# DISCRETE AND CONTINUOUS MODELS AND APPLIED COMPUTATIONAL SCIENCE


## Volume 29 Number 1 (2021)

**Founded in 1993**
**Founder: PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA**

**DOI: 10.22363/2658-4670-2021-29-1**

Edition registered by the Federal Service for Supervision of Communications, Information Technology and Mass Media
**Registration Certificate**: ПИ № ФС 77-76317, 19.07.2019

ISSN 2658-7149 (online); 2658-4670 (print)
4 issues per year.
Language: English.

Publisher: Peoples' Friendship University of Russia (RUDN University).
Indexed in Ulrich's Periodicals Directory (http://www.ulrichsweb.com), in https://elibrary.ru, EBSCOhost (https://www.ebsco.com), CyberLeninka (https://cyberleninka.ru).

## Aim and Scope

Discrete and Continuous Models and Applied Computational Science arose in 2019 as a continuation of RUDN Journal of Mathematics, Information Sciences and Physics. RUDN Journal of Mathematics, Information Sciences and Physics arose in 2006 as a merger and continuation of the series "Physics", "Mathematics", "Applied Mathematics and Computer Science", "Applied Mathematics and Computer Mathematics".

Discussed issues affecting modern problems of physics, mathematics, queuing theory, the Teletraffic theory, computer science, software and databases development.

It's an international journal regarding both the editorial board and contributing authors as well as research and topics of publications. Its authors are leading researchers possessing PhD and PhDr degrees, and PhD and MA students from Russia and abroad. Articles are indexed in the Russian and foreign databases. Each paper is reviewed by at least two reviewers, the composition of which includes PhDs, are well known in their circles. Author's part of the magazine includes both young scientists, graduate students and talented students, who publish their works, and famous giants of world science.

The Journal is published in accordance with the policies of COPE (Committee on Publication Ethics). The editors are open to thematic issue initiatives with guest editors. Further information regarding notes for contributors, subscription, and back volumes is available at http://journals.rudn.ru/miph.

E-mail: miphj@rudn.ru, dcm@sci.pfu.edu.ru.

# EDITORIAL BOARD



---


**Computer Design:** *A. V. Korolkova, D. S. Kulyabov*

**Address of editorial board:**
Ordzhonikidze St., 3, Moscow, Russia, 115419
Tel. +7 (495) 955-07-16, e-mail: publishing@rudn.ru
**Editorial office:**
Tel. +7 (495) 952-02-50, miphj@rudn.ru, dcm@sci.pfu.edu.ru
site: http://journals.rudn.ru/miph

---

Paper size 70×100/16. Offset paper. Offset printing. Typeface "Computer Modern".
Conventional printed sheet 5.80. Printing run 500 copies. Open price. The order 2.
PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA
6 Miklukho-Maklaya St., 117198 Moscow, Russia
Printed at RUDN Publishing House:
3 Ordzhonikidze St., 115419 Moscow, Russia,
Ph. +7 (495) 952-04-41; e-mail: publishing@rudn.ru

Discrete & Continuous Models
& Applied Computational Science
2021, 29 (1) 4
http://journals.rudn.ru/miph

# Contents

Discrete & Continuous Models
& Applied Computational Science
2021, 29 (1) 5–13
http://journals.rudn.ru/miph

UDC 519.6
DOI: 10.22363/2658-4670-2021-29-1-5-13

# Numerical simulation of thermal processes occurring in materials under the action of femtosecond laser pulses

**Ilkizar V. Amirkhanov, Nil R. Sarker, Ibrohim Sarkhadov**

*Laboratory of Information Technologies*
*Joint Institute for Nuclear Research*
*6, Joliot-Curie St., Dubna, Moscow Region, 141980, Russian Federation*

(received: December 26, 2020; accepted: March 12, 2021)

In this work, a numerical study of the solutions of the parabolic and hyperbolic equations of heat conduction with the same physical parameters is carried out and a comparative analysis of the results obtained is carried out. The mathematical formulation of the problem is discussed. The action of the laser is taken into account through the source function, which was chosen as a double femtosecond laser pulse. In the hyperbolic equation, in contrast to the parabolic one, there is an additional parameter that characterizes the relaxation time of the heat flux. In addition, the source of the hyperbolic equation contains an additional term — the derivative of the power density of the source of the parabolic equation. This means that the temperature of the sample is influenced not only by the power density of the source, but also by the rate of its change. The profiles of the sample temperature at different times and its dynamics at different target depths are shown. The calculations were carried out for different time delays between pulses and for different relaxation parameters.

**Key words and phrases:** parabolic and hyperbolic heat equations, femtosecond laser pulse, numerical simulation

## 1. Introduction

The study of the interaction of femtosecond laser pulses with matter is important in connection with many fundamental problems (physics of non-equilibrium processes, generation of shock waves, laser acceleration of ions, modification of the properties of the irradiated material, etc.) [1]–[3].

Currently, there is a growing need for the creation and improvement of physical models capable of describing various processes in matter. Moreover, computer modeling now occupies one of the main places in the study of such problems. There are two approaches to the study and creation of physical models — atomistic and continuous.

Atomistic approaches (molecular dynamics method) allow natural consideration of the atomic structure of the crystal lattice, the effect of impurities,

the presence of dislocations, the kinetics of phase transitions, etc. The continual approach (solving the equations of continuum mechanics) includes the parabolic and hyperbolic heat equation, the two-temperature model of heat conduction, the two-temperature hydrodynamic model, etc. [2].

The molecular dynamics (MD) method [4] can be used to describe the dynamics of fast processes that arise in a substance under the action of a laser pulse. MD is quite effective for microscopic analysis of the mechanisms of melting and evaporation [5], [6]. The appearance and propagation of pressure waves generated by laser radiation [7], [8], as well as the dynamics of laser ablation [9], are well modeled using the MD.

Each approach has its own problems. When studying transport processes within the framework of a parabolic equation, a problem that arises is the infinitely high speed of thermal perturbation propagation (a consequence of the Fourier law). Generalizing the Fourier law, taking into account the relaxation time of the heat flux, we obtain the hyperbolic equation of heat conduction. The relaxation time is a characteristic of nonequilibrium of the heat conduction process. Under exposure to femtosecond pulses, non-equilibrium heating of the material occurs. Therefore, the study of such processes may turn out to be more adequate using the hyperbolic heat equation.

In this work, we carried out a numerical study of the physical processes arising under the action of femtosecond laser pulses within the framework of the parabolic and hyperbolic equations of heat conduction and carried out a comparative analysis of the results obtained.

## 2. Setting of the problem

When simulating thermal processes arising in materials under the action of femtosecond laser pulses, we use a hyperbolic model of the heat conduction equation:

$$c\rho\left(\frac{\partial T}{\partial t}+\tau_r\frac{\partial^2 T}{\partial t^2}\right)=\lambda\frac{\partial^2 T}{\partial x^2}+A(x,t)+\tau_r\frac{\partial A(x,t)}{\partial t}. \tag{1}$$

Here $c$, $\rho$, $\lambda$ are the specific heat capacity, density, and heat conductivity of the sample material, respectively. $T(x,t)$ is the sample temperature, $A(x,t)$ is the source function, which determines the heat release power density at the point with the coordinate $x$ at the time moment $t$, $\tau_r$ is the characteristic time of energy flux relaxation.

The second term in the left-hand side of equation (1) reflects the fact that the thermal process is actually hyperbolic rather than parabolic, and this model of heat conduction is widely used in practice [1], [10]–[12].

The relaxation time $\tau_r$ of the heat flux is related to the velocity of heat propagation by the formula $v=\sqrt{\lambda/c\rho\tau_r}$. If $v\to\infty$ (i.e., $\tau_r\to 0$), then we get an equation of the parabolic type. The term $\tau_r\partial A/\partial t$ means that the temperature $T$ is affected by not only the power density of its sources, but also by the rate of its change. For metals [12] $\tau_r=10^{-11}$ $s$; for steel $v=1800$ $m/s$, for aluminum $v=2830$ $m/s$, for amorphous bodies like glass and polymers the relaxation time attains $10^{-7}-10^{-5}$ $s$; in this case $v$ can exceed the velocity of sound propagation $v_s$ in these media.

In general, the heat capacity, thermal conductivity, and material density depend on temperature. In this work, the temperature dependence of the parameters of the sample material is disregarded.

Equation (1) is solved with the following initial and boundary conditions:

$$T(x,0)=T_0, \quad T(x_{\max},t)=T_0,$$
$$\left.\frac{\partial T(x,t)}{\partial t}\right|_{t=0}=0, \quad \left.\frac{\partial T(x,t)}{\partial x}\right|_{x=0}=0. \tag{2}$$

The source function is chosen in the factorized form

$$A(x,t)=I_0[1-R(T_s)]f_1(x)f_2(t), \quad T_s=T(0,t).$$

Here $f_1(x)$, $f_2(t)$ are the spatial and temporal shape of the source, respectively, $I_0$ is the source intensity, $R(T_s)$ is the coefficient of reflection of the laser pulse from the material surface.

In the present work, $f_1(x)$ и $f_2(t)$ are chosen the same as in Ref. [13]:

$$f_1(x)=\frac{\exp(-x/L_p)}{L_p},$$

$$f_2(t)=\frac{1}{\sqrt{2\pi}}\left(\exp\left[-\frac{(t-t_0)^2}{2\sigma_t^2}\right]+\exp\left[-\frac{(t-t_0-\tau_d)^2}{2\sigma_t^2}\right]\right).$$

Here $L_p$ is the depth of penetration of laser radiation into the substance, $t_0$ is the time moment when the first pulse of the source takes the maximum value, $\tau_D$ is the time shift of the second pulse of the source with respect to the first pulse. The radiation dose is

$$\Phi=I_0\int\limits_0^\infty f_2(t)dt=2I_0\sigma_t.$$

When numerically solving equation (1) with initial and boundary conditions (2), it is convenient to replace the dimensional variables and quantities with their dimensionless counterparts. This is carried out as follows:

$$\bar{T}=\frac{T}{T_0}; \quad \bar{x}=\frac{x}{\Delta x}; \quad \bar{t}=\frac{t}{\Delta t}; \quad \bar{\sigma_t}=\frac{\sigma_t}{\Delta t}; \quad \bar{t}_0=\frac{t_0}{\Delta t}; \quad k_0=\frac{\lambda\Delta t}{c\rho\Delta x^2};$$

$$\bar{\tau}_r=\frac{\tau_r}{\Delta t}; \quad \bar{A}(\bar{x},\bar{t})=\frac{A(x,t)\Delta t}{c\rho T_0},$$

$$\frac{\partial\bar{T}}{\partial\bar{t}}+\bar{\tau}_r\frac{\partial^2\bar{T}}{\partial\bar{t}^2}=k_0\frac{\partial^2\bar{T}}{\partial\bar{x}^2}+\bar{A}(\bar{x},\bar{t})+\bar{\tau}_r\frac{\partial\bar{A}(\bar{x},\bar{t})}{\partial\bar{t}}, \tag{3}$$

$$\bar{T}(\bar{x},0)=1; \quad \frac{\partial\bar{T}(\bar{x},0)}{\partial\bar{t}}=0; \quad \frac{\partial\bar{T}(0,\bar{t})}{\partial\bar{x}}=0; \quad \bar{T}(\bar{x}_{\max},\bar{t})=1. \tag{4}$$

The dimensionless source function and the normalization conditions in this case take the form

$$\bar{A}(\bar{x},\bar{t}) = A_0 \bar{f}_1(\bar{x})\bar{f}_2(\bar{t}),$$

$$A_0 = \frac{I_0[1-R(\bar{T}_s)]\Delta t}{L_p c\rho T_0}; \quad \bar{f}_1(\bar{x}) = \exp(-\alpha\bar{x}), \quad \alpha = \Delta x/L_p,$$

$$\bar{f}_2(\bar{t}) = \frac{1}{\sqrt{2\pi}}\left(\exp\left[-\frac{(\bar{t}-\bar{t}_0)^2}{2\bar{\sigma_t}^2}\right] + \exp\left[-\frac{(\bar{t}-\bar{t}_0-\bar{\tau}_d)^2}{2\bar{\sigma_t}^2}\right]\right), \quad \Phi = 2I_0\Delta t\bar{\sigma}_t.$$

## 3. Discussion of numerical results

Numerical experiments were carried out for aluminum irradiated by the double-pulse laser with the following parameters:

$$\lambda = 236\ \frac{W}{Km}, \quad \rho = 2700\ \frac{kg}{m^3}, \quad c = 920\ \frac{J}{kgK},$$

$$x_{\max} = 3\cdot 10^{-7}\ m, \quad T_0 = 300\ K, \quad R(T_s) = 0,$$

$$\Phi = 4\cdot 10^5\ \frac{J}{m^2}, \quad \sigma_t = 5\cdot 10^{-14}\ s, \quad t_0 = 3\cdot 10^{-13}\ s,$$

$$\Delta x = 3\cdot 10^{-8}\ m, \quad \Delta t = 10^{-12}\ s.$$

The total dose $\Phi = 4\cdot 10^5\ J/m^2$ for the specified source corresponds to the intensity $I_0 \simeq 1.5957\cdot 10^{17}\ W/m^2$. Dimensionless constants $k_0$, $A_0$, $\alpha$, $\bar{t}_0$, $\bar{\sigma}_t$ take the following values:

$$k_0 \simeq 0.10556; \quad A_0 \simeq 8404.34137; \quad \alpha = 1; \quad \bar{t}_0 = 0.3; \quad \bar{\sigma}_t = 0.05.$$

Below

$$f(t) = f_2(t) + \tau_r \frac{\partial f_2(t)}{\partial t}$$

describes the time dependence of the source. For $\tau_r = 0$, we get the source for a parabolic equation.

Equation (3) with the initial and boundary conditions (4) was solved using a finite-difference three-layer explicit scheme.

Figures 1 and 3 show the time dependence of the source function, temperature profiles at different times and the dynamics of the sample temperature at different depths. The times $t_i$, $i = 1, 2, \ldots, 10$ are selected in such a way that the first five of them correspond to the action times of the source first pulse, and the rest correspond to the action times of the second pulse. The calculations were carried out until the moment the source was turned off at different times of the delay between the pulses $\tau_d$.

Figures 2 and 4 show the temperature profiles at long times, when the sources are turned off, i.e., $f(t) = 0$.

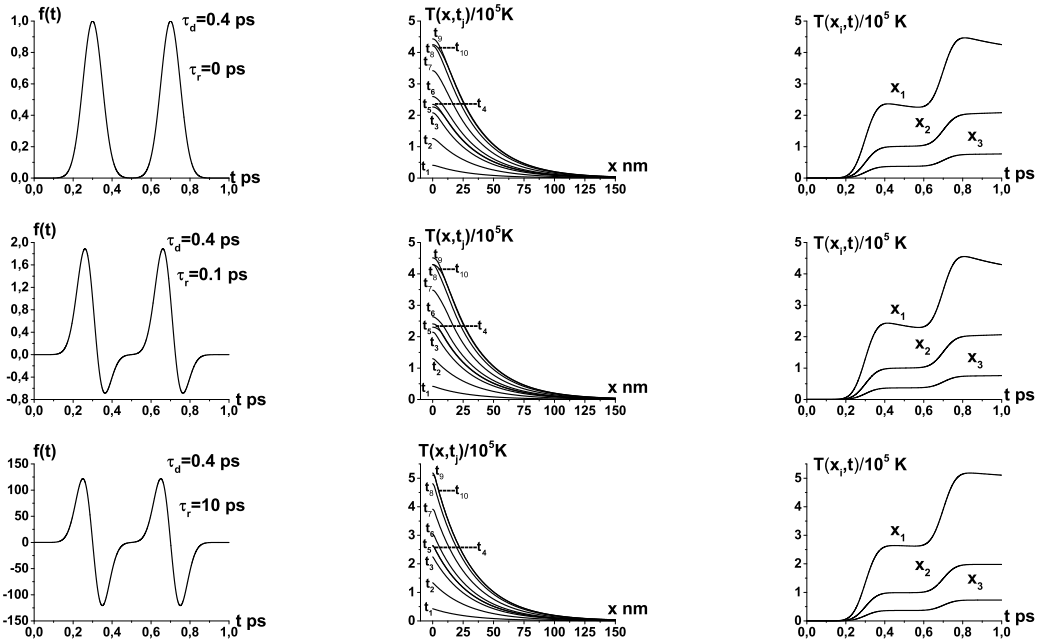


Figure 1. Time dependence of function $f(t) = f_2(t) + \tau_r \partial f_2(t)/\partial t$, temperature profiles at different time moments $T(x, t_j)$, $j = 1, 2, \ldots, 10$, $t_1 = 0.25$ ps, $t_2 = 0.3$ ps, $t_3 = 0.35$ ps,$t_4 = 0.45$ ps, $t_5 = 0.55$ ps, $t_6 = 0.65$ ps, $t_7 = 0.7$ ps, $t_8 = 0.75$ ps, $t_9 = 0.85$ ps, $t_{10} = 1$ ps, and dynamics of sample temperature at different depths ($T(x_i, t)$, $i = 1, 2, 3$, $x_1 = 0$ nm, $x_2 = 3$ nm, $x_3 = 6$ nm), obtained in the framework of the hyperbolic heat conduction equation for different values of the parameter $\tau_r$ ($\tau_r = 0$ ps, 0.1 ps, 10 ps) и $\tau_d = 0.4$ ps

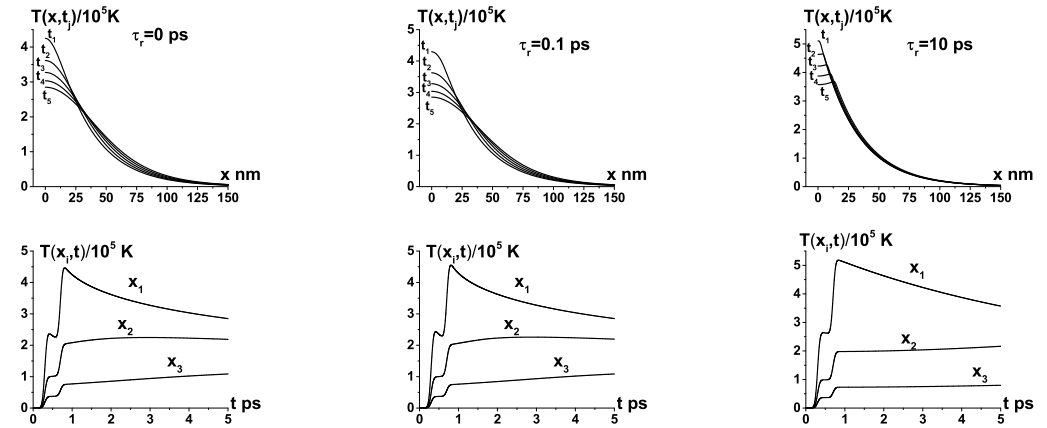


Figure 2. Temperature profiles at different time moments $T(x, t_j)$, $j = 1, 2, \ldots, 5$, $t_j = j$ ps and the sample temperature dynamics at different depths ($T(x_i, t)$, $i = 1, 2, 3$, $x_1 = 0$ nm, $x_2 = 3$ nm, $x_3 = 6$ nm), obtained in the frameworks of the hyperbolic heat conduction equation at different values of parameter $\tau_r$ ($\tau_r = 0$ ps, 0.1 ps, 10 ps) and $\tau_d = 0.4$ ps

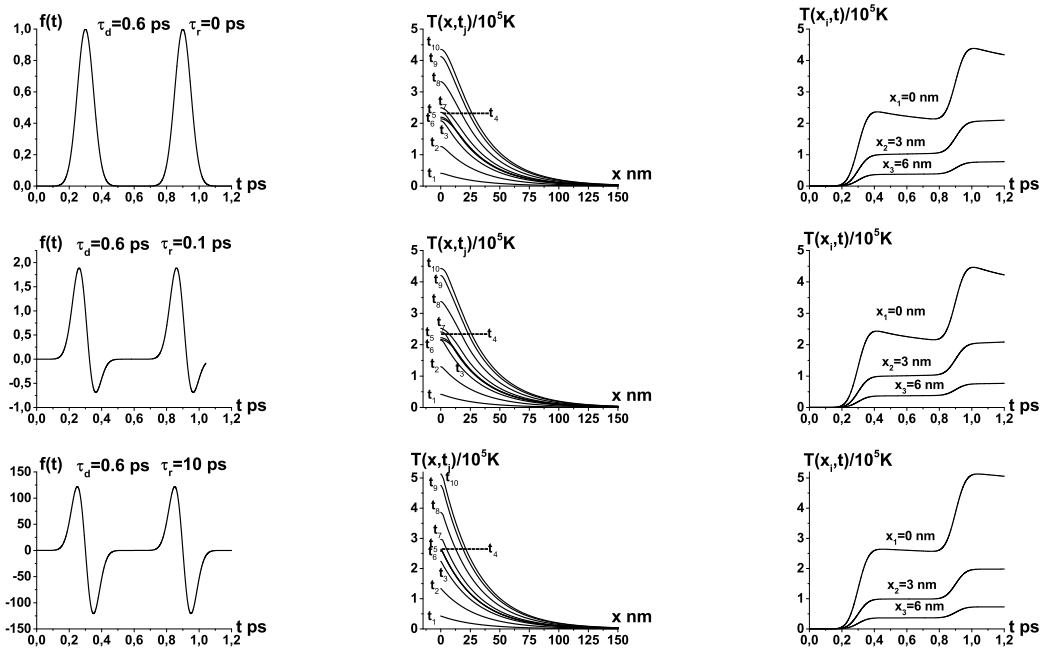


Figure 3. Time dependence of function $f(t) = f_2(t) + \tau_r \partial f_2(t)/\partial t$, temperature profiles at different time moments $T(x, t_j)$, $j = 1, 2, \ldots, 10$, $t_1 = 0.25$ ps, $t_2 = 0.3$ ps, $t_3 = 0.35$ ps,$t_4 = 0.45$ ps, $t_5 = 0.65$ ps, $t_6 = 0.75$ ps, $t_7 = 0.85$ ps, $t_8 = 0.9$ ps, $t_9 = 0.95$ ps, $t_{10} = 1.05$ ps, and the dynamics of sample temperature at different depths ($T(x_i, t)$, $i = 1, 2, 3$, $x_1 = 0$ nm, $x_2 = 3$ nm, $x_3 = 6$ nm), obtained in the frameworks of hyperbolic heat conduction equation at different values of parameter $\tau_r$ ($\tau_r = 0$ ps, 0.1 ps, 10 ps) and $\tau_d = 0.6$ ps

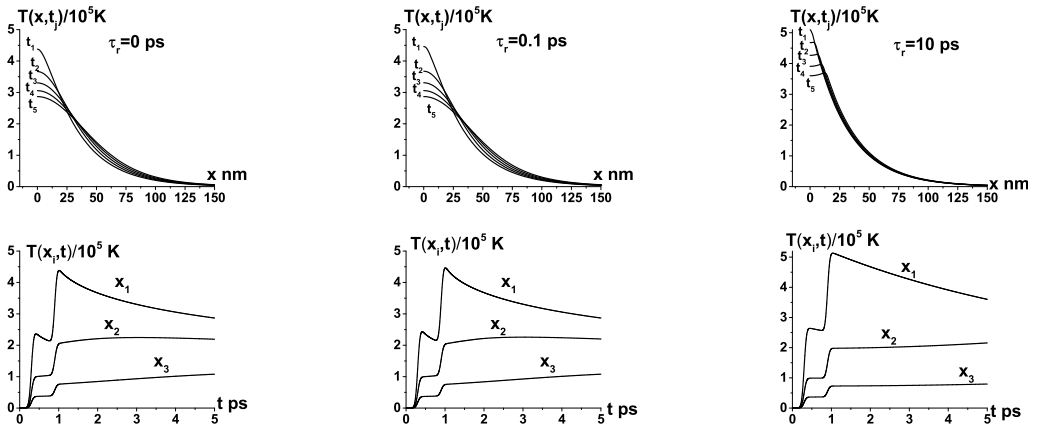


Figure 4. Temperature profiles at different time moments $T(x, t_j)$, $j = 1, 2, \ldots, 5$, $t_j = j$ ps and sample temperature dynamics at different depths ($T(x_i, t)$, $i = 1, 2, 3$, $x_1 = 0$ nm, $x_2 = 3$ nm, $x_3 = 6$ nm), obtained in the frameworks of the hyperbolic heat conduction equation at different values of parameter $\tau_r$ ($\tau_r = 0$ ps, 0.1 ps, 10 ps) and $\tau_d = 6$ ps

## 4. Conclusion

In contrast to the parabolic equation, the hyperbolic one includes an additional parameter that characterizes the heat flux relaxation time. A derivative of the power density of the source of the parabolic equation is additionally present in the source of the hyperbolic equation. This fact means that the sample temperature is affected not only by the source power density, but also by the rate of its variation. Due to this dependence, at some time moments the source takes negative values depending on the relaxation time parameter. Nevertheless, the temperature at the sample surface given by the solution of the hyperbolic equation is higher than that given by the solution of the parabolic equation.

## Acknowledgments

The work was carried out under the financial support from th Russian Foundation for Basic Research, grants No. 19-01-00645a and No. 20-51-44001 mong-a.

## References

[1] S. L. Sobolev, "Local non-equilibrium transport models," *Physics Uspekhi*, vol. 40, no. 10, pp. 1043–1053, 1997, in Russian. DOI: 10.1070/PU1997v040n10ABEH000292.

[2] S. I. Anisimov and B. S. Luk'yanchuk, "Selected problems of laser ablation theory," *Usp. Fiz. Nauk*, vol. 172, no. 3, pp. 301–333, 2002. DOI: 10.3367/UFNr.0172.200203b.0301.

[3] V. P. Veiko, M. N. Libensonm, G. G. Chervyakov, and E. B. Yakovlev, *Interaction of laser radiation with matter. Power optics [Vzaimodeystviye lazernogo izlucheniya s veshchestvom. Silovaya optika]*, V. I. Konov, Ed. Moscow: Fizmatlit, 2008, in Russian.

[4] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. Clarendon Press, 1991.

[5] Z. H. Jin, P. Gumbsch, K. Lu, and E. Ma, "Melting mechanisms at the limit of superheating," *Physical Review Letters*, vol. 87, p. 055 703, 5 Jul. 2001. DOI: 10.1103/PhysRevLett.87.055703.

[6] F. F. Abraham and J. Q. Broughton, "Pulsed melting of silicon (111) and (100) surfaces simulated by molecular dynamics," *Phys. Rev. Lett.*, vol. 56, pp. 734–737, 7 Feb. 1986. DOI: 10.1103/PhysRevLett.56.734.

[7] V. Zhigilei and B. J. Garrison, "Pressure Waves in Microscopic Simulations of Laser Ablation," in *Materials Research Society (MRS) Proceedings*, vol. 538, Cambridge University Press, 1998, pp. 491–496. DOI: 10.1557/PROC-538-491.

[8] J. I. Etcheverry and M. Mesaros, "Molecular dynamics simulation of the production of acoustic waves by pulsed laser irradiation," *Phys. Rev. B*, vol. 60, pp. 9430–9434, 13 Oct. 1999. DOI: 10.1103/PhysRevB.60.9430.

[9] L. V. Zhigilei and B. J. Garrison, "Microscopic mechanisms of laser ablation of organic solids in the thermal and stress confinement irradiation regimes," *Journal of Applied Physics*, vol. 88, no. 3, pp. 1281–1298, 2000. DOI: 10.1063/1.373816.

[10] A. V. Lykov, *Heat and Mass Transfer [Teplomassoobmen]*, 2nd. Moscow: Energiya, 1978, in Russian.

[11] P. Vernott, "Les paradoxes de la théorie continue de l'équation de la chaleur," *Comptes rendus de l'Académie des Sciences*, vol. 246, no. 22, pp. 3154–3155, 1958.

[12] E. M. Kartashov and V. A. Kudinov, *Analytical methods of the theory of heat conduction and its applications [Analiticheskiye metody teorii teploprovodnosti i yeye prilozheniy]*. Moscow: LENAND, 2018.

[13] V. B. Fokin, "Continuous-automaton model and its application for numerical calculation of the effect of single and double femtosecond laser pulses on metals [Kontinual'no-atomaticheskaya model' i yeye primeneniye dlya chislennogo rascheta vozdeystviya odinochnogo i dvoynogo femtosekundnogo lazernogo impul'sa na metally]," in Russian, Candidate of Sci. in Phys. and Math. (PhD) Thesis, Joint Institute for High Temperatures of the Russian Academy of Sciences, Moscow, 2017.

**For citation:**

I. V. Amirkhanov, N. R. Sarker, I. Sarkhadov, Numerical simulation of thermal processes occurring in materials under the action of femtosecond laser pulses, Discrete and Continuous Models and Applied Computational Science 29 (1) (2021) 5–13. DOI: 10.22363/2658-4670-2021-29-1-5-13.

**Information about the authors**:

**Amirkhanov, Ilkizar V.** — Candidate of Physical and Mathematical Sciences, Head of Sector "Scientific Division of Computational Physics". Laboratory of Information Technologies of the Joint Institute for Nuclear Research (e-mail: camir@jinr.ru, ORCID: https://orcid.org/0000-0003-2621-144X, Scopus Author ID: 6507929197)

**Sarker, Nil R.** — Candidate of Physical and Mathematical Sciences, Senior Researcher "Scientific Division of Computational Physics". Laboratory of Information Technologies of the Joint Institute for Nuclear Research (e-mail: sarker@jinr.ru, ORCID: https://orcid.org/0000-0003-0690-2534, Scopus Author ID: 14829498100)

**Sarkhadov, Ibrohim** — Candidate of Physical and Mathematical Sciences, Senior Researcher "Scientific Division of Computational Physics". Laboratory of Information Technologies of the Joint Institute for Nuclear Research (e-mail: ibrohim@jinr.ru, ORCID: https://orcid.org/0000-0001-5534-3332, Scopus Author ID: 14829718300)

УДК 519.6
DOI: 10.22363/2658-4670-2021-29-1-5-13

# Численное моделирование тепловых процессов, возникающих в материалах при воздействии фемтосекундных лазерных импульсов

**И. В. Амирханов, Н. Р. Саркер, И. Сархадов**

*Лаборатория информационных технологий*
*Объединенный институт ядерных исследований*
*ул. Жолио-Кюри, д. 6, Дубна, Московская область, 141980, Россия*

В работе проведено численное исследование решений параболического и гиперболического уравнений теплопроводности при одинаковых физических параметрах, а также сравнительный анализ полученных результатов. Обсуждена математическая постановка задачи. Действие лазера учтено через функцию источника, которую выбрали в виде двойного фемтосекундного лазерного импульса. В гиперболическом уравнении, в отличие от параболического, присутствует дополнительный параметр, который характеризует время релаксации потока тепла. Кроме этого, в источнике гиперболического уравнения присутствует дополнительное слагаемое — производная от плотности мощности источника параболического уравнения. Это означает, что на температуру образца оказывает влияние не только плотность мощности источника, но и скорости его изменения. Приведены профили температуры образца в разные моменты времени и её динамика на разных глубинах мишени. Расчёты проводились при различных временах задержки между импульсами и при различных параметрах релаксации.

**Ключевые слова:** параболическое и гиперболическое уравнения теплопроводности, фемтосекундный лазерный импульс, численное моделирование

*Research article*

UDC 519.872, 519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2021-29-1-14-21

# Normal modes of a waveguide as eigenvectors of a self-adjoint operator pencil

**Mikhail D. Malykh**

*Peoples' Friendship University of Russia (RUDN University)*
*6, Miklukho-Maklaya St., Moscow, 117198, Russian Federation*

(received: January 19, 2021; accepted: March 12, 2021)

A waveguide with a constant, simply connected section $S$ is considered under the condition that the substance filling the waveguide is characterized by permittivity and permeability that vary smoothly over the section $S$, but are constant along the waveguide axis. Ideal conductivity conditions are assumed on the walls of the waveguide. On the basis of the previously found representation of the electromagnetic field in such a waveguide using 4 scalar functions, namely, two electric and two magnetic potentials, Maxwell's equations are rewritten with respect to the potentials and longitudinal components of the field. It appears possible to exclude potentials from this system and arrive at a pair of integro-differential equations for longitudinal components alone that split into two uncoupled wave equations in the optically homogeneous case. In an optically inhomogeneous case, this approach reduces the problem of finding the normal modes of a waveguide to studying the spectrum of a quadratic self-adjoint operator pencil.

**Key words and phrases:** waveguide, normal modes, hybridization of normal modes, eigenvalue problem, quadratic operator pencils

## 1. Introduction

Consider a waveguide representing a cylinder of constant cross-section $S$ filled with an optically inhomogeneous substance, which we will characterize with a permittivity and a permeability. Hereinafter, we will make use of a Cartesian coordinate system, the $Oz$-axis of which coincides with the waveguide axis. We will assume that the permittivity and permeability do not depend on $z$, but are piecewise smooth functions of $x, y$. The normal modes of a waveguide are non-trivial solutions of Maxwell's equations of the form

$$\vec{E}(x,y)e^{ik\beta z-i\omega t}, \quad \vec{H}(x,y)e^{ik\beta z-i\omega t}, \tag{1}$$

satisfying the conditions of ideal conductivity of the waveguide walls. Here the positive parameter $\omega$ is the circular frequency of the wave, $k = \omega/c$ is the wave number, and the complex parameter $\beta$ is the phase constant.

Substitution of the expression (1) into Maxwell's equations yields 8 equations for 6 unknowns $E_x, \dots, H_z$, containing two parameters, $k$ and $\beta$. It is usually assumed that the wavenumber of the considered waves is given, and then we get an eigenvalue problem with respect to the spectral parameter $\beta$. This problem was successfully solved in the case of constant $\epsilon$ and $\mu$, thanks to the introduction of two scalar potentials, the electric and magnetic Borgnis functions $u$ and $v$ [1], [2]. In the attempt to study a general case undertaken in the beginning of 2000s [3]–[5], it was not possible to introduce potentials and the problem was investigated with respect to three randomly chosen field components. With this approach, the normal waves of the waveguide turned out to be eigenfunctions of some non-self-adjoint quadratic operator pencil acting in a space specially selected by the functional.

Not all properties of a hollow waveguide can be extended to the case of a waveguide filled with an optically inhomogeneous substance. We can confidently reject the hypothesis of the field decomposition into TE- and TM-waves, since the existence of hybrid modes has been proved analytically in half-filled waveguides [6, § 3.5]. With less confidence, one can reject the hypothesis that the propagation constants of normal modes cannot have both real and imaginary parts. In a series of numerical experiments [7]–[9], it was shown that the propagation constants of the normal modes of an axially symmetric waveguide with a dielectric core can leave the real and imaginary axes of the $\beta$ complex plane. However, to calculate these eigenvalues, we used the truncation method and standard solvers to find the eigenvalues of non-self-adjoint matrices. Our experiments in FreeFem++ [10] showed that solvers of this kind can introduce a complex addition to the spectrum of a self-adjoint problem.

We have recently succeeded in extending the theory of Borgnis functions to the case of a waveguide filled with optically inhomogeneous matter [11], [12]. In this case, we have increased the number of potentials to four. Mode hybridization makes one think that the system of equations for the potentials does not split in the general case, but we cannot exclude the fact that this system is written in a self-adjoint form. In this paper, we intend to present such a self-adjoint formulation of the problem of finding the normal modes of a waveguide.

## 2. Representation of the electromagnetic field using electric and magnetic potentials

Let for simplicity the waveguide cross section $S$ be a planar simply connected domain with smooth boundary $\partial S$, and let the permittivity $\epsilon$ and permeability $\mu$ be smooth functions of $x, y$. Denote as $Z, T$ the segments of finite of infinite length on the axes $z$ and $t$, respectively and assume $\partial_s = \frac{\partial}{\partial s}$ in all cases except $\partial t = \frac{1}{c}\frac{\partial}{\partial t}$. The unit external normal vector to the curve $\partial S$ will be denoted as $\vec{n} = (n_x, n_y, 0)^T$, and the tangent vector in the $xy$-plane as $\vec{\tau} = (-n_y, n_x, 0)^T$. Also for brevity let us assume that

$$\vec{A}_\perp = (A_x, A_y, 0)^T \quad \text{and} \quad \nabla = (\partial_x, \partial_y, 0)^T, \quad \nabla' = (-\partial_y, \partial_x, 0)^T$$

and

$$\Delta_q u = \operatorname{div}(q\nabla u) = \frac{\partial}{\partial x} q \frac{\partial u}{\partial x} + \frac{\partial}{\partial y} q \frac{\partial u}{\partial y}.$$

As in the theory of Borgnis functions, the scalar function $u$ turning into zero at the boundary will be referred to as electric potential and the scalar function satisfying the Newmann condition $\partial_n v = 0$ at the boundary — as magnetic potential $\partial S \times Z \times T$. Hereinafter electric and magnetic potentials are denoted by $u$ and $v$, respectively, with different indices.

The main result about the four potentials established by us earlier [12] is that the electromagnetic field allows a representation in terms of four potentials, namely, two electric potentials $u_e$, $u_h$ and two magnetic ones $v_e$, $v_h$:

$$\vec{E}_\perp = \nabla \partial_z u_e + \frac{1}{\epsilon} \nabla' \partial_t v_e, \quad \vec{H}_\perp = \nabla \partial_z v_h - \frac{1}{\mu} \nabla' \partial_t u_h. \tag{2}$$

Therefore, below we seek the solution of Maxwell's equations in a waveguide in the form (2) without any loss of generality.

## 3. Maxwell's equations in terms of potentials

Substituting expression (2) into Maxwell's equations, we get 8 rather than 6 independent equations. Four of these equations allow expressing the potentials in terms of the longitudinal field components $E_z$ и $H_z$. The relation is determined by classical boundary-value problems. The electric potentials can be found as solutions of Dirichlet problems

$$\begin{cases} \Delta_\epsilon u_e + \epsilon E_z = 0 & \text{in } S \times Z \times T, \\ u_e = 0 & \text{on } \partial S \times Z \times T \end{cases} \tag{3}$$

and

$$\begin{cases} \Delta_{\frac{1}{\mu}} u_h + \epsilon E_z = 0 & \text{in } S \times Z \times T, \\ u_h = 0 & \text{on } \partial S \times Z \times T. \end{cases} \tag{4}$$

The magnetic potentials can be found as solutions of Newmann problems

$$\begin{cases} \Delta_{\frac{1}{\epsilon}} v_e + \mu H_z = 0 & \text{in } S \times Z \times T, \\ \partial_n v_e = 0 & \text{on } \partial S \times Z \times T \end{cases} \tag{5}$$

and

$$\begin{cases} \Delta_\mu v_h + \mu H_z = 0 & \text{in } S \times Z \times T, \\ \partial_n v_h = 0 & \text{on } \partial S \times Z \times T. \end{cases} \tag{6}$$

In this case from Maxwell's equations it follows that

$$\partial_t \iint\limits_S \mu H_z dxdy = \partial_z \iint\limits_S \mu H_z dxdy = 0. \tag{7}$$

In problems of monochromatic wave propagation $\partial_t$ is equivalent to multiplication by the number $i\omega$, therefore from (7) the solvability of the above problems with the Neumann conditions follows.

The rest two equations can be written in the form

$$\begin{cases} \Delta_\epsilon \left(\partial_z^2 u_e - \partial_t^2 u_h - E_z\right) = \partial_z \partial_t \dfrac{\partial(v_h, \epsilon\mu)}{\partial(x,y)}, \\ \Delta_\mu \left(\partial_z^2 v_h - \partial_t^2 v_e - H_z\right) = -\partial_z \partial_t \dfrac{\partial(u_e, \epsilon\mu)}{\partial(x,y)}. \end{cases} \tag{8}$$

Substituting here the expressions for the potentials in terms of the field longitudinal components $E_z$ and $H_z$, which are obtained by solving the problems (3)–(6), we rewrite this system in the form

$$\begin{pmatrix} \hat{A}_e & 0 \\ 0 & \hat{B}_h \end{pmatrix} \partial_z^2 \vec{F} - \begin{pmatrix} \hat{A}_h & 0 \\ 0 & \hat{B}_e \end{pmatrix} \partial_t^2 \vec{F} - \begin{pmatrix} \epsilon & 0 \\ 0 & \mu \end{pmatrix} \vec{F} = \begin{pmatrix} 0 & \hat{C} \\ \hat{C}^* & 0 \end{pmatrix} \partial_z \partial_t \vec{F}, \tag{9}$$

where as an unknown we consider $\vec{F} = (E_z, H_z)^T$, composed of the field longitudinal components. Here $\hat{A}_e, \ldots, \hat{B}_h$ are symmetric positively defined integral operators acting in $L^2(S)$, and $\hat{C}$ is a non-symmetric integral operator. This operator makes impossible the separation of the problem into two independent problems, due to which the hybridization of modes occurs. We will call it a hybridization operator.

## 4. Normal modes of a waveguide

Normal mode (1) corresponds to a solution of the system (9) in the form

$$E_z = E_z(x,t) e^{ik\beta z - i\omega t}, \quad H_z = H_z(x,t) e^{ik\beta z - i\omega t}.$$

Taking the dependence on $z, t$ into account, we can formulate the problem of finding the normal modes of the waveguide as an eigenvalue problem

$$\beta^2 \begin{pmatrix} \hat{A}_e & 0 \\ 0 & \hat{B}_h \end{pmatrix} \vec{F} - \begin{pmatrix} \hat{A}_h & 0 \\ 0 & \hat{B}_e \end{pmatrix} \vec{F} + \frac{1}{k^2} \begin{pmatrix} \epsilon & 0 \\ 0 & \mu \end{pmatrix} \vec{F} = \beta \begin{pmatrix} 0 & \hat{C} \\ \hat{C}^* & 0 \end{pmatrix} \vec{F} \tag{10}$$

with respect to the spectral parameter $\beta$.

Thus the problem of finding normal modes reduces to the analysis of the spectrum of the polynomial operator pencil

$$\hat{A}_2 \beta^2 + \hat{A}_1 \beta + \hat{A}_0, \tag{11}$$

where the coefficients $\hat{A}_0$, $\hat{A}_1$, $\hat{A}_2$ are self-adjoint operators with respect to a scalar product in $L^2(S) \times L^2(S)$; the senior coefficient $\hat{A}_2$ is a positively defined and completely continuous operator; the operator $\hat{A}_1$ is completely continuous and the operator $\hat{A}_0$ is bounded and reversible. The pencils of such form arouse in the linear theory of small damped oscillations and were studied by M. G. Krein and G. K. Langer [13, §12].

## 5. Conclusion

By introducing four potentials, we were able to reduce the problem of wave propagation in a waveguide filled with an inhomogeneous substance to a linear second-order partial differential equation (9), the coefficients of which are self-adjoint operators. In this case, the problem of finding normal waves is reduced to studying the spectrum of the quadratic operator pencil (11). Thus, the formulation of the eigenvalue problem retains the symmetry characteristic of scalar eigenvalue problems.

This means, first of all, that with discretization by the truncation method, we obtain a problem for the eigenvalues of a quadratic self-adjoint matrix pencil. By means of the known procedure [13, §12] it can be reduced to the generalized eigenvalue problem $\hat{A}u = \beta\hat{B}u$, where $\hat{A}, \hat{B}$ are self-adjoint matrices. This opens up possibilities for using specialized eigenvalue solvers.

The proposed formulation is also convenient for theoretical research, since the physical meaning of its terms is clear. In particular, the linear element of the pencil describes the hybridization of modes in a waveguide filled with an optically inhomogeneous medium. A natural next step will be to study the perturbation of a hollow waveguide by a weakly inhomogeneous substance.

## Acknowledgments

The author would like to thank Prof. L. A. Sevastianov for constructive criticism of the manuscript. The publication was supported by the RUDN University Strategic Academic Leadership Program.

## References

[1] A. G. Sveshnikov and I. E. Mogilevsky, *Mathematical problems in the theory of diffraction [Matematicheskiye zadachi teorii difraktsii]*. Moscow: MSU, 2010, in Russian.

[2] K. Zhang and D. Li, *Electromagnetic theory for microwaves and optoelectronics*, 2nd ed. Berlin: Springer, 2008.

[3] A. N. Bogolyubov, A. L. Delitsyn, and A. G. Sveshnikov, "On the completeness of the set of eigen- and associated functions of a waveguide," *Computational Mathematics and Mathematical Physics*, vol. 38, no. 11, pp. 1815–1823, 1998.

[4] A. N. Bogolyubov, A. L. Delitsyn, and M. D. Malykh, "On the root vectors of a cylindrical waveguide," *Computational Mathematics and Mathematical Physics*, vol. 41, no. 1, pp. 121–124, 2001.

[5] A. L. Delitsyn, "On the completeness of the system of eigenvectors of electromagnetic waveguides," *Computational Mathematics and Mathematical Physics*, vol. 51, pp. 1771–1776, 2011. DOI: 10.1134/S0965542511100058.

[6] W. C. Chew. "Lectures on theory of microwave and optical waveguides." (2012), [Online]. Available: http://wcchew.ece.illinois.edu/chew/course/tgwAll20121211.pdf.

[7] N. A. Novoselova, S. B. Raevskii, and A. A. Titarenko, "Calculation of characteristics of symmetric modes propagating in a circular waveguide with radially-heterogeneous dielectric filling [Raschet kharakteristik rasprostraneniya simmetrichnykh voln kruglogo volnovoda s radial'no-neodnorodnym dielektricheskim zapolneniyem]," *Trudy Nizhegorodskogo gosudarstvennogo tekhnicheskogo universiteta im. R.Ye. Alekseyeva*, no. 2(81), pp. 30–38, 2010, in Russian.

[8] A. L. Delitsyn and S. I. Kruglov, "Mixed finite elements used to analyze the real and complex modes of cylindrical waveguides," *Moscow University Physics Bulletin*, vol. 66, pp. 546–560, 2011. DOI: 10.3103/S0027134911060063.

[9] A. L. Delitsyn and S. I. Kruglov, "Application of the mixed finite element method for calculating the modes of cylindrical waveguides with a variable refractive index [Primeneniye metoda smeshannykh konechnykh elementov dlya vychisleniya mod tsilindricheskikh volnovodov s peremennym pokazatelem prelomleniya]," *Zhurnal radioelektroniki*, no. 4, pp. 1–28, 2012, in Russian.

[10] F. Hecht, *Freefem++*, 3rd ed., Laboratoire Jacques-Louis Lions, Universitè Pierre et Marie Curie, Paris, 2018.

[11] M. D. Malykh, N. E. Nikolaev, L. A. Sevastianov, and A. A. Tiutiunnik, "On the representation of electromagnetic fields in closed waveguides using four scalar potentials," *Journal of Electromagnetic Waves and Applications*, vol. 32, no. 7, pp. 886–898, 2018. DOI: 10.1080/09205071.2017.1409137.

[12] M. D. Malykh and L. A. Sevast'yanov, "On the representation of electromagnetic fields in discontinuously filled closed waveguides by means of continuous potentials," *Computational Mathematics and Mathematical Physics*, vol. 59, pp. 330–342, 2019. DOI: 10.1134/S0965542519020118.

[13] I. C. Gohberg and M. G. Krein, *Introduction to the Theory of Linear Nonselfadjoint Operators in Hilbert Space*. American Mathematical Society, 1969.

**For citation:**

M. D. Malykh, Normal modes of a waveguide as eigenvectors of a self-adjoint operator pencil, Discrete and Continuous Models and Applied Computational Science 29 (1) (2021) 14–21. DOI: 10.22363/2658-4670-2021-29-1-14-21.

**Information about the authors**:

**Malykh, Mikhail D.** — Doctor of Physical and Mathematical Sciences, Assistant Professor of Department of Applied Probability and Informatics of Peoples' Friendship University of Russia (RUDN University) (e-mail: malykh_md@pfur.ru, phone: +7(495)9550927, ORCID: https://orcid.org/0000-0001-6541-6603, ResearcherID: P-8123-2016, Scopus Author ID: 6602318510)

УДК 519.872, 519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2021-29-1-14-21

# Нормальные моды волновода как собственные векторы самосопряжённого операторного пучка

**М. Д. Малых**

*Российский университет дружбы народов*
*ул. Миклухо-Маклая, д. 6, Москва, 117198, Россия*

В статье рассматривается волновод постоянного односвязного сечения $S$ при условии, что заполняющее волновод вещество характеризуется диэлектрической и магнитной проницаемостями, меняющимися плавно на сечении $S$, но постоянными вдоль оси волновода. На стенках волновода взяты условия идеальной проводимости. На основе найденного ранее представления электромагнитного поля в таком волноводе при помощи четырёх скалярных функций — двух электрических и двух магнитных потенциалов — уравнения Максвелла записаны относительно потенциалов и продольных компонент поля. Из этой системы удаётся исключить потенциалы и записать пару интегро-дифференциальных уравнений относительно одних продольных компонент, расщепляющихся на два несвязанных волновых уравнения в оптически однородном случае. В оптически неоднородном случае этот подход позволяет свести задачу об отыскании нормальных мод волновода к исследованию спектра квадратичного самосопряжённого операторного пучка.

**Ключевые слова:** волновод, нормальные моды, гибридизация нормальных мод, задача на собственные значения, квадратичные пучки

*Research article*

UDC 519.6
DOI: 10.22363/2658-4670-2021-29-1-22-35

# Study of the impact of the COVID-19 pandemic on international air transportation

**Eugeny Yu. Shchetinin**

*Financial University under the Government of Russian Federation*
*49, Leningradsky Prospect, Moscow, 125993, Russian Federation*

(received: February 20, 2021; accepted: March 12, 2021)

Time Series Forecasting has always been a very important area of research in many domains because many different types of data are stored as time series. Given the growing availability of data and computing power in the recent years, Deep Learning has become a fundamental part of the new generation of Time Series Forecasting models, obtaining excellent results.

As different time series problems are studied in many different fields, a large number of new architectures have been developed in recent years. This has also been simplified by the growing availability of open source frameworks, which make the development of new custom network components easier and faster.

In this paper three different Deep Learning Architecture for Time Series Forecasting are presented: Recurrent Neural Networks (RNNs), that are the most classical and used architecture for Time Series Forecasting problems; Long Short-Term Memory (LSTM), that are an evolution of RNNs developed in order to overcome the vanishing gradient problem; Gated Recurrent Unit (GRU), that are another evolution of RNNs, similar to LSTM.

The article is devoted to modeling and forecasting the cost of international air transportation in a pandemic using deep learning methods. The author builds time series models of the American Airlines (AAL) stock prices for a selected period using LSTM, GRU, RNN recurrent neural networks models and compare the accuracy forecast results.

**Key words and phrases:** neural networks, financial forecasting, deep learning, international air travel

## 1. Introduction

In 2020, there was a significant drop in quotations of American Airlines (AAL) associated with the COVID-19 pandemic and a record-breaking decrease in the number of air travel in the world. The generally accepted econometric methods of modeling and forecasting financial time series in these conditions turned out to be ineffective for making even short-term forecasts [1], [2]. In the present paper, methods for modeling and forecasting

international air traffic in the 2019–2020 pandemic are explored using recurrent neural networks with different architectures. As an object of research, the day quotes of the American company AAL, traded on the NASDAQ exchange, were selected; data from September 27, 2005 to September 30, 2020 from the information portal Yahoo Finance [3] were taken. The shares of this US company were selected due to its leading positions in the international air transportation market, high values of the trading turnover on the NASDAQ exchange, which in turn provides liquidity and shows investor interest in this exchange commodity [4]. Using the example of the value of AAL shares, we will try to build a reliable forecast using deep learning methods, in particular, recurrent neural networks [5]–[7].

## 2. Pre-processing of input data

As input data for the neural network model, we will take a sequence consisting of the following values:

— Opent-1 — opening price for the previous period;
— Lowt-1 — the minimum price for the previous trading day;
— Hightt-1 — the maximum price for the previous trading day;
— Volumet-1 — the amount of shares sold and bought for the previous trading day;
— Closet-1 — closing price for the previous trading day.

Based on the input data, neural networks will generate an output value that can be interpreted as the predicted value of the closing quotation today. For the correct operation of neural networks, it is necessary to normalize the data within the limits of $[0 : 1]$, as well as create training and test samples in the ratio 80:20 from the initial data having the dimension 3636. Thus, 2909 observations for the training sample and 727 observations for the test sample were obtained. The table 1 shows a fragment of the input data.

It is necessary to remove the Date and Adj Close columns from the received data. The table 2 presents descriptive statistics of input data. It is seen that the average closing price is \$27.13 and the standard deviation is \$16.74.

To study the statistical properties of the data further, let us build scatter diagrams of the profitability of the opening price and the closing price, as well as the profitability of the closing price shifted by one lag, and the closing price today. To calculate the profitability, we will use the following formula [8]–[10]:

$$R = \frac{y_t}{y_{t-1}} - 1, \tag{1}$$

where $R$ is the profitability; $y_{t-1}$ is the previous observation value; $y_t$ are the values for the current time period.

The scatter diagram of the profitability of the opening and closing prices is shown in the figure 1.

The figure 1 shows that there is no correlation between the variables under consideration. Next, we will construct a histogram of the distribution of the profitability of closing prices (figure 2). Obviously, most of the observations are in the range from $-0.1$ to 0.1. This means that in most observations, the price changed from $-10\%$ to $10\%$ in one period.

Table 1

A fragment of the input data

| Observation number | Date | Open, $ | High, $ | Low, $ | Close, $ | Adj Close[1], $ | Volume, $ |
|---|---|---|---|---|---|---|---|
| 0 | 27.09.2005 | 21.05 | 21.4 | 19.1 | 19.3 | 18.19 | 961200 |
| 1 | 28.09.2005 | 19.3 | 20.53 | 19.2 | 20.5 | 19.33 | 5747900 |
| 2 | 29.09.2005 | 20.4 | 20.58 | 20.1 | 20.21 | 19.05 | 1078200 |
| 3 | 30.09.2005 | 20.26 | 21.05 | 20.18 | 21.01 | 19.81 | 3123300 |
| 4 | 3.10.2005 | 20.9 | 21.75 | 20.9 | 21.5 | 20.27 | 1057900 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 3634 | 6.03.2020 | 15.02 | 17.12 | 14.8 | 15.97 | 15.97 | 54505000 |
| 3635 | 9.03.2020 | 14.87 | 15.79 | 14.46 | 14.75 | 14.75 | 42558000 |
| 3636 | 10.03.2020 | 15.82 | 17.67 | 14.61 | 17 | 17 | 56858200 |

Table 2

Descriptive statistics of input data

| Indicators | Open | High | Low | Close | Volume |
|---|---|---|---|---|---|
| Total number of observations | 3637 | 3637 | 3637 | 3637 | 3.637e+03 |
| Mean value, $ | 27.15 | 27.64 | 26.64 | 27.13 | 7.603118e+06 |
| Standard deviation, $ | 16.74 | 16.95 | 16.53 | 16.74 | 6.070650e+06 |
| Minimal value, $ | 1.81 | 2.03 | 1.45 | 1.76 | 1.385e+05 |
| 25% percentile, $ | 9.57 | 9.81 | 9.32 | 9.58 | 4.1782e+06 |
| 50% percentile, $ | 29.9 | 30.48 | 29.28 | 29.89 | 6.5025e+06 |
| 75% percentile, $ | 41.74 | 42.24 | 41.02 | 41.68 | 9.5455e+06 |
| Maximum value, $ | 62.7 | 63.27 | 62 | 62.95 | 1.377672e+08 |

To test the hypothesis about whether the distribution of the closing price profitability is a special case of the normal distribution, we use the Shapiro–Wilk and Jarque–Bera tests. The Jarque–Bera test rejected the null hypothesis at a significance level of $a = 0.05$. The results of the Shapiro–Wilk test and the Jarque–Bera test coincided. This means that the profitability of closing prices has a distribution that is different from the normal one.

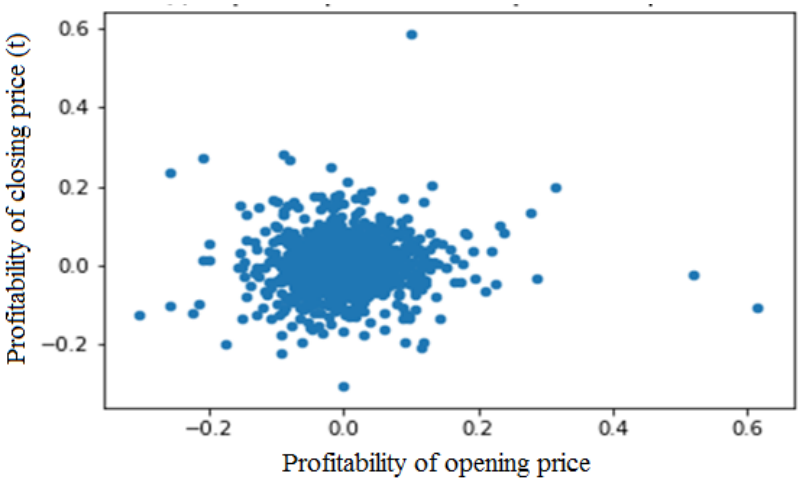


Figure 1. Scatter diagram of opening and closing prices

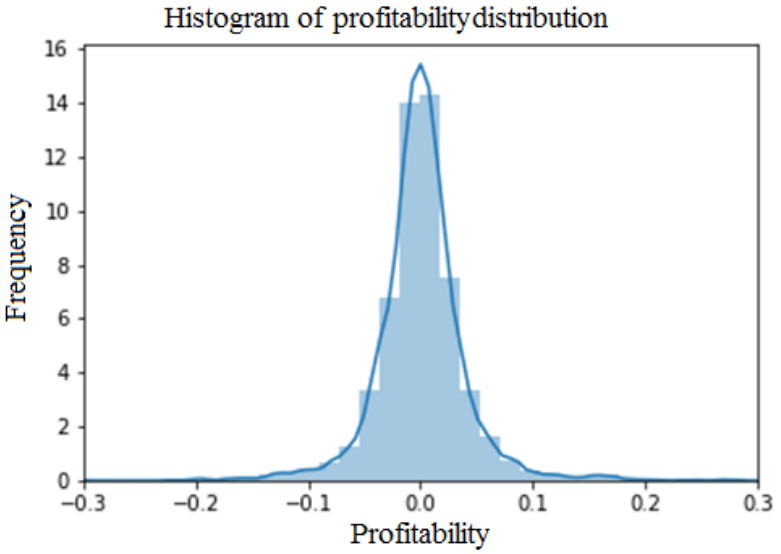


Figure 2. Distribution of closing price profitability

To check the stationarity of the profitability series, we will use the Dickey–Fuller test, which is one of the unit root tests. A time series has a unit root if its first differences form a stationary series, i.e. a series whose properties do not change over time. This condition is written as $y_t \sim I(1)$ if the series of the first differences $\Delta y_t = y_t - y_{t-1}$ is a stationary series $\Delta y \sim I(0)$ [11]. If the time series has a unit root, then it is not a stationary time series, but an integrated first-order time series [12]–[14]. As one would expect, the observed time series has no unit roots and, therefore, is stationary. For the convenience of using the input data, we will normalize them. The results are presented in the table 3.

Table 3

Normalized raw data

| Number of observation | Open | High | Low | Close | Volume |
|---|---|---|---|---|---|
| 0 | 0.31598 | 0.316297 | 0.291495 | 0.286648 | 0.005978 |
| 1 | 0.287239 | 0.30209 | 0.293146 | 0.306259 | 0.040757 |
| 2 | 0.305305 | 0.302907 | 0.30801 | 0.30152 | 0.006828 |
| 3 | 0.303005 | 0.310581 | 0.309331 | 0.314594 | 0.021687 |
| 4 | 0.313516 | 0.322012 | 0.321222 | 0.322602 | 0.00668 |
| ... | ... | ... | ... | ... | ... |
| 3634 | 0.216949 | 0.246408 | 0.220479 | 0.232227 | 0.395023 |
| 3635 | 0.214485 | 0.22469 | 0.214864 | 0.21229 | 0.308217 |
| 3636 | 0.230087 | 0.255389 | 0.217341 | 0.24906 | 0.412121 |

Next, we turn to the description of the main models of recurrent neural networks and their application in the analysis of financial time series.

## 3. Basic models of deep neural networks for simulation of financial time series

### 3.1. Basic recurrent neural network

The architecture of the proposed basic recurrent neural net (RNN) is as follows. A matrix with a dimension of 1 by 5 is fed to the input of the neural network, then the values are transferred to a recurrent layer with 25 neurons, after which the operation is repeated and the values are again fed to the recurrent layer with 25 neurons. At the penultimate step, the values are transferred to an aggregating layer with a dimension of 5 neurons, the result is displayed as a predicted value. Hidden layers have a hyperbolic tangent as an activation function. This activation function is nonlinear, which allows layers to be linked, i.e. combines them, because the combination of non-linear functions is also a non-linear function. Another advantage of the hyperbolic tangent function is that it is a smooth function, and this function is not binary and takes values in the range $(-1, 1)$, which eliminates overloading from large values. The hyperbolic tangent is very similar to the sigmoid with the difference that it has a larger gradient than the sigmoid. On the aggregate layer, a linear function is used as the activation function. The proposed neural network model, all procedures for its training and testing were implemented in the Keras library of the Python programming language [15].

The mean squared error (MSE) will be used as the loss function, and the optimization is performed using the Adam algorithm. The epoch parameter of the fit function reflects how many times the sample is passed through the neural network, in this case epoch = 150. The batch_size parameter is responsible for the size of the so-called batch. In cases where the training sample is too large, there is a need to divide it into parts. These parts are called batches. Thus, the training set with 2109 observations is divided into 210 batches with a size of 10, except for the last one with 9 observations. Thus, 210 iterations were required to pass one epoch.

Due to the tendency of recurrent neural networks to overfit, it is necessary to apply various regularization algorithms [10], [16]. As such an algorithm, the early stop method is used, which tracks the amount of losses. If during 20 epochs the improvement is less than 0.000002, then the training of the model will be stopped. The graph of the loss function on the training sample is shown in the figure 3.

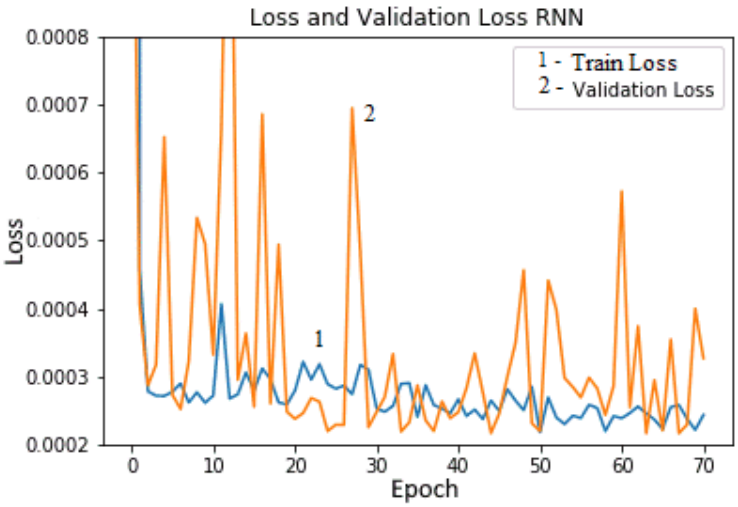


Figure 3. Plot of the RNN learning loss function:
1 — train loss; 2 — validation loss

After checking and training the neural network, we will construct a forecast of closing prices for the test sample. For a better visual appearance, the predicted values are shifted ten units up. Let us display the forecast of the last 50 observations of the test sample for a more accurate visual examination (figure 4). It can be seen from the figure that the neural network predicts closing prices closely enough.

### 3.2. Neural network with a gated recurrent unit

A recurrent neural network based on a cell architecture with a gated recurrent unit (GRU) repeats the structure of the RNN model of a recurrent network. The input layer takes the values of a matrix with a dimension of 1 by 5. Then, recurrent layers with 25 neurons and a hyperbolic tangent as an activation function are sequentially accepted and processed.

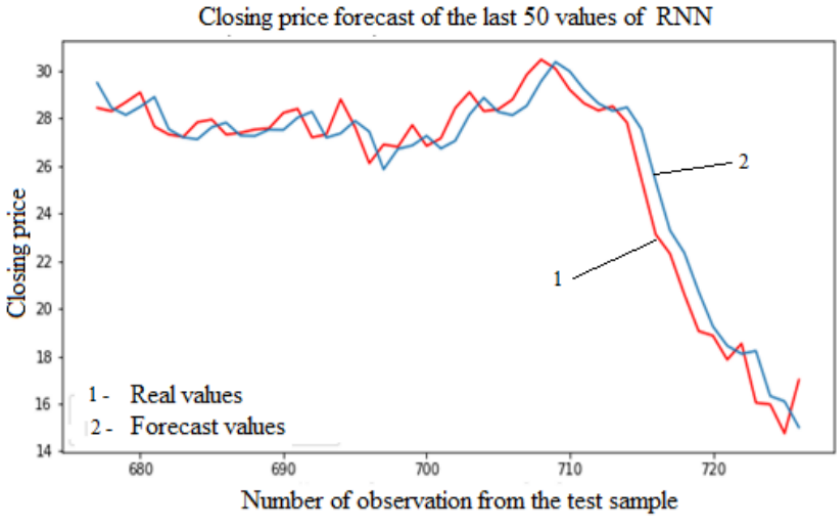


Figure 4. Forecast of the closing price for the last 50 values of RNN model network:
1 — real values stock price; 2 — forecast price

The aggregating layer has 5 neurons with a linear activation function. After processing by the last layer, the predicted value is supplied. It should be noted that the default activation function for layers with the GRU architecture is the hyperbolic tangent [16], [17]. The loss plot for the GRU recurrent neural network is shown in the figure 5.

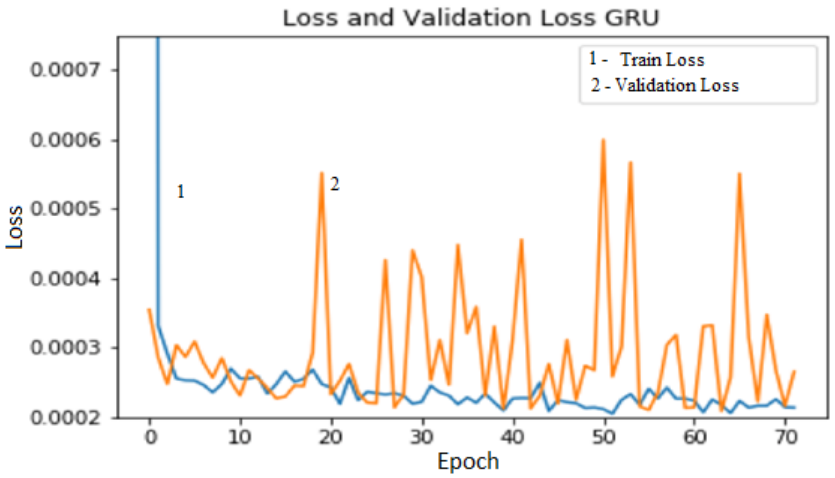


Figure 5. Loss plot for GRU model network:
1 — train loss function; 2 — validation loss function

The early stop regularization terminated the training of the neural network to prevent overfitting at epoch 72. The plot of predicted closing prices for all observations of the model of a recurrent neural network with the GRU architecture is shown in the figure 6. As in the case of the RNN, to facilitate

visualization the predicted values have been shifted ten units upward. It is also worth noting that the neural network accurately reproduced the closing price behavior. For a detailed consideration, we take the last 50 values of the test sample and display them in the figure 7.

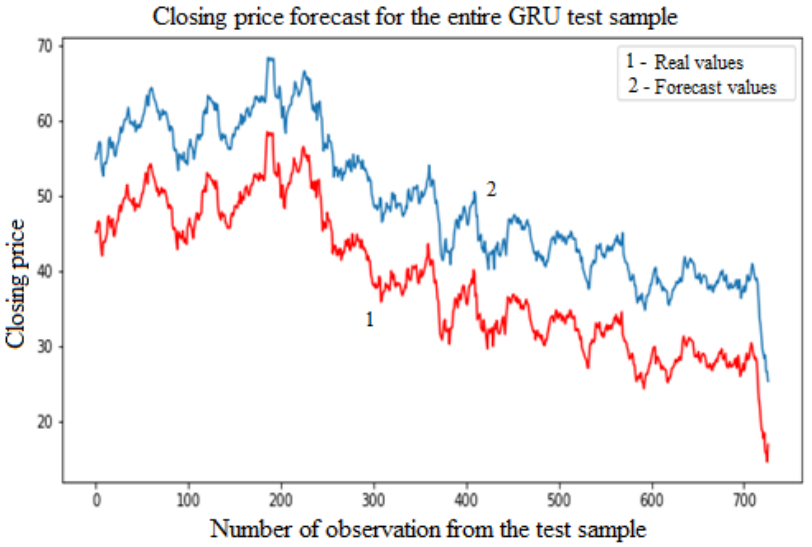


Figure 6. Closing price forecast for the entire GRU test sample:
1 — real values stock price; 2 — forecast stock price

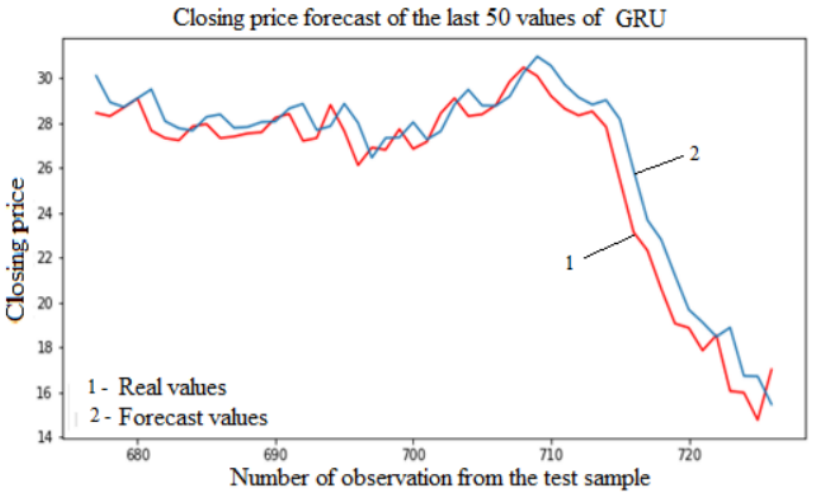


Figure 7. Closing price forecast for the last 50 values of GRU:
1 — real values of stock prices; 2 — forecast stock prices

The mean square forecast error and the R2 index have the following values: MSE = 0.9953, $R^2 = 0.9885$.

### 3.3. Neuron network with long short-term memory (LSTM)

Just like the previous networks, constructively recurrent neural network with long short-term memory (LSTM) will repeat the previous values. An input that accepts a 1-by-5 matrix transmits information to two recurrent layers with 25 neurons per layer and a hyperbolic tangent as an activation function. Then an aggregating layer of five neurons with a linear activation function passes the value to the output layer.

The closing price prediction plot calculated using a recurrent neural network with the LSTM architecture is shown in the figure 8.

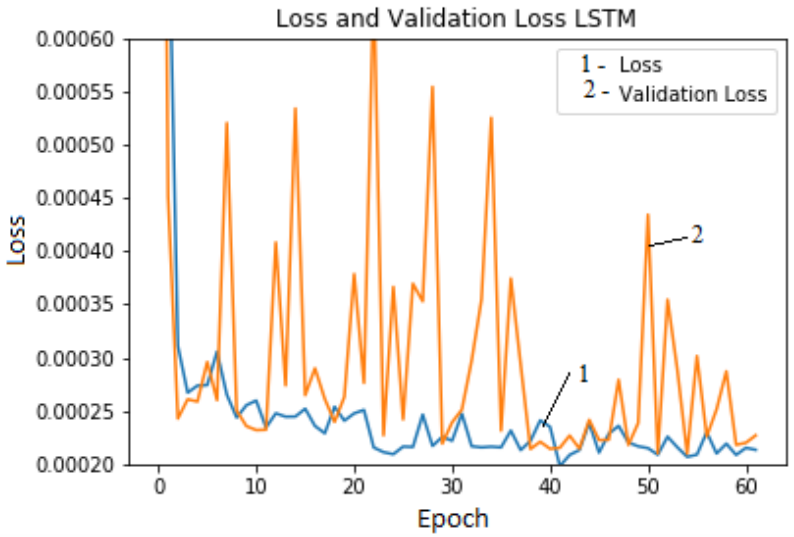


Figure 8. LSTM model network loss plot:
1 — train loss function; 2 — validation loss function

Forecasted values are shifted ten points. Based on the plot, we can conclude that the neural network under consideration predicts the required values quite accurately.

The forecast of the closing price for the entire LSTM test sample and for the last 50 values is shown in figures 9 and 10 respectively. For this recurrent neural network, MSE = 0.8508, $R^2 = 0.99$.

Let us display a comparative plot of losses during training of various constructions and architectures of the considered neural networks (figure 11). Note that the RNN recurrent neural network demonstrated the highest loss rates on the training set. Except for separately taken random epochs, its loss value was greater than that of the rest. LSTM and GRU recurrent neural networks have close values of losses on the training set. It is worth noting that the early stopping algorithm worked for all types of recurrent neural networks. For the RNN model, the algorithm stopped training at 71 epochs, for GRU — at 72. The least number of epochs – 62 — was required to train the neural network built using the LSTM architecture.

The table 4 shows the values of the mean square error and the coefficient $R^2$ for all constructed neural networks.

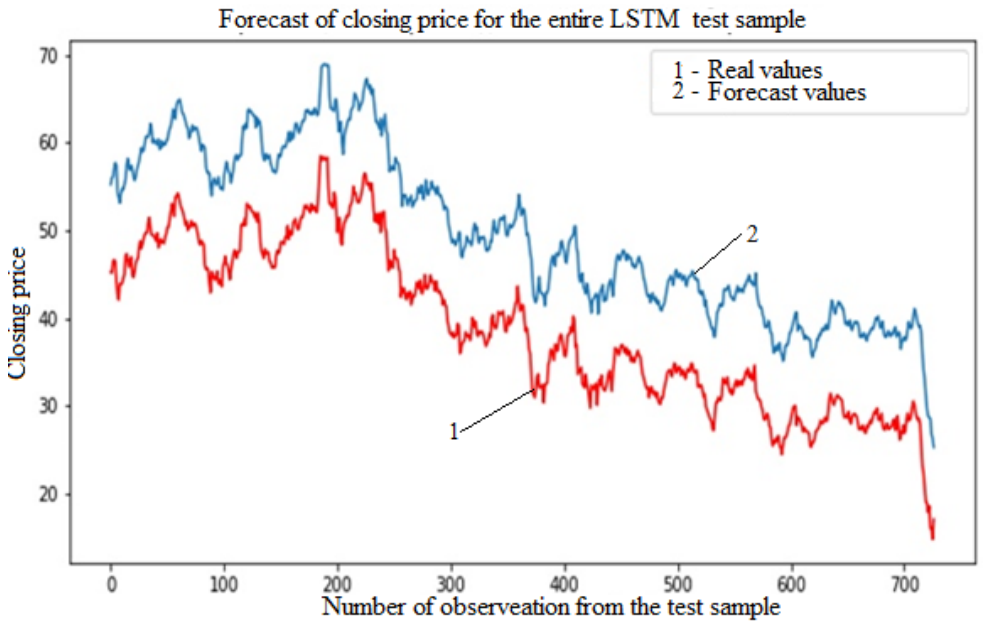


Figure 9. Forecast of the closing price for the entire LSTM model network test sample: 1 — Real values stock prices; 2 — forecast stock prices

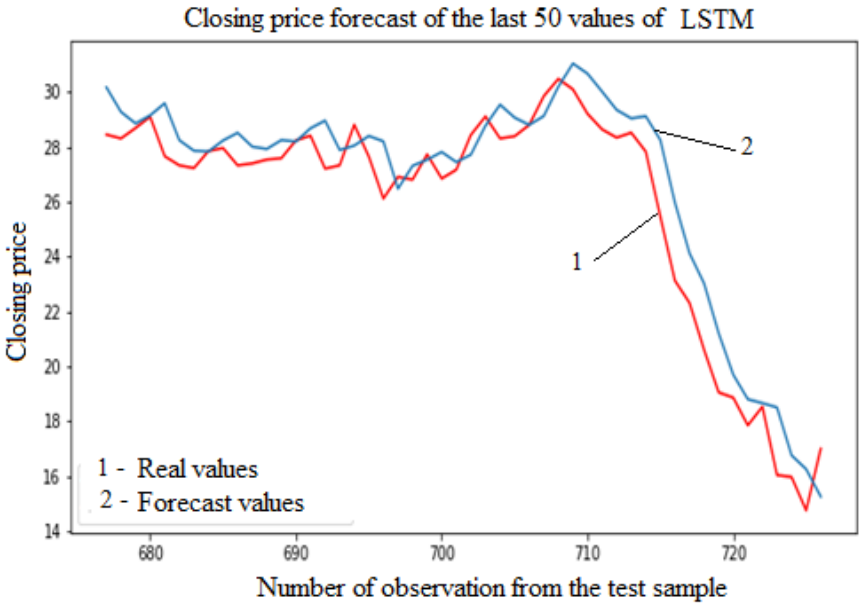


Figure 10. Forecast of the closing price of the last 50 values of LSTM model network: 1 — Real values stock prices; 2 — forecast stock prices

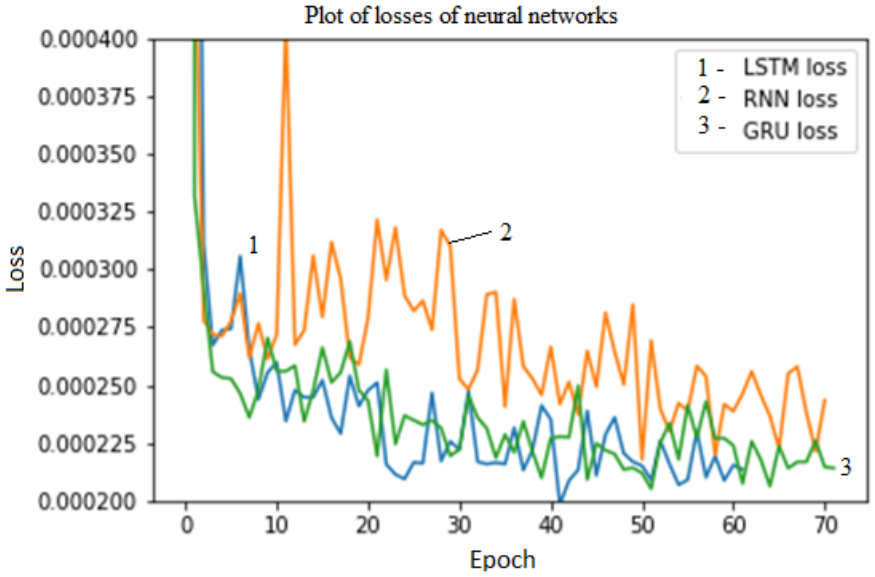


Figure 11. Plot of losses for different models of neural networks: 1 — LSTM model loss function; 2 — RNN model loss function; 3 — GRU model loss function

Table 4

Values of MSE и $R^2$ for all constructed neural networks

| Neural network | MSE | $R^2$ |
|---|---|---|
| RNN | 1.2232 | 0.9858 |
| GRU | 0.9953 | 0.9885 |
| LSTM | 0.8508 | 0.9901 |

## 4. Discussion of results of computer experiments

In the process of investigating the impact of the COVID-19 pandemic on AAL stock quotes, recurrent neural network models were built with various architectures, such as cells with long short-term memory LSTMs, cells with gated recurrent unit GRU, and a basic recurrent network. The analysis of the constructed models was carried out, as well as the comparison of the results on the training and test data. During the analysis, it was found that the neural network with long short-term memory cells (LSTM) coped best with the task of predicting the data under study.

Summing up, we can say that all networks have shown a satisfactory result, but they predict the price with a certain delay, which may entail unplanned financial losses. In view of this, it can be concluded that these models are not suitable for carrying out short-term operations in the financial market,

are not able to serve as an indicator that helps to improve the efficiency of a trading strategy and cannot be used for risk management tasks.

## 5. Conclusion

The purpose of the article was to investigate the quality of various neural network models that predict the closing price of a stock. In the course of the study, sufficiently accurate results of modeling and forecasting financial time series for the intraday closing prices of shares of the American airline ALL were obtained, which confirmed the effectiveness of using the proposed models of deep neural networks. However, in the context of the practical application of the developed models, it is necessary to take into account time delays in obtaining forecast results, as well as the horizon of financial forecasting.

## References

[1] J. D. A. Hamilton, *The time series analysis*. Princeton New Jersey: University Press, 1994.

[2] C. Brooks, *Introductory econometrics for finance*. Cambridge: Cambridge University Press, 2019.

[3] "American Airlines Group Inc. (AAL)," URL: `https://finance.yahoo.com/quote/AAL/`. Availabel: 2020-11-25, 2020.

[4] E. Y. Shchetinin, "On a structural approach to managing a company with high volatility of indicators [K analizu effektivnosti biznesa v usloviyah vysokoj izmenchivosti ego finansovyh aktivov]," *Finansy i kredit*, vol. 14, no. 218, pp. 39–41, 2006, [in Russian].

[5] J. Vander Plas, *Python Data Science Handbook*. Sebastopol, CA: O'Reilly Media, 2016.

[6] W. Richert and L. P. Coelho, *Building Machine Learning Systems with Python*. Birmingham: Packt, 2013.

[7] C. Bishop, *Pattern recognition and machine learning*. Berlin, Germany: Springer-Verlag, 2006.

[8] E. Y. Shchetinin, "Modeling the energy consumption of smart buildings using artificial intelligence," in *CEUR Workshop Proceedings*, vol. 2407, 2019, pp. 130–140.

[9] M. Mudelsee, "Trend analysis of climate time series: A review of methods," *Earth-Science Reviews*, vol. 190, pp. 310–322, 2019. DOI: `10.1016/j.earscirev.2018.12.005`.

[10] C. Chen, J. Twycross, and J. M. Garibaldi, "A new accuracy measure based on bounded relative error for time series forecasting," *PLOS ONE*, vol. 12, no. 3, pp. 1–23, Mar. 2017. DOI: `10.1371/journal.pone.0174202`.

[11] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. Melbourne, Australia: OTexts, 2018.

[12] A. Ghaderi, B. M. Sanandaji, and F. Ghaderi. "Deep forecast: deep learning-based spatio-temporal forecasting." arXiv: 1707.08110 [cs.LG]. (2017).

[13] S. B. Taieb, A. Sorjamaa, and G. Bontempi, "Multiple-output modeling for multi-step-ahead time series forecasting," *Neurocomput*, vol. 73, no. 10, pp. 1950–1957, 2010. DOI: 10.1016/j.neucom.2009.11.030.

[14] R. Sen, H.-F. Yu, and I. S. Dhillon, "Think globally, act locally: a deep neural network approach to high-dimensional time series forecasting," in *Advances in neural information processing systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32, Curran Associates, Inc., 2019.

[15] "Keras," URL: https://www.keras.io. Availabel: 2020-11-25, 2020.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Cambridge: The MIT Press, 2016.

[17] S. Galeshchuk and S. Mukherjee, "Deep networks for predicting direction of change in foreign exchange rates," *Intelligent Systems in Accounting, Finance and Management*, vol. 24, no. 4, pp. 100–110, 2017. DOI: 10.1002/isaf.1404.

**For citation:**

E. Yu. Shchetinin, Study of the impact of the COVID-19 pandemic on international air transportation, Discrete and Continuous Models and Applied Computational Science 29 (1) (2021) 22–35. DOI: 10.22363/2658-4670-2021-29-1-22-35.

**Information about the authors**:

**Shchetinin, Eugeny Yu.** — Doctor of Physical and Mathematical Sciences, Lecturer of Department of Mathematics (e-mail: riviera-molto@mail.ru, ORCID: https://orcid.org/0000-0003-3651-7629)

УДК 519.6
DOI: 10.22363/2658-4670-2021-29-1-22-35

# Исследование влияния пандемии COVID-19 на международные авиаперевозки

**Е. Ю. Щетинин**

*Финансовый университет при Правительстве Российской Федерации*
*Ленинградский проспект, д. 49, Москва, 125993, Россия*

Прогнозирование временных рядов играет важную роль во многих областях исследований. Вследствие растущей доступности данных и вычислительных мощностей в последние годы глубокое обучение стало фундаментальной частью нового поколения моделей прогнозирования временных рядов, получающих отличные результаты.

В данной работе представлены три различные архитектуры глубокого обучения для прогнозирования временных рядов: рекуррентные нейронные сети (RNN), которые являются наиболее известной и используемой архитектурой для задач прогнозирования временных рядов; долгая краткосрочная память (LSTM), которая представляет собой обобщённую и развитую РНС, разработанную для преодоления проблемы исчезающего градиента; закрытый рекуррентный блок (GRU), который является ещё одной эволюционной моделью РНС.

Статья посвящена моделированию и прогнозированию стоимости международных авиаперевозок в условиях пандемии с использованием методов глубокого обучения и моделей рекуррентных сетей. В работе построены модели временных рядов цен акций American Airlines (AAL) с использованием моделей рекуррентных нейронных сетей LSTM, GRU, RNN и проведён сравнительный анализ результатов точности прогноза на выбранный период. Его результаты показали эффективность применения алгоритмов глубокого обучения для оценивания точности прогнозирования временных рядов.

**Ключевые слова:** нейронные сети, финансовое прогнозирование, глубокое обучение, международные авиаперевозки

*Research article*

UDC 519.872, 519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2021-29-1-36-52

# A simulator for analyzing a network slicing policy with SLA-based performance isolation of slices

**Nikita A. Polyakov**[1],
**Natalia V. Yarkina**[1], **Konstantin E. Samouylov**[1, 2]

[1] *Peoples' Friendship University of Russia (RUDN University)*
*6, Miklukho-Maklaya St., Moscow, 117198, Russian Federation*
[2] *Federal Research Center "Computer Science and Control"*
*of the Russian Academy of Sciences (FRC CSC RAS)*
*44-2, Vavilov St., Moscow, 119333, Russian Federation*

(received: February 24, 2021; accepted: March 12, 2021)

Efficient allocation of radio access network (RAN) resources remains an important challenge with the introduction of 5G networks. RAN virtualization and division into logical subnetworks – slices – puts this task into a new perspective. In the paper we present a software tool based on the OMNeT++ platform and developed for performance analysis of a network slicing policy with SLA-based slice performance isolation. The tool is designed using the object-oriented approach, which provides flexibility and extensibility of the simulation model. The paper briefly presents the slicing policy under study and focuses on the simulator's architecture and design. Numerical results are provided for illustration.

**Key words and phrases:** queuing system, resource allocation, network slicing, simulation, optimization

## 1. Introduction

Network slicing is a key next-generation networking technology that allows multiple virtual subnetworks to be built over a shared physical infrastructure. The virtual subnetworks are then configured to meet the specific needs of applications, services, devices, customers, or virtual network operators. This approach makes it possible to implement in practice flexible configuration and infrastructure management, which make part of the requirements for new generation networks [1]. This concept allows the infrastructure provider to lease network slices to tenants. These relationships are governed by the Service Level Agreements (SLA). Efficient use of network bandwidth and adherence to the terms of these agreements provides economic benefits to all parties. Guaranteeing slice isolation when allocating RAN radio resources makes the problem of efficient resource allocation even more challenging.

The emerging fifth generation (5G) telecommunication networks are envisioned to offer a large number of end-to-end network services for various applications. These stem not only from traditional mobile services, but also from vertical market segments such as automatic driving, unmanned aerial vehicles, telemedicine, massive Internet of Things (mIoT), etc. To provide services with so different requirements for the quality of service (QoS), it is crucial to be able to implement specific virtual subnetworks by using network slicing, since fourth generation (4G) networks with their one-fits-all paradigm are no longer fitted for the task [2], [3].

In this paper, we propose a simulation model as a reusable, versatile tool for evaluating slicing policies for next-generation network resource sharing. The rest of the article is structured as follows. Section 2 presents the system model. In Section 3 we briefly present the slicing policy under study, which was initially proposed by the authors in [4]. Further, it is considered in terms of queuing theory in Section 4. Section 5 explains the architecture of the simulator. The experimental results are discussed in Section 6. Finally, in Section 7, conclusions are drawn and future work is outlined.

## 2. System model and notation

Following [4], [5], we consider the downlink transmission of a 5G base station (BS) with a virtualized RAN and network slicing. We assume that there are $S$ instantiated slices at the BS and denote their set by $\mathscr{S}$, $|\mathscr{S}| = S$. Let $C_{s[Gbps]} \geqslant 0$ denote the capacity of slice $s \in \mathscr{S}$, so that

$$\sum_{s \in \mathscr{S}} C_s \leqslant C, \tag{1}$$

where $C_{[Gbps]}$ is the total BS capacity. Let $N_s$ denote the number of users in slice $s \in \mathscr{S}$, and let $\mathbf{N} = (N_s)_{s \in \mathscr{S}}$. We assume that each slice is intended for one type of services (e.g., for video streaming, video conferencing, gaming, file transfer, web browsing), and hence the traffic in each slice is homogeneous in terms of characteristics and QoS requirements. Let $R_{s[Gbps]}$ denote the average user data rate in slice $s$, i.e.,

$$R_s = \frac{C_s}{N_s}, \ s \in \mathscr{S}. \tag{2}$$

The column vector of data rates is denoted by $\mathbf{R}_{[S \times 1]} = (R_s)_{s \in \mathscr{S}}$.

It is assumed that the infrastructure provider (InP) leases parts of its infrastructure in the form of slices to tenants. A Service Level Agreement (SLA) between the InP and the tenant includes the following slice characteristics:

— a minimum average user data rate $0 < R_s^{\min} \leqslant R_s$,
— a maximum average user data rate $R_s \leqslant R_s^{\max} \leqslant C$,
— a guaranteed capacity share $\gamma_s$ or contracted number of users $N_s^{\text{cont}}$.

We assume that performance isolation of slice $s$ is provided as long as

$$N_s \leqslant N_s^{\text{cont}}, \text{ or equivalently, } \frac{N_s R_s^{\min}}{C} \leqslant \gamma_s, \quad 0 \leqslant \gamma_s \leqslant 1. \tag{3}$$

By performance isolation we understand that traffic fluctuation in one slice does not negatively affect performance in other slices.

## 3. Slicing scheme

The calculation of slice capacities is performed according to the slicing scheme with SLA-based isolation [4].

Let us partition $\Omega = \mathbb{N}^S$ as

$$\Omega = \Omega^{\max} \cup \Omega^{\text{opt}} \cup \Omega^{\text{cong}}. \tag{4}$$

Now, for $\mathbf{N} \in \Omega^{\max} \stackrel{\text{def}}{=} \{\mathbf{N} \in \Omega : \mathbf{N}\mathbf{R}^{\max} \leqslant C\}$ we set

$$R_s(\mathbf{N}) = R_s^{\max}, \ s \in \mathscr{S} \quad \Longrightarrow \quad C_s(\mathbf{N}) = N_s R_s^{\max}, \ s \in \mathscr{S}, \mathbf{N} \in \Omega^{\max}. \tag{5}$$

For $\mathbf{N} \in \Omega^{\text{opt}} \stackrel{\text{def}}{=} \{\mathbf{N} \in \Omega : \mathbf{N}\mathbf{R}^{\min} \leqslant C < \mathbf{N}\mathbf{R}^{\max}\}$ we determine the data rates as the solution to the convex programming problem

$$\text{maximize } U(\mathbf{R}) = \sum_{s \in \mathscr{S}} W_s(N_s) N_s \ln(R_s), \tag{6}$$

$$\text{subject to } \mathbf{N}\mathbf{R} = C, \tag{7}$$

$$\text{over } \mathbf{R} \in \mathbb{R}_+^S : R_s^{\min} \leqslant R_s \leqslant R_s^{\max}, \tag{8}$$

where $W_s(N_s)$ is given by

$$W_s(N_s) = \begin{cases} 1, & N_s \leqslant N_s^{\text{cont}} \\ N_s^{\text{cont}}/N_s, & N_s > N_s^{\text{cont}} \end{cases} \tag{9}$$

The objective function (6) is differentiable and strictly concave by assumption and the feasible region (7), (8) is compact and convex, there exists hence a unique maximum for the data rate vector $R_s$, which can be found by Lagrangian methods.

Now consider $\mathbf{N} \in \Omega^{\text{cong}} \stackrel{\text{def}}{=} \{\mathbf{N} \in \Omega : \mathbf{N}\mathbf{R}^{\min} > C\}$. Denote $N_s^{\min}(\mathbf{N}) \stackrel{\text{def}}{=} \min\{N_s, N_s^{\text{cont}}\}_{s \in \mathscr{S}}$. Thus $\mathbf{N}^{\min}\mathbf{R}^{\min}$ is a due capacity. If $\mathbf{N}^{\min}\mathbf{R}^{\min} \geqslant C$, we set

$$C_s(\mathbf{N}) = \frac{N_s^{\min} R_s^{\min}}{\mathbf{N}^{\min}\mathbf{R}^{\min}} C. \tag{10}$$

If, conversely, $\mathbf{N}^{\min}\mathbf{R}^{\min} < C$, then

$$C_s(\mathbf{N}) = N_s^{\min} R_s^{\min} + \frac{(N_s - N_s^{\min}) R_s^{\min}}{(\mathbf{N} - \mathbf{N}^{\min})\mathbf{R}^{\min}} (C - \mathbf{N}^{\min}\mathbf{R}^{\min}). \tag{11}$$

To solve the problem (6)–(8) numerically, we use the gradient projection method (Algorithm 1).

**Algorithm 1:** Numerical solution of (6)–(8) using the Gradient Projection Method

**input** : $C, S, \mathbf{N}, \mathbf{R}^{\min}, \mathbf{R}^{\max}, \mathbf{N}^{cont}$
**output**: $\mathbf{R}$

initialization
$\mathbf{W} := [W_1(N_1), ..., W_S(N_S)]$
$\mathbf{X^{stat}} := \mathbf{W}C(\mathbf{WN})^{-1}$ `// stationary point`
**if** $R_i^{\min} \leqslant X_i^{stat} \leqslant R_i^{\max}, i = \overline{1, S}$ **then**
  **return** $\mathbf{X^{stat}}$
$\mathbf{M}_{[1\times S]} := \mathbf{N}$
$\mathbf{P}_{[S\times S]} := \mathbf{I} - \mathbf{N}^T(\mathbf{NN}^T)^{-1}\mathbf{N}$
$\mathbf{X^0} := \mathbf{R}^{\min} + (C - \mathbf{NR}^{\min})(\mathbf{N}(\mathbf{R}^{\max} - \mathbf{R}^{\min}))^{-1}(\mathbf{R}^{\max} - \mathbf{R}^{\min})$
$\tau := ||\mathbf{X^0} - \mathbf{X^{stat}}||; \delta := 1$
**while** $\delta > 0.0001$ **do**
  $\mathbf{X^1} := \mathbf{X^0} + \tau\mathbf{P}$ `div(`$\mathbf{N}^T\mathbf{W}, \mathbf{X^0}$`)` `// div(A, B) is element-wise division of vector` $A$ `by` $B$
  $t_{\text{bound}} := 2; t_{\text{coord}} := -1; \delta_+ = 0$
  **for** $i = \overline{1, S}$ **do**
    **if** $N_i > 0$ **then**
      **if** $X_i^1 < R_i^{\min}$ **then**
        **if** $t_{bound} > (R_i^{\min} - X_i^0)(X_i^1 - X_i^0)^{-1}$ **then**
          $t_{\text{bound}} := (R_i^{\min} - X_i^0)(X_i^1 - X_i^0)^{-1}; t_{\text{coord}} := i$
      **if** $X_i^1 > R_i^{\max}$ **then**
        **if** $t_{bound} > (R_i^{\max} - X_i^0)(X_i^1 - X_i^0)^{-1}$ **then**
          $t_{\text{bound}} := (R_i^{\max} - X_i^0)(X_i^1 - X_i^0)^{-1}; t_{\text{coord}} := i$
  **if** $t_{bound} < 2$ **then**
    $\mathbf{X^1} := \mathbf{X^0} + t_{\text{bound}}(\mathbf{X^1} - \mathbf{X^0})$
    **if** *Row number of* $\mathbf{M} < S - 1$ **then**
      Add empty row to $\mathbf{M}$
      $\delta_+ := 1$
    Last row of $\mathbf{M} := \mathbf{I}[t_{\text{coord}}]$
    **if** $||\mathbf{MM}^T|| > 0.0000001$ **then**
      $\mathbf{P} := \mathbf{I} - \mathbf{M}^T(\mathbf{MM}^T)^{-1}\mathbf{M}$
  $\delta := \delta_+ + ||\mathbf{X^0} - \mathbf{X^1}||; \mathbf{X^0} := \mathbf{X^1}$
**return** $\mathbf{X^0}$

The gradient projection method is a well-known algorithm for solving optimization problems with linear constraints. It is specified by a standard

iterative procedure [6]: $\mathbf{X^{k+1}} = \mathbf{X^k} + \tau \mathbf{d^k}$, where $\mathbf{X^k}$ is the point at which the algorithm arrived at the $k$-th iteration, $\tau$ — the stepsize, $\mathbf{d^k}$ — the increment vector, which is found as the projection of the target function gradient on the constraints: $\mathbf{d^k} = P\nabla U(\mathbf{X^k})$, where the projection matrix is initially given by $P = \mathbf{I} - \mathbf{N}^T(\mathbf{N}\mathbf{N}^T)^{-1}\mathbf{N}$.

## 4. Queuing system model

We use queuing theory to model the system described in the Section 2. Each slice is modeled as a separate queuing systems (QS). The types of QS must be selected in such a way as to adequately reflect the nature of the service provided. Jobs in QSs correspond to user sessions in slices. Since in the system model the slices are part of a single network of a total capacity $C$, the $S$ queuing systems share a total resource (capacity) $C$, which is partitioned so that the resource share available to QS $s$ equals $C_s$.

At the moment, we have implemented a slice of a Best Effort (BE) type without admission control and with maximum user data rate, which we denote by $BE^{\max}$. It is represented by a QS with the EPS (egalitarian processor sharing) service discipline. The job service rate $R_s$ of all jobs is equal and inversely proportional to their number $N_s$, but cannot exceed $R_s^{\max}$. Serving jobs in such a QS can be interpreted as downloading files.

Network slicing from this perspective corresponds to a repeated redistribution (re-slicing) of the capacity $C$ among otherwise independent QSs. The considered model is shown in figure 1, where $A_s(x)$ is the distribution law of the interarrival times, $B_s(x)$ is the distribution law of the job lengths (service time on one resource unit) for $s \in \mathscr{S}$.

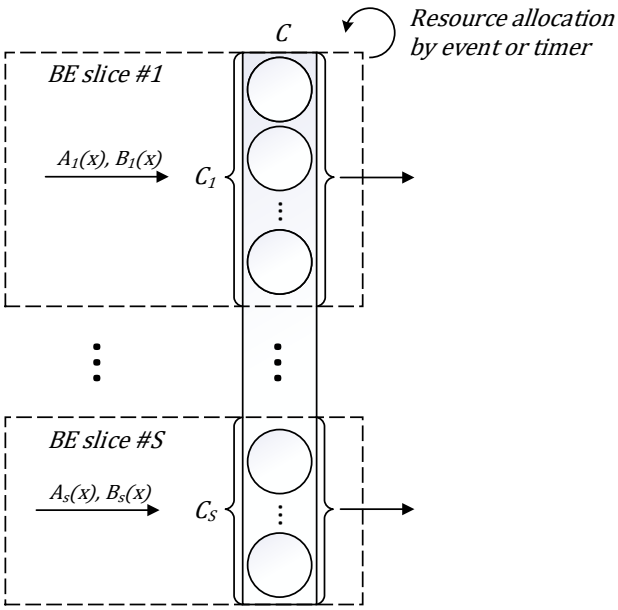


Figure 1. A system with S slices of type $BE^{\max}$

It should be noted that in our model admission control and resource allocation within a slice are individual characteristics for each type of slice. For the $BE^{\max}$ considered in this work, we assume the same service rate for all users (jobs) and unlimited admission (any number of jobs in service). Since this type of slice lacks admission control and queue, it makes sense to introduce a service level degradation threshold ($0 \leqslant R_s^d \leqslant C$) to assess the efficiency of the slicing scheme. This parameter sets the threshold for job service rate in the slice, below which degradation of service occurs, the service is provided poorly. Slice degradation can occur as a result of user arrival and/or redistribution of capacity.

For simplicity, in what follows, the terms slice and queuing system will be used interchangeably.

## 5. Simulator architecture

### 5.1. Modules

The discrete event simulator is written on the OMNeT++ platform using the queuinglib standard library. The implementation of the algorithm for solving the optimization problem for the slicing scheme required the inclusion of Boost library for operations with matrices. The construction of a simulation model in OMNeT++ assumes a modular structure, and also allows the use of both standard and modified modules (figure 2).

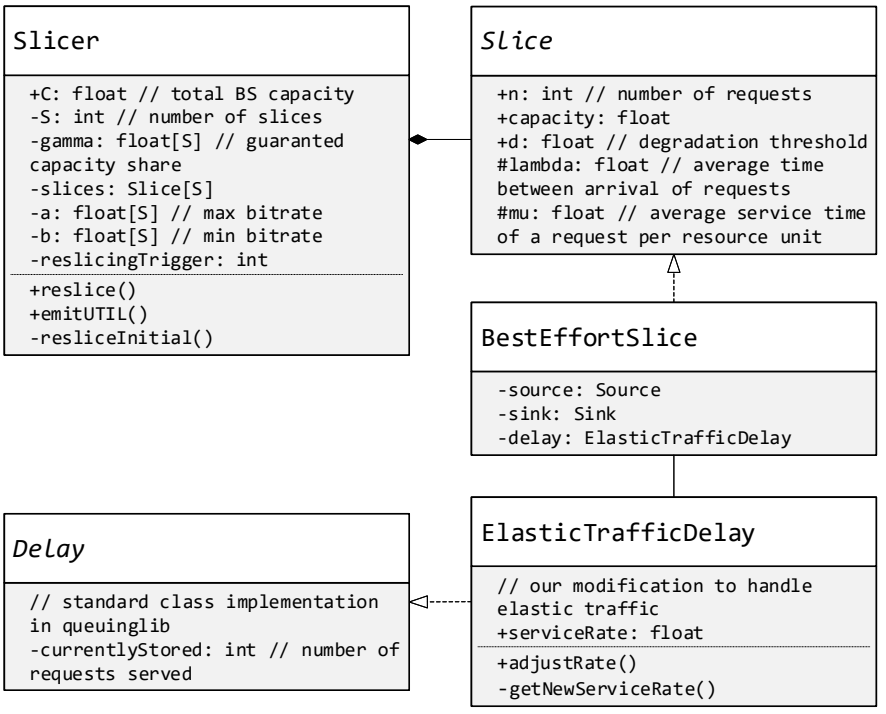


Figure 2. UML diagram of classes developed for the simulator based on the queuinglib standard library

To achieve the required level of abstraction, we have developed two modules:

— is a container consisting of simple modules inside that form a QS. For different types of slices, the way of servicing jobs (users), and as a consequence, the internal structure is not strictly defined and can vary greatly. However, all types of slices are inherited from a common ancestor, which defines the required external parameters that are used to receive the initial data of the model (table 1), and the characteristics that are passed to the slicer as re-slicing parameters, which collected to vectors: $\mathbf{N}, \mathbf{R}^{\min}, \mathbf{R}^{\max}$.

— is a simple module that handles requests for capacity re-slicing from slices. The slicer also performs initial re-slicing of the capacity by formula (12). Slices and slicers communicate via channels — standard OMNeT++ technology.

Let us take a closer look at the table 1. First, the structural characteristics of the model are determined, such as the number of slices and the type of each one. Further, the distribution laws for the arrival of requests and their service time are established, the parameters of the slices are selected, etc.

Table 1

Input data structure

| **Slicer** | |
|---|---|
| Total capacity $C$ | **float** $> 0$ |
| $S$ | **int** $> 0$ |
| Re-slicing trigger | { All events, Arrivals, Degradation, Timer, Static} |
| Timer interval $t_{timer}$ | **float** $> 0$ |
| **Slice** $i, i = \overline{1, S}$ | |
| $R_i^{\min}$ | $0 \leqslant$ **float** $\leqslant R_i^{\max}$ |
| $R_i^{\max}$ | $R_i^{\min} \leqslant$ **float** $\leqslant C$ |
| $R_i^d$ | $0 \leqslant$ **float** $\leqslant C$ |
| $\gamma_i$ | $0 \leqslant$ **float** $\leqslant 1$ |
| Distribution $A_i(x)$ | $\{U(a,b), Exp(\lambda), N(a,\sigma^2), \Gamma(\alpha,\beta),$ |
| Distribution $B_i(x)$ | $W(k,\lambda), Beta(\alpha,\beta), Cauchy(\theta), Pareto(\alpha), ...\}$ |

One of the parameters of the initial data is the way of invoking the re-slicing — this is an event or message that occurs periodically during the simulation, which is a condition for invoking the capacity re-allocation algorithm. We consider re-slicing triggered by

— events:
  – all events, i.e., job arrivals and departures (in our case this corresponds to optimal real-time slicing),
  – arrivals only,
  – degradation in any slice;

— timer (every $t_{timer}$ s);
— static slicing (no re-slicing, corresponds to complete partitioning), where the capacity of slice $i$ equals

$$C_i = \frac{\gamma_i}{\sum_{j=1}^{S} \gamma_j} C, i \in \mathscr{S}. \tag{12}$$

Consider the implementation of $BE^{\max}$ slice type. Figure 3 shows a diagram of the correspondence of the QS elements with software modules in a slice, which include the Delay modification — **ElasticTrafficDelay**, and the **Source** and **Sink** modules from the standard set provided by the OMNeT++ and queuinglib bundle.

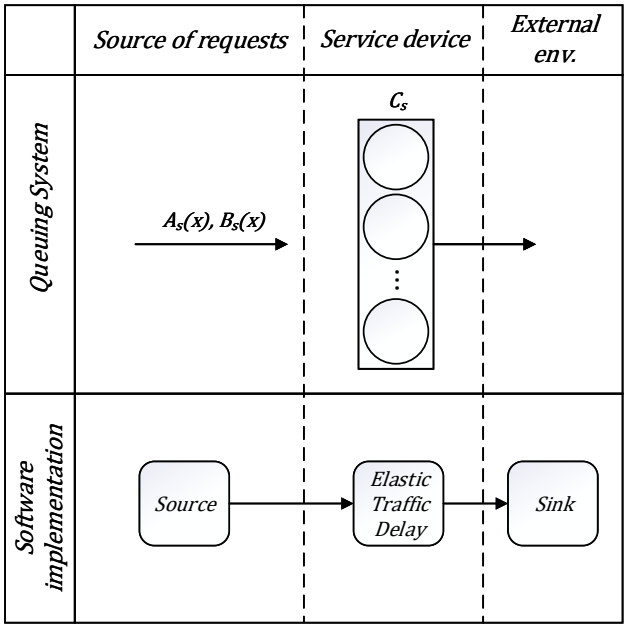


Figure 3. Scheme of logical correspondence of program classes with elements of the $BE^{\max}$ QS

Consider them:

— **Source** is a basic generator of requests that correspond to users' requests for the provision of a service, according to specified distributions.
— **Sink** is a module that receives serviced jobs and destroys them. The Sink collects all the primitive statistics on jobs, such as average, maximum, minimum time spent in the system, average time in queue, etc.
— **ElasticTrafficDelay** extension was written for the standard Delay module. This modification is intended to simulate the service of "elastic" traffic, as the name of the module implies. With the help of standard Delay, you can simulate the service of traffic on discrete devices: after the arrival, the job is in the system for a certain time, and then goes to the drain. ElasticTrafficDelay takes into account the presence of all jobs on the device and equally distributes the available resource between

them. Therefore, the standard module was extended with mechanisms for recalculating the service rate (13) and departure time (Algorithm 2):

$$R_i = \begin{cases} \min(\frac{C_i}{N_i}, R_i^{\max}), & N_i > 0, \\ 0, & N_i = 0, \end{cases} \quad i \in \mathscr{S}. \tag{13}$$

**Algorithm 2:** Service rate recalculation into $BE^{\max}$ slice $i$.

```
class Job {
    float t_arr   // is arrival time
    float t_dep   // is departure time
    ...
}
input: R_i, Job[N_i] jobs   // set of jobs in slice i
R_prev := R_i
R_i := getNewServiceRate()   // formula (13)
foreach job in jobs do
    Delete job from event queue
    // t_cur is model current time
    t_serv := |job.t_arr − t_cur|   // how much is already served
    t_serv^new := t_serv * R_prev / R_i
    job.t_dep := t_cur + t_serv^new   // set to job new service end time
    Add job in event queue
```

### 5.2. Simulation algorithm

Slices, in their essence, function independently of each other, however, as mentioned earlier, the simulator is built on a discrete-event basis, so there is a common queue of events. It contains all the events generated by the model and is executed in the occurrence.

Depending on their type, slices, can generate many different events, but all will be characterized by the following:

— arrival of a job in a slice;
— departure of a job from a slice;
— slice degradation;
— arrival of a job in a slice $s$ with zero resource $C_s, s \in \mathscr{S}$.

Only the events of the model cause a change in the state of the system, which we designated as **N**. Therefore, re-slicing for all events is reduced to tracking the events of arrival and departure of jobs. In our system, the slices themselves notify the slicer of these events (figure 4). After capacity $C$ allocation, the slicer notifies the slices that their available resource $C_s, s \in \mathscr{S}$ has changed. On these notifications, the slices adjust the end time of servicing their jobs in the event queue (Algorithm 2). If there are no jobs $N_s = 0$ in the slice $s$, then after re-slicing it can be assigned a zero resource value $C_s = 0$, which means that when the first request arrives, it will be necessary to activate the slice, in other words, call re-slicing again.

In the case when re-slicing is triggered by timer (figure 5), the slicer sends messages to itself with the required delay $t_{\text{timer}}$ s. Since there is a chance that the slice can receive zero resource, it became necessary to enter the activation of the slice upon the arrival of the request in this case.

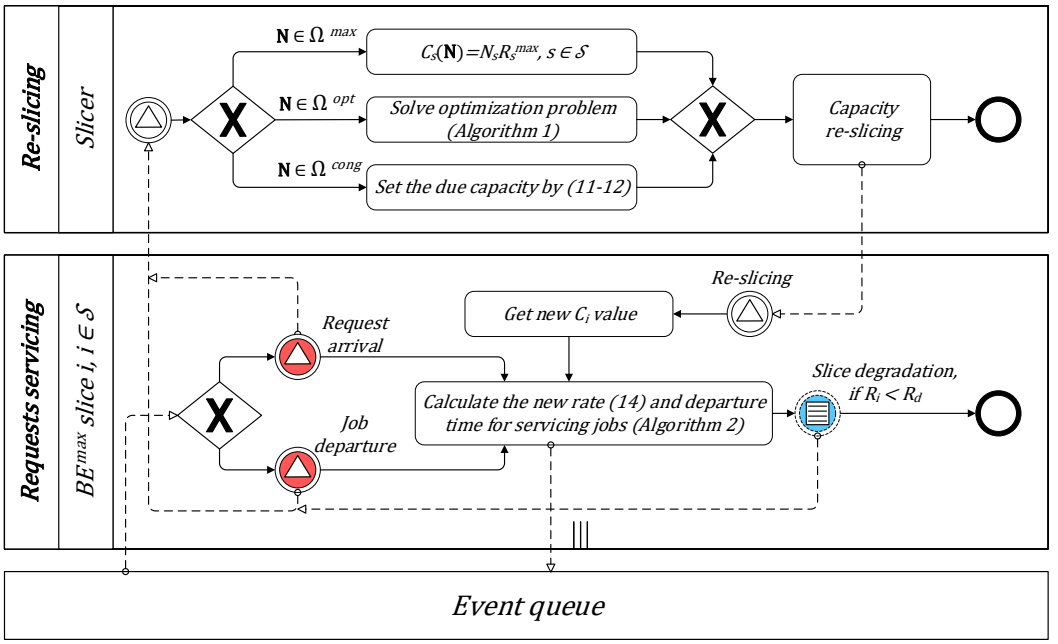


Figure 4. Interaction of slicer and slices when re-slicing triggered by all events (red) or degradation (blue). Re-slicing is called only by a group of events of the same color

With static slicing, the slices receive resource proportionally, in accordance with the values of $\gamma$ by equation (12).

### 5.3. Metrics

The built simulator allows you to take indicators in various forms using the built-in OMNeT++ tools, and more specifically using signals and statistics. The signal (@signal) transmits information at the right moments in the form of values of primitive types: bool, int, float, etc., or more complex data objects [7]. Statistics (@statistic) is a signal processing mechanism that allows you to accumulate vectors of original data transmitted by signals and scalars calculated by these vectors: sum, quantity, average, time average, maximum, minimum, etc. Preset simulator settings allow you to take such indicators like:

— average time spent in each module of the constructed QS inside slices and in the network as a whole;
— average number of jobs in each module of the constructed QS within slices and in the network as a whole;
— average service rate in slice;
— average number of jobs in slice, etc.

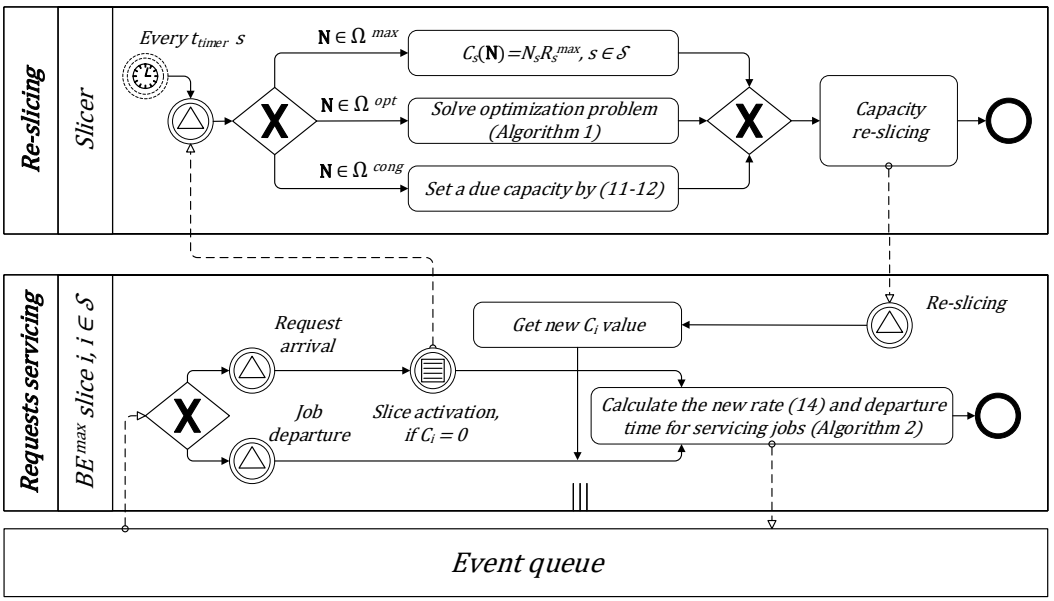


Figure 5. Interaction of slicer and slices when re-slicing triggered by timer

As part of assessing the effectiveness of slicing, the following additional indicators were taken:

— Slice degradation probability,

$$P_s^{\deg} = P\{R_s < R_s^d\} = \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{D_s(T)} (d_{s,i} - d_{s,i-1}) \mathcal{H}\{R_{s,i} < R_s^d\}, \qquad (14)$$

where $s \in \mathscr{S}$, $T$ is model time, $D_s(T)$ — counter of slice $s$ degradation threshold $R_s^d$ crossing (in any direction), $R_{s,i}$ — time of the $i$-th rate change, $d_{s,i}$ — time of the $i$-th degradation threshold $R_s^d$ crossing, and $\mathcal{H}$ is Heaviside step function.

— Average slice resource,

$$\overline{C_s} = \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{L_s(T)} (c_{s,i} - c_{s,i-1}) C_{s,i}, \quad s \in \mathscr{S}, \qquad (15)$$

where $L_s(T)$ — counter of slice $s$ resource changes, $c_{s,i}$ — moment $i$ of changing resource $C_s$.

— Average duration of slice degradation period,

$$\overline{t_s^{\deg}} = \lim_{T\to\infty} \frac{1}{D_s(T)+1} \sum_{i=1}^{D_s(T)} (d_{s,i} - d_{s,i-1}) \mathcal{H}\{R_{s,i} < R_s^d\}, \quad s \in \mathscr{S}. \qquad (16)$$

— Capacity utilization,

$$\text{UTIL} = \frac{1}{C}\sum_{s\in\mathscr{S}}\lim_{T\to\infty}\frac{1}{T}\sum_{i=1}^{Y_s(T)}(y_{s,i}-y_{s,i-1})N_{s,i}R_{s,i}, \quad s\in\mathscr{S}, \tag{17}$$

where $Y_s(T)$ — counter of slice $s$ service rate $R_s$ and number of jobs $N_s$ changes, $y_{s,i}$ — moment $i$ of changing $R_s$ or $N_s$.

— Re-slicing frequency.

— Average duration of the re-slicing operation.

## 6. Numerical results

To illustrate the performance of the simulator, we consider five slices with the parameters given in the table 2.

Table 2

Parameters' values for the numerical example

| **Slicer** | | | | | |
|---|---|---|---|---|---|
| Total capacity $C$ | 8000 | | | | |
| $S$ | 5 | | | | |
| Timer interval $t_{timer}$ | 100s | | | | |
| **Slice** $i$ | **1** | **2** | **3** | **4** | **5** |
| $R_s^{\min} = R_i^d$, Mbps | 2 | 5 | 25 | 50 | 30 |
| $R_s^{\max}$, Mbps | 2.2 | 8 | 30 | 75 | 8000 |
| $\gamma_i$ | 0.075 | 0.075 | 0.35 | 0.25 | 0.25 |
| $A_i(x)$ | $\exp(\lambda)$ | | | | |
| Request interarrival time $\lambda^{-1}$, s | 1.65 | 7.25 | 16 | 19 | 5 |
| $B_i(x)$ | $\exp(\theta)$ | | | | |
| Mean file size $\theta^{-1}$, GB | 0.3 | 1.2 | 2.5 | 5 | 1 |

Scenario is intended to demonstrate a system with an increased workload in slices 1 and 2. The guarantees are selected in such a way, that slices 3 and 4 are the main donors of capacity.

Figure 6 illustrates how the degradation probability $P_s^{\text{deg}}$ varies depending on the re-slicing triggers for the cases under study. Static re-slicing gives a high degradation probability in slice 1. For event triggers, we observe low degradation probability ($\sim 1\%$) for slices 1, 2, 4 and insignificant degradation probability in 5. When re-slicing is triggered by timer, the slicer reacts to the state of the system with a long delay, so there is an unacceptably high probability of degradation in donor slices 3 and 4.

The capacity utilization metric in figure 7 indicates that re-slicing upon all events and arrivals provide the highest resource utilization and the lowest waste of resources. This would be good if it were not for the fact that at a much lower system utilization, re-slicing upon degradation yields the same efficiency in terms of degradation probability.

Let us take a look at such an important indicator as the frequency of re-slicing calls. Figure 8 additionally confirms the efficiency of re-slicing upon degradation compared to re-slicing upon all events and arrivals, and even by timer. For all triggers, as expected, slicing takes roughly the same amount of time, averaging 0.04 ms.

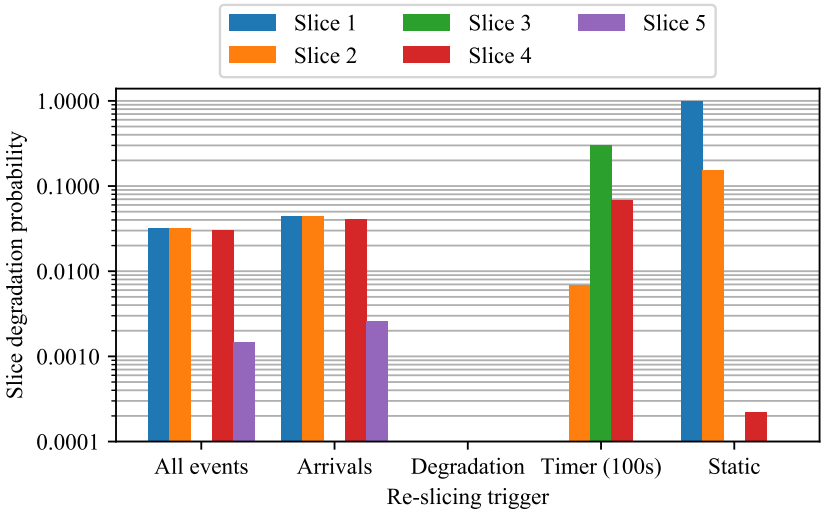


Figure 6. Slice degradation probability for different re-slicing triggers

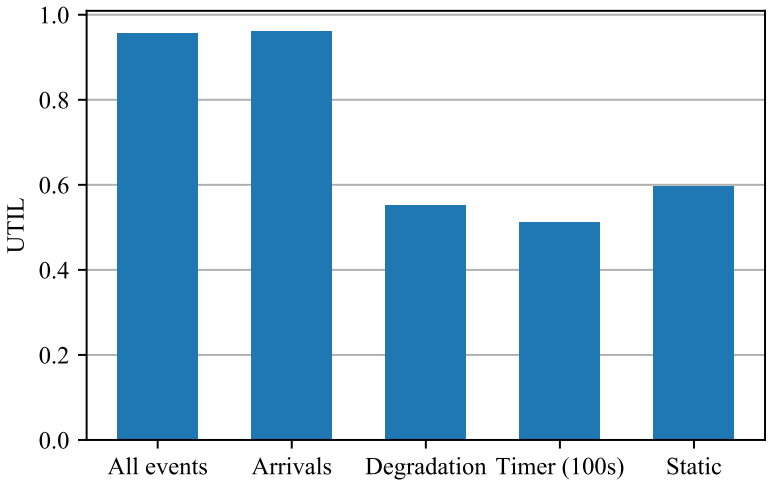


Figure 7. System utilization for different re-slicing triggers

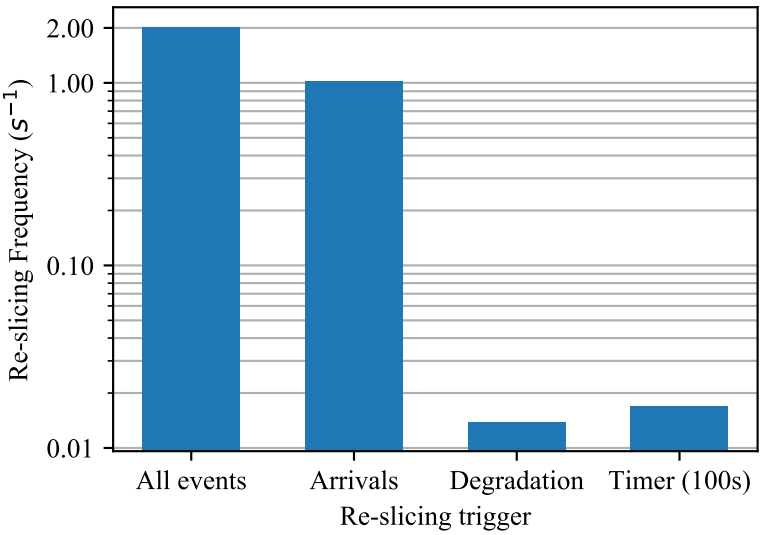


Figure 8. Re-slicing frequency for different re-slicing triggers

Let us consider the average share of capacity $\overline{C_s}$ allocated to each slice depending on the re-slicing trigger (figure 9). As we see, slices 1 and 2 receive significantly more capacity with frequent re-slicing than indicated in the SLA — the scheme allows this.

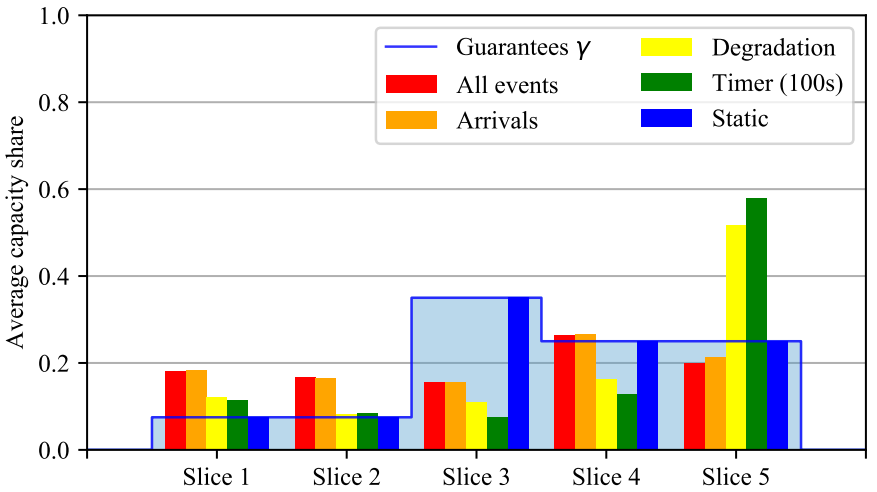


Figure 9. Average slice capacity share for different re-slicing triggers compared with the contracted share $\gamma_s$

## 7. Conclusion

A simulation model of network slicing with SLA-based isolation has been developed. By using the Object-Oriented Programming paradigm [8], as well as the built-in functionality of OMNeT++ and queuinglib, the following principles have been achieved:

— **Modularity of the system**: model elements (slicer, slice, queue, source of requests, delay, etc.) implemented as objects are logically separated, and the interaction among them occurs by transmitting global signals or messages through special channels.
— **Polymorphism, inheritance and encapsulation of slices**: all types of slices have a common ancestor which specifies all the mechanisms necessary for communicating with the slicer, so each descendant class describing a new slice type can replace their implementation with their own without breaking the interaction structure. In connection with the same principle, the QS describing the way of processing users (jobs) within a slice can take any form and be designed at the discretion of the developer. Thus, any slice is characterized only by its type and unified set of parameters.
— **Homogeneity of the structure of the input data**: an important characteristic for any simulator is the ease of use, in particular, the way of specifying the input data. In our implementation, based on the previous principle, the initial conditions for any slice are set in the same way using a configuration file.

Compliance with the indicated principles leads to scalability and extensibility of the simulation model.

Further research objectives:

— taking into account the state of the radio channel;
— adding and analyzing other re-slicing triggers;
— adding other types of slices;
— extensive numerical analysis.

## Acknowledgments

This paper has been supported by the RUDN University Strategic Academic Leadership Program. The reported study was funded by RFBR, project number 19-07-00933, 20-07-01052.

## References

[1] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: a survey on principles, enabling technologies, and solutions," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2429–2453, 2018. DOI: 10.1109/COMST.2018.2815638.

[2] R. Su, D. Zhang, R. Venkatesan, Z. Gong, C. Li, F. Ding, F. Jiang, and Z. Zhu, "Resource allocation for network slicing in 5G telecommunication networks: a survey of principles and models," *IEEE Network*, vol. 33, no. 6, pp. 172–179, 2019. DOI: 10.1109/MNET.2019.1900024.

[3] H. Yu, F. Musumeci, J. Zhang, M. Tornatore, and Y. Ji, "Isolation-aware 5G RAN slice mapping over WDM metro-aggregation networks," *Journal of Lightwave Technology*, vol. 38, no. 6, pp. 1125–1137, 2020. DOI: 10.1109/JLT.2020.2973311.

[4] N. Yarkina, Y. Gaidamaka, L. M. Correia, and K. Samouylov, "An analytical model for 5G network resource sharing with flexible SLA-oriented slice isolation," *Mathematics*, vol. 8, 2020. DOI: 10.3390/math8071177.

[5] B. Rouzbehani, L. M. Correia, and L. Caeiro, "A service-oriented approach for radio resource management in virtual RANs," *Hindawi Wireless Communications and Mobile Computing*, 2018. DOI: 10.1155/2018/4163612.

[6] B. Gladkih, *Optimization techniques and operations research for computer science bachelors. Part 2. Non-linear and dynamic programming [Metody optimizacii i issledovanie operacij dlya bakalavrov informatiki. CH. 2. Nelinejnoe i dinamicheskoe programmirovanie].* Tomsk: Izd-vo NTL, 2011, in Russian.

[7] A. Virdis and M. Kirsche, *Recent Advances in Network Simulation.* 2019. DOI: 10.1007/978-3-030-12842-5.

[8] R. Lafore, *Object-Oriented Programming in C++, 4th Edition.* CourseSams Publishing, 2001.

**For citation:**

N. A. Polyakov, N. V. Yarkina, K. E. Samouylov, A simulator for analyzing a network slicing policy with SLA-based performance isolation of slices, Discrete and Continuous Models and Applied Computational Science 29 (1) (2021) 36–52. DOI: 10.22363/2658-4670-2021-29-1-36-52.

**Information about the authors**:

**Polyakov, Nikita A.** — Bachelor of Science, Master student (e-mail: goto97@mail.ru, phone: +7(916)5858743, ORCID: https://orcid.org/0000-0003-0152-9646, Scopus Author ID: 57211203793)

**Yarkina, Natalia V.** — Candidate of Sciences, ORCID: https://orcid.org/0000-0003-3197-2737

**Samouylov, Konstantin E.** — Doctor of Technical Sciences, Professor, applied Mathematics & Communications Technology Institute (e-mail: ksam@sci.pfu.edu.ru, ORCID: https://orcid.org/0000-0002-6368-9680, ResearcherID: E-9966-2014, Scopus Author ID: 14009785000)

УДК 519.872, 519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2021-29-1-36-52

# Имитационное моделирование разделения ресурсов с изоляцией слайсов на базе SLA

**Н. А. Поляков**[1], **Н. В. Яркина**[1], **К. Е. Самуйлов**[1,2]

[1] *Российский университет дружбы народов*
*ул. Миклухо-Маклая, д. 6, Москва, 117198, Россия*
[2] *Федеральный исследовательский центр «Информатика и управление» РАН*
*ул. Вавилова, д. 44, кор. 2, Москва, 119333, Россия*

В настоящее время, несмотря на ввод в эксплуатацию сетей мобильной связи 5-го поколения, эффективное разделение ресурсов сети радиодоступа по-прежнему остаётся актуальной задачей. Свои коррективы в её постановку вносят технологии виртуализации и нарезки сети (network slicing), позволяющие разделять сеть доступа на логические подсети. В статье предложен инструмент имитационного моделирования, разработанный на платформе OMNeT++ для анализа эффективности схемы разделения ресурсов с изоляцией слайсов на базе соглашений об уровне обслуживания. Объектно-ориентированный подход к построению симулятора обеспечивает гибкость и расширяемость модели. В статье кратко изложена исследуемая схема слайсинга, подробно описана архитектура программного средства и особенности построения имитационой модели, приведены результаты численного анализа.

**Ключевые слова:** система массового обслуживания, разделение ресурсов, нарезка сети, имитационное моделирование, оптимизация

*Research article*

UDC 004.75
PACS 02.50.Fz, 02.60.Pn,
DOI: 10.22363/2658-4670-2021-29-1-53-62

# Optimization of mobile device energy consumption in a fog-based mobile computing offloading mechanism

**Anastasia V. Daraseliya**[1], **Eduard S. Sopin**[1, 2]

[1] *Peoples' Friendship University of Russia (RUDN University)*
*6, Miklukho-Maklaya St., Moscow, 117198, Russian Federation*
[2] *Federal Research Center "Computer Science and Control"*
*of the Russian Academy of Sciences (FRC CSC RAS)*
*44-2, Vavilova St., Moscow 119333, Russian Federation*

(received: March 5, 2021; accepted: March 12, 2021)

The offloading of computing tasks to the fog computing system is a promising approach to reduce the response time of resource-greedy real-time mobile applications. Besides the decreasing of the response time, the offloading mechanisms may reduce the energy consumption of mobile devices. In the paper, we focused on the analysis of the energy consumption of mobile devices that use fog computing infrastructure to increase the overall system performance and to improve the battery life. We consider a three-layer computing architecture, which consists of the mobile device itself, a fog node, and a remote cloud. The tasks are processed locally or offloaded according to the threshold-based offloading criterion. We have formulated an optimization problem that minimizes the energy consumption under the constraints on the average response time and the probability that the response time is lower than a certain threshold. We also provide the numerical solution to the optimization problem and discuss the numerical results.

**Key words and phrases:** queuing system, fog computing, cloud computing, queuing theory, optimization, Laplace–Stieltjes transform

## 1. Introduction

In recent years, fog computing has received attention from the scientific and industrial community. Many papers were related to opportunities and challenges of fog, focusing primarily on the networking context of the Internet of Things (IoT) [1]. Another one of the most popular topics and pressing research issue is the compromise between the energy-efficiency and the response time in offloading of mobile application tasks to fog computing infrastructure. The paper [2] presents the results of a study on energy consumption, execution delay and payment cost of offloading processes in a fog computing network in terms of queuing theory. Research in [3] focuses on energy-efficient task

offloading, whose main idea is taking into account both energy consumption and schedule delay under fog devices. Energy efficient offloading is also a vital task in the context of the Internet of Things concept [4].

In our previous paper [5], we developed an analytical framework for response time analysis that takes into account the variation of tasks in terms of processing volume. Then in the paper [6], we analyze the two-parameter offloading mechanism that takes into account both the computing complexity and the data size to be transferred in case of offloading. In [7] we derived the cumulative distribution function of the response time in terms of Laplace-Stieltjes Transform. In the current work, we solve the optimization problem by minimizing energy efficiency, subject to the average time constraint and taking into account the probability that the time exceeds a given threshold.

## 2. Mathematical model

We consider a distributed computing system that consist of mobile devices (MDs), a fog node and a remote cloud. MDs run real-time applications that require significant amount of computational resources. For each task, a MD makes a decision, whether it will be offloaded to the fog node or processed locally. The capacity of the fog node is limited, which means if there are too many tasks offloaded, then some of the offloaded tasks are redirected to the remote cloud to prevent the fog node congestion. In terms of queuing theory, the considered system can be represented as shown in the figure 1.

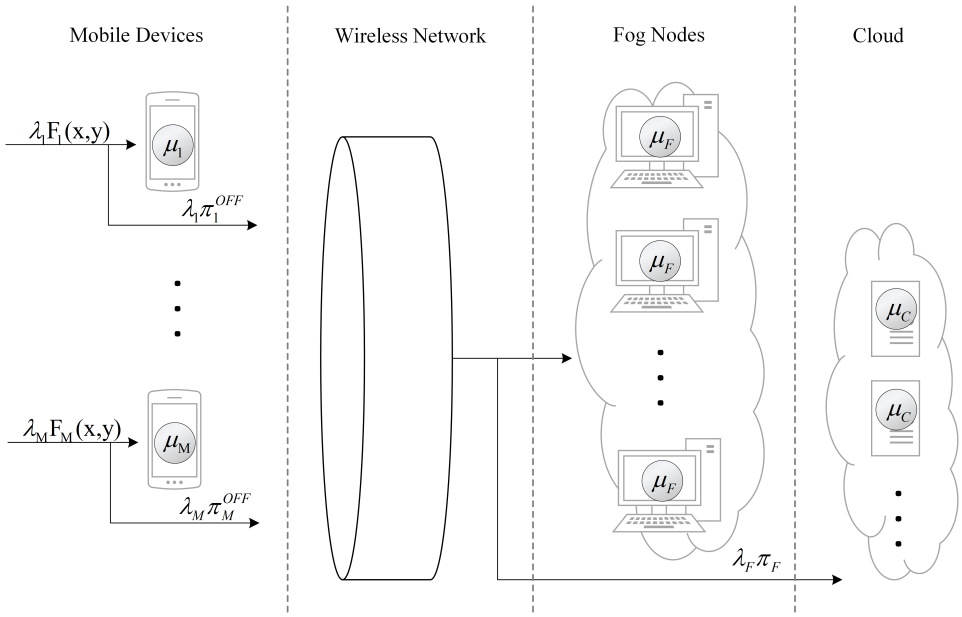


Figure 1. Mathematical model in terms of queuing network

Assume there are $M$ MDs, each of them generating a flow of tasks with exponentially distributed interarrival times according to Poisson's law with intensity $\lambda_i$, $i = 1, \ldots, M$. Each task is characterized by the amount of processing volume required and the data size to be transferred in case of

offloading. We assume that the processing volume (measured in millions of instructions, MI) and the data size (measured in MB) are independent random variables with CDFs $W_i(x)$ and $S_i(x)$, probability density functions (PDF) $w_i(x)$ and $s_i(x)$ respectively. MDs process locally served tasks in the FCFS mode with constant serving rate $\mu_i$, $i = 1, \dots, M$ (measured in MIPS).

We propose the offloading mechanism that implies offloading tasks that are "heavy" in terms of processing volume and "light" in terms of data size. Splitting to "heavy" and "light" tasks are done by the threshold $O_w$ on the processing volume and the threshold $O_s$ on the data size. Hence, the offloading probability $\pi_{i,O}$ on the $i$-th MD is evaluated [6], [7] according to the following formula

$$\pi_{i,O} = \int_{O_w}^{\infty} w_i(x)\,dx \int_0^{O_s} s_i(y)dy = (1 - W_i(O_w))S_i(O_s). \tag{1}$$

If a task is processed locally, then the response time consists of processing time on an MD only. If a task is offloaded to the fog node, then the total response time is the sum of task transmission time to the fog node through wireless network, the processing time on the fog node and the transmission time back to the MD. If the fog node is overloaded and an offloaded task is sent to the remote cloud, then the processing time at the fog is replaced by the transmission time between the fog node and the remote cloud, the processing time on the cloud and the transmission time back to the fog node.

We assume that the wireless network provides total bitrate $R$, which is used to transmit the data of tasks one-by-one in FCFS order, so the transmission time is obtained as the fraction of the data size of a task and total bitrate $R$. On the other side, the transmission time between the fog node and the cloud is assumed constant.

The fog node provide computational resources by means of virtual machines (VMs), each of them having the constant serving rate $\lambda_F$. The total number of VMs at the fog node is $N$. The constant serving rate $\mu_C$ of VMs at the cloud is greater than $\mu_F$, and amount of computational resources (VMs) at the remote cloud is assumed to be large enough, so that it cannot be overloaded.

## 3. The response time analysis

### 3.1. CDFs of the response time components

The service process at MD $i$ is modeled in terms of a queuing system $M/G/1$ with arrival intensity $\lambda_i$, $\sum\limits_{i=1}^{M} \lambda_i = \lambda$. The distribution function of the processing volume on a MD can be determined by conditional CDF $W_{MD,i}(x)$ as follows

$$W_{MD,i}(x) = \begin{cases} \dfrac{W_i(O_w) + (W_i(x) - W_i(O_w))(1 - S_i(O_s))}{1 - \pi_{i,O}}, & x > O_w, \\ \dfrac{W_i(x)}{1 - \pi_{i,O}}, & x \leqslant O_w. \end{cases} \tag{2}$$

Having obtained the distribution function of the processing volume $W_{MD,i}(\mu_i x)$, we can find the serving time at a MD. By virtue of the fact that the serving rate on the $i$-th MD is constant and being $\mu_i$, its CDF is easily obtained as $T_{MD,i}(x) = W_{MD,i}(\mu_i x)$. The average serving time at MD $i$ can be found through integration using the CDF $T_{MD,i}$.

If a task is offloaded to the distributed computing infrastructure, it is first transferred through the wireless network to the fog node. The delays in wireless networks are obtained analogously, by employing $M/G/1$ queue. The arrival intensity $\lambda_F$ is the sum of the offloading intensities from all MDs $\lambda_F = \sum_{i=1}^{M} \lambda_i \pi_{i,O}$. The CDF $S_{tr,i}(x)$ of the file size to be transmitted is

$$S_{tr,i}(x) = \begin{cases} \dfrac{1}{\pi_{i,O}(1 - W_i(O_w))S_i(x)}, & x \leqslant O_s, \\ 1, \quad x > O_s, \end{cases} \tag{3}$$

and the service time distribution in the wireless network is $T_{tr,i}(x) = S_{tr,i}(Rx)$.

At the fog node, there are $N$ VMs to serve offloaded tasks, so the service process may be modeled by $M/G/N/0$ queue, where the blocked customers are redirected to the next layer — remote cloud. The arrival intensity is the same as for wireless network — $\lambda_F$. The service time is determined by

$$W_{F,i}(x) = \begin{cases} \dfrac{1}{\pi_{i,O}}(W_i(x) - W_i(O_w))S_i(O_s), & x > O_w, \\ 0, \quad x \leqslant O_w. \end{cases} \tag{4}$$

The service time is simply processing volume divided by the service rate $\mu_F$, so the CDF of the service time at the fog node is

$$T_{F,i}(x) = W_{F,i}(\mu_F x), \tag{5}$$

$$T_F(x) = \frac{\lambda_i \pi_{i,O}}{\lambda_F} T_{F,i}(x). \tag{6}$$

The probability that a task is redirected to the remote cloud $\pi_F$ is obtained from Erlang formula for $M/G/N/0$ queues as

$$\pi_F = \frac{(\lambda_F \tau_F)^N}{N!} \left( \sum_{k=0}^{N} \frac{(\lambda_F \tau_F)^k}{k!} \right), \tag{7}$$

where $\tau_F$ is the average serving time at the fog node, which can be easily evaluated from the CDF $TF(x)$. The service time distribution $T_C(x)$ at the cloud is

$$T_{C,i}(x) = W_{C,i}(\mu_C x). \tag{8}$$

### 3.2. The average response time

The total response time is the conditional sum of processing and transmission delays. A task from $i$-th MD is processed locally with probability $1-\pi_{i,O}$ on the fog node with probability $\pi_{i,O}(1-\pi_F)$ and on the cloud with probability $\pi_{i,O}\pi_F$. In [7], we derived the Laplace-Stieltjes Transforms (LST) for all delay components for the case of Gamma distribution of both processing volume and data size, and obtain the LST of the total response time.

First, we derived the LST $\tilde{T}_{MD,i}(s)$ of the service time at the MD $i$ as

$$\tilde{T}_{MD,i}(s) = \int_0^\infty e^{-sx}\, d(T_{MD,i}(x)) = \frac{1}{1-\pi_{i,O}}\left[\frac{\mu_i^2}{(s\delta_W+\mu_i)^2} - \right.$$
$$\left. - \left(1 - e^{-\frac{O_s}{\delta_s}}\left(1+\frac{O_s}{\delta_s}\right)\right) e^{-\left(\frac{sO_w}{\mu_i}+\frac{O_w}{\delta_w}\right)} \frac{\mu_i O_w(s\delta_w+\mu_i)+\mu_i^2\delta_w}{\delta_w(s\delta_w+\mu_i)^2}\right] \quad (9)$$

with the the LST of $\phi_{MD,i}$ the sojourn time distribution on mobile device $i$

$$\omega_{MD,i}(s) = \frac{s(1-\rho_i)}{s-\lambda_i+\lambda_i\tilde{T}_{MD,i}(s)}, \quad (10)$$

$$\phi_{MD,i}(s) = \tilde{T}_{MD,i}(s)\omega_{MD,i}(s). \quad (11)$$

Then we obtained the LST $\tilde{T}_{F,i}(s)$ and $\tilde{T}_{C,i}(s)$ of the service time distribution at the fog node and cloud, respectively. LST $\tilde{T}_{F,i}(s)$ of the service time distribution at the fog node is derived from CDF

$$\tilde{T}_{F,i}(s) = \int_0^\infty e^{-sx}\, d(T_{F,i}(x)) = e^{-\left(\frac{sO_w}{\mu_F}\right)} \frac{\mu_F O_w(s\delta_w+\mu_F)+\mu_i^2\delta_w}{\delta_w(s\delta_w+\mu_i)^2}. \quad (12)$$

LST of the service time distribution in the cloud is obtained by analogy with $\tilde{T}_{F,i}(s)$.

The LST $\omega_{tr,i}(s)$ of the waiting time distribution and the LST $\phi_{tr,i}(s)$ of sojourn time in the wireless network are:

$$\omega_{tr,i}(s) = \frac{s(1-\rho_{tr})}{s-\lambda_F+\lambda_F\tilde{T}_{tr,i}(s)}, \quad \phi_{tr,i}(s) = \tilde{T}_{tr,i}(s)\omega_{tr,i}(s). \quad (13)$$

Having obtained the LST of all these delay component distributions, we made use of the convolution formula and obtain the LST $\tilde{\tau}(s)$ of the response time distribution of a task from MD $i$:

$$\tilde{\tau}_i(s) = (1-\pi_{i,0})\phi_{MD,i}(s) + \pi_{i,0}(1-\pi_F)\tilde{T}_{F,i}(s)\phi_{tr,i}^2(s) +$$
$$+ \pi_{i,0}\pi_F\tilde{T}_{C,i}(s)\phi_{tr,i}^2(s)\tilde{T}_{FC}^2(s). \quad (14)$$

After this we used numerical Reverse LST $\tilde{\tau}(s)$ to evaluate the CDF $\tau(s)$ of the response time.

Actually, the average response time can be calculated as

$$\tau = \sum_{i=0}^{M} \frac{\lambda_i}{\left(\sum\limits_{j=1}^{M} \lambda_j\right)} \tau_i. \tag{15}$$

The resulting expressions allow to get the probability $\Pi(T)$ that the response time is lower than a threshold $T$

$$\Pi(T) = \tau(T). \tag{16}$$

## 4. The energy consumption analysis

In this section, we present the formulas for the average power consumption of MDs obtained at an earlier stage of research [6].

The energy consumption for tasks processed on MD is proportional to the processing volumes of tasks, therefore the average energy consumption $E_{pr,i}$ during locally executing on $i$-th MD can be evaluated as follows:

$$E_{pr,i} = P_{pr,i} t_{MD,i}, \tag{17}$$

where $P_{pr,i}$ is the power consumption (W) during the processing of the $i$-th MD, which is considered constant for simplicity of calculations.

The average file size transmitted by $i$-th MD, can be calculated through integration using CDF $S_{tr,i}(x)$ from the previous section.

The energy consumption during transmitting is also proportional to the transmission time, so the average energy consumption $E_{tr,i}$ of the $i$-th VD during task transmission is

$$E_{tr,i} = P_{tr,i} \frac{\theta_i}{R}. \tag{18}$$

Then the average energy consumption for any $i$-th MD is the weighted sum of processing and transmission energies:

$$E_i = (1 - \pi_{i,0}) E_{pr,i} + \pi_{i,0} E_{tr,i}. \tag{19}$$

At the end, we can evaluate the average energy consumption for a task from an arbitrary MD as follows

$$E = \sum_{i=0}^{M} \frac{\lambda_i}{\left(\sum\limits_{j=1}^{M} \lambda_j\right)} E_i. \tag{20}$$

## 5. Optimization problem

In order to find the minimum energy consumption $E$ under constraints on the average response time and the probability $\Pi(T)$ that the response time is lower than a threshold $T$, we formulate the optimization problem as follows:

$$\begin{cases} E = \sum\limits_{i=0}^{M} \dfrac{\lambda_i}{\left(\sum\limits_{j=1}^{M} \lambda_j\right)} E_i \to \min, \\ \tau = \sum\limits_{i=0}^{M} \dfrac{\lambda_i}{\left(\sum\limits_{j=1}^{M} \lambda_j\right)} \tau_i \leqslant T, \\ \Pi(T) \leqslant \Pi^*. \end{cases} \tag{21}$$

## 6. Numerical results

In this section, we presented the numerical results of our study. The main metric of interest here is the minimum power consumption $E$ for a task from an arbitrary MD under the constraints from the optimization problem.

We consider a system with $M = 20$ homogeneous MDs that run the same applications, so the distributions of processing volume and data size of tasks are also the same. The fog nodes can run maximum $N = 8$ VMs. All values of parameters used in the section are gathered in table 1.

Table 1

Parameter values for the numerical analysis

| Parameter | Value |
|---|---|
| $M$ | 20 |
| $N$ | 8 |
| $R$ | 150 Mbps |
| $\lambda_i$ | 2 tasks/s |
| $\mu_i$ | 4 MIPS |
| $\mu_F$ | 6 MIPS |
| $\mu_C$ | 10 MIPS |
| $\delta_s$ | 0.25 |
| $\delta_\omega$ | 0.75 |
| $t_{FC}$ | 0.5 s |
| $P_{pr}$ | 16 W |
| $P_{tr}$ | 0.2 W |

Figure 2 shows the probability $\Pi(T)$ as a function of processing volume threshold with the response time threshold $T = 0.5$. The power consumption graph begins to descend only at values of $\Pi^* = 0.94$ and above. This shows that only with a very high threshold value of the probability $\Pi(T)$ that the response time is lower than a threshold $T$, there will be a gain in terms of energy costs.

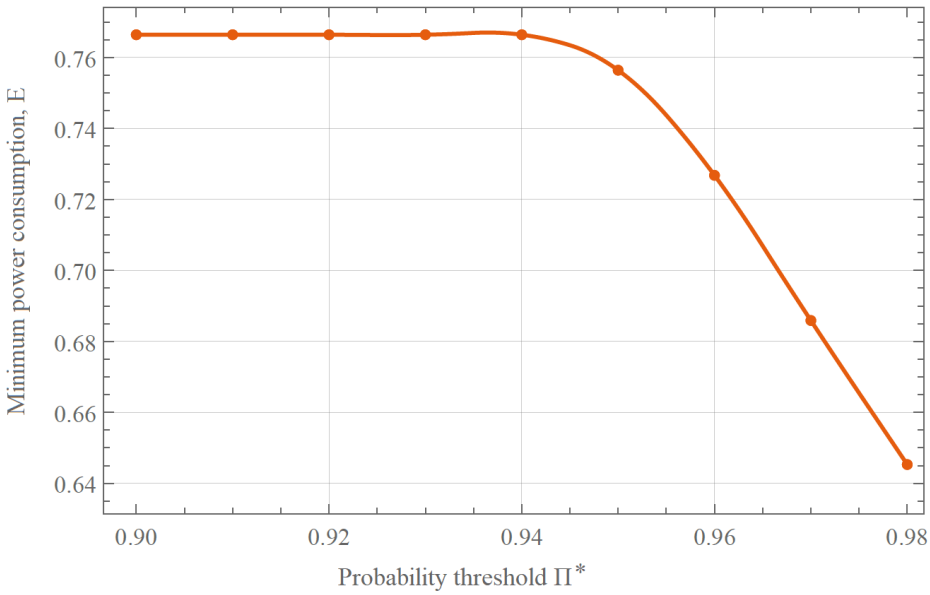


Figure 2. Dependence of the minimum power consumption $E \to \min$ on the threshold value $\Pi^*$, $T = 0.5$

## 7. Conclusions

In the paper, we focused on the analysis of the reducing energy consumption of MD's that use fog computing infrastructure to increase the performance and to improve the battery life of mobile devices. We have formulated and solved the problem of energy consumption optimization using constraints on the average response time and the probability that the response time is lower than a certain threshold, on the basis of which we offer some recommendations for offloading the system.

## Acknowledgments

This paper has been supported by the RUDN University Strategic Academic Leadership Program (recipient Sopin E., mathematical model development). The reported study was funded by RFBR, project number 20-07-01052 (recipient Daraseliya A., optimization problem). The reported study was funded by RFBR, project number 19-07-00933 (recipient Sopin E., numerical analysis).

## References

[1] M. Chiang and T. Zhang, "Fog and IoT: an overview of research opportunities," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 854–864, 2016. DOI: 10.1109/JIOT.2016.2584538.

[2] Z. Chang, Z. Zhou, T. Ristaniemi, and Z. Niu, "Energy efficient optimization for computation offloading in fog computing system," in *GLOBECOM 2017 — 2017 IEEE Global Communications Conference*, 2017, pp. 1–6. DOI: 10.1109/GLOCOM.2017.8254207.

[3] Y. Jiang, Y. Chen, S. Yang, and C. Wu, "Energy-efficient task offloading for time-sensitive applications in fog computing," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2930–2941, 2019. DOI: 10.1109/JSYST.2018.2877850.

[4] Q. Li, J. Zhao, Y. Gong, and Q. Zhang, "Energy-efficient computation offloading and resource allocation in fog computing for Internet of Everything," *China Communications*, vol. 16, no. 3, pp. 32–41, 2019. DOI: 10.12676/j.cc.2019.03.004.

[5] E. S. Sopin, A. V. Daraseliya, and L. M. Correia, "Performance analysis of the offloading scheme in a fog computing system," in *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2018, pp. 1–5. DOI: 10.1109/ICUMT.2018.8631245.

[6] E. Sopin, K. Samouylov, and S. Shorgin, "The analysis of the computation offloading scheme with two-parameter offloading criterion in fog computing," pp. 11–20, 2019. DOI: 10.1007/978-3-030-34914-1_2.

[7] E. Sopin, N. Zolotous, K. Ageev, and S. Shorgin, "Analysis of the response time characteristics of the fog computing enabled real-time mobile applications," *Lecture Notes in Computer Science*, vol. 12525, pp. 764–779, 2020. DOI: 10.1007/978-3-030-65726-0_9.

**For citation:**

A. V. Daraseliya, E. S. Sopin, Optimization of mobile device energy consumption in a fog-based mobile computing offloading mechanism, Discrete and Continuous Models and Applied Computational Science 29 (1) (2021) 53–62. DOI: 10.22363/2658-4670-2021-29-1-53-62.

**Information about the authors**:

**Daraseliya, Anastasia V.** — PhD student of Department of Applied Probability and Informatics of Peoples' Friendship University of Russia (RUDN University) (e-mail: avdaraseliya@sci.pfu.edu.ru, phone: +7(495)9550927, ORCID: https://orcid.org/0000-0002-6603-2596)

**Sopin, Eduard S.** — Candidate of Physical and Mathematical Sciences, Assistant professor of Department of Applied Probability and Informatics of Peoples' Friendship University of Russia (RUDN University); Senior Researcher of Institute of Informatics Problems of Federal Research Center "Computer Science and Control" Russian Academy of Sciences (e-mail: sopin-es@rudn.ru, phone: +7(495)9550927, ORCID: https://orcid.org/0000-0001-9082-2152)

УДК 004.75
PACS 02.50.Fz, 02.60.Pn,
DOI: 10.22363/2658-4670-2021-29-1-53-62

# Оптимизация энергопотребления мобильных устройств в системе туманных вычислений

**А. В. Дараселия**[1], **Э. С. Сопин**[1, 2]

[1] *Российский университет дружбы народов*
*ул. Миклухо-Маклая, д. 6, Москва, 117198, Россия*
[2] *Федеральный исследовательский центр «Информатика и управление» РАН*
*ул. Вавилова, д. 44, кор. 2, Москва, 119333, Россия*

Выгрузка задач мобильных вычислений в систему туманных вычислений представляется многообещающим подходом для снижения времени отклика ресурсоёмких мобильных приложений, функционирующих в режиме реального времени. Помимо снижения времени отклика, механизмы выгрузки вычислений помогут также снизить энергопотребление мобильных устройств. В этой статье мы проводим анализ энергопотребления мобильных устройств, которые используют инфраструктуру туманных вычислений для повышения производительности и увеличения времени их автономной работы. Рассматривается трёхуровневая вычислительная система, состоящая из непосредственно мобильного устройства, узла системы туманных вычислений и удалённого облака. Задачи мобильных вычислений могут быть обработаны локально на устройстве или быть выгружены в соответствии с пороговым критерием выгрузки. Сформулирована и решена задача оптимизации энергопотребления при наличии ограничений на среднее время отклика и на вероятность того, что время отклика ниже определённого порога.

**Ключевые слова:** система массового обслуживания, туманные вычисления, облачные вычисления, оптимизация, преобразование Лапласа–Стилтьеса

*Research article*

UDC 519.872, 519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2021-29-1-63-72

# On conjugate difference schemes: the midpoint scheme and the trapezoidal scheme

**Yu Ying**[1], **Mikhail D. Malykh**[2]

[1] *Kaili University*
*3, Kaiyuan Road, Kaili, 556011, China*
[2] *Peoples' Friendship University of Russia (RUDN University)*
*6, Miklukho-Maklaya St., Moscow, 117198, Russian Federation*

(received: February 10, 2021; accepted: March 12, 2021)

The preservation of quadratic integrals on approximate solutions of autonomous systems of ordinary differential equations $\dot{x} = f(x)$, found by the trapezoidal scheme, is investigated. For this purpose, a relation has been established between the trapezoidal scheme and the midpoint scheme, which preserves all quadratic integrals of motion by virtue of Cooper's theorem. This relation allows considering the trapezoidal scheme as dual to the midpoint scheme and to find a dual analogue for Cooper's theorem by analogy with the duality principle in projective geometry. It is proved that on the approximate solution found by the trapezoidal scheme, not the quadratic integral itself is preserved, but a more complicated expression, which turns into an integral in the limit as $\Delta t \to 0$. Thus the concept of conjugate difference schemes is investigated in pure algebraic way. The results are illustrated by examples of linear and elliptic oscillators. In both cases, expressions preserved by the trapezoidal scheme are presented explicitly.

**Key words and phrases:** dynamical systems, quadratic integrals, difference schemes, conservation laws, midpoint scheme, trapezoidal scheme

## 1. Introduction

Dynamical systems are the most important mathematical models in mechanics and physics. Only a few of these models are integrated in a closed form [1], therefore, they have to be investigated using numerical methods, of which the most important is the finite difference method.

Let $x$ be a point in an $m$-dimensional affine space. Any difference scheme that approximates differential equation

$$\frac{dx}{\Delta t} = f(x) \tag{1}$$

describes a transition from the value $x$ at some initial moment of time to the value of $\hat{x}$ at the moment of time shifted from the initial value by the quantity $\Delta t$, called the step. We will consider algebraic schemes, i.e., those in which the above correspondence is specified using a system of algebraic equations

$$F(x, \hat{x}, \Delta t) = 0. \tag{2}$$

If the original equation has an algebraic integral $g(x) = C$, and it follows from the equations (2) that

$$g(\hat{x}) = g(x),$$

then this difference scheme is said to preserve this integral.

If we use explicit difference schemes for integrating dynamical systems, then the values of the integrals of motion will change monotonically step by step. At the turn of the 1980s and 1990s, the first difference schemes were constructed that preserve exactly the algebraic integrals of dynamical systems. For example, the scheme constructed by D. Greenspan preserves all classical integrals of N-body problems [2]–[5], the symplectic Runge–Kutta schemes, including the simplest of them, the midpoint scheme

$$\hat{x} - x = f\left(\frac{\hat{x} + x}{2}\right) \Delta t, \tag{3}$$

preserve linear and quadratic integrals in virtue Cooper's theorem [6]–[9]. This circuit has a whole bunch of wonderful properties inherited from the original differential equation [10].

This seems to be a simple consequence of the $t$-symmetry of the midpoint circuit: the equation (3) is invariant under the transformation

$$\Delta t \to -\Delta t, \quad \hat{x} \to x, \quad x \to \hat{x}.$$

The trapezoidal scheme has the same property

$$\hat{x} - x = (f(\hat{x}) + f(x)) \frac{\Delta t}{2}, \tag{4}$$

however, in experiments with an elliptic oscillator performed by Yu. A. Blinkov for PCA'2019 [11], the quadratic integrals oscillated, although they did not increase monotonically. The absence of monotonicity in the variation of the values of the integrals of motion on approximate solutions is extremely important from the physical point of view, since, on average, all fundamental conservation laws are satisfied on solutions of this type.

The noted behavior of the approximate solutions found by the trapezoidal scheme can be explained by the fact that it is conjugated to the midpoint scheme and therefore some more complex expression is retained on it [9, §VI.8.1-2].

The very concept of conjugate difference schemes [9, def 8.1] is formulated locally in terms of power series. The implicit function theorem can be applied to the system of algebraic equations (2) and, under certain conditions, we can assert that

$$\hat{x} = x + f(x)\Delta t + ... = \Phi_{\Delta t}(x),$$

where $\Phi$ is a series in powers of $\Delta t$, the coefficients of which are rational functions of $x$. Difference schemes

$$\hat{x} = \Phi_{\Delta t}(x) \text{ and } \hat{x} = \Psi_{\Delta t}(x)$$

are referred to as mutually conjugate, if there exists a change of coordinates

$$x = \chi_{\Delta t}(y) = y + ...,$$

such that

$$\Phi_{\Delta t} = \chi_{\Delta t}^{-1} \circ \Psi_{\Delta t} \circ \chi_{\Delta t}.$$

It is clear from this definition that the exact preservation of the expression of one of these schemes entails the preservation of some expression by the other scheme. Say, if the scheme $\Psi_{\Delta t}$ preserves the integral $g(x)$ exactly, then the scheme $\Phi_{\Delta t}$ preserves the expression

$$g(\chi_{\Delta t}(x)) = g(x) + g_1(x)\Delta t + ...,$$

depending on $\Delta t$ [9, §VI.8.2].

In this article, we will clarify the geometric meaning of the conjugacy of the two above schemes and write down explicitly the expression that preserves the trapezoidal scheme.

## 2. Relationship between trapezoidal and midpoint schemes

The approximate solution of the system (1), found by the scheme (2) with constant step $\Delta t$, is a finite or infinite sequence of points

$$x_0, x_1, x_2, ... \tag{5}$$

the first element of which is chosen in an arbitrary way, and all the others are defined recursively: $x_{n+1}$ is the root $\hat{x}$ of the equation

$$F(x_n, \hat{x}, \Delta t) = 0,$$

tending to $x_n$ at $\Delta t \to 0$.

**Theorem 1.** *Let $x_0, x_1, x_2, ...$ be an approximate solutions of equation* (1)*, calculated using the midpoint scheme* (3)*. Then coordinates $x'_0, x'_1, x'_2, ...$ of middles of links of broken line $x_0x_1x_2...$ (figure 1) yield another approximate solution of the same equation, calculated by the trapezoidal scheme* (4)*.*

**Proof.** The middle of link $x_nx_{n+1}$ of the solution found by the midpoint scheme is given by the formula

$$x'_n = \frac{x_{n+1} + x_n}{2}.$$

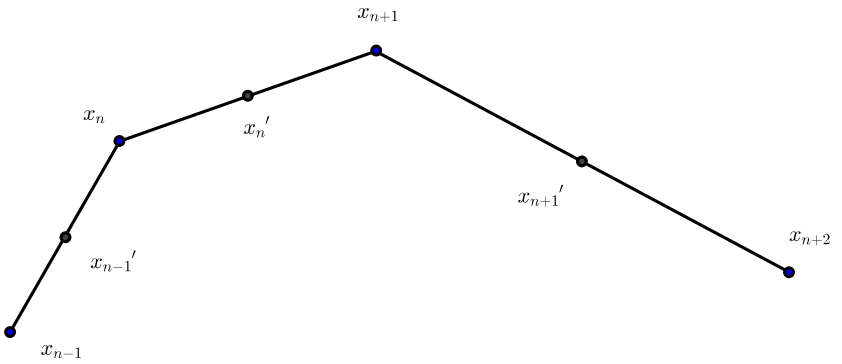


Figure 1. The solutions $x_0, x_1, x_2, ...$ and $x'_0, x'_1, x'_2, ...$ found by the midpoint schemes and by the trapezoidal scheme

In this case, the ends of the link are unambiguously reconstructed from its given midpoint:

$$x_{n+1} = \frac{x_{n+1} + x_n}{2} + \frac{x_{n+1} - x_n}{2} = x'_n + f(x'_n)\frac{\Delta t}{2}$$

and

$$x_n = \frac{x_{n+1} + x_n}{2} - \frac{x_{n+1} - x_n}{2} = x'_n - f(x'_n)\frac{\Delta t}{2}.$$

Since $x_{n+1}$ belongs to two links $x_n x_{n+1}$ and $x_{n+1}x_{n+2}$, we have

$$x_{n+1} = x'_n + f(x'_n)\frac{\Delta t}{2} = x'_{n+1} - f(x'_{n+1})\frac{\Delta t}{2}$$

from where it immediately follows that

$$x'_{n+1} - x'_n = \left(f(x'_{n+1}) + f(x'_n)\right)\frac{\Delta t}{2}.$$

Thus, the midpoints can be calculated using the trapezoidal difference scheme (4). □

By virtue of the theorem 1, the solutions $x_0x_1x_2 ...$ and $x'_0x'_1x'_2 ...$, found by the midpoint schemes (3) and by the trapezoidal scheme (4) turn out to be coupled with each other. By analogy with the duality principle in projective geometry [12] it is hoped that any statement about the midpoint scheme should have a 'twin' in the trapezoidal scheme.

One of the most interesting properties of the midpoint scheme is Cooper's theorem [9, th. 2.2], according to which this scheme preserves any quadratic integral of motion.

Equality

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

is easily rewritten by expressing $x_n$ though $x'_n$, and $x_{n+1}$ though $x'_{n+1}$:

$$g\left(x'_n - f(x'_n)\frac{\Delta t}{2}\right) = g\left(x'_{n+1} - f(x'_{n+1})\frac{\Delta t}{2}\right).$$

Therefore, for dual scheme (4) the conservation law takes the form

$$g\left(\hat{x} - f(\hat{x})\frac{\Delta t}{2}\right) = g\left(x - f(x)\frac{\Delta t}{2}\right). \tag{6}$$

Thus, a quadratic integral is also inherited by the trapezoidal scheme (4), but the expression for the conserved quantity coincides with $g$ only in the limit $\Delta t \to 0$. This circumstance made complicated finding it.

**Definition 1.** Let us say that a difference scheme inherits the integral $g(x) = C$ if there exists a rational function $G(x, \Delta t)$ such that

1) from the equations that specify the scheme it follows that

$$G(\hat{x}, \Delta t) = G(x, \Delta t),$$

2) in the limit $\Delta t \to 0$ expression $G(x, \Delta t)$ turns into $g(x)$

Function $G$ itself will be referred to as the difference analog of the integral $g$.

**Theorem 2 (Cooper's dual theorem).** *The trapezoidal scheme inherits all linear and quadratic integrals of motion, and the difference analogue of the integral $g$ will be*

$$g\left(x - f(x)\frac{\Delta t}{2}\right).$$

## 3. Examples

Consider several examples.

### 3.1. Linear oscillator

In the case of a linear dynamical system, the midpoint scheme and the trapezoidal scheme are the same, so the midpoint scheme becomes self-conjugate. This circumstance greatly simplifies the study of the midpoint scheme for a linear oscillator.

Consider a dynamic system

$$\dot{x} = -y, \quad \dot{y} = x \tag{7}$$

which has a quadratic integral

$$x^2 + y^2 = C.$$

The midpoint scheme will give points $(x_0, y_0), (x_1, y_1), ...$ lying on the circle

$$x^2 + y^2 = R^2,$$

the radius of which is determined by the initial point

$$R = \sqrt{x_0^2 + y_0^2}.$$

The midpoints of the links lying on the circle

$$x^2 + y^2 = r^2,$$

the radius of which can be determined by the first link

$$r = \sqrt{\left(\frac{x_1 + x_0}{2}\right)^2 + \left(\frac{y_1 + y_0}{2}\right)^2} = \frac{R}{\sqrt{1 + \Delta t^2/4}}.$$

Thus, the trajectory on the phase plane turns out to be a broken line, the vertices of which lie on a circle of radius $R$, and the links touch a concentric circle, the radius of which is $\sqrt{1 + \Delta t^2/4}$ times less than $R$. In particular, the trajectory will be closed, and the solution will be periodic if $R$ and $r$ are the radii of the circumscribed and inscribed circle in the $N$-gon, that is, if

$$r/R = \cos\frac{\pi}{N}.$$

This immediately gives the formula for choosing a step

$$\sqrt{1 + \Delta t^2/4} = \cos\frac{\pi}{N}.$$

This formula was previously obtained by us analytically [10].

### 3.2. Elliptic oscillator

By the definition of Jacobi functions [13],

$$p = \operatorname{sn} t, \quad q = \operatorname{cn} t, \quad r = \operatorname{dn} t$$

is a particular solution of the autonomous system of differential equations

$$\dot{p} = qr, \quad \dot{q} = -pr, \quad \dot{r} = -k^2 pq \tag{8}$$

with the initial conditions

$$p = 0, \quad q = r = 1 \quad \text{at } t = 0.$$

The midpoint scheme preserves both intergals

$$p^2 + q^2 = \text{const} \quad \text{and} \quad k^2 p^2 + r^2 = \text{const} \tag{9}$$

of this system. Now the trapezoidal scheme

$$\hat{p} - p = (\hat{\hat{q}}\hat{r} + qr)\,\frac{\Delta t}{2}, \ldots$$

does not coincide with the midpoint scheme and, therefore, its invariants are more complicated.

Nevertheless, the integral

$$p^2 + q^2 = \text{const}$$

corresponds to the integral

$$\left(p - qr\frac{\Delta t}{2}\right)^2 + \left(q + pr\frac{\Delta t}{2}\right)^2 = p^2 + q^2 + q^2r^2\frac{\Delta t^2}{4} + p^2r^2\frac{\Delta t^2}{4}$$

or

$$(p^2 + q^2)\left(1 + \frac{r^2\Delta t^2}{4}\right).$$

The integral

$$k^2p^2 + r^2 = \text{const}$$

correspods to

$$k^2\left(p - qr\frac{\Delta t}{2}\right)^2 + \left(r + k^2pq\frac{\Delta t}{2}\right)^2 = k^2p^2 + r^2 + k^2q^2r^2\frac{\Delta t^2}{4} + k^4p^2q^2\frac{\Delta t^2}{4}$$

or

$$(k^2p^2 + r^2)\left(1 + k^2q^2\frac{\Delta t^2}{4}\right).$$

Thus, in the space $pqr$ the vertices of the trajectory lie on the elliptic curve (9), and the midpoints of the links of the broken line lie on a more complex curve

$$(p^2 + q^2)\left(1 + \frac{r^2\Delta t^2}{4}\right) = C_1, \quad (k^2p^2 + r^2)\left(1 + k^2q^2\frac{\Delta t^2}{4}\right) = C_2. \quad (10)$$

This means that the trapezoidal scheme for an elliptic oscillator inherits both quadratic integrals, and their difference counterparts are the expressions (10).

If we follow the change in $p^2 + q^2$ on the approximate solution found by the trapezoidal scheme, then we will see a deviation from a constant value equal to

$$(p^2 + q^2)\frac{r^2\Delta t^2}{4}.$$

The exact solution is periodic, so in the plots these deviations appear as periodic fluctuations.

## 4. Conclusion

If you do not use specially developed difference schemes, discretization of continuous models by the method of finite differences introduces completely new properties into these models: calculations lead to a monotonic change in quantities, which, from physical considerations, must remain constant. For example, in computer experiments, dissipation appears even in those cases when energy was conserved in the original continuous model. In calculations for sufficiently long time intervals, this dissipation becomes very noticeable, and the parameters of the dynamical system are significantly distorted.

The number of schemes that preserve algebraic integrals of motion exactly is small and their drawbacks are well known. Difference schemes, in which the integrals of motion fluctuate around their initial values, significantly expand this set. However, the noted property is usually accepted without explanation and even more rigorous proof. Theorem 2, which is dual to Cooper's theorem, allows us to fill in this gap for the trapezoidal scheme (4) by explicitly specifying expressions that coincide in the limit $\Delta t \to 0$ with exact integrals and at the same time are preserved on approximate solutions exactly.

It would be very interesting to generalize this result to other schemes, the use of which does not lead to a monotonic increment of the integrals of motion. For such a generalization, in our opinion, it is necessary to investigate in more detail the question of schemes that are, in a sense, dual to the symplectic Runge-Kutta schemes.

## Acknowledgments

The publication was supported by the RUDN University Strategic Academic Leadership Program.

## References

[1] A. Goriely, "Integrability and nonintegrability of dynamical systems," in *Advanced Series in Nonlinear Dynamics*. Singapore; River Edge, NJ: World Scientific, 2001, vol. 19. DOI: 10.1142/3846.

[2] D. Greenspan, "Completely conservative, covariant numerical methodology," *Computers & Mathematics with Applications*, vol. 29, no. 4, pp. 37–43, 1995. DOI: 10.1016/0898-1221(94)00236-E.

[3] D. Greenspan, "Completely conservative, covariant numerical solution of systems of ordinary differential equations with applications," *Rendiconti del Seminario Matematico e Fisico di Milano*, vol. 65, pp. 63–87, 1995. DOI: 10.1007/BF02925253.

[4] J. C. Simo and M. A. González, "Assessment of Energy-momentum and Symplectic Schemes for Stiff Dynamical Systems," in *American Society of Mechanical Engineers. ASME Winter Annual Meeting*, New Orleans, Louisiana, 1993.

[5] E. Graham, G. Jelenić, and M. A. Crisfield, "A note on the equivalence of two recent time-integration schemes for N-body problems," *Communications in Numerical Methods in Engineering*, vol. 18, pp. 615–620, 2002. DOI: 10.1002/cnm.520.

[6] G. J. Cooper, "Stability of Runge–Kutta methods for trajectory problems," *IMA Journal of Numerical Analysis*, vol. 7, pp. 1–13, 1 1987. DOI: 10.1093/imanum/7.1.1.

[7] Y. B. Suris, "Hamiltonian methods of Runge–Kutta type and their variational interpretation [Gamil'tonovy metody tipa Runge–Kutty i ikh variatsionnaya traktovka]," *Matematicheskoe modelirovaniye*, vol. 2, no. 4, pp. 78–87, 1990, in Russian.

[8] J. M. Sanz-Serna, "Symplectic Runge–Kutta schemes for adjoint equations, automatic differentiation, optimal control, and more," *SIAM review*, vol. 58, pp. 3–33, 2016. DOI: 10.1137/151002769.

[9] E. Hairer, G. Wanner, and C. Lubich, *Geometric numerical integration. Structure-preserving algorithms for ordinary differential equations*. Berlin Heidelberg New York: Springer, 2000.

[10] V. P. Gerdt, M. D. Malykh, L. A. Sevastianov, and Yu Ying, "On the properties of numerical solutions of dynamical systems obtained using the midpoint method," *Discrete & Continuous Models & Applied Computational Science*, vol. 27, no. 3, pp. 242–262, 2019. DOI: 10.22363/2658-4670-2019-27-3-242-262.

[11] Y. A. Blinkov and V. P. Gerdt, "On computer algebra aided numerical solution of ODE by finite difference method," in *International Conference Polynomial Computer Algebra'2019; St. Petersburg, April 15–20, 2019*, N. N. Vassiliev, Ed., SPb: VVM Publishing, 2019, pp. 29–31.

[12] F. Klein, *Vorlesungen über Nicht-Euklidische Geometrie*. Springer, 1967. DOI: 10.1007/978-3-642-95026-1.

[13] P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists*. Springer, 1971. DOI: 10.1007/978-3-642-65138-0.

**For citation:**

Y. Ying, M. D. Malykh, On conjugate difference schemes: the midpoint scheme and the trapezoidal scheme, Discrete and Continuous Models and Applied Computational Science 29 (1) (2021) 63–72. DOI: 10.22363/2658-4670-2021-29-1-63-72.

**Information about the authors**:

**Ying, Yu** — Candidate of Physical and Mathematical Sciences, Assistant Professor of Department of Algebra and Geometry, Kaili University (e-mail: 45384377@qq.com, ORCID: https://orcid.org/0000-0002-4105-2566)

**Malykh, Mikhail D.** — Doctor of Physical and Mathematical Sciences, Assistant Professor of Department of Applied Probability and Informatics of Peoples' Friendship University of Russia (RUDN University); (e-mail: malykh_md@pfur.ru, phone: +7(495)9550927, ORCID: https://orcid.org/0000-0001-6541-6603, ResearcherID: P-8123-2016, Scopus Author ID: 6602318510)

УДК 519.872, 519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2021-29-1-63-72

# О сопряжённых разностных схемах: схема средней точки и схема трапеций

**Юй Ин**[1], **М. Д. Малых**[2]

[1] *Университет Кайли*
*Kaiyuan Road 3, Кайли, 556011, Китай*
[2] *Российский университет дружбы народов*
*ул. Миклухо-Маклая, д. 6, Москва, 117198, Россия*

В статье исследован вопрос о сохранении квадратичных интегралов на приближённых решениях автономных систем обыкновенных дифференциальных уравнений $\dot{x} = f(x)$, найденных по схеме трапеций. Установлена связь между схемой трапеции и схемой средней точки, которая сохраняет все квадратичные интегралы движения в силу теоремы Купера. Эта связь позволяет рассматривать схему трапеций как двойственную к схеме средней точки и отыскать двойственный аналог для теоремы Купера. Доказано, что на приближённом решении, найденном по симметрической схеме, сохраняется не сам квадратичный интеграл, а более сложное выражение, которое переходит в интеграл в пределе при $\Delta t \to 0$. Результаты проиллюстрированы примерами — линейным и эллиптическим осцилляторами. В обоих случаях в явном виде выписаны выражения, которые сохраняет схема трапеций.

**Ключевые слова:** динамические системы, квадратичные интегралы, разностные схемы, законы сохранения, схема средней точки, схема трапеций