 systems are closed commercial products, we managed to get the same or similar images for all scenes, which indicates the viability of our proposed approach.

1 ВВЕДЕНИЕ

Сравнение систем расчёта глобального освещения по многим причинам является нетривиальной задачей. На конференции SIGGRAPH 2018 известный учёный Ярослав Криванек выделил следующие фундаментальные проблемы, возникающие при проведении такого сравнения [1]:

1. В современном фотореалистичном рендеринге не существует стандартов. Все рендер-системы вычисляют в пределах разные эталонные изображения.

2. Коммерческие продукты зачастую реализуют огромное количество “физически некорректных” функций. Причём, каждая рендер-система делает это по-своему. Пример подобной функциональности — карты нормалей (normal mapping) [2].

3. Наконец, метрика качества для изображений, определяющая качество изображения в настоящий момент является открытой научной проблемой [3–5].

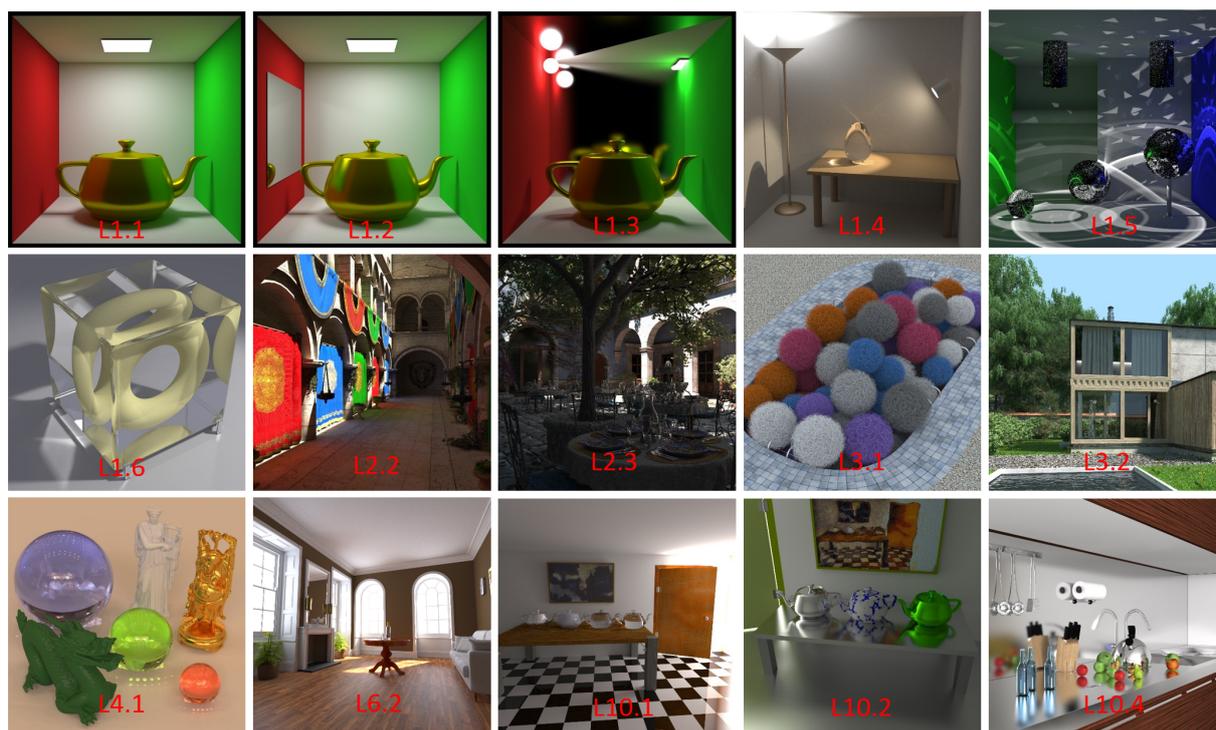


Рис. 1: Изображения некоторых сцен из созданного набора.

Краткий обзор существующих работ свидетельствует о том, что сравнения зачастую проводятся на нескольких сценах, выбранных случайным образом [6–8]. Это не позволяет проанализировать характеристики устойчивости рендера к различным сценариям: массивная геометрия, оптически сложные материалы, сложное освещение, наличие трудновычисляемых феноменов освещённости [9-11]. Время расчёта, которое при этом пишется в подобных сравнениях, в действительности не демонстрирует ровным счётом ничего, поскольку качество изображения фиксируется “на глаз” лишь приблизительно. В то же время, едва заметная для глаза разница в шуме на изображении может означать разницу в скорости в разы. Это особенно верно при приближении к “финальному качеству”, когда изображение становится максимально чистым, поскольку качество в физически корректных рендерах, как правило, пропорционально квадрату времени расчёта.

Открытые наборы сцен [12,13], с другой стороны, как правило, сохраняют сцены в разрозненном виде. В лучшем случае из них можно получить информацию о геометрии и некоторое изображение, полученное неизвестной рендер-системой в прошлом.

Материалы и условия освещения, при которых возникают, например, трудновычисляемые феномены освещённости при этом безнадежно теряются.

В 2014 году впервые был представлен [14] более точный анализ производительности для нескольких рендер-систем под Autodesk 3ds Max на 7 различных упрощённых сценариях. Для того чтобы решить проблему отличающихся эталонов, в этой работе предлагается использовать для каждой системы свой собственный эталон, получаемый именно этой системой за достаточно большое время и не содержащий шума [14]. При этом считается, что изображения эталонов визуальны совпадают. То есть рендер-системы отличаются лишь в “реализации отдельных незначительных моментов”. Общий уровень яркости, геометрическая форма объектов на изображениях и характер освещения должны совпадать. Далее вводился некоторый индекс производительности, зависящий от среднеквадратичной ошибки (MSE) между изображением за некоторое фиксированное время и эталоном. Таким образом, этой в работе фактически измеряется то, насколько быстро алгоритм/реализация/рендер-система сходится к своему собственному эталону [14].

Хотя в этом сравнении сцен было существенно больше, чем в существующих ранее сравнениях, данный набор нельзя считать объективным. Кроме того, среди использованных сцен не было задействовано ни одной популярной в научном сообществе сцены. Это делает такое сравнение менее интересным для специалистов, т. к. не позволяет хотя бы приблизительно сопоставить существующие реализации в статьях и индустриальных программах.

В данной работе мы ставим задачу создания т. н. “бенчмарка” — способа объективного сравнения производительности различных алгоритмов интегрирования освещённости и их реализаций. Наша глобальная цель заключается в том, чтобы предоставить возможность сравнения производительности всех существующих рендер-систем в мире, а также отдельных алгоритмов интегрирования освещённости и их реализаций, решающих эту проблему численно и с высокой точностью (т. е. мы не рассматриваем приближённые алгоритмы глобального освещения).

2 ПОСТАНОВКА ЗАДАЧИ И ОБОСНОВАНИЕ ВЫБОРА РЕНДЕР СИСТЕМ

2.1 Постановка задачи

С математической точки зрения все рендер-системы в той или иной мере решают проблему глобальной освещённости [15] (формула 1). Эта проблема сводится к вычислению многомерного интеграла, называемого интегралом освещённости (соответственно решение этой проблемы называют “интегрированием освещённости”). Размерность интеграла зависит от количества переотражений света, которые необходимо учитывать, т. к. значение функций L в одних точках трёхмерной сцены зависит от значения самого интеграла освещённости в других точках (формула 1).

$$I(\phi_r, \theta_r) = \int_{\phi_i, \theta_i} L(\phi_i, \theta_i) R(\phi_i, \theta_i, \phi_r, \theta_r) \cos(n, l_{\phi_i, \theta_i}) d\phi_i d\theta_i \quad (1)$$

Формула 1. Уравнение рендеринга или т. н. интеграл освещённости [15]. Пара (ϕ_r, θ_r) задаёт направление на наблюдателя, пара (ϕ_i, θ_i) — направление на источник освещения. Функция L — яркость падающего освещения, R — двунаправленная функция отражательной способности (ДФО). Важно отметить, что значение функций L в одних точках трёхмерной сцены зависит от значения самого интеграла освещённости в других точках сцены. Поэтому данная проблема трудная, и эквивалентна решению интегрального уравнения Фредгольма 2 рода.

Системы Corona [16] и V-Ray [17] являются адептами “старой школы”. Они работают на CPU (Central Processing Unit или центральный процессор) и для ускорения используют так называемые алгоритмы со смещённым решением [18–22]. В этих алгоритмах численная точность не ставится во главу угла. Кроме того, они не обладают способностью достигать правильного решения в пределе, что можно считать недостатком. С другой стороны, типичный пользователь рендер-системы в Autodesk 3ds Max беспокоится не столько о правильности решения, сколько о реалистичности и красоте. Однако здесь есть нюанс: смещённые алгоритмы зачастую дают артефакты, неприемлемые для пользователя [23].

Ocane [24] и Hydra [25] во многом противоположны рассмотренным ранее программам Corona и V-Ray. Они ориентированы на расчёт без допущений (несмещённая оценка) и работают полностью на GPU. Кроме того, обе системы используют методы рендеринга на основе Марковских цепей. Программа Ocane использует алгоритм PMC/ERPT [26,27] (Population Monte-Carlo/Energy Redistribution Path Tracing), Hydra – MMLT [18] (Multiplexed Metropolis Light Transport).

Отдельно следует сказать несколько слов про рендер-систему Hydra Renderer. Эта система, как и Ocane, позиционирует скорость в качестве своего основного преимущества. Однако, в отличие от остальных является полностью открытым программным обеспечением. Поэтому, хотя она и не является столь распространённой как три оставшиеся программы, для наших целей она подходит. Открытость системы позволяет нам анализировать её исходные коды и используемые алгоритмы для того, чтобы создавать “предельные случаи” – сцены, на которых определённые алгоритмы работают плохо, либо, наоборот хорошо, т. к. они создавались именно для таких сцен.

2.2 Используемые рендер-системы в деталях

Так как V-Ray, Corona и Ocane являются закрытым программным обеспечением, мы не можем со 100% уверенностью утверждать что-либо об их внутренней реализации. Тем не менее, обладая опытом разработки аналогичных систем, анализируя публикации и артефакты, производимые различными алгоритмами в этих системах, можно с достаточной степенью достоверности делать предположения об использовании тех или иных методов рендеринга.

2.2.1 Система V-Ray

Система V-Ray впервые стала доступной для широкой общественности в 2000 году. Ключевыми средствами повышения производительности являются кэш освещенности (называемый в V-Ray термином "Irradiance Maps") и некоторый закрытый алгоритм, называемый в V-Ray термином "Light Cache". Каустики в V-Ray могут рассчитываться при помощи фотонных карт с накоплением [20,21]. Таким образом, алгоритмы, используемые в данной системе, были разработаны в 80–90 годах прошлого века и в настоящий момент могут считаться морально устаревшими (что, тем не менее, не означает, что они не могут быть с успехом использованы).

Начиная с версии 3.5, V-Ray изменила направление развития и перешла в сторону GPU и несмещённых алгоритмов (V-Ray Next). В настоящий момент система проходит бета-тестирование и информации по ней не так много. По этой причине для тестирования GPU за точку отсчёта была выбрана система Octane.

2.2.2 Система Corona

Corona набрала популярность в течение 2012 года. Эта программная система использует такие алгоритмы как Radiance Cache [18,19] и фотонные карты [20,21] но, наряду с этим, реализует более современный метод — Vertex Connection Merging (VCM [22]), являющийся гибридом фотонных карт и двунаправленной трассировки путей (Bidirectional Path Tracing [28]). По сравнению с V-Ray используемые подходы являются более современными (Radiance Cache разрабатывался в 2000-ых годах, а VCM был представлен в 2012 на SIGGRAPH Asia).

2.2.3 Система Octane

Octane ориентирован на расчёт без допущений (unbiased rendering), позиционируется как самая быстрая рендер-система в мире и работает полностью на GPU. Она также впервые появилась в 2012 году. В настоящий момент поддерживаются только GPU с Nvidia CUDA, несмотря на неоднократные заявления о поддержке OpenCL. Одним из существенных преимуществ в плане алгоритмов является использование методов рендеринга на основе Марковских цепей — PMC/ERPT [26,27]. Однако разработчики ограничились лишь однонаправленной версией PMC на основе обратной трассировки путей.

2.2.4 Система Hydra

Hydra Renderer, также как и Octane, ориентирована на расчёт без допущений. Алгоритмы расчёта глобального освещения реализованы целиком на GPU посредством технологии OpenCL. Отличительной особенностью является использование двунаправленной трассировки путей и двунаправленного алгоритма Multiplexed MLT (MMLT), впервые представленного в 2014 году [29]. В отличие от рассмотренных ранее систем Hydra Renderer является полностью открытым программным обеспечением. Таким образом, она позиционируется как некоторое ПО, соединяющее научные разработки в области рендеринга и параллельных вычислений на GPU с индустрией. При этом данная система не обладает таким же большим объёмом функциональности, как упомянутые ранее коммерческие системы. Но в нашем сравнении этого и не

требуется.

3 РАЗРАБОТАННЫЙ НАБОР СЦЕН И ОГРАНИЧЕНИЙ

Одним из вызовов при подготовке сравнения является создание объективного набора тестовых сцен. Здесь существует несколько трудностей, которые необходимо решать.

3.1 Стандартизация, спецификация

Первая проблема, как уже было отмечено, это отсутствие стандарта для физически-корректных рендер-систем. Например, в некоторых системах могут быть реализованы математические модели материалов или источники света, отсутствующие в других. Чтобы минимизировать различие в изображениях и чтобы каждая рендер-система была способна в принципе отобразить эти сцены так, как они должны выглядеть, мы вводим набор ограничений при создании сцен (т. е. стремимся минимизировать используемый набор функций):

1. Материал может быть только следующих типов:
 - i. Полностью диффузный (Ламбертов) материал;
 - ii. идеально зеркальный материал;
 - iii. матовые (glossy) отражения с использованием т. н. “исправленной модели Фонга”;
 - iv. простое стекло (идеально-зеркальное, т.е. не матовое) с использованием формул Френеля для диэлектриков;
 - v. материал, обладающей суммой компонент: (i и ii) или (i и iii);
 - vi. материал, смешивающий компоненты (i и ii) или (i и iii) при помощи формул Френеля для диэлектриков в одной-двух сценах;
 - vii. Допускается использование техники bump mapping в нескольких сценах. При этом разрешается использовать только готовые карты нормалей (а не карты высот, из которых рендер-системы самостоятельно оценивают нормаль).
 - viii. Материал с транслюценцией – пропусканием света под поверхность. Такой материал разрешается использовать только на тонких объектах вроде листьев и штор. Отношение отражённой энергии к пропущенной должно быть строго 1 к 1.
2. Источник света может быть только следующих типов:
 - i. Площадный источник света прямоугольной формы не нулевого размера;
 - ii. сферический источник света не нулевого размера;
 - iii. направленный источник света;
 - iv. точечный источник-прожектор (spot распределение);
 - v. окружение, заданное константой, т.е. цветной фон без текстуры;
 - vi. окружение в виде HDR панорамы в одной-двух сценах.

Выбор подобных ограничений – это компромисс между упрощением для исследователей и интересностью полученных результатов для индустрии. Изначально мы планировали оставить в материалах только пункты i – iv, а источники ограничить пунктами i и ii. Это бы значительно упростило реализацию алгоритмов для будущих исследователей (т.к. такой набор позволяет однозначно и легко реализовать физически корректный рендеринг) и позволило бы более легко добиться совпадения изображений в нашем сравнении. Однако в действительности это было бы слишком сильным упрощением задачи глобального освещения по сути, поскольку:

1. Материалы, смешивающие несколько элементарных компонент, используются повсеместно. Кроме того, многокомпонентные материалы увеличивают размерность пространства интегрирования, поскольку интегратору необходимо делать выбор между компонентами и использовать под это ещё одно случайное число, ещё одну размерность. Поэтому запрет такой функциональности был бы слишком сильным упрощением. С другой стороны, полноценное дерево компонент выглядит чрезмерно сложным для оценки производительности. Поэтому мы ограничили такое смешивание всего двумя заранее известными компонентами.

2. Формулы Френеля для смешивания компонент тоже используются слишком часто, чтобы их не рассматривать. Тем не менее, мы старались сократить число сцен, в которых данный вид смешивания был применён, чтобы можно было оценить производительность на большинстве сцен без реализации этой функциональности (Fresnel Blend в терминах PBRT).

3. Bump Mapping и карты нормалей [2] существенно повышают реалистичность и повсеместно используются. Кроме того, такие алгоритмы как фотонные карты и однонаправленный Kelemen Metropolis Light Transport [30] теряют свои преимущества при наличии микрорельефа в том или ином виде. Это делает их значительно менее полезными на практике, и отчасти именно поэтому они не часто используются в индустрии. Поэтому с нашей точки зрения системы и алгоритмы, которые будут не вполне справляться с микрорельефом, нужно штрафовать, а значит, эта функциональность должна быть включена. Тем не менее, мы старались не использовать карты нормалей слишком часто, т. к. они представляют собой отдельную функциональность рендера и, строго говоря, нарушают физическую корректность интерпретации сцены, поскольку изменяют нормаль к поверхности, не меняя при этом геометрию поверхности.

4. Транслюценция (просвечивание) необходима для реалистичной имитации сцен с растительностью. Заменять их полностью диффузными материалами было бы неправильно с точки зрения производительности, т.к. это существенно меняет траектории путей света в сцене.

5. Направленный источник света легко реализуется и часто встречается в научных работах. Используется в различных популярных сценах. Поэтому мы решили добавить его в базовый набор.

6. Прожектор (spot) необходим для имитации освещения от фонарных столбов и других типичных уличных светильников. Его отличительная особенность в том, что он может быть эффективно отбракован при вычислении освещения, если область находится вне освещаемого конуса.

7. Окружение в виде панорамы с текстурой также часто встречается, но уже в индустрии. Отдельный вопрос – это HDR текстура, поскольку её нужно сэмплировать пропорционально яркости пикселей. Мы решили ограничиться константным цветом в большинстве сцен, чтобы избежать ошибок с ориентацией систем координат при переводе из декартовой в сферическую. Кроме ситуации с HDR панорамой, когда именно в этом состоит смысл теста.

3.2 Покрытие

Следующая проблема заключается в составлении объективного набора сцен,

наиболее полно покрывающего как типичные случаи из индустрии, так и известные проблемы, и вызовы в науке. Мы постарались расширить подход 2014 года [14] и сделать наш набор сцен более полным, уделяя внимание не только сценам, привычным для индустрии, но и сценам, часто используемым в научных работах и демонстрирующим известные вызовы в области интегрирования освещения (т. н. Light transport challenges). Мы сгруппировали сцены по номерам. Каждому номеру соответствует некоторая группа сценариев, встречающаяся в индустрии либо некоторому известному вызову в науке. Сами группы мы выбирали не случайным образом, а формировали их в соответствии с известными вызовами и тестовыми сценами, встречающимися в научных работах по компьютерной графике.

1. Cornell Box [31] с несколькими вариациями. Геометрически простые сцены, содержащие тем не менее почти все основные эффекты: шумное первичное освещение и мягкие тени, зеркальные блики от источника освещения, отраженные каустики.

1.1 Базовый вариант, содержащий все эффекты более-менее поровну.

1.2 Базовый вариант + зеркало, создающее широкий каустик, плюс SDS (Specular-Diffuse-Specular) каустик, видимый в отражении. Тривиальное изменение сцены, кардинально меняющее, тем не менее, характер освещения.

1.3 Вариант, ориентированный на многократные глянцевые отражения.

1.4 Оригинальный Cornell Box Эрика Вича, используемый им для демонстрации Metropolis Light Transport [32]. Ярко выраженные трудновычислимые феномены освещённости. SDS каустики присутствуют в небольшом количестве (в стеклянном яйце).

2. Сильное вторичное диффузное освещение и геометрия средней тяжести.

2.1 Dabrovic Sponza [12] и вторичный свет от направленного источника. Часто используется в статьях по компьютерной графике. Недостаток оригинальной сцены в том, что все текстуры в ней однотонные, и вторичное освещение получается в значительной степени равномерным. Поэтому практически любой сколь угодно плохой алгоритм глобального освещения показывает на этой сцене высокую точность. По этой причине в итоговой версии бенчмарка мы исключили данную сцену.

2.2 Crytek Sponza [12] – усложнённый вариант предыдущей сцены, с более детальной геометрией, треугольниками с текстурой непрозрачности на листьях и (opacity для имитации растительности) и разноцветными шторами. В различных демонстрационных программах и статьях шторы (как и растительность) могут частично пропускать свет или полностью отражать его. Мы выбрали второй вариант, т.к. он является более распространённым.

2.3 San Miguel [12] – более тяжёлая геометрическая сцена с растительностью, пропускающей свет. По-прежнему яркое солнце и сильное вторичное освещение. Более сложная с геометрической точки зрения (частично из-за opacity, частично просто из-за более сложной геометрии).

2.4 Ночной вариант сцены San Miguel. Несколько источников света, более сложное первичное освещение.

3. Сцены на исключительно сложную геометрию. Но простые в плане освещения.

4. Тестирование освещения от HDR панорамы. Геометрия простая. Сложность этого сценария в сильно неравномерном первичном освещении от карты окружения (HDR панорамы).

5. Тест на небесные порталы – источники света в окнах, имитирующие рассеянный

свет с улицы. Данный сценарий проверяет систему на наличие реализации многократной выборки по значимости (Multiple Importance Sampling [28,32]), поскольку при наличии больших площадных источников явная стратегия сэмплирования (теневые лучи) становится мало эффективной.

6. Тесты на вторичное освещение от слонца, образованное относительно узкой, но яркой полоской света.

7. Трудновычислимые феномены освещённости (light transport challenge), вызванные многократными отражениями от поверхностей с различными свойствами.

8. Трудновычислимые феномены освещённости (light transport challenge), вызванные SDS каустиками (бассейн с водой).

9. Трудновычислимые феномены освещённости (light transport challenge), вызванные отражениями от глянцевых поверхностей в сцене, содержащей карты нормалей. Данный вид сцен может представлять проблему для однонаправленных алгоритмов на основе Марковских цепей, работающих в первичном пространстве путей.

10. Трудновычислимые феномены освещённости, оригинальная сцена Эрика Вича с приоткрытой дверью и ещё одна сцена, её вариация с глянцевыми отражениями и некоторые менее известные сцены, проверяющие трудновычислимые феномены освещённости.

11. Сцена с самосветящимися поверхностями.

12. Тест на большое количество источников света.

13. Известная по курсу SIGGRAPH 2014 [33] сцена красной кухни на трудновычислимые феномены освещённости, содержащая в себе элемент “зеркального коридора”.

4 МЕТОДИКА ТЕСТИРОВАНИЯ

Следующая серьёзная проблема, которую необходимо решить – это отсутствие совпадения эталонов. Различные программные системы в действительности вычисляют различные изображения. Отличия вызваны отсутствием общего стандарта для фотореалистичного рендеринга и огромным количеством деталей в реализации. Более того, даже одна и та же программная система может давать отличающиеся эталонные изображения при использовании разных алгоритмов. Это может являться как следствием ошибок в реализации (которые мы, вообще говоря, не всегда хотим учитывать, если нас интересует только скорость интегрирования), так и некоторыми объективными факторами. Такими как невозможность расчёта определённых феноменов в одном из алгоритмов либо намеренное упрощение вычислений (например, так называемый “clamping”, когда выбросы ограничиваются некоторым максимальным значением).

В нашей работе мы выбрали следующий компромиссный критерий оценивания:

3. Мы требуем лишь визуального совпадения эталонов для различных систем и алгоритмов. Форма освещения и приблизительный уровень яркости должны совпадать. Допускаются незначительные отличия, не влияющие на суть вычисляемых феноменов освещённости.

4. Если пункт 1 выполнен, то для каждой системы и каждого алгоритма мы вычисляем свой эталон, получающийся при достаточно большом времени расчёта этим

алгоритмом в этой самой системе. Единственно исключение – для алгоритмов со смещённой оценкой мы выбираем в качестве эталона Path Tracing [15] (но в той же самой системе).

5. Если у какой-либо системы пункт 1 не выполнен для разных алгоритмов, тогда для этой системы за её эталон мы принимаем трассировку путей (Path Tracing [15]) реализованную в этой системе. Path Tracing, как базовый алгоритм, реализован везде, мы можем использовать его. Смысл данного пункта в том, что мы хотим штрафовать алгоритм или систему, если реализация не может рассматриваться как корректная даже с натяжкой.

6. Если по какой-то причине в одной из систем эталонное изображение значительно отличается от других, тогда на основе анализа вычисляемых феноменов освещённости выносится вердикт о некорректной реализации, и за данный тест система получает ноль баллов. Впрочем, в наших тестах таких прецедентов не было.

Для оценки ошибки и последующего вычисления баллов мы использовали известную метрику – MSE. Плюс этой метрики для нашего сравнения заключается в линейной зависимости MSE (квадрат ошибки, формула 2) от времени расчёта для систем с несмещённой оценкой. Таким образом, чтобы уменьшить в 2 раза MSE, необходимо рассчитывать изображение в 2 раза дольше. Тестирование производилось на следующем оборудовании: для GPU рендер-систем была использована видеокарта Nvidia GTX2070, для CPU рендер-систем аналогичный по стоимости процессор Intel Xeon E5 2690v2 3Ghz (10 ядер, 20 потоков).

$$MSE = \frac{1}{W * H} \sum_{i=1}^{W*H} ((R_i - R_i^{ref})^2 + (G_i - G_i^{ref})^2 + (B_i - B_i^{ref})^2) \quad (2)$$

Формула 2. Вычисление MSE между измеряемым изображением (R,G,B) и эталоном ($R^{ref}, G^{ref}, B^{ref}$), получаемым за большое время. W — размер изображения по горизонтали. H — по вертикали. Тройки (R,G,B) обозначают значения в красном, зелёном и синем каналах изображений.

$$Winner = \underset{t}{Maxcount} \left\{ \underset{render}{argmin} MSE(render, t) \right\} \quad (3)$$

Формула 3. Выбор победителя по метрике MSE (получающего 3 балла) происходит по нескольким точкам на графике, соответствующим существенно различающимся временам расчёта (*maxcount* по *t*). Далее (соревнование за 2 балла) проводится аналогично. *render* — одна из четырёх тестируемых систем.

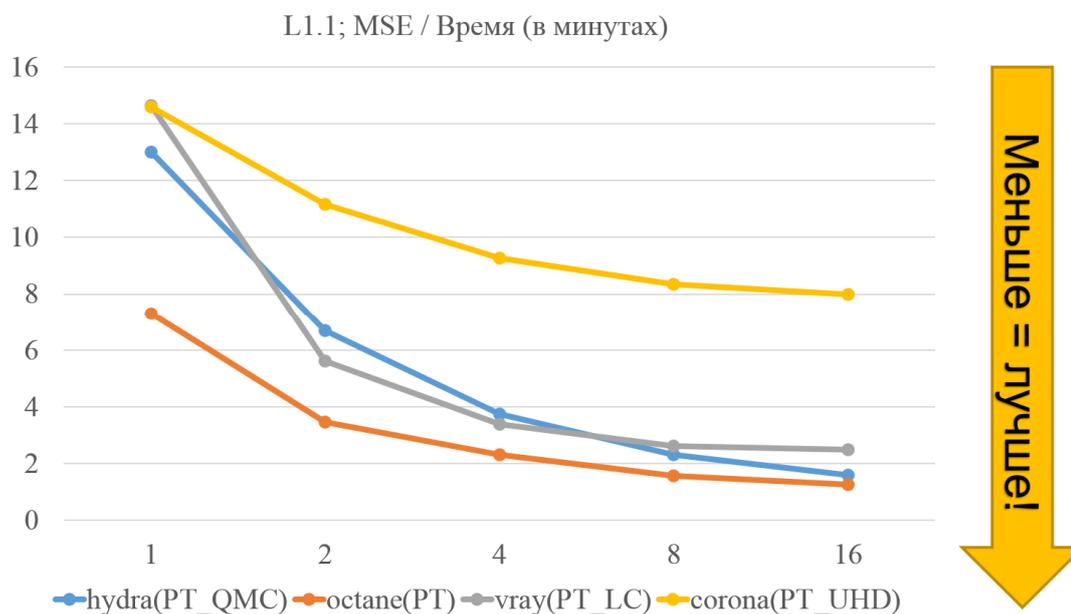


Рис. 2. Зависимость квадрата ошибки (MSE) от времени на сцене L1.1. На графике каждая следующая позиция соответствует увеличению времени в 2 раза. Можно считать, что MSE и скорость связаны линейным соотношением. Уменьшение MSE в 2 раза соответствует увеличению времени расчёта в 2 раза.

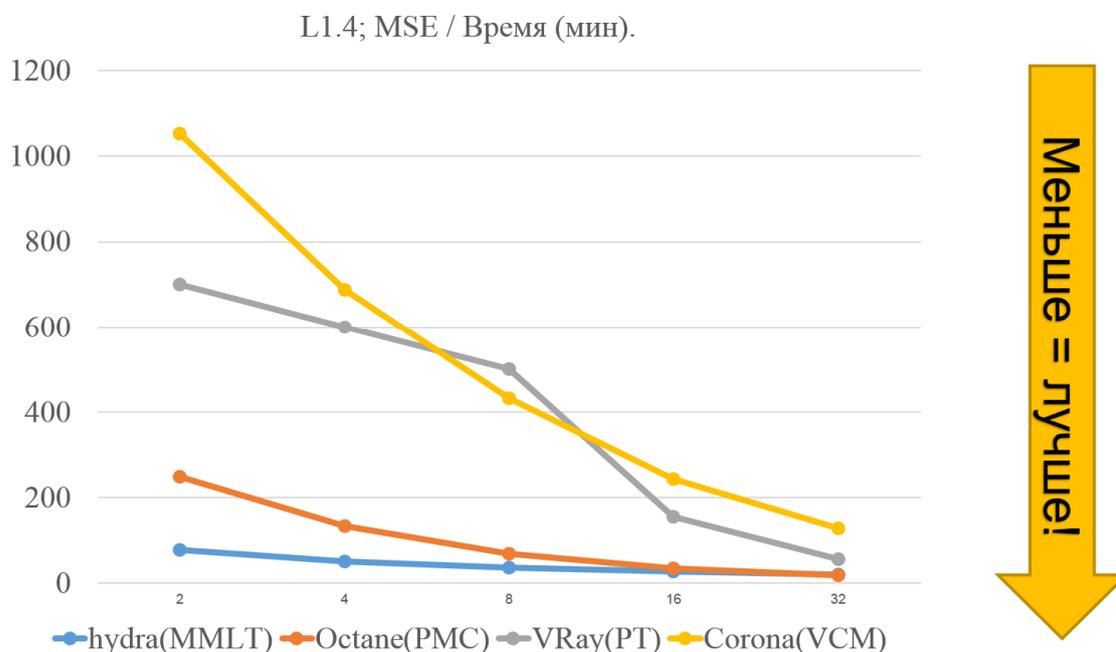


Рис. 3. Зависимость квадрата ошибки (MSE) от времени на сцене L1.4. На графике каждая следующая позиция соответствует увеличению времени в 2 раза. Можно считать, что MSE и скорость связаны линейным соотношением. Уменьшение MSE в 2 раза соответствует увеличению времени расчёта в 2 раза.

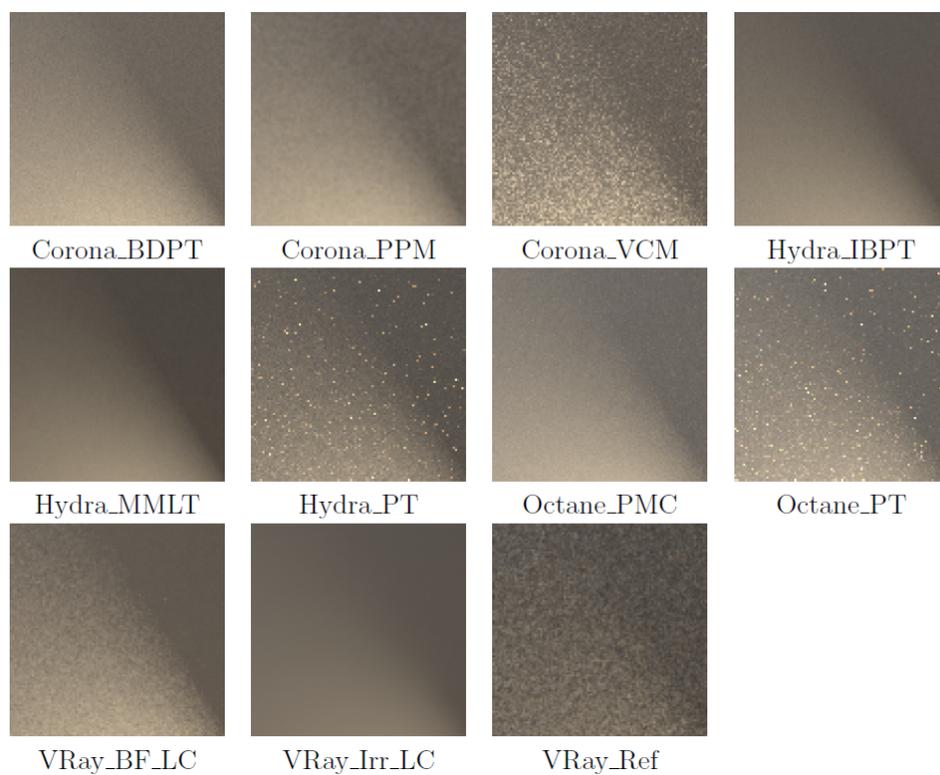


Рис. 4: Пример сопоставления рендер-систем и их алгоритмов на сцене ORB L1.4. 16 минут.

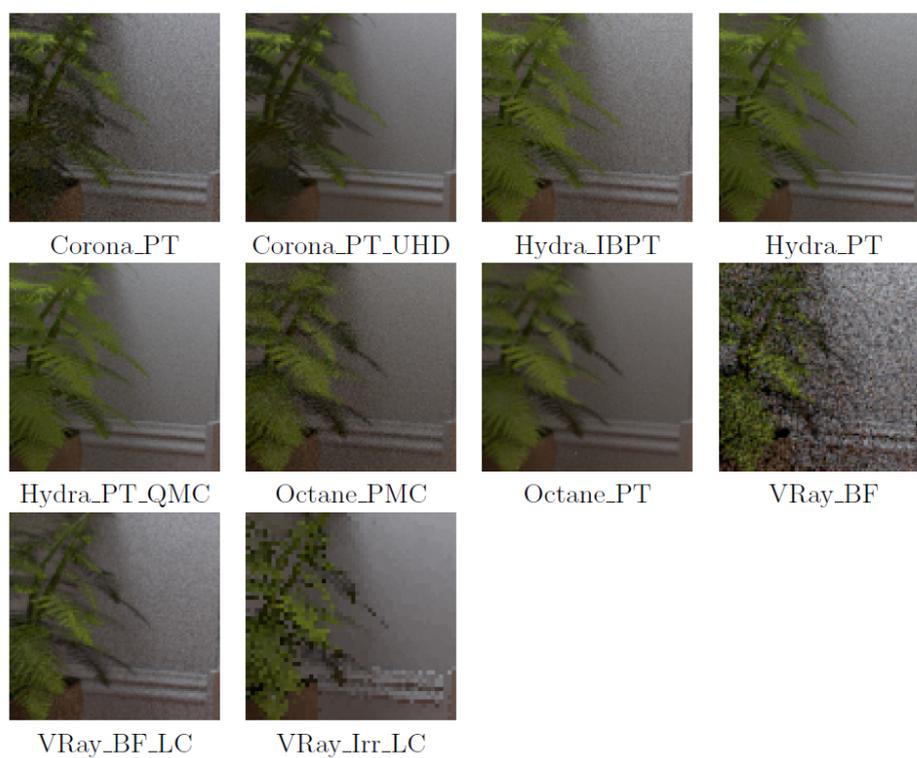


Рис. 5: Пример сопоставления рендер-систем и их алгоритмов на сцене ORB L6.2. 8 минут.

5 АНАЛИЗ РЕЗУЛЬТАТОВ

В экспертном тестировании участвовали 10 профессиональных 3D художников. Примеры визуального сравнения, по которому производилась экспертная оценка, показаны на рис 4,5. Каждому эксперту было показано 17 сопоставлений отдельных фрагментов изображений и было предложено отранжировать системы по местам – 1, 2 и 3. За первое место мы ставили 3 балла, за второе 2, а за третье 1. При этом для каждой системы выбирался только наиболее удачный алгоритм. Эксперт может поставить ноль баллов системе (то есть не назначать никакое место), если он считает, что изображение некорректно либо результат, по его мнению, “не тянет” на призовое место. Результаты экспертного ранжирования приведены на рис. 6 слева.

Примеры численного сравнения (метрика MSE) можно увидеть на рис. 2 и 3. Однако они не наглядны, особенно если оценивать их для всех сцен (17 графиков). Поэтому для того чтобы поставить такое сравнение в один ряд с экспертным, мы использовали ту же соревновательную форму (формула 3), в которой единственным “экспертом” была метрика MSE, а баллы ставились тем же образом от 3 за первое место до 1 или 0 за последнее (рис. 6 справа).

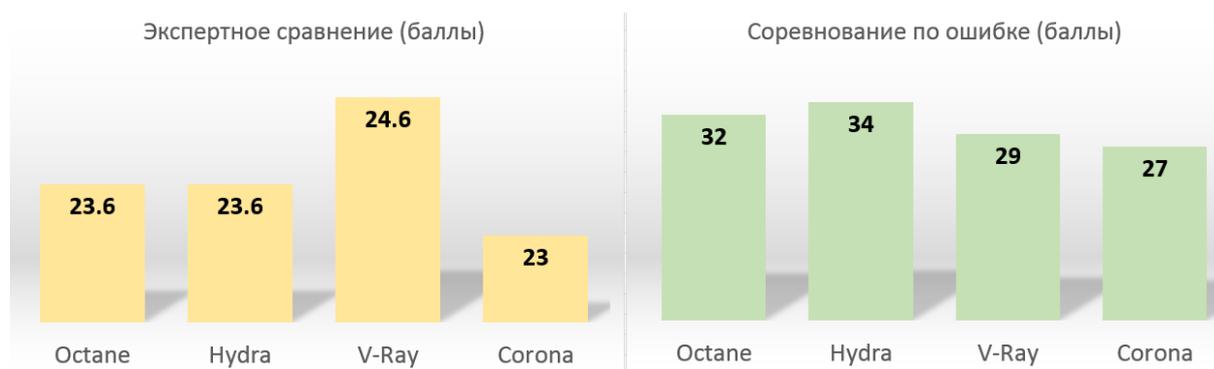


Рис. 6: Экспертное сравнение в “баллах” (слева). Соревнование по точности/ошибке (справа).

Результаты сравнения стали во многом для нас неожиданными. С одной стороны, с точки зрения ошибки победили системы с расчётом без допущений на GPU (Octane и Hydra). Это, в принципе, ожидаемо, т. к. эти системы ориентированы на точный расчёт. С другой стороны, наиболее “чистые” изображения давала система V-Ray, которая победила в экспертном сравнении. Далее, если внимательно посмотреть на цифры, видно, что разница между победителем и проигравшим в экспертном сравнении очень незначительна (1.5 балла). А в сравнении по ошибке она достигает 7 баллов. Из этого можно заключить, что измерение MSE – более чувствительный способ сравнения, чем экспертный. И более понятный, поскольку, в отличие от экспертной оценки, ошибка может быть измерена непосредственно.

Неожиданным оказалось также, что по оценке экспертов (эксперты оценивали качество изображений за одно и то же время, но мы называем это также оценкой производительности) большой разницы в производительности этих рендер-систем нет. Используя разные подходы, разные алгоритмы и разные вычислители (CPU и GPU) все продукты, за исключением отдельных сцен, достигают паритета между собой. Из этого

можно сделать как минимум 2 вывода. Первый — борьба между разными подходами, методами и системами будет продолжаться, т. к. ярко выраженного победителя нет. Второй — расчёт без допущений на GPU вплотную подобрался по скорости к “приближённым” методам на CPU, использующим смещённую или состоятельную оценку. Это означает, что за исключением небольшого количества случаев (вроде искусственного примера L1.5, рис. 1) “приближённые” методы теряют свою актуальность в фотореалистичном рендеринге, поскольку методы расчёта без допущений, как правило, универсальней и проще в поддержке (добавление новых материалов, источников и т. д.). Кроме того, расчёт без допущений делает осмысленным усложнение моделей взаимодействия света с веществом для всё большего погружения рендер-систем в физику и углубления моделирования вплоть до квантового уровня (такие работы существуют [34,35]).

При сравнении коммерческой и открытой системы на GPU мы видим также, что существенной разницы в производительности между ними нет. Причём, это показывает и ошибка, и эксперты. В одних сценариях побеждает Hydra, в других Octane. Это лишний раз подчёркивает необходимость тесного взаимодействия науки и индустрии посредством разработки открытого программного обеспечения, т. к. закрытые научные разработки (Octane) не показали преимущества над открытой реализацией хорошо известных методов.

Сравнивая Octane и Hydra, можно косвенно сравнить технологии CUDA и OpenCL, поскольку Octane реализован на CUDA, а Hydra на OpenCL. На сцене с тяжёлой геометрией, но простым освещением видно (сцена L2.2, рис. 1), что ощутимой разницы в производительности трассировки лучей нет, поскольку на такой сцене скорость интегрирования будет зависеть исключительно от скорости трассировки лучей. Следовательно, закрытым решениям, использующим аппаратные особенности конкретных GPU, вполне можно предпочесть открытую и аппаратно-независимую технологию OpenCL без существенного ущерба производительности.

6 ЗАКЛЮЧЕНИЕ

Предложенный способ сравнения позволяет провести сопоставление различных алгоритмов и различных рендер систем, даже если их эталоны не совпадают в силу различий в реализациях моделей материалов и источников освещения. Стандартизация такого “бенчмарка” в будущем позволит избежать т. н. “cherry-picking” в исследованиях по рендерингу [1] и сделает эти исследования более достоверными. Хотя, возможно, они станут менее зрелищными в силу того, что сцены всегда будут одни и те же.

Мы продолжаем расширять набор сцен и ставим своей целью создать объективное сравнение для всех доступных рендер-систем в мире. Открытость исходных кодов нашей системы (Hydra) позволяет будущим исследователям получить совпадающие изображения, поскольку все формулы, по которым ведутся вычисления в нашей системе, известны. Это даёт возможность в будущем проводить более точное сопоставление между нашими результатами и результатами других исследователей.

REFERENCES

- [1] J. Křivánek, “Open problems and research directions”, *Proceedings of SIGGRAPH-2018 conference*,

- Vancouver, Canada (2018). <https://cgg.mff.cuni.cz/~jaroslav/papers/2018-archvizcourse/slides/07-krivanek-research-notes.pdf>
- [2] Vincent Schüssler, Eric Heitz, Johannes Hanika, and Carsten Dachsbacher, “Microfacet-based normal mapping for robust Monte Carlo path tracing”, *ACM Trans. Graph.*, **36**(6), Article 205, (2017). doi://doi.org/10.1145/3130800.3130806
- [3] Robert Herzog, Martin Čadík, Tunç O. Aydıçin, Kwang In Kim, Karol Myszkowski, and Hans-P. Seidel, “NoRM: No-Reference Image Quality Metric for Realistic Image Synthesis”, *Comp. Graph. Forum*, 545-554 (2012). doi: 10.1111/j.1467-8659.2012.03055.x
<http://dx.doi.org/10.1111/j.1467-8659.2012.03055.x>
- [4] Martin Čadík, Robert Herzog, Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel, “New measurements reveal weaknesses of image quality metrics in evaluating graphics artifacts”, *ACM Trans. Graph.*, **31**(6), Article 147, (2012). doi: <https://doi.org/10.1145/2366145.2366166>
- [5] Qingbo Wu, Hongliang Li, Fanman Meng, King N. Ngan, *A Perceptually Weighted Rank Correlation Indicator for Objective Image Quality Assessment*, ArXiv preprint, (2017).
<https://arxiv.org/abs/1705.05126>
- [6] Render Engine Comparison: Cycles vs The Rest (2015).
<https://www.blenderguru.com/articles/render-engine-comparison-cycles-vs-giants> (accessed July 06, 2019).
- [7] Render Tests, Comparisons and Guides (2019). <https://www.antonibosi.com/maya-render-tests-reviews-comparisons> (accessed July 06, 2019).
- [8] Yuichiro Yama, “Render Comparison Test 4 (Volume) - Redshift, Arnold, V-Ray, Mantra”, *Youtube.com* (2018). <https://www.youtube.com/watch?v=K0Wn6lhGNAQ> (accessed July 06, 2019).
- [9] V.A. Frolov, V.A. Galaktionov, “Memory-compact Metropolis light transport on GPUs”, *Programming and Computer Software*, **43**(3), 196-203 (2017). doi: 10.1134/S0361768817030057
- [10] S.V. Ershov, A.A. Garbul, S.G. Pozdnyakov, V.G. Sokolov, A.G. Voloboy, “Lighting simulation in anisotropic media”, *Mathematica Montisnigri*, **39**, 42-56 (2017).
- [11] S.V. Ershov, D.D. Zhdanov, A.G. Voloboy, “Calculation of luminance of scattering medium by MCRT using multiple integration spheres”, *Mathematica Montisnigri*, **44**, 122-134 (2019).
- [12] Morgan McGuire, “Computer Graphics Archive” (2017). <https://casual-effects.com/data> (accessed July 06, 2019).
- [13] Lighting Challenges (2015). <http://www.3drender.com/challenges/> (accessed July 06, 2019).
- [14] V. Frolov, V. Galaktionov, M. Trofimov, “Sravnenie industrialnyh sistem raschyota osvesheniya”, *Privolzhskij nauchnyj zhurnal*, **14**, 79-85 (2014).
- [15] James T. Kajiya, “The rendering equation”, *Proceedings of the 13th annual conference on Computer graphics and interactive techniques (SIGGRAPH '86)*, David C. Evans and Russell J. Athay (Eds.). ACM, New York, NY, USA, 143-150 (1986).
- [16] Corona Renderer, (2019). <https://corona-renderer.com/> (accessed July 06, 2019).
- [17] Chaos Group. V-Ray Home Page, (2019). <https://www.chaosgroup.com/> (accessed July 06, 2019).
- [18] J. Krivanek, et al, “Radiance caching for efficient global illumination computation”, *IEEE Transactions on Visualization and Computer Graphics*, **11**(5), 550-561 (2005).
- [19] J. Křivánek, K. Bouatouch, S. Pattanaik, and J. Žára, “Making radiance and irradiance caching practical: adaptive caching and neighbor clamping”, *Rendering Techniques*, 127-138 (2006).
- [20] H. Jensen, *Realistic image synthesis using photon mapping*, AK Peters/CRC Press, (2001).
- [21] H. Jensen and T. Hachisuka, “Stochastic progressive photon mapping”, *ACM Transactions on Graphics*, **28**(5), Article 14, (2009).
- [22] I. Georgiev, J. Křivánek, T. Davidovič and P. Slusallek, “Light transport simulation with vertex connection and merging”, *ACM Transactions on Graphics*, **31**(6), Article 192-1 (2012).
- [23] J. Schwarzhaupt, H.Jensen and W. Jarosz, “Practical Hessian-based error control for irradiance

- caching”, *ACM Transactions on Graphics*, **31**(6), Article 193 (2012). doi:10.1145/2366145.2366212
- [24] OTOY Octane Renderer, (2019). <https://home.otoy.com/render/octane-render/> (accessed July 06, 2019).
- [25] Ray Tracing Systems. Hydra Renderer, (2019). <https://github.com/Ray-Tracing-Systems/HydraAPI>(accessed July 06, 2019)
- [26] David Cline, Justin Talbot, and Parris Egbert, “Energy redistribution path tracing”, *Proceedings of ACM SIGGRAPH-2005 conf. (SIGGRAPH '05)*, New York, USA, 1186-1195 (2005). doi: //doi.org/10.1145/1186822.1073330
- [27] Yu-Chi Lai, Shao Hua Fan, Stephen Chenney, and Charcle Dyer, “Photorealistic image rendering with population monte carlo energy redistribution”, *Proceedings of the 18th Eurographics conference on Rendering Techniques (EGSR'07)*, Jan Kautz and Sumanta Pattanaik (Eds.). Eurographics Association, Aire-la-Ville, Switzerland, 287-295 (2007). doi://dx.doi.org/10.2312/EGWR/EGSR07/287-295
- [28] E. Veach and L. Guibas, “Bidirectional estimators for light transport”, *Photorealistic Rendering Techniques*, Springer, 145-167 (1995).
- [29] T. Hachisuka , A. Kaplanyan and C. Dachsbacher, “Multiplexed metropolis light transport”, *ACM Transactions on Graphics*, **33**(4), Article 100 (2014).
- [30] C. Kelemen, et al, “A simple and robust mutation strategy for the metropolis light transport algorithm”, *Computer Graphics Forum*, **21**(3) (2002). doi: 10.1111/1467-8659.t01-1-00703
- [31] Cornell Box Data. Cornell University Program of Computer Graphics (2019). <https://www.graphics.cornell.edu/online/box/data.html> (accessed July 06, 2019).
- [32] E. Veach and L. Guibas, “Metropolis light transport”, *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley, 65-76 (1997).
- [33] Recent Advances in Light -Transport Simulation: Some Theory and a Lot of Practice, *Courses of 41th International conference on Computer Graphics and Interactive Techniques, Vancouver, Canada* (2014). <http://s2014.siggraph.org/attendees/courses/events/recent-advances-light-transport-simulation-some-theory-and-lot-practice.html>.
- [34] A.G. Voloboy, S.V. Ershov, S.G. Pozdnyakov “Interaktivnoe modelirovanie avtomobilnykh krasok”, *Proceedings of the 22nd International Conference on Computer Graphics and Vision (GraphiCon 2012)*, 242-247 (2012). <https://www.graphicon.ru/html/2012/conference/RU6%20-%20Graphics/gc2012Voloboy.pdf>
- [35] D. D. Zhdanov, I. S. Potemin, V. A. Galaktionov, B. Kh. Barladyan, K. A. Vostryakov, and L. Z. Shapiro ”Spectral Ray Tracing in Problems of Photorealistic Imagery Construction”, *Programming and Computer Software*, **37**(5), 236–244 (2011). doi:10.1134/S0361768811050069

Received May 15, 2019

A NOTE ON THE PERPETUAL AMERICAN STRADDLE

L. OBRADOVIĆ

Faculty of Science and Mathematics, University of Montenegro
E-mail: lazaro@ucg.ac.me

DOI: 10.20948/mathmontis-2019-45-10

Summary. In this paper we derive a single parameter characterization of the value and the optimal exercise time of the perpetual American straddle (the portfolio consisting of a put and a call option on the same underlying asset with the same price) in the classical Black-Scholes-Samuelson model. The parameter is the unique solution of a single non-linear equation with one unknown variable; this is the first time that the single equation characterization has been obtained for the perpetual American straddle. The equation is derived after multiple transformations of the defining optimal stopping problem in continuous time using a combination of classical techniques: Hamilton-Jacobi-Bellman equation, reduction to a Cauchy-Euler first order differential equation, smooth pasting conditions, and, finally, verification theorem for optimal stopping problems.

1 INTRODUCTION

Pricing of derivatives in the classical Black-Scholes-Merton model of a financial market is a classical topic in financial mathematics. Given that the stock price is modeled by a geometric Brownian motion, pricing problems can often be formulated as problems of optimal stopping in continuous time. Arguably the best known models involve pricing of perpetual American options: options without expiration date. Although the financial derivatives of this kind are not actively traded they represent an important theoretical concept and, due to their diminishing value, a valuable first approximation of the value of American derivatives with expiration dates. When considering these kind of problems of interest are the value function of the problem, which gives information about the price of the derivative, as well as the optimal stopping time, which gives information about the optimal exercise time of the derivative under consideration.

In this paper we consider the perpetual *American straddle*: a classical portfolio consisting of a put option and a call option on the same underlying asset with the same strike price. The pricing of the perpetual American straddle has been studied using different approaches and tools: in [1] by applying the theory of Laplace transforms, in [5] by transforming the problem to a "generalized parking problem", in [6] by exploiting "an analogy with asymmetric rebates of double knock-out barrier options", in [7] "by means of the Esscher transform and the optional sampling theorem", and, more recently, by using a combination of several optimization techniques [3] and [4]. The characterizations obtained in these papers are often cumbersome: indeed, in all of these papers the value function and the optimal exercise time are characterized by a solution of a non-linear system of equations consisting of (at least) two equations.

2010 Mathematics Subject Classification: 91G20, 62P05, 97M30

Keywords and Phrases: Derivative pricing; Perpetual American options; American straddle; Optimal stopping; Black-Scholes model; Mathematical finance

In this note, we show that the value function and the optimal exercise time of the perpetual American straddle can be characterized via a unique solution of a *single* one-variable equation; the solution lies in the interval $(0, 1)$. We do so by using one of the classical optimal stopping theory approaches: the Hamilton-Jacobi-Bellman (HJB) equation and the smooth-fit principle in combination with a verification theorem. In particular, we begin with an optimal stopping problem in continuous time and assume that its optimal stopping time is the first exit time of a bounded interval. The HJB equation for the value function of the problem on the continuation region is a partial differential equation that can be reduced to an ordinary differential equation (the Cauchy-Euler equation), as is customary for the problems of this nature. After solving the equation and exploiting the assumption about the continuation region we are able to explicitly write down the form of the value function: it is a piecewise function with several unknown parameters. As the value function is expected to be continuous and differentiable we are able to apply what is known as smooth pasting conditions to obtain a nonlinear system of equations the solution of which will give us the unknown parameters. Finally, after applying the verification theorem and some theoretical considerations about the uniqueness of the solution of the system, we reduce the system to a single equation and prove that its solution is unique and in the $(0, 1)$ interval. To the best of our knowledge this is the first time that such one-equation characterization of the value and the optimal exercise time of the perpetual American straddle is obtained. In the next section we present our result in full detail.

2 RESULT

Let the price process S_t be a geometric Brownian motion,

$$dS_t = \alpha S_t dt + \sigma S_t dB_t,$$

where $\alpha \in \mathbb{R}$ and $\sigma \in \mathbb{R}$ are known constants. The American straddle yields a payoff

$$f(t, S_t) = e^{-rt} |S_t - I|$$

when exercised at time t , where $I > 0$ is the strike price and $r \leq \alpha$ is a given discount rate (the inequality $r \leq \alpha$ is a standard assumption; see for example [10]).

The value of the perpetual American straddle at time t is given by

$$V_t = \operatorname{ess\,sup}_{\tau \in \mathcal{T}_t} E[e^{-r\tau} |S_\tau - I|], \tag{1}$$

where \mathcal{T}_t is a set of all stopping times $\tau \geq t$. Our goal is to find a value function $v(t, x)$ such that $v(t, S_t) = V_t$ and an optimal stopping time τ^* such that $V_{\tau^*} = E[e^{-r\tau^*} |S_{\tau^*} - I|]$. Hamilton-Jacobi-Bellman (HJB) equation related to this problem is:

$$\max_{(t,x) \in [0, +\infty] \times \mathbb{R}} \{f(t, x) - v(t, x), v_t(t, x) + \mathcal{L}v(t, x)\} = 0, \tag{2}$$

where $\mathcal{L} := \alpha \partial_x + \frac{1}{2} \sigma^2 \partial_{xx}$ is a differential operator related to Ito's lemma (see e.g. [8, ch.11]).

A well known approach when dealing with time-discounted optimal stopping problems is to assume that the value function is of the form

$$v(t, x) = e^{-rt} \varphi(x);$$

this will later be confirmed using a verification theorem. The equality

$$v_t(t, x) + \mathcal{L}v(t, x) = 0$$

holds on the continuation region (due to the HJB equation). After canceling out e^{-rt} this gives:

$$r\varphi(x) - \alpha x \varphi'(x) - \frac{1}{2} \sigma^2 x^2 \varphi''(x) = 0.$$

The last equation is a well known Cauchy-Euler ordinary differential equation and its solution is

$$\varphi(x) = Ax^\lambda + Bx^\mu,$$

where A and B are two unknown constants and λ and μ are the solutions of the characteristic equation

$$r - \alpha m - \frac{1}{2} \sigma^2 m(m-1) = 0.$$

It can be easily verified that inequalities $\lambda > 1$ and $\mu < 0$ hold.

It is known that the optimal stopping time will be the first exit time from the interval $(x_1, x_2) \ni I$: it is optimal to exercise the put (call) option when the value of S_t goes beneath x_1 (above x_2). Furthermore, on the stopping region, the HJB equation implies that $f = v$. Thus, we assume that the function v should be of the form:

$$v(t, x) = \begin{cases} e^{-rt}(I - x), & 0 < x < x_1 \\ e^{-rt}(Ax^\lambda + Bx^\mu), & x_1 \leq x \leq x_2 \\ e^{-rt}(x - I), & x > x_2 \end{cases} \quad (3)$$

where A, B, x_1, x_2 are constants chosen in a way that makes the function v differentiable (smooth pasting conditions). In particular, we require continuity and differentiability in x_1 and x_2 .

It is already clear that, should we find such constants, the above function $v(t, x)$ will be a value function. Indeed, conditions of any of the well known verification theorems for the optimal stopping of diffusions (e.g. ch. 3 in [9] or ch. 10 in [8]) are easily satisfied for functions that coincide, piecewise, with (discounted) linear combinations of power functions. Furthermore, since the functions v and f coincide outside the interval (x_1, x_2) , if v is indeed the value function, then the optimal stopping time is:

$$\tau^* = \inf\{t \geq 0 | S_t \notin (x_1, x_2)\}.$$

Smooth pasting conditions lead to a highly non-linear system of equations. We show that it can be reduced to a single equation:

Theorem 1. *The value process of the perpetual American Straddle V_t defined in (1) satisfies the equality $V_t = v(t, S_t)$ for the function v as defined in (3) where*

$$\begin{aligned} A &= \frac{1}{\mu - \lambda} ((1 - \mu)x_1^{1-\lambda} + \mu x_1^{-\lambda}); & x_2 &= \frac{\mu I}{\mu - 1} \frac{1 + \gamma^{-\lambda}}{1 + \gamma^{1-\lambda}} \\ B &= \frac{1}{\lambda - \mu} ((1 - \lambda)x_1^{1-\mu} + \lambda x_1^{-\mu}); & x_1 &= \gamma x_2 \end{aligned}$$

and $\gamma \in (0, 1)$ is the unique number satisfying:

$$\frac{\mu}{\mu - 1} \frac{1 + \gamma^{-\lambda}}{1 + \gamma^{1-\lambda}} - \frac{\lambda}{\lambda - 1} \frac{1 + \gamma^{-\mu}}{1 + \gamma^{1-\mu}} = 0. \quad (4)$$

Proof. Smooth pasting conditions, after cancelling out e^{-rt} , can be written as:

$$\begin{aligned} I - x_1 &= Ax_1^\lambda + Bx_1^\mu, & -x_1 &= A\lambda x_1^\lambda + B\mu x_1^\mu, \\ x_2 - I &= Ax_2^\lambda + Bx_2^\mu, & x_2 &= A\lambda x_2^\lambda + B\mu x_2^\mu. \end{aligned} \quad (5)$$

In order to prove the theorem it is, by construction of the value function v , sufficient to prove that unique solution of the system (5) is the one given in the formulation of the theorem. The proof consists of reducing the system to equation (4), and proving that the solution of the latter is unique on the interval $(0, 1)$.

First we comment on the uniqueness of the solution of the system of equations (5). Due to the uniqueness of the value function of the optimal stopping problems the solution of the system above must be unique. Indeed, two different solutions of the system (5) would lead to two functions v_1 and v_2 both of which would satisfy the verification theorem and the equation $v_1(t, S_t) = v_2(t, S_t)$ would holds almost surely, which is clearly impossible.

We now turn to proving the existence. We can eliminate variables A and B in the two equations containing x_1 by treating them as a two dimensional linear system. Since determinant of that system is $D = x_1^{\lambda+\mu}(\mu - \lambda) \neq 0$, A and B are uniquely determined by it. We can do the same for the two equations containing x_2 . If we introduce, for notational purposes, the function $Q(x; \mu, \lambda) = (\mu - \lambda)^{-1}(1 - \mu)x^{1-\lambda} + \mu x^{-\lambda}$, we can write the solutions of those two systems as:

$$A = Q(x_1; \mu, \lambda); B = Q(x_1; \lambda, \mu); A = -Q(x_2; \mu, \lambda); B = -Q(x_2; \lambda, \mu).$$

Equating the expressions for A and B we obtain the following nonlinear system with two equations and two variables, x_1 and x_2 :

$$Q(x_1; \mu, \lambda) + Q(x_2; \mu, \lambda) = 0 \quad Q(x_1; \lambda, \mu) + Q(x_2; \lambda, \mu) = 0 \quad (6)$$

Due to the nice form of the above system, we immediately see that if (x_1, x_2) is its solution so is (x_2, x_1) . This means that there is a unique solution pair satisfying $x_1 < x_2$, and it will be the

unique solution that we are looking for. We introduce a variable γ such that $x_1 = x_2\gamma$; since inequality $0 < x_1 < x_2$ holds, we have $\gamma \in (0, 1)$. The right hand side of the first equation of the system (6) can now, after some simple calculations, be written as:

$$Q(x_2\gamma; \mu, \lambda) + Q(x_2; \mu, \lambda) = (1 - \mu)x_2^{1-\lambda}(1 + \gamma^{1-\lambda}) + \mu x_2^{-\lambda}(1 + \gamma^{-\lambda}).$$

from which we obtain:

$$x_2 = \frac{\mu I}{\mu - 1} \frac{1 + \gamma^{-\lambda}}{1 + \gamma^{1-\lambda}}$$

Similarly, by changing $x_1 = x_2\gamma$ in $Q(x_1; \lambda, \mu) + Q(x_2; \lambda, \mu) = 0$ after multiplication with $x_1^{-\mu}$ we obtain:

$$x_2 = \frac{\lambda I}{\lambda - 1} \frac{1 + \gamma^{-\mu}}{1 + \gamma^{1-\mu}}.$$

Equating the two obtained expressions for x_2 , after rearranging and cancelling out parameter I , we obtain the one-dimensional equation (4), stated in the formulation of the theorem.

It remains to prove that there exists a unique solution of equation (4) in the interval $(0, 1)$. Let us denote the left hand side of the equation with $h(\gamma)$. It is obvious that function h is continuous on $(0, 1)$ and, since $\lambda > 1$ and $\mu < 0$, it is easy to check that $h(1) < 0$ and $\lim_{\gamma \rightarrow 0^+} h(\gamma) = +\infty$. We can thus conclude that a solution exists on the interval $(0, 1)$, and its uniqueness is a consequence of the argument from the beginning of the proof. □

3 CONCLUSION

We have demonstrated that the perpetual American straddle, a classical and well studied portfolio of options, can be priced and fully characterized using a unique solution of a single non-linear equation on the unit interval. Our contribution is technical, and it's value lies in it's elegance as well as the fact that the solution itself gives a direct relation between two exercise boundaries of the American straddle. The result represents an represents a rare advancement in a well studied field, showing that even in classical literature on derivative pricing there can be space for contributions to the theory; the contributions are likely to be of the technical kind, as demonstrated in this material.

As the result we presented belongs to a wide field of financial mathematics, we conclude the paper with two comments that hopefully address the relative significance of the results within finance and mathematics, respectively.

Curiously, the equation the solution of which is the single parameter that characterizes the optimal exercise time and the price of the perpetual American straddle does not depend on the strike price I . This technical curiosity might have a deeper economic interpretation within the field of finance; this however lies beyond the mathematical aspects studied in this work.

From the perspective of the mathematics of the theory of optimal stopping in continuous time, the result we presented gives rise to a natural question for future research: which optimal stopping problems with bounded continuation regions can be characterized by a single parameter? The solution we presented exploited the "symmetry" of the function f and one can naturally assume that one needs to formalize this condition in the most general terms in order to prove the general version of the result presented here.

REFERENCES

- [1] G. Alobaidi and R. Mallier, "Laplace transforms and the American straddle", *Journal of Applied Mathematics*, **2** (3), 121–129 (2002)
- [2] F. Black and M. Scholes, "The pricing of options and corporate liabilities", *Journal of political economy*, **81** (3), 637–654 (1973)
- [3] J. Lempa, "A note on optimal stopping of diffusions with a two-sided optimal rule", *Operations Research Letters*, **38** (1), 11–16 (2010)
- [4] D. Lamberton and M. Zervos, "On the optimal stopping of a one-dimensional diffusion", *Electronic Journal of Probability*, **18** (2013)
- [5] M. Beibel and H. R. Lerche, "A new look at optimal stopping problems related to mathematical finance", *Statistica Sinica*, **7** (1), 93–108 (1997)
- [6] F. Moraux, "On perpetual American strangles", *The Journal of Derivatives*, **16** (4), 82–97 (2009)
- [7] H. Gerber and E. Shiu, "Martingale approach to pricing perpetual American options", *ASTIN Bulletin: The Journal of the IAA*, **24** (2), 195–220 (1994)
- [8] B. Øksendal, *Stochastic differential equations*, Springer (2003)
- [9] N. V. Krylov, *Controlled diffusion processes*, Springer Science & Business Media (2008)
- [10] A. N. Shiryaev, *Essentials of stochastic finance: facts, models, theory*, World scientific (1999)

Received July 25, 2019

ETHICAL NORMS AND LEGAL FRAMEWORK OF SCIENTIFIC PUBLICATION

T.A. POLILOVA*

Keldysh Institute of Applied Mathematics of RAS
*Corresponding author. E-mail: polilova@keldysh.ru

DOI: 10.20948/mathmontis-2019-45-11

Summary. Scientific research is conducted, as a rule, for a long period of time. At certain points, researchers prepare publications that reflect the results obtained at the current stage. For example, suppose a multi-year research project has several reporting periods, at each of which researchers present to the customer publications created in the reporting period. Then, since the publications are related to a common theme, their texts inevitably have intersections associated with the description of the conditions of the problem, the context of research, the methods used, etc. How will publishers react to such publications? Codes of ethics for scientific publications not encourage duplication of text fragments in several articles. Do ethical codes restrict the author's freedom to create a derivative work that uses fragments of text from articles he has previously written? Can the author free to publish his article in multiple publications?

According to the Civil code of the Russian Federation, the creation of derivative works is one of the ways to realize the exclusive right of the author to the work. The Civil code does not regulate specific permissible methods and volumes of processing of a scientific work. The feasibility of creating a derivative work is entirely within the competence and responsibility of the author.

The Civil code of the Russian Federation establishes the mechanism of simple (non-exclusive) licenses: any publisher can obtain from the copyright holder a simple license to publish the article, including without its processing. Publication of an article in several journals is one of the ways enshrined in the law to realize the right of the author (copyright holder) to publicize the work. The discussion in the scientific community on whether it is ethical to publish an article in several journals often makes no sense. The ethical sphere can only include the obligation of the author to inform the publisher about the held or planned publication of the article in another journal.

1 INTRODUCTION

Scientific research is usually conducted over a long period of time. At certain points, researchers prepare publications that reflect the results obtained at the current stage. For example, suppose a multi-year research project has several reporting periods, at each of which researchers present to the customer publications created in the reporting period. Because the publications are related to a common theme, their texts inevitably have got intersections associated with the description of the conditions of the problem, the context of the research, the methods used, etc. How will publishers react to such publications? Codes of ethics for

2010 Mathematics Subject Classification: 00-02, 00A99.

Key words and Phrases: Scientific publication, Text recycling, Redundant publication, Exclusive rights, Derivative work.

scientific publications now tend to assess cases of repeated use of the text in several articles of the author as a significant violation.

Do ethical codes restrict the author's freedom to create a derivative work that uses fragments of text from articles he has previously written? Can the author free to publish his article in multiple publications? How does the Civil code of the Russian Federation (hereinafter — the Civil code) [1] protect the intellectual rights of authors?

2 CODES OF ETHICS OF SCIENTIFIC PUBLICATIONS

Consider the well-known code of ethics developed by the Committee on Publication Ethics (COPE) [2]. Ethical principles of COPE were formulated in 1997 to address ethical concerns in research and publication. Major companies and associations operating in the field of pharmacology and medicine, publishers of medical journals participated in the development of the COPE code of ethics. The code addresses the issues of research ethics, reliability of the data obtained, correctness of the data analysis methods used, completeness and openness of the results publication. The code also includes topics of conflict of interest, falsification of research, biased evaluation of results, etc. In medicine and pharmacology these problems are extremely important.

Other topics of the code cover issues common to many areas of knowledge: authorship, peer review, illegal actions of participants in the publishing process, plagiarism, re-use of the text in several articles of the author (self-plagiarism, text recycling), excessive publication (redundant publication).

Reuse of the text or "self-plagiarism" is a very urgent problem, which is currently actively discussed among the editors of scientific publications. The facts of reuse of texts are now easily established: overlapping texts in several articles are identified with the help of programs such as "Anti-plagiarism".

Editors can make certain decisions regarding the publication of the article depending on the extent to which the text of the article intersects with the author's previous publications, in which sections of the article fragments of the text from earlier publications are found. In accordance with the recommendations of COPE, editors should consider cases of text re-use on an individual basis, since the adequacy of the editor's actions will depend on many factors.

In some cases, the revision of the text is not a serious violation, and the authors do not have to correct such parts of their article. If the overlap of the text is unjustified, but insignificant, the COPE instruction recommend to the editors to suggest the authors to rewrite the overlapping sections or to make the text in the form of a quotation from the previous articles. A larger overlap may result in rejection of the manuscript.

Note how the COPE code defines the term "redundant publication". The term refers to the case when several articles without reciprocal link is devoted to the description of the results of one study put the same tasks, formulates the same conclusion and so on. Ethical standards declared by the code are the following:

- 1) published research results should not be repeated, unless it is associated with a special case of additional confirmation;
- 2) publication of abstracts at the conference (other event) does not exclude further publication of the full version of the article;
- 3) publication of the article in another language is acceptable provided that the full meaning and adequate translation;

4) at the time of submission of the article to the journal, authors should indicate all related articles, including translations into other languages.

The first point in the above list requires editors to reject the previously published article, but the rigidity of such a decision is understandable in the context of the characteristics of the subject area (pharmacology, medical science). If the authors publish the same results in several journals, the medical professionals may be deformed idea of the scale, complexity and importance of the study. Such distortions can eventually lead to negative consequences in practical medicine.

The following paragraphs formulate perfectly acceptable, natural ethical rules to be followed by an author working in any scientific field.

The COPE code is not a directive document binding on all editors, reviewers and authors. The code is a kind of guide that offers the participants of the publishing process ways to solve certain ethical problems. At the same time, the COPE website provides a forum where you can express your opinion on a specific case of application of the rules of the COPE code. The developers of the code express their readiness to change some of its provisions based on the results of an open discussion.

Participants in the discussion on the COPE website [3] often express opposing views on some ethical rules, in particular on the reuse of text (text recycling). Someone annoyed by the constant repetition of the same text in several articles by the author ([3], Charlotte Seidman post, 22/2/2013 7.42 pm). Someone refers to other examples where reuse of the text is necessary. Thus, there is often a need for introductory paragraphs that describe a larger project, thus creating the context of the work in the article on one of the directions of this large project. Forcing authors to revise such paragraphs does not improve the content of the article. Generally, paraphrasing degrades the quality of the text or leads to lengthening of the article ([3], Justin Starren post, 28/2/2013 2.03 pm).

In the discussion on the COPE website there is an opinion that there is no ethical flaw in the reuse of the text ([3], Soumitra Kumar post, 12/3/2013 8.31 am). The author of this post develops an extravagant idea that the article could be constructed using existing text modules. As the author notes, such a constructive approach to writing an English-language article could benefit, in particular, researchers for whom English is not a native language.

The expressed design idea corresponds to the technology of reuse of program code, when previously written program blocks are included in the program. Code reuse is widely used in programming. Including a previously written and debugged block in his new program, the programmer knows that such a modernization of the project will expand the functionality of the system and most likely will not lead to errors.

This is not the case with reuse of the text. There are a lot of contradictions. The inclusion to the new article the fragment of the text from the previously written paper entails a violation of ethical norms. At the same time, a link to a previously published article, placed instead of the desired text fragment, does not allow the reader to directly get acquainted with the context or conditions of the project. And if you follow the editor's recommendation to "rephrase the text fragment", such reformations can cause a lot of questions from the reader. Seeing the new lexical set of words and expressions in a familiar text, the reader will be puzzled: has the author's position changed, has there been any new data on the previously discussed issue?

The COPE code of ethics for scientific publications is very popular. This code is widely implemented in Russia [4-6]. The ethical rules of many Russian journals are based on the COPE recommendations. Sometimes the provisions of the COPE code are reworked, often in

the direction of tightening. For example, in the code of ethics of scientific publications on the portal psychological publications PsyJournals.ru [7] ethical standards relating to the re-publication, are collapsed into a single rule:

"Authors should not submit to the journal a manuscript that has been sent to another journal and is under consideration, as well as an article already published in another journal."

Less categorically expressed the final document of the conference "Problems of quality of scientific work and academic plagiarism", held on September 26, 2018 [8]:

"Multiple publication of an article (replication, self-plagiarism) means reprint by the author (s) of his / her own works, which is not justified by any objective reasons (changes in the text of the article, translation into another language, appeal to another readership, inclusion of the text of the article in a thematic selection or anthology) and without specifying the source of the initial publication."

Not a bad formulation, which, in particular, covers the case of the publication of an article in the journal after the preprint publishing: there is definitely an appeal to a different, generally speaking, wider readership.

In our opinion, the current restrictions on the publishing and spreading of the article may infringe the rights of the author (copyright holder) to promulgate his work, i.e. to bring it to the general public. Often, the distorted interpretation of ethical norms hides the routine market competition of journal publishers. In this way the publisher wants to gain a competitive advantage from the exclusive publication. After publishing the article in the journal, many publishers offer authors to re-publish the article in thematic collection (see, for example, the proposal of the publishing house "Science" [9]), thus obtaining additional income from spreading and sale of previously published articles.

At the same time, ethical considerations about the inadmissibility of multiple publication of the same work cannot be completely discarded. But ethical norms only come into effect when multiple publication takes ugly forms. It is hardly a normal situation when the same article appears in dozens of magazines. Here, indeed, there is a violation of ethics, as the reader is forced over and over again to stumble upon identical texts. In this case, it is usually easy to see the attempt of an unscrupulous author to harshly increase his publication figures.

3 THE CIVIL CODE OF THE RUSSIAN FEDERATION ABOUT PROTECTION OF INTELLECTUAL PROPERTY RIGHTS

Part IV of the Civil code is devoted to the protection of intellectual rights. There, in section VII "Rights to results of intellectual activity and means of individualization", article 1225 defines the types of results of intellectual activity to which state protection is granted. The result of intellectual activity, in particular, is a work of science. One of the types of works of science is a scientific paper written by the author on the results of scientific research. The paper is created for the purpose of its further publishing. Relations between the author and the publisher are regulated by the provisions of the Civil code.

According to the Civil code, the author (or employer) has the exclusive right to the created work. What exactly does the exclusive right to the work include? According to paragraph 1 of article 1270 of the Civil Code, the copyright holder may use the work in any form and in any manner not contrary to the law.

Paragraph 2 of article 1270 lists the following ways of using the work:

- reproduction of the work, i.e. making one or more copies of the work or part of it in any material form;
- spreading (distributing) the work;
- translation or other processing the work;

et al.

Let us return to the ethical question of reusing the text of the paper discussed above. In terms of the Civil code we are talking about the creation of derivative works as a result of processing the original work.

Processing is one of the ways to realize the exclusive right of the author (copyright holder) to the work. The Civil code of the Russian Federation does not regulate specific permissible ways and volumes of processing of a scientific work. The expediency of creating a derivative work is entirely within the competence and responsibility of the author (copyright holder) of the work.

The author has no restrictions on the creation of derivative works based on the original work. No license agreements or even an agreement on the alienation of the exclusive right to the work deprive the author of the right to create derivative works. Paragraph 4 of article 1233 of the Civil code reads as follows:

"The conditions of the agreement on alienation of the exclusive right or the license agreement limiting the right of the citizen to create results of intellectual activity of a certain kind or in a certain area of intellectual activity or to alienate the exclusive right to such results to other persons are void".

Thus, with regard to future works, the Civil code protects the author from contracts that infringe the interests of the author. In the Russian Federation, the publisher in the contract may not require the author not to create derivative works in the future.

Derivative works should occupy a legal place in the space of scientific publications.

The author has the right to create derivative works, focusing on different audiences of readers. For example, a publication describing the results of work on the creation of a computational model, addressed to specialists in the field of mathematics, can reveal the details of the formal mathematical apparatus and applied computational methods. In the derivative publication on the same subject, but focused on IT-specialists, the emphasis will be shifted towards the description of the architecture and functionality of the software complex, serving the constructed mathematical model.

Derivative works may reflect the author's new views on the subject of research, clarify the tasks, adjust or supplement the previously obtained results, etc.

The author can create his work in the technology of live publications. The author of the live publication takes the responsibility not only to constantly improve his work, but also to monitor development in the research area and systematically reflect everything new in his online text [10]. At some point, the author can publish an article in the journal — a fixed time slice of a live publication — and continue to develop his work. Having accumulated new, quite interesting facts and results, the author can re-publish another time slice of his continuously developing work. Thus, the original work may give rise to a family of derivative works intended for publication in journals.

The derivative work can be created by the author specifically for the functioning in the Internet. Such work is implemented not in the form of a text file, but on a special

technological platform. In comparison with the usual text, an online scientific article can have new qualities: flexible text visualization in the technique of adaptive design, multimedia illustrations, online computing, etc.

Finally, the author can write a monograph, which will include the texts of several articles of the author. The monograph is a traditional type of scientific production. Despite the fact that in today's conditions writing a monograph is not encouraged by officials from science, leading scientists continue this vital scientific tradition [11].

Does the author have the right to publish an article in several journals? Publication of the work in several editions is a legal way to implement the exclusive right specified in paragraph 11 of article 1270 of the Civil code:

"(11) making the work available to the public in such a way that any person may access the work from any place and at any time of his own choice (making it available to the public)."

From the point of view of the Civil code, it is also permissible to publish the same article in different journals without processing on the terms of a simple (non-exclusive) license — in this way the right holder has the opportunity to expand its readership, to acquaint the general scientific community with the results of the research.

Does the publisher have the right to reject an article on the grounds that it has already been published in another journal? Certainly, yes. Moreover, the publisher has the right to reject the article and for any other absurd reason, as well as without explanation. The only thing that the publisher does not seem to need to mention here is that he is guided by high ethical considerations in this questionable decision.

4 NEW RELATIONS BETWEEN AUTHOR AND PUBLISHER

Technologically multiple publication, in our opinion, is a phenomenon of yesterday, accompanying the transition stage in the development of means of representation of scientific knowledge. Already now there is a new model of relationship between the publisher and the author, which does not lead to multiple publications.

The paper [12] presents an open platform for the organization of review by the publisher of the journal of open access preprints of articles previously placed in the popular open archive arXiv.org. The author submits his / her article to the journal directly from arXiv.org. The journal reviews the article by its experts. After successful peer review, author corrects his article, and the article is published again in arXiv.org. The journal assigns a DOI to the article and publishes a link to such article.

This kind of relationship can, generally speaking, arise in the author with several magazines. If all these journals, after successful evaluations of the article, publish only links on the article, that is placed in arXiv.org, then duplication of the text of the article is not happening. At the same time, the editors of any of the journals may, of course, decide to publish the full text of the article.

The publication of an article in one form or another in several journals should not have a negative connotation. On the contrary, the author can be deservedly proud of the fact that his article received a kind of quality mark from several magazines.

About these new trends in "Green OA" told, in particular, the representative of the company Clarivate Analytics in Russia Oleg Utkin on held in may 2019 public hearing in the

Presidium of the Russian Academy of Sciences, devoted to the development in Russia open access to scientific publications [13]. The essence of the "Green OA" model is that the author, after internal review, places his article (preprint) in an open archive deployed on the website of his organization. Only then the article is sent for review to specialized scientific journals. After a positive review link to the article is placed on the website of the journal.

5 CONCLUSION

The legality of creating a new work on the basis of previously published works of the author is actively discussed among the editors of scientific publications. As already mentioned, the facts of re-use of texts are now easily established with the help of programs such as "Anti-plagiarism". Editors can make certain decisions regarding the publication of the article depending on the extent to which the text of the article intersects with the author's previous publications, in which sections of the article reused fragments are found, etc. In accordance with the norms of publication ethics, editors should consider cases of text processing on an individual basis.

The publication by the author of derivative works containing fragments of the text from his previous articles should not refer to violations of publishing ethics. The Civil code of the Russian Federation assigns to the author the right to create derivative works. Derivative works can be oriented to different readership, reflect new views of the author on the subject of research, adjust or supplement the results obtained earlier, etc. Derivative works should occupy a legitimate place in the space of scientific publications.

The discussion in the scientific community on whether it is ethical to publish an article in several journals makes no sense. The ethical sphere can only include the obligation of the author to inform the publisher about the held or planned publication of the article in another edition. The Civil code of the Russian Federation establishes the mechanism of simple (non-exclusive) licenses: any publisher can obtain from the copyright holder a simple license to publish the article, including without its processing. The publication of articles in several journals — this is one of the enshrined in the law of natural methods of implementation of the rights of the author (copyright holder) on a wide publication of a work.

REFERENCES

- [1] "Grazhdanskii kodeks Rossiiskoi Federatsii", *ConsultantPlus*, website: URL: http://www.consultant.ru/document/cons_doc_LAW_5142/ (accessed August 08, 2019).
- [2] "Guidelines on good publication practice": Committee on Publication Ethics (COPE), *Library of Financial University under the Government of Russian Federation*, website: URL: <http://www.library.fa.ru/files/GuidelinesGPP.pdf> (accessed August 08, 2019).
- [3] "How to deal with text recycling": Text Recycling Guidelines, *COPE*, website: URL: <https://publicationethics.org/text-recycling-guidelines> (accessed August 08, 2019).
- [4] "Podgotovka i izdanie nauchnogo zhurnala. Mezhdunarodnaia praktika po etike redaktirovaniia, retsenzirovaniia, izdaniia i avtorstva nauchnykh publikatsii: Sbornik perevodov", O.V. Kirillova (Ed), *Akademiya ANRI*, website: URL: https://academy.rasep.ru/files/documents/____3_1.pdf (accessed August 08, 2019).
- [5] "Eticheskie printsipy nauchnykh publikatsii": Declaratsiia, *Assotsiatsiya nauchnykh redaktorov i izdatelei*, website: URL: <https://rasep.ru/sovet-po-etike/deklaratsiya> (accessed August 08, 2019).

- August 08, 2019)
- [6] “Spisok i obieiasnenie priznakov nekorrektnoi redaktsionnoi politiki”, *Dissernet*, website: URL: <http://biblio.dissernet.org/prizn> (accessed August 08, 2019).
 - [7] “Kodeks etiki nauchnykh publikatsii”, *Portal psikhologicheskikh izdaniy PsyJournals.ru*, website: URL: http://psyjournals.ru/info/code_of_ethics.shtml (accessed August 08, 2019).
 - [8] “Itogovi dokument konferentsii ‘Problemy kachestva nauchnoi raboty i akademicheskii plagiat’“, *Novosti Sibirskoi Nauki*, website: URL: <http://www.sib-science.info/ru/conferences/itogovyy-dokument-09102018> (accessed August 08, 2019).
 - [9] “Novye vozmozhnosti: izdanie sbornikov”, *Izdatelstvo Nauka*, website: URL: <https://naukapublishers.ru/history/novosti/novosti-izdatelstva/229-novye-vozmozhnosti-izdanie-sbornikov> (accessed August 08, 2019).
 - [10] M.M. Gorbunov-Posadov, “Zhivaia publikatsiia”, *Keldysh Institute of Applied Mathematics of RAS*, website: URL: <https://keldysh.ru/gorbunov/live.htm> (accessed August 08, 2019).
 - [11] M.P. Galanin, M.M. Gorbunov-Posadov, V.P. Kandidov, T.A. Polilova, “Publication of the scientific monograph in Russia”, *Mathematica Montisnigri*, **39**, 146-160 (2017), website: URL: <http://www.montis.pmf.ac.me/vol39/13.pdf> (accessed August 08, 2019).
 - [12] E. Gibney, “Open journals that piggyback on arXiv gather momentum”, *Nature*, website: URL: <https://www.nature.com/news/open-journals-that-piggyback-on-arxiv-gather-momentum-1.19102> (accessed August 08, 2019).
 - [13] S. Beliaeva, “Tsena otkrytosti: Vo chto oboidetsia perekhod k Open Access?”, *Poisk*, website: URL: <https://www.poisknews.ru/skript/czena-otkrytosti-vo-chto-obojdetsya-perehod-k-open-access/> (accessed August 08, 2019).

Received June 17, 2019

RUSSIAN EXPLORATION OF VENUS: PAST AND PROSPECTS

G.K. BOROVIN*, A.V. GRUSHEVSKII, A.G. TUCHIN, D.A. TUCHIN

Keldysh Institute of applied mathematics RAS Moscow, Russia

*G.K.Borovin. E-mail: borovin@keldysh.ru

DOI: 10.20948/mathmontis-2019-45-12

Summary. Secure methods of the cost-effective flights to Venus and the passage of its gravity sphere are necessary to increase the amount of scientific and service equipment on board spacecraft and effectively exploration of the solar system. The results of the modern methods of ballistics design using for interplanetary flights to Venus are given with the purpose of entering the Venusian satellite orbit and for the landing on its surface. The main attention devotes to the construction of the launch windows and to the calculation of the reachability areas of the descent vehicle on the surface of Venus.

1 INTRODUCTION

Domestic space missions to Venus traditionally dominated in the world research of Venus, and the bulk of the fundamental knowledge about the planet by the Soviet missions in 1961-1985 was obtained. In those years, Venus was launched 18 automatic stations and made 10 landings – all successful [1]. NASA also carried out two successful orbital projects, the “Pioneer-Venus” and later - the “Magellan”.

After a long break, Venus was studied only from orbit, for example, by the European Venus Express project (2005-2015).

The Venus-D project is the next step after the successful series of Venus and VEGA missions in the 1970s and 1980s. The qualitative difference between modern landing craft from the programs "VEGA" is equipped with the knowledge of the geology of the surface according to the results of radar studies of the spacecrafts "Venera 15,16" and the "Magellan" spacecraft. The lander will land not blindly, as before, but in an area with a known geological context.

The development of cost-effective ballistic schemes of spacecraft’s flights to Venus and the passage of its scope is relevant and necessary to increase the payload of the spacecraft by reducing the mass of fuel [2, 3]. Ballistic design of such schemes, in particular, is an essential part of the promising domestic project "Venus-D", providing for the landing of the descent module on its surface in a given area.

The paper presents the results of the use of modern methods of ballistic design of interplanetary flights for planning flights to Venus (the mission of delivery and, in particular, to land the spacecraft on its surface in a given area), analyzes the methods of implementation of Hesperian projects and improve their efficiency. The main attention is paid to the construction of the launch windows and the method of construction and direct calculation of the reachability areas of the descent vehicle on the surface of Venus.

2010 Mathematics Subject Classification: 70E55, 70M20, 70Q05.

Key words and Phrases: mission design, Venus, descent module, target region of accessibility, gravity assist, “Venera-D” mission

2 THE SHORT HISTORY OF THE HESPERIAN MISSIONS

The Soviet automatic interplanetary probe Venera-3 was the first to reach the Venus surface on March 1, 1966. The USA accomplished the successful mission Pioneer Venus in 1978. This mission included the launch of two spacecraft—the first one was an orbital module and the second one included one large and three small descent modules. They acquired a lot of data concerning the atmosphere, but the landing was not planned, and the descent modules stopped making measurements at the altitude of 12 km. Thus, only Soviet probes have successfully worked on the Venus surface. A number of Soviet space projects aimed at approaching Venus and landing on its surface have been accomplished (Tab. 1 and Fig. 1).

By contrast with the preceding missions, the modern projects of Venus exploration using landing modules plan to land at a given landing point on the Venus surface, which is impossible without using a high-quality Venus maps.

In the middle of October of 1983, Venera-15 and Venera-16 were inserted into orbits around Venus. These satellites were equipped with a radar system for mapping the Venus surface. Later, an atlas of Venus [4] was created based on these data. The radar mapping started by Veneras was continued by the NASA Magellan mission, which did not plan to land on the Venus surface. To help plan the survey, the NASA researchers used the data acquired by Venera-15 and Venera-16. In 1990–1993, Magellan mapped the Venus surface at a higher

Automatic interplanetary probe	Date		Landing coordinates (degrees)		Remark
	launch	arrival	latitude	longitude	
Venera-1	1961/02/12	1961/05/19	—	—	Flyby at a distance of ~100 thousands of km
Venera-2	1965/11/12	1966/02/27	—	—	Flyby at a distance of ~100 thousands of km
Venera-3	1965/11/16	1966/03/01	—	—	Reached the surface
Venera-4	1967/06/12	1967/10/18	19.0	38.0	
Venera-5	1969/01/05	1969/05/16	-3.0	18.0	
Venera-6	1969/01/10	1969/05/17	-5.0	23.0	
Venera-7	1970/08/17	1970/12/15	-5.0	351.0	
Venera-8	1972/03/27	1972/07/22	-10.0	335.0	
Venera-9	1975/06/08	1975/10/22	31.7	290.8	Venus satellite
Venera-10	1975/06/14	1975/10/25	16.0	291.0	Venus satellite
Venera-11	1978/09/09	1978/12/25	-14.0	299.0	
Venera-12	1978/09/14	1978/12/21	-7.0	294.0	
Venera-13	1981/10/30	1982/03/01	-7.5	303.5	
Venera-14	1981/11/04	1982/03/05	-13.0	310.0	
Venera-15	1983/06/02	1983/10/10	—	—	Venus satellite
Venera-16	1983/06/07	1983/10/14	—	—	Venus satellite
Vega-1	1984/12/15	1985/06/11	7.9	176.7	Halley's Comet

Table 1. Soviet projects aimed at Venus exploration.

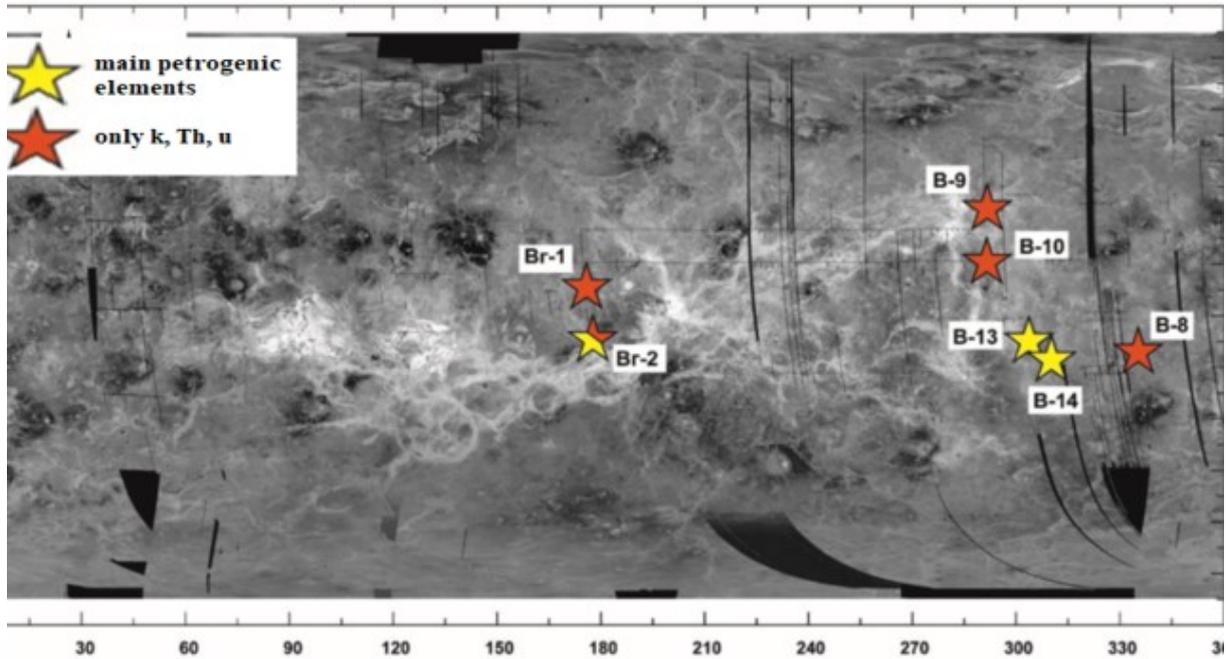


Figure 1. Landing regions of Soviet descent modules of the series Venera (B) and Vega (Br) on the Venus surface (borrowed from [5]).

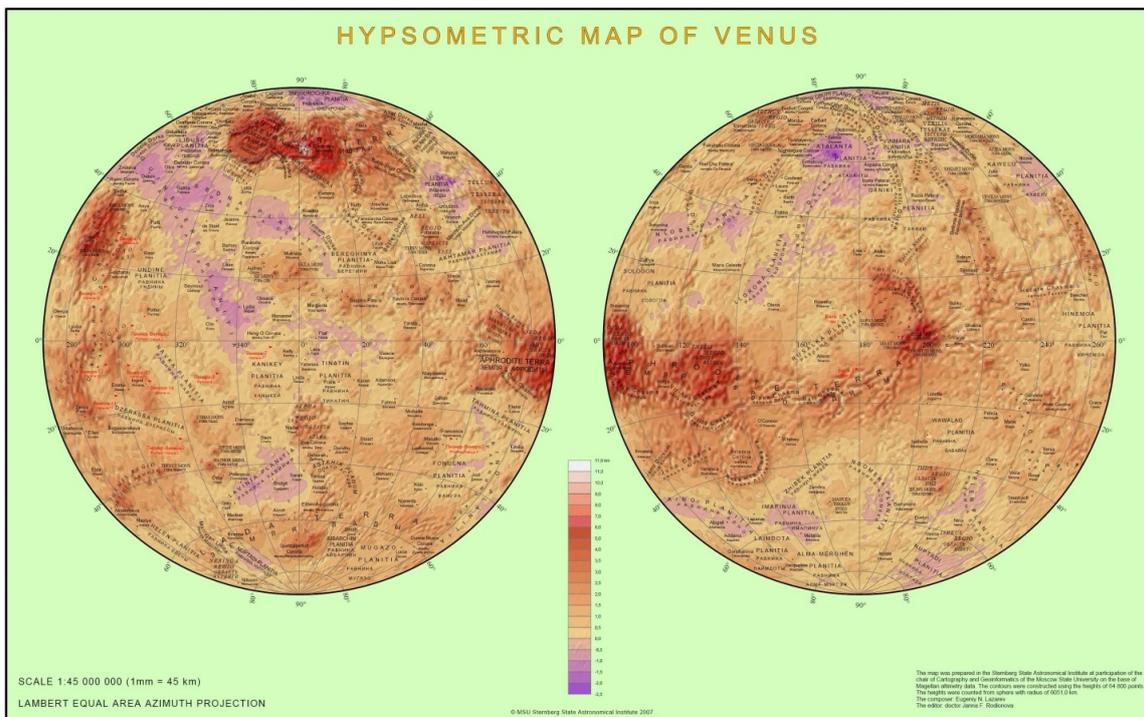


Figure 2. Hypsometrical topographic map of Venus.

spatial resolution.

Later, more accurate hypsometrical topographic and geological maps of Venus were created based on the data collected by Magellan (in particular, the hypsometrical topographic map of Venus (Sternberg State Astronomical Institute, 2006), see Fig. 2.

Nowadays, Venusian landing missions are planned based on the analysis of ballistic reachability regions for specific launch dates; the reachability regions are put on maps of terrain priority types, which intensively use the available morphological and geological data (Fig. 3, [5, 6]). The reachability regions are shown by grey and white circles in this figure. The lander is able to land in a region whose geological characteristics are known in advance [7].

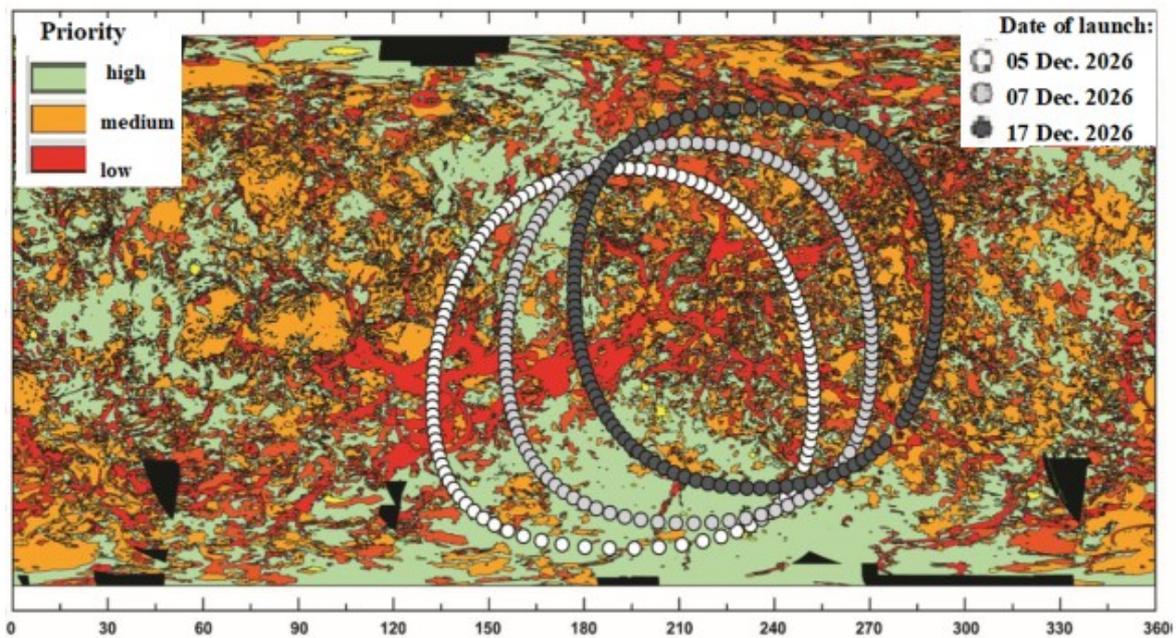


Figure 3. Priority map of Venus corresponding to the geological map of Venus with ballistic reachability regions. Borrowed from [5, 6].

2. DESIGNING LAUNCH TIME WINDOWS

Procedures for calculating launch windows for interplanetary flights are thoroughly described in the literature [8-11]. Since the flight to Venus is energy consuming, the priority in designing flight trajectories is to minimize the characteristic velocity.

The possibility to reach a planet with the minimum possible energy consumption repeats periodically. First of all, the cyclicity is determined by the synodic period T_{syn} of repeated configuration of mutual arrangement of two planets. The period T_{syn} is almost fixed relative to the stars [9], and it is expressed in terms of the revolution periods of these planets T_1 and T_2 by the rule

$$T_{\text{syn}} = T_1 T_2 / |T_2 - T_1|.$$

In the classical astronomy, oppositions are the configurations in which Earth is on the same straight line with Sun and another planet, and the direct ascent of Sun and the second planet differ by 180° . The time interval between two events when Earth and an outer planet are closest to each other is called the period of their favorable oppositions. In modern astrodynamics [9], this concept is generalized to the case of two planets; the period of their mutual configuration in the heliocentric frame of reference is approximately determined as the least common multiple of the sidereal periods of these planets and their synodic period.

The opportunity of transfer to Venus with minimum energy consumption occurs every eight years at the time of favorable oppositions. A slightly greater characteristic velocity for the transfer to Venus is required every 1.6 years.

The preliminary approximate evaluation of the optimal launch dates is determined using the Hohmann transfer orbit in the circular model of planetary motion in the Solar system. The initial approximation of the launch dates is then improved taking into account the eccentricity of the planets' orbits and their noncoplanarity. A grid of departure and arrival dates is formed, the Lambert problem about the transfer in the central field of Sun is solved for them, and the departure and arrival velocities are found. The sum of these velocities gives the characteristic velocity required for the transfer. Based on these calculations, the isolines of the characteristic velocity in the coordinate plane are constructed. The departure dates are plotted on the horizontal axis, and the transfer time is plotted on the vertical axis [3].

Next the distance of the pericenter the hemisphere through which the asymptotic axis of the arrival hyperbola passes, and the inclination to the equatorial Venus plane are specified. The transfer from the near-Earth orbit to the trajectory ensuring the arrival with the prescribed parameters is then determined. The grid of departure and arrival dates and the range of transfer durations were calculated with the accuracy up to one earth day.

3. OPTIMAL LAUNCH DATES FOR THE VENUSIAN MISSIONS DELIVERING PAYLOAD

Tab. 2 summarizes the launch dates that are best from the viewpoint of characteristic velocity expenditure for the period 2021–2028. The best launch dates to Venus are in May 2023 and in December 2024.

Date		Flight duration, days	Velocity, km/s		
launch	arrival		departure	arrival	sum
2021/10/27	2022/04/05	160	2.80	4.76	7.56
2023/05/26	2023/10/27	154	2.56	3.71	6.27
2024/12/06	2025/05/15	160	3.27	2.70	5.97
2026/06/09	2026/12/09	183	3.86	2.98	6.84
2028/01/11	2028/07/24	195	4.63	3.49	8.11

Table 2. The best launch dates to Venus in 2021–2028.

The launch window in 2024 requires the minimum characteristic velocity expenditure, and the launch window in 2028 requires the maximum characteristic velocity expenditure. The difference between the maximum and minimum values is about 2 km/s. Fig.4 illustrates the characteristic velocity expenditure for Earth–Venus transfer in the launch windows 2021–2028. The characteristic velocity expenditure is plotted on the vertical axis in km/s for the central point of the window.

In Fig. 4,5 we give the plots of launch windows for the period 2021–2028 for interplanetary Earth–Venus transfers.

Each figure shows the level lines of the characteristic velocity needed for the transfer (they are called porkchop plots in English and seashells in Russian [12]). The launch dates (in days) are plotted on the horizontal axis. The zero point corresponds to the first of January of the year indicated in the figure caption. For example, for Fig. 4 this is January 1, 2021. The transfer time is plotted on the vertical axis (in days). The maximum value of the characteristic velocity for these figures is 10.5 km/s. The cursor (in the shape of a cross) is at the point with the minimum total characteristic velocity (its value is printed nearby). The cross bars are extended to the intersection with the coordinate axes. Near these intersection points, the departure date and time (on the horizontal axis) and the arrival date and time (near the vertical axis) are shown.

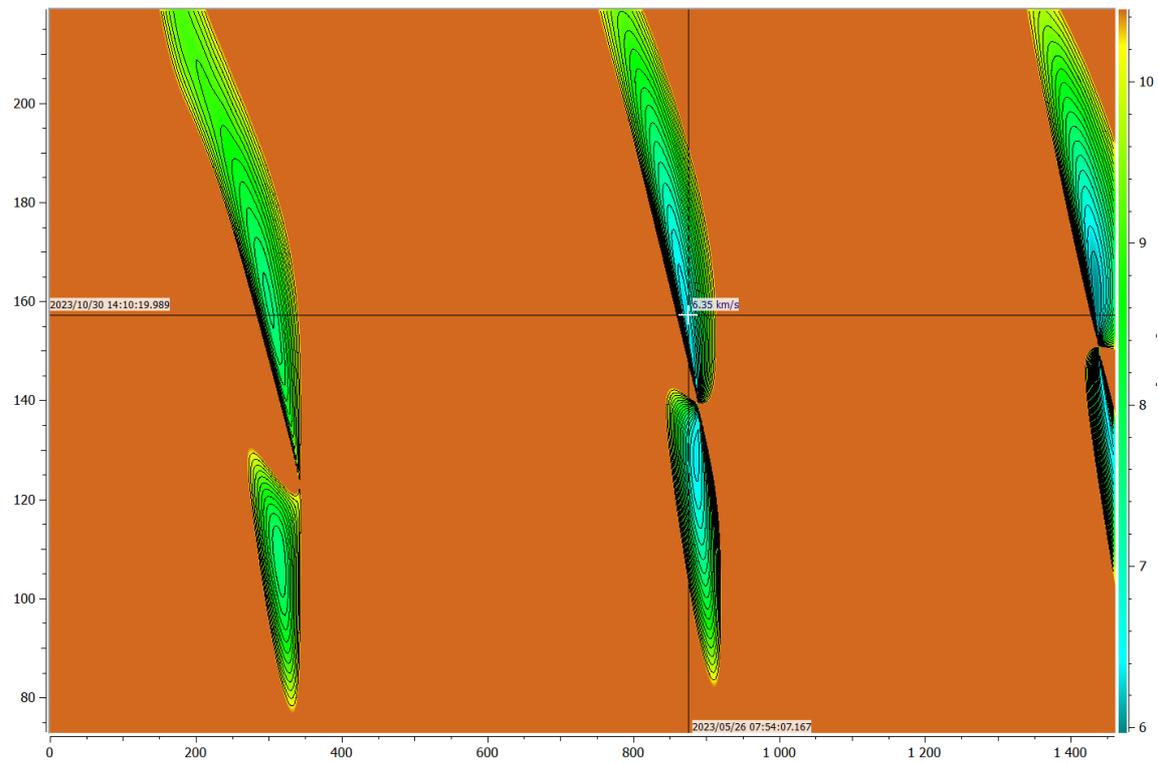


Figure 4. Earth–Venus transfer for 2021–2024. Total characteristic velocity.

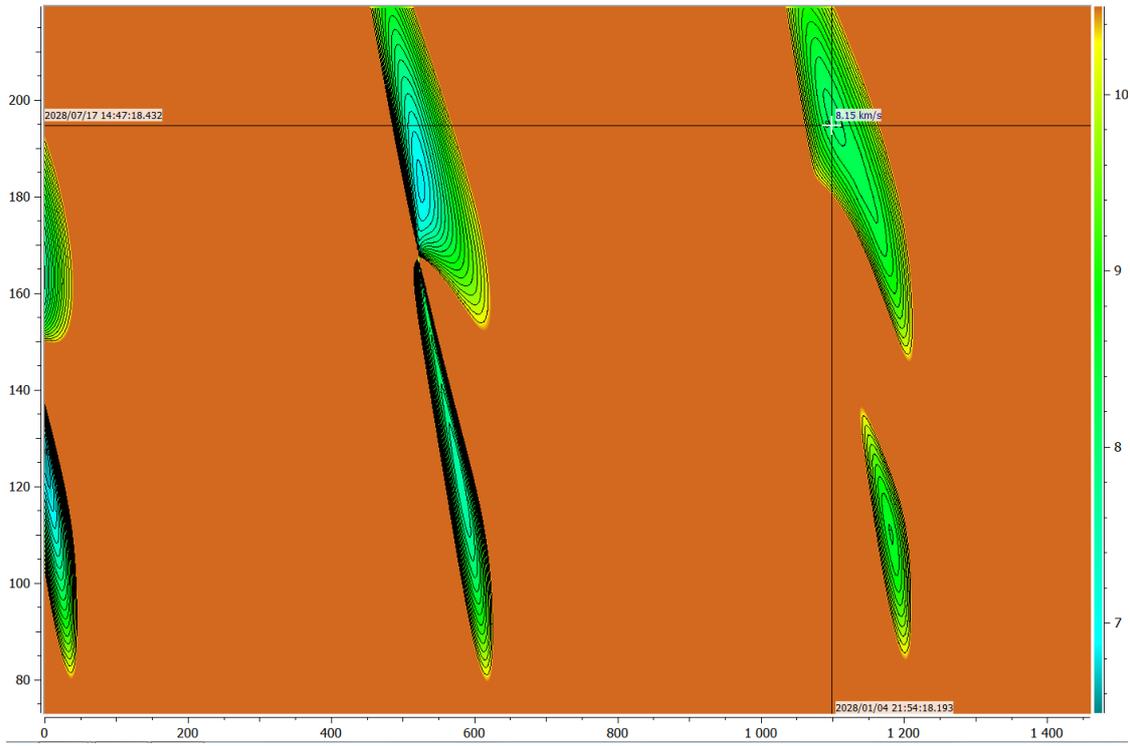


Figure 5. Earth–Venus transfer for 2025–2028. Total characteristic velocity.

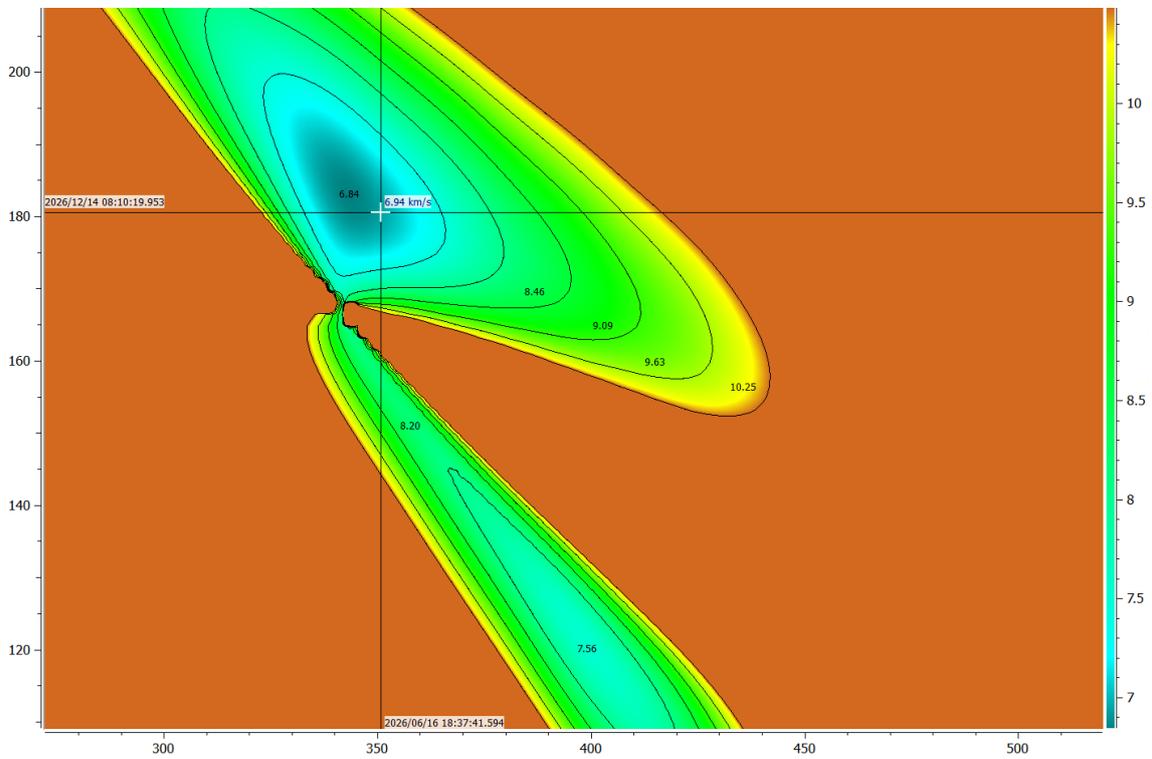


Figure 6. Isolines of characteristic velocity for the launch window in 2026.

4. OPTIMAL LAUNCH DATES FOR THE VENUSIAN DELIVERING PAYLOAD: TECHNIQUE

The numerical scale is represented by the fraction $1/M$, in which the denominator M is the number showing the factor by which the dimensions in the map are decreased (1: M). When different numerical scales are compared, the scale with a greater denominator M is smaller.

Let us describe in more detail the launch dates found above. Fig. 6,7 show the isolines of the total characteristic velocity (*porkchop plots*) for the launch dates presented in Table 3 on a greater scale [12]. The launch dates are plotted on the horizontal axis, and the transfer time is plotted on the vertical axis (in days). The numbers show the total characteristic velocity in km/s. For the second part of the eight-period cycle (including the launch windows in 2026, which is most probable for the implementation of space missions), the detailed isolines of the characteristic velocity are shown. They were constructed interactively using the software package BalCalc developed in the Ballistic Center of the Keldysh Institute of Applied Mathematics.

The probable early launch date for the Venera-D project is 2026 (the launch window is 2026/05/13–2026/07/08). However, the dates up to 2032 are also considered. In Fig. 8 show the isolines of the total characteristic velocity for the additional launch windows 2029–2032 of the Venera-D project.

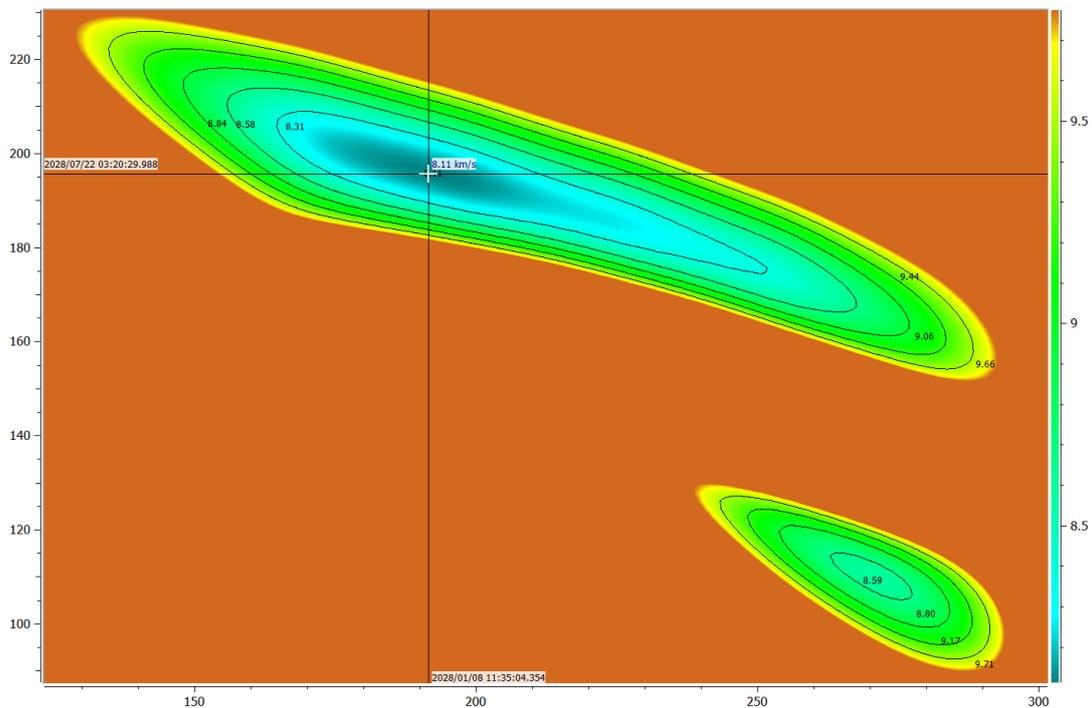


Figure 7. Isolines of characteristic velocity for the launch window in 2028.

5. SPECIFIC FEATURES OF THE PROBLEM OF GUIDING DESCENT MODULES TO THE PRESCRIBED VENUSIAN REGIONS

The problem of guiding descent modules to the prescribed regions of Venus and the problem of forming Venus-centric orbits with given parameters was studied, e.g., in [13], where Pioneer–Venus 1, 2 ballistic missions was designed; this mission was launched on August 8, 1978 with the goal of examining the Venus atmosphere. The solutions found in that research are still useful for designing new missions. As a preparation for designing this mission, reachability regions on the Venus surface depending on the coordinates and the entry angle of the descent module into the Venus atmosphere were mapped.

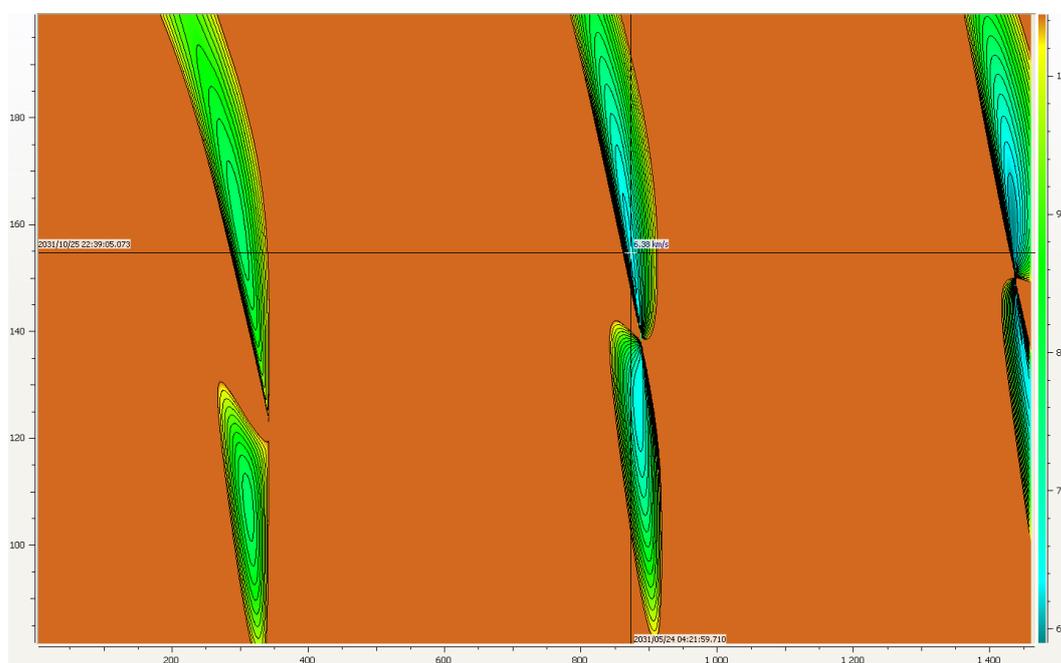


Figure 8. Earth–Venus transfer for 2029–2032. Total characteristic velocity.

More detailed exploration of Venus (e.g., within the Russian project Venera-D [5]) require not only the transfer to the planet but also the formation of an orbital subsatellite and a descent module at the given point of the flyby trajectory. The orbital subsatellite is designed to fly in the Venus atmosphere at an altitude of 50 km, and the descent module must investigate the causes of water loss from Venus. In one version of the project, the subsatellite separates from the main SC after the braking impulse accomplished when the spacecraft approaches Venus on a hyperbolic trajectory. After the subsatellite separation, the SC goes into a circular orbit around Venus. On this orbit, the descent module separates from the SC and lands on the Venus surface. Another project version assumes landing without injecting the SC into a Venus orbit.

The Venera-D project concept is based on the experience gained in preceding successful missions to Venus [5, 12]. The descent module, which includes a lander and maybe other

atmospheric modules, separates from the SC near Venus. The SC goes into a Venus orbit and becomes its satellite. It is used for communication with Earth, with other modules, receives data from the lander, coordinates the investigations, and transmits data to Earth.

6. CALCULATION OF THE REACHABILITY VENUSIAN SURFACE FOR THE DECENT MODULE

In this section, the procedure for constructing reachability regions on the Venus surface for the landing mission scheme is described in which the landing is planned from the transfer trajectory without the SC becoming a Hesperian satellite. If the entry angle into the atmosphere is too small, the SC can bounce from the atmosphere; if the entry angle is large, the SC can burn. In this paper, we use the nominal entry angle of $-15^{\circ} \pm 1^{\circ}$. Different points on the Venus surface can be reached by varying the SC orbit inclination.

For calculations, we use the picture plane (ξ, η) [10, 12]. The initial data for finding the reachability regions on the Venus surface are the departure date and time, the starting velocity V_{∞} , the inclination of the transfer orbit, and the distance of the pericenter for the arrival hyperbola. For the given orbit inclination, the coordinates (ξ, η) are used to solve the corresponding boundary value problem for ensuring the prescribed atmosphere entry angle. As a result of the calculation, the launch time and V_{∞} are refined, and the coordinates (ξ, η) that unambiguously correspond to the given entry angle are found. Additionally, for the time when the descent module enters the Venus atmosphere, the angles Earth–descent module–Venus and Sun–descent module–Venus are determined. By varying the orbit inclination, we can obtain a band of reachability points on the Venus surface. The parameter of this band is the launch date. Thus, we find the reachability region for landing on Venus without creating an intermediate Venus satellite.

In [3], for each launch date given in Tab. 2, a combination of according figures for the case when the asymptotic axis of the arrival hyperbola passes through the northern hemisphere is presented (for the southern hemisphere, the configuration is symmetric). The first figure shows the locus of landing points in the plane longitude–latitude. The same locus of points is marked at the second figure on the sphere depicting the Venus surface. The arrow shows the Venus rotation axis (the arrow is directed to the North); the equator and the meridians for 0° and 180° are also shown. The segment in this figure connects the Venus center with the point $(0^{\circ}, 0^{\circ})$.

For launch window in 2026, which is the most favorable for space projects, Fig. 9 depicts the reachability regions on the Venus surface on a greater scale. They were constructed interactively using the software package BalCalc mentioned above. The union of these regions (the "reachability region" on the Venus surface) in the entire arrival window 2026/05/14 – 2026/07/03 is also shown.

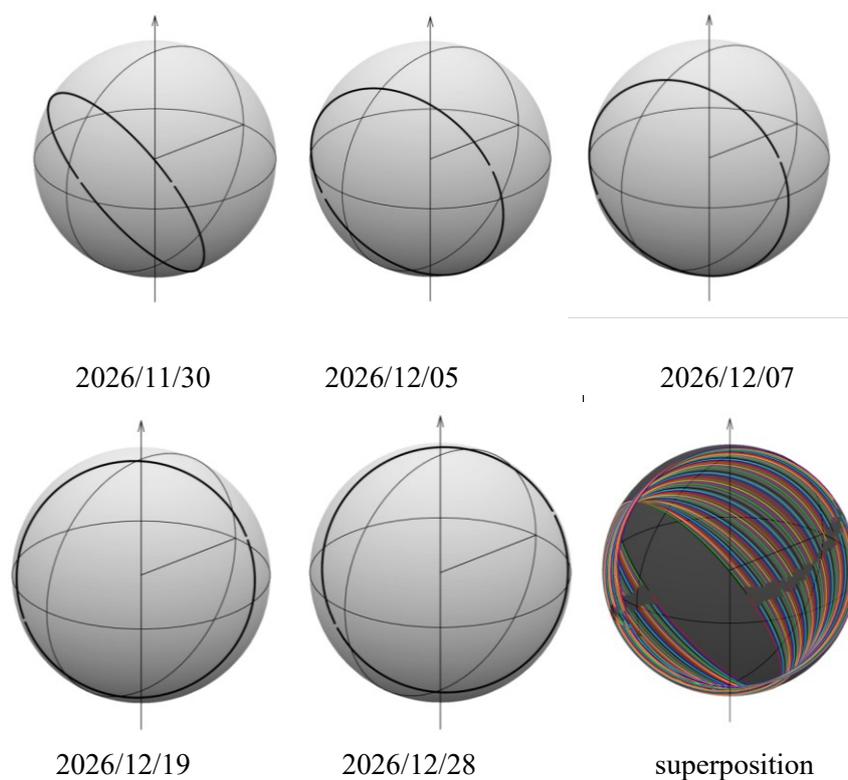


Figure 9. Dependence of the latitude and longitude of the landing point (deg.) for the launch window in 2026 for different launch dates for the arrival through the northern or southern hemisphere and their union over the entire arrival window 2026/05/14 – 2026/07/03.

CONCLUSIONS

Methods of ballistic mission design adapted for future exploration of the near-solar space and Venus are presented, and the following main results are obtained.

For planning Venus missions, which requires trajectories for the delivery (in particular, for landing) spacecraft on the Venus surface in a prescribed region, launch windows in the period 2021–2032 are found. For these launch windows, the total characteristic velocity, the departure (from Earth) asymptotic velocity, and the arrival (to Venus) asymptotic velocity are calculated.

An efficient procedure for constructing the reachability regions corresponding to these launch dates for the descent module on the Venus surface for the landing transfer scheme is described.

It is found that the launch window in 2024 requires the minimum expenditure of characteristic velocity. The launch window in 2028 for the orbital– landing transfer ballistic scheme requires the maximum expenditure of characteristic velocity, which is over 8 km/s. The difference between the maximum and minimum expenditures is about 2 km/s.

The implementation of the transfer trajectory requiring the maximum expenditure of characteristic velocity is questionable for the available Russian rockets. For this reason, some

reachability regions on the Venus surface can be actually unreachable. The reachability regions evolve with time. To overcome the difficulty due to large energy expenditure, missions to Venus can be planned once in eight years or more energy efficient ballistic schemes can be used [14].

REFERENCES

- [1] G.K. Borovin, A.V. Grushevskii, M.V. Zakhvatkin, G.S. Zaslavsky, V.A. Stepanyantz, A.G. Tuchin, D.A. Tuchin, V.S. Yaroshevsky, "Space researches in Keldysh Institute of Applied Mathematics of RAS: Past, Present, Future", *Mathematica Montisnigri*, **XLIII**, 101-127 (2018).
- [2] Yu.F. Golubev, A.V. Grushevskii, V.V. Koryanov, A.G. Tuchin, D.A. Tuchin, "A technique for designing highly inclined spacecraft orbits using gravity-assist maneuvers", *Dokl. Phys.*, **62**(2), 76-79 (2017).
- [3] Yu.F. Golubev, A.V. Grushevskii, V.V. Koryanov, A.G. Tuchin, D.A. Tuchin, "Mission Design of Multipurpose Flights to Venus," *J. Comput. Syst. Sci. Int.*, **57**(6), 970-988 (2018).
- [4] V.A. Kotel'nikov, V.L. Barsukov, E.L. Akim, et al., *Atlas poverxnosti Venery*, Glav. Upravl. Geodezii Kartogr. pri Sovete Ministrov SSSR, Moscow, (1989).
- [5] *Zemlya i Venera: Razny'e sud'by` sosednix planet. poyasnitel'naya zapiska. Sovmestny`j Rossijsko-Amerikanskij Issledovatel'skij Proekt "Venera"*, L.M.Zelenyi (Ed.), IKI RAN, Moscow, (2017).
- [6] M.A. Ivanov, L.V. Zasova, M.V. Gerasimov, O.I. Korablev, M.Ya. Marov, L.M. Zelenyi, N.I. Ignat'ev, A.G. Tuchin, "The nature of terrains of different types on the surface of Venus and selection of potential landing sites for a descent probe of the Venera-D mission," *Solar Syst. Res.* **51**(1), 1-19 (2017).
- [7] Institut kosmicheskix issledovaniy Rossijskoj akademii nauk, Nauchnaya i nauchno-organizatorskaya deyatel'nost', otchet za 2013 god.
http://www.iki.rssi.ru/annual/2013/R32_OKR_NIR-13.pdf
- [8] D.E. Okhotsimskii, Yu.G. Sikharulidze, *Osnovy` mexaniki kosmicheskogo polyota*, Uch. Posobie, Moscow, Nauka. Gl. red. fiz.-mat.-lit. (1990).
- [9] Yu.G. Sikharulidze, *Ballistika i navedenie letatelnykh apparatov*, Moscow, BINOM, Laboratoriya Znaniy Publ. (2011).
- [10] Ts.V. Solov'ev and E.V. Tarasov, *Prognozirovaniye mezhplanetnykh poletov*, Moscow, Mashinostroenie (1973).
- [11] E.V. Tarasov. *Kosmonavtika. Mexanika poleta i ballisticheskoe proektirovaniye KLA*. Moscow, Mashinostroenie (1977).
- [12] M.V. Keldysh, V.G. Ershov, D.E. Okhotsimskii, and T.M. Eneev, *Teoreticheskie issledovaniya dinamiki poleta na Mars i Veneru, Selected Works of M. V. Keldysh, Rocket Technology and Astronautics*, Moscow, Nauka (1988).
- [13] J.W. Dyer, R.R. Nunamaker, J. R. Cowley, and R. W. Jackson, "Pioneer Venus mission plan for atmospheric probes and an orbiter", *AIJA J. Spacecr. Rockets* **11**(10), 710-715 (1974).
- [14] A. Castillo, M. Bellò-Mora, J. Gonzalez, G. Janin, F. Graziani, P. Teofilatto, C. Circi, "Use of weak stability boundary trajectories for planetary capture", *Proc. of 54th International Astronautical Congress*, 766-774 (2003).

Received June 6, 2019