DISCRETE AND CONTINUOUS MODELS AND APPLIED COMPUTATIONAL SCIENCE

Volume 32 Number 2 (2024)

Founded in 1993

Founder: PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA NAMED AFTER PATRICE LUMUMBA

DOI: 10.22363/2658-4670-2024-32-2

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 named after Patrice Lumumba (RUDN University).

Indexed by

- Scopus (https://www.scopus.com),
- Ulrich's Periodicals Directory (http://www.ulrichsweb.com),
- Directory of Open Access Journals (DOAJ) (https://doaj.org),
- Russian Index of Science Citation (https://elibrary.ru),
- 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: *Anna V. Korolkova*, *Dmitry S. Kulyabov*

English Text Editors: *Nikolay E. Nikolaev*, *Ivan S. Zaryadov*, *Konstantin P. Lovetskiy*

Address of editorial board:
3 Ordzhonikidze St, 115419 Moscow, Russia
Ph. +7 (495) 955-07-16, e-mail: publishing@rudn.ru

Editorial office:
Ph. +7 (495) 952-02-50, miphj@rudn.ru, dcm@sci.pfu.edu.ru,
site: http://journals.rudn.ru/miph

Paper size 70×108/16. Offset paper. Offset printing. Typeface “Adobe Source”.
Conventional printed sheet 9.80. Printing run 500 copies. Open price. The order 604.
PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA NAMED AFTER PATRICE LUMUMBA
6 Miklukho-Maklaya St, Moscow, 117198, Russia

Printed at RUDN Publishing House:
3 Ordzhonikidze St, Moscow, 115419, Russia
Ph. +7 (495) 952-04-41; e-mail: publishing@rudn.ru

Discrete & Continuous Models
& Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (2)
http://journals.rudn.ru/miph

Contents

Discrete & Continuous Models & Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (2) 135–139
http://journals.rudn.ru/miph

Editorial

DOI: 10.22363/2658-4670-2024-32-2-135–139

EDN: FPZJNY

Author's ethics

Dmitry S. Kulyabov[1,2], Leonid A. Sevastianov[1,2]

[1] *RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation*
[2] *Joint Institute for Nuclear Research, 6 Joliot-Curie St, Dubna, Moscow region, 141980, Russian Federation*

Abstract. We describe the system of ethical disclaimers being introduced in the journal.

Key words and phrases: author ethics, CRediT taxonomy, ethical disclaimer

Citation: Kulyabov D. S., Sevastianov L. A., Author's ethics. *Discrete and Continuous Models and Applied Computational Science* **32** (2), 135–139. doi: 10.22363/2658-4670-2024-32-2-135–139. edn: FPZJNY (2024).

1. Ethical disclaimers

The editorial board has decided to introduce a section with ethical disclaimers in the journal.

A disclaimer is a note of disclaimer of responsibility. For example, authors' statements about conflicts of interest, author contributions, acknowledgements, etc. Some disclaimers are integrated as metadata in international citation databases. If the full text of the manuscript is not available in the public domain, the disclaimers are usually placed on the journal's website together with other public bibliographic data of the article.

Disclaimers are usually placed at the end of the manuscript before the reference list. In a situation where a particular disclamer is not applicable to the research, the author should point out the irrelevance of its use.

We suggest using a mandatory standardized template of the journal's disclaimer. Let's take a closer look at the disclaimers used.

2. Author contributions

Researchers always face the problem of authorship [1]:

- the list of authors includes very different people from those who have made a principal contribution to the work;
- people are added against their wishes or without their knowledge;
- the order does not reflect the contribution of each author.

Contribution to authorship is determined by the overall involvement of the research participant in the study. The Committee on Publication Ethics (COPE) draws attention to the problem of authorship [2]. The CRediT system (https://credit.niso.org/) is proposed to formalize author roles. The CRediT (Contributor Roles Taxonomy) offers 14 possible author roles [3]. This is not really a taxonomy, but a faceted classification. Author roles are not always independent in themselves.

2.1. Contributor roles

1. *Conceptualization* Ideas; formulation or evolution of overarching research goals and aims.
2. *Data curation* Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.
3. *Formal analysis* Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.
4. *Funding acquisition* Acquisition of the financial support for the project leading to this publication.
5. *Investigation* Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.
6. *Methodology* Development or design of methodology; creation of models.
7. *Project administration* Management and coordination responsibility for the research activity planning and execution.
8. *Resources* Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.
9. *Software* Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.
10. *Supervision* Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.
11. *Validation* Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.
12. *Visualization* Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.
13. *Writing—original draft* Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).
14. *Writing—review & editing* Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision—including pre- or post-publication stages.

3. Acknowledgments

Acknowledgments should be addressed to specific individuals for specific contributions to the study and manuscript. Acknowledgments should refrain from thanking individuals who abstractly inspired or anonymously participated in the review. Acknowledgments are made to:

- people who contributed to the research and preparation of the manuscript of the article, but who do not meet all the criteria for authorship;
- people and organizations that have supported the research and preparation of the manuscript to some extent, or partially funded certain stages at the initiative of the authors.

It is advisable for authors to assure the journal (e.g., in a cover letter) that they have obtained permission to mention certain people in the acknowledgments section.

4. Funding

Disclaimer *funding* refers primarily to external funding if the research was externally initiated. If the research is entirely the initiative of the author's team, it is better to indicate gratitude for partial funding of some of the stages of the research in the *Acknowledgments* section. The fact that the

author's team has received external funding should be recorded in the disclaimer as a matter of course.

When mentioning the sponsor, its exact data (name of the organization, grant number, etc.) and the country of its location should be specified.

If there is any support, it is recommended to clarify in the *Conflicts of interest* section at which stages of the research and how the support was used.

If there is no external funding, it is written: *This research received no external funding*.

If it is impossible to obtain information from the authors about the source of funding, then write: *Not specified*.

5. Data availability statement

Data are particularly important in reproducible researches. In particular, when statistical methods are used. The data availability statement tells the reader where the research data related to the article are located and under what conditions the data can be accessed. References to the dataset are also provided.

Basic information provided to the reader:

- where the data can be accessed (data repository);
- a persistent identifier, such as a digital object identifier (doi) or access number, or a link to a permanent record of the dataset;
- details of any restrictions on access to the data and a reasonable explanation (e.g. for ethical, legal or commercial reasons).

5.1. Possible options for accessing the data and examples of description

- *The data are publicly available in the repository*. "The data supporting this study are publicly available in *repository name* at (doi, url)."
- *Data are available in the repository but are embargoed*. "Data supporting this study will be available from *repository name* at (doi, url) after a 6-month embargo."
- *Data are available from the repository but access is restricted due to legal, ethical or commercial reasons*. "The data supporting this study are available in *repository name* at (doi, url). Access to the data is subject to approval and data sharing agreement due to *reason*."
- *Secondary analysis of third-party data subject to restrictions.* "This study used third-party data provided under license, which the author does not have permission to disclose. Requests for access to the data should be directed to *third party* at *contacts*."
- *Data available as supplementary information.* "Data supporting this study are included in the article and/or supporting materials."
- *Data cannot be shared due to ethical, legal, or commercial restrictions*. "Data supporting this study cannot be made available due to *what reason*."
- *No new data is created or analyzed*. "No new data were created or analyzed in this study. Data sharing is not applicable."

6. Conflicts of interest

This disclaimer must be included.

Conflicts of interest can comment on various aspects, but usually the author's past or current employment is indicated. Grants (especially from for-profit companies) received not only by the author but also by the organization for which he or she works are indicated. If the author is associated with a sponsor, it is indicated where the research was conducted.

If there is no conflict of interest, then the corresponding statement should also be included (*The authors declare no conflict of interest*).

Author Contributions: The contributions of the authors are equal. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analysed during this study. Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Politzer, D. Nobel Lecture: The dilemma of attribution. *Reviews of Modern Physics* **77,** 851–856. doi:10.1103/revmodphys.77.851 (Sept. 2005).
2. *COPE Council. COPE Discussion Document: Authorship* 2nd ed. doi:10.24318/cope.2019.3.3 (Sept. 2019).
3. Holcombe, A. O. Contributorship, Not Authorship: Use CRediT to Indicate Who Did What. *Publications* **7,** 48.1–48.11. doi:10.3390/publications7030048 (July 2019).

Information about the authors

Dmitry S. Kulyabov (Russian Federation)—Docent, Doctor of Sciences in Physics and Mathematics, Professor of Department of Probability Theory and Cyber Security of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University); Senior Researcher of Laboratory of Information Technologies, Joint Institute for Nuclear Research (e-mail: kulyabov-ds@rudn.ru, phone: +7 (495) 952-02-50, ORCID: 0000-0002-0877-7063, ResearcherID: I-3183-2013, Scopus Author ID: 35194130800)

Leonid A. Sevastianov (Russian Federation)—Professor, Doctor of Sciences in Physics and Mathematics, Professor of Department of Computational Mathematics and Artificial Intelligence of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University) (e-mail: sevastianov-la@rudn.ru, phone: +7 (495) 955-07-83, ORCID: 0000-0002-1856-4643, ResearcherID: B-8497-2016, Scopus Author ID: 8783969400)

DOI: 10.22363/2658-4670-2024-32-2-135–139 EDN: FPZJNY

Авторская этика

Д. С. Кулябов[1,2], Л. А. Севастьянов[1,2]

[1] *Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация*
[2] *Объединённый институт ядерных исследований, ул. Жолио-Кюри, д. 6, Дубна, 141980, Российская Федерация*

Аннотация. Описывается вводимая в журнале система этических дисклеймеров.

Ключевые слова: авторская этика, таксономия CRediT, этические дисклеймеры

Discrete & Continuous Models
& Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (2) 140–153
http://journals.rudn.ru/miph

Research article

UDC 519.872
PACS 07.05.Tp
DOI: 10.22363/2658-4670-2024-32-2-140–153
EDN: CRYMNN

Marginal asymptotic diffusion analysis of two-class retrial queueing system with probabilistic priority as a model of two-modal communication networks

Anatoly A. Nazarov, Ekaterina A. Fedorova, Yana E. Izmailova

National Research Tomsk State University, 36 Lenina Ave, Tomsk, 634050, Russian Federation

(received: March 24, 2024; revised: April 12, 2024; accepted: April 20, 2024)

Abstract. In the paper, a retrial queueing system of $M_2/M_2/1$ type with probabilistic priority and interruptions is considered as a model of a two-modal communication network. Two classes of customers come to the system according Poisson arrival processes. There is one service device (or channel). If a customer finds the server occupying by a customer of the same class, it goes to an orbit and makes a repeated attempt after a random delay. If an arrival customer finds the other class customer on the server, it can interrupt its service with the given probability and start servicing itself. Customers from the orbit behave the same way. There is a multiply access for customers in the orbit. Service times and inter-retrial times have exponential distributions. Customers are assumed heterogeneous, so the parameters of the distributions are different for each class. In the paper, we propose the original marginal asymptotic-diffusion method for finding of the stationary probability distributions of the number of each class customers under the long delays condition.

Key words and phrases: two-class retrial queueing system, probabilistic priority, interruptions, marginal asymptotic-diffusion analysis

Citation: Nazarov A. A., Fedorova E. A., Izmailova Y. E., Marginal asymptotic diffusion analysis of two-class retrial queueing system with probabilistic priority as a model of two-modal communication networks. *Discrete and Continuous Models and Applied Computational Science* **32** (2), 140–153. doi: 10.22363/2658-4670-2024-32-2-140–153. edn: CRYMNN (2024).

1. Introduction

Heterogeneous of information is characteristic of modern telecommunication networks. Transmitted data may contain text, sound, image, service information, etc. Thus, we observe several arrival processes in networks with different characteristics (required job, quality of service, permissible latency, possibility of losses). In the field of robotics and telemedicine, multimedia networks are called as multimodal communication [1]. Interest in multimodal systems is increasing with the development of multimodal interfaces. By modality it is called physically recorded elements of communication (human-machine and/or human-human), including both the transmitted data (message) and individual information. The set of multimodal data and their size may vary depending on the task. So, in speech recognition systems based on audio recordings, it is sufficient 70–80 Kb for a speech modality [2], while for a sign modality (i.e. in Russian sign language), one modality record

can require 125 Mb [3]. One of solutions to ensure the quality of data transmission is to distribute user application data into several sub-streams and to provide multi-stream data transmission using several available communication interfaces. Often some modalities may be different by priority [4]. It means that some information is more significant than other in a particular task (i.e. urgent emergency messages must be delivered immediately that leads to interruption of lower priority information transmission).

Most of the works devoted to the problems of multi-stream data transmission are based on the analysis of data transmission quality parameters using simulation models. Here, the mathematical modelling is applied for analyzing and optimization of multimodal data transmission systems, taking into account the priorities, interruptions and the existence of repeated attempts of information transmission after fails. In this way, we propose the mathematical model of multimodal communication system in the form of a retrial queueing system with two classes of arrivals and opportunity of interruptions in the case of probabilistic priority.

Retrial queueing systems (or queueing system with repeated calls) [5, 6] are new class of queuing models widely applied in various communication systems (call-centers, cellular networks, LANs, etc. [7, 8]). In retrial queues, there is a some virtual place (the orbit) for repeated calls, where unserved calls wait during random time before an attempt to receive service again.

In spite of the large number of studies of retrial queueing systems of various configurations, heterogeneous models are weak investigated. Retrial queues with several types of customers (and several orbits too) are called as multiclass RQs and considered in [9–15]. Most of cited papers are devoted only stability analysis, while probability distributions or even means of processes under study are hardly investigated. Queuing models with interruption are proposed in [16, 17]. Different types of service interruptions are described in [18]. Queuing systems with probabilistic priority are presented in [19, 20]. Retrial queues with different types of priority are studied in [8, 21–23]. The most close study of retrial queues with two classes and priority are considered in [24, 25].

The rest of the paper is organized as follows. In Section 2, the mathematical model is described, the process under study is denoted and a system of differential Kolmogorov equations is written. In Section 3, we propose the original marginal asymptotic-diffusion analysis method for the two-class retrial queueing system studying. We derived the formula for the marginal asymptotic stationary probability distribution of number of each class calls in the orbit under the long delays limit condition. Section 5 consists some conclusions.

2. Mathematical model

Let us consider a retrial queueing system with two classes of customers. A customer of the n-th class comes to the system according Poisson arrival process with parameter λ_n, where $n = 1, 2$. There is one server. If the server is idle, the n-th class customer starts its servicing during the exponentially distributed random time with rate μ_n. We assume that customers have probabilistic priority. If an arrival customer finds the servicing customer of the same class, it goes to the orbit. If a customer finds the other class customer on the server, it can: a) with probability s_n interrupt the servicing and starts servicing itself (and the displaced customer goes to the orbit); b) with probability $1 - s_n$ joins to the orbit. After a random time distributed exponentially with rate σ_n, a customer from the orbit makes a repeated attempt to get service. The customers from the orbit behave in the same way (and with same probabilities). So, in the model we assume that there are no loses of customers.

Note that it does not matter to consider one common orbit for both classes of customers or two orbits for each class. It is important to distinguish a number of customers of each class in the system at some time moment. The model structure is presented on Figure 2.

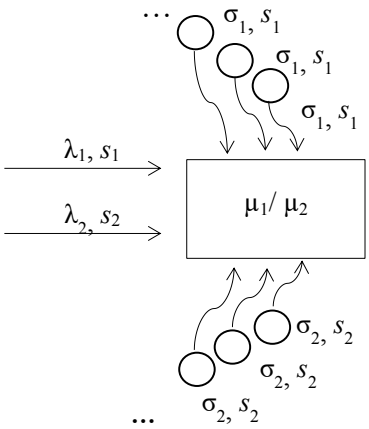


Figure 1. Two-class retrial queueing system with probabilistic priority

Note that in real life, the most common situation is when one class of customers is prioritized ($s_1 = 1$ and $s_2 = 0$). The considered mathematical model and obtained further results are general and can be applied for different cases.

Let us denote random processes of the number of the n-th class calls in the orbit by $i_n(t)$, where $n = 1, 2$. Process $k(t)$ determines states of the server as follows:

$$k(t) = \begin{cases} 0, & \text{if the server is free,} \\ 1, & \text{if the 1-st class customer is servicing,} \\ 2, & \text{if the 2-nd class customer is servicing.} \end{cases}$$

Denote by $P\{k(t) = k, i_1(t) = i_1, i_2(t) = i_2\} = P(k, i_1, i_2, t)$ the probability that the server has state k and there are i_1 customers of the first class and i_2 customers of the second class in the orbit at time t. Process $\{k(t), i_1(t), i_2(t)\}$ is three-dimensional continuous-time Markov chain. Let us write the following system of Kolmogorov equations for probability distribution $P(k, i_1, i_2, t)$:

$$\begin{cases} \dfrac{\partial P(0,i_1,i_2,t)}{\partial t} = -(\lambda_1+\lambda_2+i_1\sigma_1+i_1\sigma_2)P(0,i_1,i_2,t) + \\ \qquad\qquad + \mu_1 P(1,i_1,i_2,t) + \mu_2 P(2,i_1,i_2,t), \\ \dfrac{\partial P(1,i_1,i_2,t)}{\partial t} = -(\lambda_1+\lambda_2+\mu_1+i_2 s_2\sigma_2)P(1,i_1,i_2,t) + \lambda_1 P(0,i_1,i_2,t) + \\ \qquad + (i_1+1)\sigma_1 P(0,i_1+1,i_2,t) + \lambda_1 s_1 P(2,i_1,i_2-1,t) + \\ \qquad + \lambda_2(1-s_2)P(1,i_1,i_2-1,t) + \lambda_1 P(1,i_1-1,i_2,t) + \\ \qquad\qquad + (i_1+1)\sigma_1 s_1 P(2,i_1+1,i_2-1,t), \\ \dfrac{\partial P(2,i_1,i_2,t)}{\partial t} = -(\lambda_1+\lambda_2+\mu_2+i_1 s_1\sigma_1)P(2,i_1,i_2,t) + \lambda_2 P(0,i_1,i_2,t) + \\ \qquad + (i_2+1)\sigma_2 P(0,i_1,i_2+1,t) + \lambda_2 s_2 P(1,i_1-1,i_2,t) + \\ \qquad + \lambda_1(1-s_1)P(2,i_1-1,i_2,t) + \lambda_2 P(2,i_1,i_2-1,t) + \\ \qquad\qquad + (i_2+1)\sigma_2 s_2 P(1,i_1-1,i_2+1,t). \end{cases} \tag{1}$$

Let us introduce the partial characteristic functions:

$$H(k,u_1,u_2,t)=\sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}e^{ju_1i_1}\cdot e^{ju_2i_2}P(k,i_1,i_2,t).$$

Then we rewrite Equations (1) for the characteristic functions as follows

$$\begin{cases}\dfrac{\partial H(0,u_1,u_2,t)}{\partial t}=-(\lambda_1+\lambda_2)H(0,u_1,u_2,t)+j\sigma_1\dfrac{\partial H(0,u_1,u_2,t)}{\partial u_1}+\\ \quad+j\sigma_2\dfrac{\partial H(0,u_1,u_2,t)}{\partial u_2}+\mu_1H(1,u_1,u_2,t)+\mu_2H(2,u_1,u_2,t),\\ \dfrac{\partial H(1,u_1,u_2,t)}{\partial t}=-(\lambda_1+\lambda_2+\mu_1)H(1,u_1,u_2,t)+\lambda_1H(0,u_1,u_2,t)+\\ \quad+j\sigma_2s_2\dfrac{\partial H(1,u_1,u_2,t)}{\partial u_2}-j\sigma_1e^{-ju_1}\dfrac{\partial H(0,u_1,u_2,t)}{\partial u_1}+\\ \quad+\lambda_1s_1e^{ju_2}H(2,u_1,u_2,t)+\lambda_2(1-s_2)e^{ju_2}H(1,u_1,u_2,t)+\\ \quad+\lambda_1e^{ju_1}H(1,u_1,u_2,t)-j\sigma_1s_1e^{-ju_1}e^{ju_2}\dfrac{\partial H(2,u_1,u_2,t)}{\partial u_1},\\ \dfrac{\partial H(2,u_1,u_2,t)}{\partial t}=-(\lambda_1+\lambda_2+\mu_2)H(2,u_1,u_2,t)+\lambda_2H(0,u_1,u_2,t)+\\ \quad+j\sigma_1s_1\dfrac{\partial H(2,u_1,u_2,t)}{\partial u_1}-j\sigma_2e^{-ju_2}\dfrac{\partial H(0,u_1,u_2,t)}{\partial u_2}+\\ \quad+\lambda_2s_2e^{ju_1}H(1,u_1,u_2,t)+\lambda_1(1-s_1)e^{ju_1}H(2,u_1,u_2,t)+\\ \quad+\lambda_2e^{ju_2}H(2,u_1,u_2,t)-j\sigma_2s_2e^{-ju_2}e^{ju_1}\dfrac{\partial H(1,u_1,u_2,t)}{\partial u_2}.\end{cases}\tag{2}$$

Introduce matrix form of the characteristic functions:

$$\mathbf{H}(u_1,u_2,t)=\{H(0,u_1,u_2,t),H(1,u_1,u_2,t),H(2,u_1,u_2,t)\}.$$

So Equations (2) are transformed in the following matrix equation:

$$\begin{aligned}\frac{\partial\mathbf{H}(u_1,u_2,t)}{\partial t}&=\mathbf{H}(u_1,u_2,t)(\mathbf{A}+e^{ju_1}\mathbf{B}_1+e^{ju_2}\mathbf{B}_2)+\\&\quad+j\sigma_1\frac{\partial\mathbf{H}(u_1,u_2,t)}{\partial u_1}(\mathbf{I}_1-e^{-ju_1}\mathbf{I}_2-e^{-ju_1}e^{ju_2}\mathbf{I}_3)+\\&\quad+j\sigma_2\frac{\partial\mathbf{H}(u_1,u_2,t)}{\partial u_2}(\mathbf{I}_4-e^{-ju_2}\mathbf{I}_5-e^{ju_1}e^{-ju_2}\mathbf{I}_6),\end{aligned}\tag{3}$$

where

$$\mathbf{A}=\begin{pmatrix}-(\lambda_1+\lambda_2)&\lambda_1&\lambda_2\\ \mu_1&-(\lambda_1+\lambda_2+\mu_1)&0\\ \mu_2&0&-(\lambda_1+\lambda_2+\mu_2)\end{pmatrix},$$

$$\mathbf{B}_1=\begin{pmatrix}0&0&0\\0&\lambda_1&\lambda_2s_2\\0&0&\lambda_1(1-s_1)\end{pmatrix},\quad\mathbf{B}_2=\begin{pmatrix}0&0&0\\0&\lambda_2(1-s_2)&0\\0&\lambda_1s_1&\lambda_2\end{pmatrix},$$

$$\mathbf{I}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & s_1 \end{pmatrix}, \quad \mathbf{I}_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{I}_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & s_1 & 0 \end{pmatrix},$$

$$\mathbf{I}_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{I}_5 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{I}_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & s_2 \\ 0 & 0 & 0 \end{pmatrix}.$$

3. Asymptotic Analysis

The further analysis can be devided on several stages:

1. Derivation of "marginal" asymptotic equations for each process $i_1(t)$ and $i_2(t)$ under a long delay limit condition.
2. Finding of the asymptotic stationary means of numbers of calls of each class and stationary probabilities of the server states.
3. Implementation of asymptotic-diffusion analysis for "marginal" asymptotic equations. Finding of the asymptotic stationary probability distribution of numbers of customers of each class.

3.1. Marginal asymptotic equations

Let us derive marginal asymptotic equations for process $i_1(t)$ when parameter $\sigma_2 \to 0$. First of all, we introduce infinitesimal parameter ε and substitutions

$$\sigma_2 = \gamma_2 \sigma, \quad \sigma = \varepsilon, \quad u_2 = \varepsilon w,$$

$$\mathbf{H}(u_1, u_2, t) = \mathbf{F}(u_1, w, t, \varepsilon).$$

From System (3), we obtain the following asymptotic matrix equation:

$$\begin{aligned} \frac{\partial \mathbf{F}(u_1, w, t, \varepsilon)}{\partial t} &= \mathbf{F}(u_1, w, t, \varepsilon)(\mathbf{A} + e^{ju_1}\mathbf{B}_1 + e^{jw\varepsilon}\mathbf{B}_2) + \\ &+ j\sigma_1 \frac{\partial \mathbf{F}(u_1, w, t, \varepsilon)}{\partial u_1}(\mathbf{I}_1 - e^{-ju_1}\mathbf{I}_2 - e^{-ju_1}e^{jw\varepsilon}\mathbf{I}_3) + \\ &+ j\gamma_2 \frac{\partial \mathbf{F}(u_1, w, t, \varepsilon)}{\partial w}(\mathbf{I}_4 - e^{-jw\varepsilon}\mathbf{I}_5 - e^{ju_1}e^{-jw\varepsilon}\mathbf{I}_6). \end{aligned}$$

Under limit $\varepsilon \to 0$, we have

$$\begin{aligned} \frac{\partial \mathbf{F}(u_1, w, t)}{\partial t} &= \mathbf{F}(u_1, w, t)(\mathbf{A} + e^{ju_1}\mathbf{B}_1 + \mathbf{B}_2) + \\ &+ j\sigma_1 \frac{\partial \mathbf{F}(u_1, w, t)}{\partial u_1}(\mathbf{I}_1 - e^{-ju_1}\mathbf{I}_2 - e^{-ju_1}\mathbf{I}_3) + \\ &+ j\gamma_2 \frac{\partial \mathbf{F}(u_1, w, t)}{\partial w}(\mathbf{I}_4 - \mathbf{I}_5 - e^{ju_1}\mathbf{I}_6). \end{aligned} \tag{4}$$

Let the solution have the following form:

$$\mathbf{F}(u_1, w, t) = \mathbf{H}_1(u_1, t)e^{jw \cdot x_2}.$$

By substituting in Equation (4), we have

$$\frac{\partial \mathbf{H}_1(u_1,t)}{\partial t} = \mathbf{H}_1(u_1,t)(\mathbf{A}+\mathbf{B}_2+\gamma_2 x_2(\mathbf{I}_5-\mathbf{I}_4)+e^{ju_1}(\mathbf{B}_1+\gamma_2 x_2\mathbf{I}_6))+ \\ +j\sigma_1\frac{\partial \mathbf{H}_1(u_1,t)}{\partial u_1}(\mathbf{I}_1-e^{-ju_1}(\mathbf{I}_2+\mathbf{I}_3)). \tag{5}$$

In the same way, the following marginal asymptotic equation for process $i_2(t)$ (when $\sigma_1 \to 0$) can be derived:

$$\frac{\partial \mathbf{H}_2(u_2,t)}{\partial t} = \mathbf{H}_2(u_2,t)(\mathbf{A}+\mathbf{B}_1+\gamma_1 x_1(\mathbf{I}_2-\mathbf{I}_1)+e^{ju_2}(\mathbf{B}_2+\gamma_1 x_1\mathbf{I}_3))+ \\ +j\sigma_2\frac{\partial \mathbf{H}_2(u_2,t)}{\partial u_2}(\mathbf{I}_4-e^{-ju_2}(\mathbf{I}_5+\mathbf{I}_6)). \tag{6}$$

3.2. Asymptotic means

The next step of the study is finding of parameters x_n, $n=1,2$. Let us write Equation (3) in the steady state

$$\mathbf{H}(u_1,u_2)(\mathbf{A}+e^{ju_1}\mathbf{B}_1+e^{ju_2}\mathbf{B}_2)+j\sigma_1\frac{\partial \mathbf{H}(u_1,u_2)}{\partial u_1}(\mathbf{I}_1-e^{-ju_1}\mathbf{I}_2-e^{-ju_1}e^{ju_2}\mathbf{I}_3)+ \\ +j\sigma_2\frac{\partial \mathbf{H}(u_1,u_2)}{\partial u_2}(\mathbf{I}_4-e^{-ju_2}\mathbf{I}_5-e^{ju_1}e^{-ju_2}\mathbf{I}_6)=\mathbf{0}. \tag{7}$$

For obtaining an additional scalar equation, we multiply (7) by unit vector $\mathbf{e}$. Taking into account the form of matrix $\mathbf{A}$, $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{I}_\upsilon$ ($\upsilon=\overline{1,6}$), we obtain

$$\mathbf{H}(u_1,u_2)((e^{ju_1}-1)\mathbf{B}_1+(e^{ju_2}-1)\mathbf{B}_2)\mathbf{e}+ \\ +j\sigma_1\frac{\partial \mathbf{H}(u_1,u_2)}{\partial u_1}((1-e^{-ju_1})\mathbf{I}_2+(1-e^{-ju_1}e^{ju_2})\mathbf{I}_3)\mathbf{e}+ \\ +j\sigma_2\frac{\partial \mathbf{H}(u_1,u_2)}{\partial u_2}((1-e^{-ju_2})\mathbf{I}_5+(1-e^{ju_1}e^{-ju_2})\mathbf{I}_6)\mathbf{e}=0. \tag{8}$$

Let us use the following substitutions:

$$\sigma_n=\gamma_n\sigma,\ \sigma=\varepsilon,\ u_n=\varepsilon w_n,\ \mathbf{H}(u_1,u_2)=F(w_1,w_2,\varepsilon),\ n=1,2.$$

From Equations (7)–(8), we obtain the following system:

$$\begin{cases} \mathbf{F}(w_1,w_2,\varepsilon)(\mathbf{A}+e^{jw_1\varepsilon}\mathbf{B}_1+e^{jw_2\varepsilon}\mathbf{B}_2)+ \\ +j\gamma_1\dfrac{\partial \mathbf{F}(w_1,w_2,\varepsilon)}{\partial w_1}(\mathbf{I}_1-e^{-jw_1\varepsilon}\mathbf{I}_2-e^{-jw_1\varepsilon}e^{jw_2\varepsilon}\mathbf{I}_3)+ \\ +j\gamma_2\dfrac{\partial \mathbf{F}(w_1,w_2,\varepsilon)}{\partial w_2}(\mathbf{I}_4-e^{-jw_2\varepsilon}\mathbf{I}_5-e^{jw_1\varepsilon}e^{-jw_2\varepsilon}\mathbf{I}_6)=\mathbf{0}, \\ \mathbf{F}(w_1,w_2,\varepsilon)((e^{jw_1\varepsilon}-1)\mathbf{B}_1+(e^{jw_2\varepsilon}-1)\mathbf{B}_2)\mathbf{e}+ \\ +j\gamma_1\dfrac{\partial \mathbf{F}(w_1,w_2,\varepsilon)}{\partial w_1}((1-e^{-jw_1\varepsilon})\mathbf{I}_2+(1-e^{-jw_1\varepsilon}e^{jw_2\varepsilon})\mathbf{I}_3)\mathbf{e}+ \\ +j\gamma_2\dfrac{\partial \mathbf{F}(w_1,w_2,\varepsilon)}{\partial w_2}((1-e^{-jw_2\varepsilon})\mathbf{I}_5+(1-e^{jw_1\varepsilon}e^{-jw_2\varepsilon})\mathbf{I}_6)\mathbf{e}=0. \end{cases} \tag{9}$$

Using Maclaurin series, Equations (9) are written under $\varepsilon \to 0$ as follows

$$\begin{cases} \mathbf{F}(w_1, w_2)(\mathbf{A} + \mathbf{B}_1 + \mathbf{B}_2) + j\gamma_1 \dfrac{\partial \mathbf{F}(w_1, w_2)}{\partial w_1}(\mathbf{I}_1 - \mathbf{I}_2 - \mathbf{I}_3) + \\ + j\gamma_2 \dfrac{\partial \mathbf{F}(w_1, w_2)}{\partial w_2}(\mathbf{I}_4 - \mathbf{I}_5 - \mathbf{I}_6) = \mathbf{0}, \\ \mathbf{F}(w_1, w_2)(w_1\mathbf{B}_1 + w_2\mathbf{B}_2)\mathbf{e} + j\gamma_1 \dfrac{\partial \mathbf{F}(w_1, w_2)}{\partial w_1}(w_1\mathbf{I}_2 + (w_1 - w_2)\mathbf{I}_3)\mathbf{e} + \\ + j\gamma_2 \dfrac{\partial \mathbf{F}(w_1, w_2)}{\partial w_2}(w_2\mathbf{I}_5 + (w_2 - w_1)\mathbf{I}_6)\mathbf{e} = 0. \end{cases} \tag{10}$$

Let us find the solution in the following form:

$$\mathbf{F}(w_1, w_2) = \mathbf{R} \cdot \exp\{jw_1x_1 + jw_2x_2\}.$$

By substituting into (10), we get

$$\begin{cases} \mathbf{R}(\mathbf{A} + \mathbf{B}_1 + \mathbf{B}_2) - \gamma_1 x_1 \mathbf{R}(\mathbf{I}_1 - \mathbf{I}_2 - \mathbf{I}_3)\mathbf{e} - \gamma_2 x_2 \mathbf{R}(\mathbf{I}_4 - \mathbf{I}_5 - \mathbf{I}_6) = \mathbf{0}, \\ \mathbf{R}\mathbf{B}_1\mathbf{e} - \gamma_1 x_1 \mathbf{R}(\mathbf{I}_2 + \mathbf{I}_3)\mathbf{e} + \gamma_2 x_2 \mathbf{R}\mathbf{I}_6\mathbf{e} = 0, \\ \mathbf{R}\mathbf{B}_2\mathbf{e} + \gamma_1 x_1 \mathbf{R}\mathbf{I}_3\mathbf{e} - \gamma_2 x_2 \mathbf{R}(\mathbf{I}_5 + \mathbf{I}_6)\mathbf{e} = 0. \end{cases}$$

By solving the system above and taking into account the normalization condition $\mathbf{Re} = 1$, we can obtain the values of stationary probabilities of server states R_k and asymptotic stationary means $M\{i_1(t)\} = \gamma_1 x_1/\sigma_1$ and $M\{i_2(t)\} = \gamma_2 x_2/\sigma_2$.

3.3. Asymptotic-diffusion analysis

Let us solve asymptotic Equations (5)–(6). First of all, we write this equations in general form as follows

$$\frac{\partial \mathbf{H}(u,t)}{\partial t} = \mathbf{H}(u,t)(\mathbf{A}_1 + e^{ju}\mathbf{A}_2) + j\sigma \frac{\partial \mathbf{H}(u,t)}{\partial u}(\mathbf{J}_1 - e^{-ju}\mathbf{J}_2), \tag{11}$$

where matrices $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{J}_1$, $\mathbf{J}_2$ have the following values for corresponding classes of customers:

– for the first class of customers (process $i_1(t)$):

$$\begin{aligned} \mathbf{A}_1^{(1)} &= \mathbf{A} + \mathbf{B}_2 + \gamma_2 x_2(\mathbf{I}_5 - \mathbf{I}_4), \\ \mathbf{A}_2^{(1)} &= \mathbf{B}_1 + \gamma_2 x_2 \mathbf{I}_6, \\ \mathbf{J}_1^{(1)} &= \mathbf{I}_1, \\ \mathbf{J}_2^{(1)} &= \mathbf{I}_2 + \mathbf{I}_4; \end{aligned} \tag{12}$$

– for the second class of customers (process $i_2(t)$):

$$\begin{aligned} \mathbf{A}_1^{(2)} &= \mathbf{A} + \mathbf{B}_1 + \gamma_1 x_1(\mathbf{I}_2 - \mathbf{I}_1), \\ \mathbf{A}_2^{(2)} &= \mathbf{B}_2 + \gamma_1 x_1 \mathbf{I}_3, \\ \mathbf{J}_1^{(2)} &= \mathbf{I}_4, \\ \mathbf{J}_2^{(2)} &= \mathbf{I}_5 + \mathbf{I}_6. \end{aligned} \tag{13}$$

By multiplying Equation (11) by unit vector $\mathbf{e}$, we have an additional equation:

$$\frac{\partial \mathbf{H}(u,t)\mathbf{e}}{\partial t} = (e^{ju} - 1)\left(\mathbf{H}(u,t)\mathbf{A}_2 + j\sigma \frac{\partial \mathbf{H}(u,t)}{\partial u} e^{-ju}\mathbf{J}_2\right)\mathbf{e}. \tag{14}$$

3.3.1. First asymptotics

Let us denote

$$\sigma = \varepsilon,\ \sigma t = \varepsilon t = \tau,\ u = \varepsilon w,\ \mathbf{H}(u,t) = \mathbf{F}(w,\tau,\varepsilon).$$

Then equations (11) and (14) are rewritten as

$$\varepsilon\frac{\partial \mathbf{F}(w,\tau,\varepsilon)}{\partial \tau} = \mathbf{F}(w,\tau,\varepsilon)(\mathbf{A}_1 + e^{jw\varepsilon}\mathbf{A}_2) + j\frac{\partial \mathbf{F}(w,\tau,\varepsilon)}{\partial w}(\mathbf{J}_1 - e^{-jw\varepsilon}\mathbf{J}_2), \tag{15}$$

$$\varepsilon\frac{\partial \mathbf{F}(w,\tau,\varepsilon)\mathbf{e}}{\partial \tau} = (e^{jw\varepsilon} - 1)\left(\mathbf{F}(w,\tau,\varepsilon)\mathbf{A}_2 + j\frac{\partial \mathbf{F}(w,\tau,\varepsilon)}{\partial w}e^{-jw\varepsilon}\mathbf{J}_2\right)\mathbf{e}. \tag{16}$$

Under $\varepsilon \to 0$, Equation (15) has the form

$$\mathbf{F}(w,\tau)(\mathbf{A}_1 + \mathbf{A}_2) + j\frac{\partial \mathbf{F}(w,\tau)}{\partial w}(\mathbf{J}_1 - \mathbf{J}_2) = \mathbf{0}. \tag{17}$$

We assume that the solution has the following form

$$\mathbf{F}(w,\tau) = \mathbf{R}\cdot e^{jwx(\tau)}, \tag{18}$$

where values of vector $\mathbf{R}$ depend on $x(\tau)$ too.

By substituting (18) into Equation (17) and taking into account normalization condition, we obtain the following system for $\mathbf{R}$ finding:

$$\begin{cases} \mathbf{R}(\mathbf{A}_1 + \mathbf{A}_2) - x(\tau)\mathbf{R}(\mathbf{J}_1 - \mathbf{J}_2) = \mathbf{0}, \\ \mathbf{R}\mathbf{e} = 1. \end{cases} \tag{19}$$

By using Maclaurin series in Equation (16), we have the following equation under $\varepsilon \to 0$

$$\frac{\partial \mathbf{F}(w,\tau)\mathbf{e}}{\partial \tau} = jw\left(\mathbf{F}(w,\tau)\mathbf{A}_2 + j\frac{\partial \mathbf{F}(w,\tau)}{\partial w}\mathbf{J}_2\right)\mathbf{e}. \tag{20}$$

Substituting (18) into (20), we obtain that

$$\frac{dx(\tau)}{d\tau} = (\mathbf{R}\mathbf{A}_2 - x(\tau)\mathbf{R}\mathbf{J}_2)\,\mathbf{e}.$$

A derived equation is a differential equation for $x(\tau)$ with a transfer coefficient:

$$a(x) = \mathbf{R}\mathbf{A}_2\mathbf{e} - x(\tau)\mathbf{R}\mathbf{J}_2\mathbf{e}. \tag{21}$$

3.3.2. Second asymptotics

Let us make the following substitution:

$$\mathbf{H}(u,t) = \mathbf{H}^{(2)}(u,t)\exp\left\{\frac{ju}{\sigma}x(\sigma t)\right\}. \tag{22}$$

By substituting (22) into Equations (11) and (14), we have:

$$\begin{aligned}\frac{\partial \mathbf{H}^{(2)}(u,t)}{\partial t} + jua(x)\mathbf{H}^{(2)}(u,t) = j\sigma\frac{\partial \mathbf{H}^{(2)}(u,t)}{\partial u}(\mathbf{J}_1 - e^{-ju}\mathbf{J}_2) + \\ +\mathbf{H}^{(2)}(u,t)\left(\mathbf{A}_1 + e^{ju}\mathbf{A}_2 - x(\sigma t)(\mathbf{J}_1 - e^{-ju}\mathbf{J}_2)\right),\end{aligned} \tag{23}$$

$$\frac{\partial \mathbf{H}^{(2)}(u,t)\mathbf{e}}{\partial t} + jua(x)\mathbf{H}^{(2)}(u,t)\mathbf{e} = \\ = (e^{ju}-1)\left(\mathbf{H}^{(2)}(u,t)(\mathbf{A}_2 - e^{-ju}x(\sigma t)\mathbf{J}_2) + j\sigma e^{-ju}\frac{\partial \mathbf{H}^{(2)}(u,t)}{\partial u}\mathbf{J}_2\right)\mathbf{e}. \tag{24}$$

Let us denote

$$\sigma = \varepsilon^2, \quad \sigma t = \varepsilon^2 t = \tau, \quad u = \varepsilon w, \quad \mathbf{H}^{(2)}(u,t) = \mathbf{F}^{(2)}(w,\tau,\varepsilon).$$

Thus, Equations (23)–(24) are rewritten as follows

$$\varepsilon^2\frac{\partial \mathbf{F}^{(2)}(w,\tau,\varepsilon)}{\partial \tau} + jw\varepsilon a(x)\mathbf{F}^{(2)}(w,\tau,\varepsilon) = j\varepsilon\frac{\partial \mathbf{F}^{(2)}(w,\tau,\varepsilon)}{\partial w}(\mathbf{J}_1 - e^{-jw\varepsilon}\mathbf{J}_2) + \\ +\mathbf{F}^{(2)}(w,\tau,\varepsilon)\left(\mathbf{A}_1 + e^{jw\varepsilon}\mathbf{A}_2 - x(\tau)(\mathbf{J}_1 - e^{-jw\varepsilon}\mathbf{J}_2)\right), \tag{25}$$

$$\varepsilon^2\frac{\partial \mathbf{F}^{(2)}(w,\tau,\varepsilon)\mathbf{e}}{\partial \tau} + jw\varepsilon a(x)\mathbf{F}^{(2)}(w,\tau,\varepsilon)\mathbf{e} = \\ = (e^{jw\varepsilon}-1)\left(\mathbf{F}^{(2)}(w,\tau,\varepsilon)(\mathbf{A}_2 - e^{-jw\varepsilon}x(\tau)\mathbf{J}_2) + j\varepsilon e^{-jw\varepsilon}\frac{\partial \mathbf{F}^{(2)}(w,\tau,\varepsilon)}{\partial w}\mathbf{J}_2\right)\mathbf{e}. \tag{26}$$

For simplicity of notation, in further expressions we will write $x(\tau) = x$.

Let the solution be in the following form

$$\mathbf{F}^{(2)}(w,\tau,\varepsilon) = \Phi(w,\tau)(\mathbf{R} + jw\varepsilon\mathbf{f}) + \mathbf{O}(\varepsilon^2). \tag{27}$$

By substituting (27) into Equation (25), after some transforms we obtain

$$\varepsilon\frac{\partial \Phi(w,\tau)}{\partial \tau}(\mathbf{R} + jw\varepsilon\mathbf{f}) + jw\varepsilon a(x)\Phi(w,\tau)(\mathbf{R} + jw\varepsilon\mathbf{f}) = \\ = \Phi(w,\tau)\mathbf{R}(jw\mathbf{A}_2 - jwx\mathbf{J}_2) + jw\Phi(w,\tau)\mathbf{f}(\mathbf{A}_1 + \mathbf{A}_2 + x(\mathbf{J}_2 - \mathbf{J}_1)) + \\ +j\frac{\partial \Phi(w,\tau)}{\partial w}(\mathbf{R} + jw\varepsilon\mathbf{f})(\mathbf{J}_1 - (1 - jw\varepsilon)\mathbf{J}_2) - \\ -\varepsilon\Phi(w,\tau)\mathbf{f}(\mathbf{J}_1 - (1 - jw\varepsilon)\mathbf{J}_2) + \mathbf{O}(\varepsilon^2).$$

Under $\varepsilon \to 0$, we have

$$\mathbf{f}(\mathbf{A}_1 + \mathbf{A}_2 + x(\mathbf{J}_2 - \mathbf{J}_1)) = a(x)\mathbf{R} - \mathbf{R}(\mathbf{A}_2 - x\mathbf{J}_2) - \frac{\partial \Phi(w,\tau)/\partial w}{w\Phi(w,\tau)}\mathbf{R}(\mathbf{J}_1 - \mathbf{J}_2).$$

It is obvious that vector $\mathbf{f}$ can be written as the sum

$$\mathbf{f} = C\mathbf{R} + \mathbf{g} - \mathbf{q}\frac{\partial \Phi(w,\tau)/\partial w}{w\Phi(w,\tau)}, \tag{28}$$

where $C = const$, vectors $\mathbf{g}$ and $\mathbf{q}$ are defined by the following equations:

$$\begin{cases} \mathbf{g}(\mathbf{A}_1 + \mathbf{A}_2 + x(\mathbf{J}_2 - \mathbf{J}_1)) = a(x)\mathbf{R} - \mathbf{R}(\mathbf{A}_2 - x\mathbf{J}_2), \\ \mathbf{q}(\mathbf{A}_1 + \mathbf{A}_2 + x(\mathbf{J}_2 - \mathbf{J}_1)) = \mathbf{R}(\mathbf{J}_1 - \mathbf{J}_2), \\ \mathbf{g}\mathbf{e} = 0, \\ \mathbf{q}\mathbf{e} = 0. \end{cases}$$

By comparing the second equation with (19), we can note that $\mathbf{q} = \mathbf{R}'(x)$.

For finding function $\Phi_n(w,\tau)$, we substitute Expression (27) into Equation (26)

$$\varepsilon^2\frac{\partial\Phi(w,\tau)}{\partial\tau} + jw\varepsilon a(x)\Phi(w,\tau)(1+jw\varepsilon\mathbf{fe}) =$$
$$= jw\varepsilon\Phi(w,\tau)(\mathbf{R}+jw\varepsilon\mathbf{f})(\mathbf{A}_2-(1-jw\varepsilon)x(\tau)\mathbf{J}_2)\mathbf{e}+$$
$$+\frac{(jw\varepsilon)^2}{2}\Phi(w,\tau)\mathbf{R}(\mathbf{A}_2-x(\tau)\mathbf{J}_2)\mathbf{e} + (j\varepsilon)^2 w\frac{\partial\Phi(w,\tau)}{\partial w}\mathbf{R}\mathbf{J}_2\mathbf{e} + O(\varepsilon^3).$$

After some transforms, we obtain

$$\frac{\partial\Phi(w,\tau)}{\partial\tau} = (jw)^2\Phi(w,\tau)\times$$
$$\times\left(-a(x)\mathbf{fe}+\mathbf{f}(\mathbf{A}_2-x\mathbf{J}_2)\mathbf{e}+x\mathbf{R}\mathbf{J}_2\mathbf{e}+\frac{\partial\Phi(w,\tau)/\partial w}{w\Phi(w,\tau)}\mathbf{R}\mathbf{J}_2\mathbf{e}+\frac{a(x)}{2}\right)+O(\varepsilon).$$

Taking into account (28), under $\varepsilon\to 0$ we finally have

$$\frac{\partial\Phi(w,\tau)}{\partial\tau} = (jw)^2\Phi(w,\tau)\times$$
$$\times\left(\mathbf{g}(\mathbf{A}_2-x\mathbf{J}_2)\mathbf{e}+x\mathbf{R}\mathbf{J}_2\mathbf{e}+\frac{\partial\Phi(w,\tau)/\partial w}{w\Phi(w,\tau)}(\mathbf{R}\mathbf{J}_2\mathbf{e}-\mathbf{q}(\mathbf{A}_2-x\mathbf{J}_2)\mathbf{e})+\frac{a(x)}{2}\right).$$

So, we obtain the following equation

$$\frac{\partial\Phi(w,\tau)}{\partial\tau} = w\frac{\partial\Phi(w,\tau)}{\partial w}a'(x)+\frac{(jw)^2}{2}\Phi(w,\tau)b(x),$$

where

$$b(x) = a(x) + 2\left[\mathbf{g}(\mathbf{A}_2-x\mathbf{J}_2)+x\mathbf{R}\mathbf{J}_2\right]\mathbf{e}. \tag{29}$$

Let us introduce probability distribution density function

$$P(y,\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-jwy}\Phi(w,\tau)dw$$

of diffusion process $dy(\tau) = y(\tau)a'(x)d\tau + \sqrt{b(x)}dw(\tau)$, where $w(\tau)$ is a Wiener process. We can write the following Fokker-Planck equation for $P(y,\tau)$:

$$\frac{\partial P(y,\tau)}{\partial\tau} = -\frac{\partial}{\partial y}\left(P(y,\tau)ya'(x)\right)+\frac{1}{2}\frac{\partial^2}{\partial y^2}(P(y,\tau)b(x)).$$

Combining the results of two asymptotics, we introduce process $z(\tau) = x(\tau)+\varepsilon y(\tau)$, such as $dz(\tau) = a(z)d\tau + \sqrt{\sigma b(z)}dw(\tau)$. which is diffusion random process satisfying the following Fokker-Planck equation

$$\frac{\partial P(z,\tau)}{\partial\tau} = -\frac{\partial\left(P(z,\tau)a(z)\right)}{\partial z}+\frac{1}{2}\frac{\partial^2(P(z,\tau)\sigma b(z))}{\partial z^2}.$$

In steady state, it is easy to obtain the following expression for probability distribution density of process $z(\tau)$:

$$P(z) = \frac{C}{b(z)}\exp\left\{\frac{\sigma}{2}\int_0^z\frac{a(x)}{b(x)}dx\right\}.$$

Returning to processes $i_n(t)$, we conclude that stationary probabilities of numbers of each class customers in the orbit are calculated as follows:

$$P_n(i_n) = \frac{C}{b(\sigma_n i_n)}\exp\left\{\frac{\sigma_n}{2}\int_0^{\sigma_n i_n}\frac{a_n(x)}{b_n(x)}dx\right\}, \tag{30}$$

where $C = \text{const}$ obtained from the normalization condition, parameters $a_n(x)$ and $b_n(x)$ are defined by expressions (21) and (29) with corresponding matrices (12)–(13).

4. Results and discussion

In this way, it was shown that asymptotic marginal distributions of the number of calls of each class in the orbit have the form (30). Undoubtedly, in real stochastic processes of the numbers of k-class calls $i_k(t)$ (where $k = 1, 2$) are correlated, but the obtaining a two-dimensional distribution is difficult (or may be impossible). The asymptotic marginal method let us analyze marginal distributions with enough high accuracy. In addition we obtained expression for transfer coefficients (21), which can help us to define stability conditions of the system (if all $a_k(\infty) < 0$) or the condition of partial stability (if only $a_1(\infty) < 0$ or only $a_2(\infty) < 0$).

5. Conclusions

In the paper, we have proposed an original method of the marginal asymptotic-diffusion analysis of the two-class retrial queueing systems studying. We have considered non-classical model of retrial queues – systems with interruptions of servicing, furthermore, customers have probabilistic priority. Such model has not yet been studied analytically. The proposed model and analytical results allow to evaluate the effectiveness of various scenarios of multimodal data transmission with possibility priority e.g. by setting of interruption probabilities or delay rates.

Note that in real life (i.e. in communication networks) the most common situation is when one class of customers is prioritized ($s_1 = 1$ and $s_2 = 0$). The considered mathematical model is more general and the obtained results can be applied in different cases.

In further research, we plan to study more complex models of multimodal comunication networks, such as multiclass retrial queues with probabilistic priority, multiclass retrial queues with constant retrial rate and models with non-Poisson arrivals.

Author Contributions: Conceptualization, A.Nazarov and E.Fedorova; methodology, A.Nazarov and E.Fedorova; software, E.Fedorova; validation, Y.Izmailova and E.Fedorova; formal analysis, E.Fedorova; investigation, A.Nazarov, Y.Izmailova and E.Fedorova; writing—original draft preparation, Y.Izmailova and E.Fedorova; writing—review and editing, A.Nazarov. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Funding: The research is supported by Russian Science Foundation according to the research project No. 24-21-00454.

References

1. Al Jaafreh, M. Multimodal systems, experiences, and communications: A review toward the tactile internet vision. *Recent Trends in Computer Applications,* 191–220 (2018).
2. Matveev, Y. N. Technologies for biometric personal identification by voice and other modalities (in Russian). Russian. *Engineering Journal: Science and Innovation* **3,** 46–61 (2012).
3. Kagirov, I. A. Multimedia database of Russian sign language gestures in three-dimensional format (in Russian). Russian. *Questions of Linguistics* **1,** 104–123 (2020).
4. Ryndin, A., Pakulova, E., Basov, O. & Veselov, G. *Modelling of multi-path transmission system of various priority multimodal information* in *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)* (2020), 1–5. doi:10.1109/AICT50176.2020.9368802.
5. Artalejo, J. R. & Gomez-Corral, A. *Retrial Queueing Systems* 318 pp. (Springer Berlin, 2008).
6. Falin, G. & Templeton, J. *Retrial Queues* 320 pp. (Taylor & Francis, 1997).

7. Phung-Duc, T. Retrial Queueing Models: A Survey on Theory and Applications. *Stochastic Operations Research in Business and Industry,* 1–26 (May 2017).
8. Makeeva, E., Kochetkova, I. & Alkanhel, R. Retrial Queueing System for Analyzing Impact of Priority Ultra-Reliable Low-Latency Communication Transmission on Enhanced Mobile Broadband Quality of Service Degradation in 5G Networks. *Mathematics, MDPI* **11,** 1–23. doi:10.3390/math11183878 (2023).
9. Avrachenkov, K., Morozov, E. & Nekrasova, R. *Optimal and Equilibrium Retrial Rates in Single-Server Multi-orbit Retrial Systems* in *Lecture Notes in Computer Science* **9305** (Springer, Cham, 2015), 135–146. doi:10.1007/978-3-319-23440-3_11.
10. Morozov, E., Rumyantsev, A., Dey, S. & Deepak, T. G. Performance analysis and stability of multiclass orbit queue with constant retrial rates and balking. *Performance Evaluation* **134,** 102005. doi:10.1016/j.peva.2019.102005 (2019).
11. Krishnamoorthy, A., Joshua, V. & Mathew, A. *A Retrial Queueing System with Multiple Hierarchial Orbits and Orbital Search* in *Communications in Computer and Information Science, vol. 919* (Springer, Cham, 2018), 224–233. doi:10.1007/978-3-319-99447-5_19.
12. Kim, B. & Kim, J. Proof of the conjecture on the stability of a multi-class retrial queue with constant retrial rates. *Queueing System* **104,** 175–185. doi:10.1007/s11134-023-09881-z (2023).
13. Kim, B. & Kim, J. Stability of a multi-class multi-server retrial queueing system with service times depending on classes and servers. *Queueing System* **94,** 129–146. doi:10.1007/s11134-019-09634-x (2020).
14. Avrachenkov, K. Stability and partial instability of multi-class retrial queues. *Queueing Systems* **100,** 177–179. doi:10.1007/s11134-022-09814-2hal-03767703 (2022).
15. Shin, Y. W. & Moon, D. H. M/M/c Retrial Queue with Multiclass of Customers. *Methodology and Computing in Applied Probability* **16,** 931–949. doi:10.1007/s11009-013-9340-0 (2014).
16. White, H. C. & Christie, L. S. Queueing with preemptive priorities or with breakdown. *Operations Research* **6,** 79–95. doi:10.1287/opre.6.1.79 (1958).
17. Gaver, D. A waiting line with interrupted service including priority. *J. Roy. Stat. Soc. B24* **24,** 73–90. doi:10.3390/sym11030419 (1962).
18. Fiems, D. & Bruneel, H. Queueing systems with different types of server interruptions. *Eur. J. Oper. Res.* **188,** 838–845 (2008).
19. Razumchik, R. *Two-priority queueing system with LCFS service, probabilistic priority and batch arrivals* in *AIP Conference Proceedings* **2116** (July 2019), 090011. doi:10.1063/1.5114076.
20. Yin, M., Yan, M., Guo, Y. & Liu, M. Analysis of a Pre-Emptive Two-Priority Queuing System with Impatient Customers and Heterogeneous Servers. *Mathematics* **11.** doi:10.3390/math11183878 (2023).
21. Babu, D., Krishnamoorthy, A. & Joshua, V. C. *Retrial Queue with Search of Interrupted Customers from the Finite Orbit* in *Information Technologies and Mathematical Modelling. Queueing Theory and Applications* **912** (Springer International Publishing, Cham, 2018), 360–371. doi:10.1007/978-3-319-97595-5_28.
22. Ammar, S. I. & Rajadurai, P. Performance Analysis of Preemptive Priority Retrial Queueing System with Disaster under Working Breakdown Services. *Symmetry* **11.** doi:10 . 3390 / sym11030419 (2019).
23. Atencia, I. A Geo/G/1 retrial queueing system with priority services. *Eur. J. Oper. Res.* **256,** 178–186. doi:10.1016/j.ejor.2016.07.011 (2017).
24. Jain, M., Bhagat, A. & Shekhar, C. Double orbit finite retrial queues with priority customers and service interruptions. *Appl. Math. Comput.* **253,** 324–344 (2015).

25. Nazarov, A. & Izmailova, Y. Study of the RQ-system M(2)/B(x)(2)/1 with R-persistent displacement of alternative customers (in Russian). Russian. *Bulletin of the Siberian State Aerospace University named after academician M.F. Reshetneva* **17,** 328–334 (2016).

Information about the authors

Nazarov, Anatoly A.—Doctor of Technical Sciences, professor of Department of Probability Theory and Mathematical Statistic of National Research Tomsk State University (e-mail: nazarov.tsu@gmail.com, phone: +7 (913) 879-70-33, ORCID: 0000-0002-5097-5629, ResearcherID: O-5862-2014, Scopus Author ID: 7201780364)

Fedorova, Ekaterina A.—PhD in Physical and Mathematical Sciences, associate professor of Department of Probability Theory and Mathematical Statistic of National Research Tomsk State University (e-mail: moiskate@mail.ru, phone: +7 (952) 801-93-94, ORCID: 0000-0001-8933-5322, ResearcherID: E-3161-2017, Scopus Author ID: 56439120600)

Izmailova, Yana E.—PhD in Physical and Mathematical Sciences, associate professor of Department of Probability Theory and Mathematical Statistic of National Research Tomsk State University (e-mail: evgenevna.92@mail.ru, phone: +7 (923) 430-89-47, ORCID: 0000-0002-9132-0127, ResearcherID: T-6377-2017, Scopus Author ID: 57191051392)

UDC 519.872
PACS 07.05.Tp
DOI: 10.22363/2658-4670-2024-32-2-140–153 EDN: CRYMNN

Маргинальный асимптотически-диффузионный анализ двуклассовой RQ-системы с вероятностным приоритетом как математической модели сети связи с двумодальной информацией

А. А. Назаров, Е. А. Фёдорова, Я. Е. Измайлова

Национальный исследовательский Томский государственный университет, пр-т. Ленина, д. 36, Томск, 634050, Российская Федерация

Аннотация. В работе исследуется RQ-система $M_2/M_2/1$ с вероятностным приоритетом и вытеснением заявок как модель двумодальной сети связи. На вход системы поступает два класса заявок, т.е. два потока. В системе имеется одно обслуживающее устройство (канал связи). Если входящая заявка застает прибор занятым заявкой того же класса, она идет на орбиту и осуществляет случайную задержку через экспоненциально распределенное случайное время. Если же на приборе находится заявка другого типа, то с некоторой вероятностью возможно прерывание обслуживания (вытеснение заявки). Необслуженная заявка уходит на орбиту. Обращаясь к прибору с орбиты, заявки действуют тем же образом. Время обслуживания каждой заявки распредлено экспоненциально. На орбите реализован протокол множественного доступа. В статье предложен оригинальный метод маргинального асимптотически-диффузионного анализа в условии большой задержки заявок на орбите для нахождения стационарных распределений вероятностей числа заявок каждого типа в системе.

Ключевые слова: RQ-системы, теория массового обслуживания, вероятностный приоритет, вытеснение, асимптотически-диффузионный анализ

Discrete & Continuous Models
& Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (2) 154–171
http://journals.rudn.ru/miph

Review

UDC 519.872
DOI: 10.22363/2658-4670-2024-32-2-154–171 EDN: CMFCZB

Chronology of the development of active queue management algorithms of RED family. Part 3: from 2016 up to 2024

Ivan S. Zaryadov[1,2], Viana C.C. Hilquias[1], Anna V. Korolkova[1], Tatiana A. Milovanova[1]

[1] *RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation*
[2] *Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, 44 Vavilova St, bldg 2, Moscow, 119333, Russian Federation*

(received: February 12, 2024; revised: March 10, 2024; accepted: March 12, 2024)

Abstract. This work is the first part of a large bibliographic review of active queue management algorithms of the Random Early Detection (RED) family, presented in the scientific press from 1993 to 2023. The third part will provide data on algorithms published from 2016 to 2023.

Key words and phrases: active queue management, AQM, random early detection, RED, congestion control

Citation: Zaryadov I. S., Hilquias V. C., Korolkova A. V., Milovanova T. A., Chronology of the development of active queue management algorithms of RED family. Part 3: from 2016 up to 2024. *Discrete and Continuous Models and Applied Computational Science* **32** (2), 154–171. doi: 10.22363/2658-4670-2024-32-2-154–171. edn: CMFCZB (2024).

1. Introduction

This work is the third (the last) part of the brief bibliographic review of algorithms of the Random Early Detection (RED) family, compiled according to the dates of publication of scientific works (articles and conference proceedings) in which the algorithms in question were presented to the public. The previous parts were presented in [1, 2]

The authors do not claim that the prepared review includes all existing algorithms, but is the most complete of those published previously, since it includes bibliographic data on more then 240 algorithms.

The characteristics of the RED algorithm, as the reasons for its modifications were presented and described in [1, 2].

The review is structured as follows. Each subsequent section is dedicated to one year, and it presents algorithms of the RED family, scientific publications (articles in scientific journals, conference proceedings, technical reports, etc.) on which were presented this year. In Section 11 the authors discussed the results and the future research directions are highlighted.

2. 2016

The new AQM based congestion control mechanism for 4G/LTE networks with a minimal adjustment to classical RED algorithm [3] was proposed in [4] and named Smart RED (SmRED). In this algorithm in order to achieve the optimal end-to-end performance by regulating the queue size the packet drop probability function was divided into two sections for distinguishing of different network loads and dynamical adaptation to the levels of network load.

The novel version of RED, based on fuzzy logic, was introduced in [5] and named Fuzzy logic-based RED (FL-RED). In FL-RED the thresholds are dynamically increased or decreased according to fuzzy logic rules in order to utilize the resource effectively.

The combined modified version of NLRED [6] and REDwM [7] algorithms, Non-linear RED with Weighted Moving Average (NLREDwM) was proposed and analyzed in [8]. In this algorithm the average queue length was defined by the first order difference equation.

The new version of RED algorithm based on the optimized link state routing protocol (OLSR), named Optimized Link RED (OLRED), was described in [9]. OLRED algorithm selects optimal routing link periodically and calculates the average of queue size one time.

The version of Weighted RED (WRED) [10] with implementation of the optimized link state routing protocol (OLSR) was also presented in [9] and named as Optimized Link WRED (OLWRED). OLWRED algorithm selects optimal routing link and calculates the average of queue size periodically. Thus OLWRED is more sensitive to average queue size. The OLWRED algorithm was presented in more detail in [11].

The improved variant of RED algorithm to stability and thus named Stability RED (S-RED) was described in [12]. S-RED was focused on improving linear structure of packet drop probability function (the quadratic function was proposed) and reducing the quantity of setting parameters (the parameters $p_{\max}$ and $Q_{\max}$ were removed).

The modified version of Adaptive Gentle RED (AGRED) [13], designed for the detection of the congestion in early stage before the router buffer overflows and employing a dynamic setting of queue weight parameter w_q for the congestion control, was introduced in [14] and named Enhanced Adaptive GRED (EAGRED).

3. 2017

In order to guarantee the QoS requirements for real-time applications and to prevent starvation of the best effort traffic (TCP-based applications) in the presence of non-TCP-based applications (real-time traffics) the new modification of RED was introduced in [15] and named Dynamic Queue RED (DQRED). Incoming packets were classified by the DQRED algorithm into three class types, each of which was handled by one of three queues in the internet router. The queued packets were scheduled dynamically based on the number of active packets already existing in the queues during a predefined time interval.

In [16, 17] the Adaptive Queue Random Early Detection (AQMRD) algorithm was described. In this algorithm the information not just about the average queue size but also the rate of change of it was incorporated. Also the adaptively changing threshold level between $Q_{\min}$ and $Q_{\max}$ thresholds was introduced. For dynamic tuning of AQMRD algorithm parameters stochastic approximation based optimization scheme was proposed in [18].

The modified version of the Effective RED (ERED) [19] algorithm based on fuzzy logic (Fuzzy Inference Process) was introduced in [20] and named Fuzzy ERED (FERED).

The new version of RED algorithm with two virtual queues for TCP and UD traffic in order to minimize the packet loss of sensitive traffic flows was proposed in [21] and named MultiRED (MRED). The proposed algorithm takes into consideration the specific characteristics of the real-time traffic and provides a solution that keeps all the advantages of the existing active queue management QoS schemes and reduces as much as possible dropped packets due to buffer overflow.

The modified version of the Multi-level RED (MRED) [22] algorithm, highly depending on the weight w_q parameter and maximum drop probability $p_{\max}$, named Fair Weighted Multi-level RED (FWMRED), was described in [23].

The new RED algorithm with nonlinear probability drop function, based on the Hemi-rise cloud membership model, was proposed in [24] and named Cloud RED (CRED).

The extension of Smart RED algorithm [12] with probability drop function for different traffic-load scenarios was presented in [25] and called SmRED-i.

In [26] the Three-section RED (TRED) algorithm was proposed. The drop probability function in TRED is nonlinear and divided into three sections to distinguish between light, moderate, and high loads to achieve a tradeoff in the delay and the throughput between low and high traffic loads.

The new RED algorithm, which takes into account instantaneous and average queue sizes for the estimation of the queue occupancy, was described in [27] and named Adaptive Threshold RED (AT-RED).

4. 2018

The modified version of the Dynamic RED (DRED) algorithm [28] with fuzzy proportional integral derivative (FuzzyPID, FPID) controller, named DRED-FPID (Dynamic RED with Fuzzy Proportional Integral Derivative (FPID) controller) was proposed in [29] for congestion control by adjusting the transmission rate and packet loss probability calculation.

The modification of Adaptive Gentle RED (AGRED) [13] and Adaptive RED (ARED) [30] algorithms, was proposed in [31] and named Adaptive Gentle Adaptive RED (AGARED). The AGARED was aimed to provide better congestion control over the network taking advantages of AGRED and ARED algorithms by modifying the value of maximum drop probability $p_{\max}$ and enhancing the parameter setting of average queue size $\hat{Q}$.

In order to resolve this problem of parameter setting in Adaptive Queue Management with Random Dropping (AQMRD) algorithm [17] it was proposed in [18] to use the optimization technique based on stochastic approximation and thus the new version of the AQMRD was introduced — the Optimized AQMRD (OAQMRD) algorithm.

The new version of RED algorithm, based on Modified Gaussian Function based RED (MGF-RED) [32] and Modified RED (MDRED) [33] algorithms was introduced in [34] for tuning the values of RED thresholds $Q_{\min}$ and $Q_{\max}$ and packet drop probability and named as Hybrid Modified RED (HMDRED).

The modification of RED algorithm, based on monitoring of the average queue size $\hat{Q}$ for every single output queue, was named Probability based RED (P-RED) and introduced in [35] in order to enhance fairness by isolating ill-behaved flows and protecting the bursty and low speed flows in a larger amount.

The new version of Dynamic GRED (DGRED) algorithm [36], based on Markov-Modulated Bernoulli process (MMBP) and therefore named Markov-Modulated Bernoulli Dynamic GRED (MMBDGRED), was presented in [37]. The reason to use MMBP was to implement DGRED with multiple traffic classes with different priorities for each class and to enhance the mechanism of DGRED in stabilizing the average queue size $\hat{Q}$ between the given values of $Q_{\min}$ and $Q_{\max}$ thresholds (using variable calculating parameters stored in the MMBP states).

In [38] the new improved RED algorithm, named LTRED (L means the queue length and T means the thresholds), which incorporated the impact of load variation in early congestion notification along with tuning of threshold parameter of RED, was presented.

One more variant of RED algorithm using fuzzy logic to overcome problems of classical RED was proposed in [39] and called fuzzy logic RED (FLRED). Tn FLRED algorithm to predict and avoid congestion at an early stage the average queue size $\hat{Q}$ and the new parameter, delay D_{Spec}, were used.

In [40] the version of RED, where the packet drop probability function was divided into three sections (three dropping probabilities for different type of incoming traffic) with using Additive Increase Multiplicative Decrease (AIMD) technique, was introduced and named Modified RED (ModRED).

The new modification of RED, based on Q-learning algorithm for adjustment of the maximum drop probability $p_{\max}$ according to the network situation and named as Q-learning RED (QRED), was presented in [41].

The extended version of the Weighted RED (WRED) [10, 42] algorithm with the delay as one of the parameters of the algorithm (the estimation of which was based on the load factor) and with a type-2 fuzzy logic system, used to handle the nonlinear parameter uncertainties and to obtain more satisfactory results in terms of packet loss in the event of heavy congestion, was proposed in [43, 44].

The modification of DRED-FPID [29] with Deep Neural Network (DNN) was introduced in [45] and named as Dynamic RED with Fuzzy-Deep Neural Network Proportional Integral Derivative (DRED-FDNNPID). In DRED-FDNNPID the network traffic estimation was performed by using modified additive increase decrease scheme for congestion resolving.

For the congestion control in complex heterogeneous networks the version of Adaptive RED algorithm [30] with Bell type Fuzzy membership function (as the probability drop function) was proposed in [46] and named as BARED.

5. 2019

The modification of the RED [3] algorithm with a nonlinear probability drop function at the midpoint between the minimum $Q_{\min}$ and maximum $Q_{\max}$ thresholds was presented in [47] and named Half-Way RED (HRED).

The improved RED [3] algorithm for multimedia traffic was described in [48] by using reconfigurable approach to redefining the maximum drop probability $p_{\max}$ parameter by using the reconfiguration factor γ depending on current average queue size $\hat{Q}$, thresholds $Q_{\min}$ and $Q_{\max}$, target average queue size. The proposed algorithm was named as RED with Reconfigurable Maximum Dropping Probability (RRMDP).

The modified version of the Gentle RED algorithm [49] was proposed in [50] and named as Improved Gentle RED (IGRED). In IGRED the number of parameters used for definition of the probability drop function was reduced. The evaluation of the IGRED performance and comparison with other algorithms was conducted in [51].

The development of the Dynamic Gentle RED (DGRED) algorithm [36] with dynamic change of thresholds values in order to stabilize the average queue size, named as Stabilised Dynamic GRED (SDGRED), was presented in [52].

In [53] the Enhanced RED (EnRED) algorithm was presented. The current queue size q is used to determine the reaction scenario for packet dropping for each moment of a packet arrival, and the probability drop function was defined by the average queue size $\hat{Q}$.

Also in [53] the another version of RED algorithm, named as Time-Window Augmented RED (Windowed-RED), was described, where the average queue length on a limited time window $\hat{Q}_w$ was introduced instead of $\hat{Q}$ to determine the reaction scenario and to calculate drop probabilities.

The Novel RED algorithm was introduced in [54], where only the formula for the probability drop function was changed (the drop function remains linear, but when calculating the drop probability the threshold values $Q_{\min}$ and $Q_{\max}$ were squared, also the average queue size $\hat{Q}$ was squared).

The new congestion control mechanism for Constrained Application Protocol (CoAP) Observe Group Communication, based on the modification of the RED algorithm and a Fibonacci Pre-Increment Backoff (FPB) (to obtain retransmission timeout (RTO) estimations for the transmission of the CoAP messages) and named Congestion Control Random Early Detection (CoCo-RED), was proposed in [55].

In [56] the delay parameter was introduced to RED, so the proposed new algorithm was named Delay-Controller RED (DcRED). The delay parameter was calculated for each packet arrival, and the resulted value was used as a reference for the packet dropping. Because the delay conflicted with the throughput and the dropping rate, the balance between throughput and delay was maintained.

In order to reduce the packet loss and to improve the throughput the extension to the classic RED algorithm [3] was developed in [57] and named DyRED, The average queue size $\hat{Q}$ and the instance queue size q were used in DyRED as congestion indicators and the maximum threshold $Q_{\max}$ was dynamically changed to control the congestion in the router buffer at the early stage.

6. 2020

In [58], the RED-Exponential (RED_E) algorithm was introduced, and it was proposed to drop arriving packets in an exponential manner (the exponential form of the probability drop function) without utilizing the maximum packet dropping probability.

The new RED algorithm with current traffic load as a congestion indicator, was proposed [59] and named Flexible Random Early Detection (FXRED). In order to maintain stable performance FXRED tuned its drop probability suitable to the observed load situation (increasing if load becomes high in order to avoid overflow and congestion, decreasing as load becomes low in order to maximize link utilization and throughput).

In [60] the new congestion control method, named Double Index RED (DI-RED) and based on usage of a cache state with a dual threshold in the router buffer, for wireless sensor networks (WSNs) was introduced. In DI-RED a channel transmission state was implemented as a new indicator to control the packet drop probability, based on the control of the average queue size.

The three nonlinear modifications of the Gentle RED (GRED) algorithm [49] were presented in [61]. The first version was the nonlinear NLGRED_1 algorithm with fixed value of the $Q_{\max}$ threshold and variable value of the maximum drop probability $p_{\max}$. The second version was the nonlinear NLGRED_2 algorithm with variable value of the $Q_{\max}$ threshold and fixed value of the maximum drop probability $p_{\max}$. And the third modification was the Reconfigurable Nonlinear GRED (RNLGRED) algorithm, switching between the NLGRED_1 and NLGRED_2 algorithms.

The new version of the Adaptive RED (ARED) algorithm [30] with modified probability drop function (the nonlinear quadratic function instead of the linear) was described in [62] and named Quadratic Adaptive RED (QARED).

The Weight Queue Dynamic Active Queue Management algorithm (WQDAQM), based on the Stabilized Dynamic GRED (SDGRED) algorithm [52], was proposed in [63]. WQDAQM was designed to control and manage the packet drop probability by maintaining the queue weight w_q and the thresholds dynamically.

7. 2021

Based on the S-RED [12] the Quadratic Random Early Detection (QRED) algorithm was developed in [64] with a quadratic probability drop function (instead of linear one) and a new parameter, called target queue q_t for better use of buffer space. This parameter was defined as the difference between the current queue size and the average of $Q_{\min}$ and $Q_{\max}$ thresholds.

In [65] the Self-Adaptive RED (SARED) algorithm was proposed. In this modification of the Adaptive RED algorithm [30] not only average queue size $\hat{Q}$ was considered as a congestion indicator, but also the ratio of the total capacity demand of the current traffic to the available output (bottleneck) link's capacity (traffic load). In SARED, the maximum drop probability $p_{\max}$ was adapted based on an observed network's traffic load. Two different probability drop functions were introduced: linear for a high load in order to be aggressive for avoiding congestion and forced packet drops and nonlinear for low and moderate loads with increasing value of the degree of nonlinearity for decreasing load in order to be very gentle so as to avoid link under-utilization.

The Enhanced Congestion Control RED based mechanism (EnCoCo-RED) for CoAP observe group communication as modification of the CoCo-RED algorithm [55] was presented in [66]. The version of RED algorithm, which was used for CoCo-RED, was improved with dynamic buffer management in order to provide a selective backoff algorithm (SBA) and include a congestion measurement process.

The new version of the RED algorithm with dynamic tuning of RED parameters ($Q_{\min}$ and $Q_{\max}$ thresholds, the maximum drop probability $p_{\max}$), named Congestion Control Algorithm using Buffer Occupancy RED (CCA-BO-RED), was introduced in [67]. In the CCA-BO-RED algorithm the rate of occupancy of the queue, called Buffer Occupancy (BOC), was proposed as a new congestion parameter. This new parameter infers network conditions from the rapidity of the buffer occupancy in the router. Also it was proposed to use the additive-increase and multiplicative-decrease approach for the maximum drop probability $p_{\max}$ instead of multiplicative increase and decrease.

In [68] several modifications of the RED algorithm were described. The first model, called Application of Dynamic Weight with Distance to Improve the Performance of RED (ADWD-RED-IP), was based on dynamic setting queue parameters, especially weight parameter w_q. The second model was proposed for reducing the rate of packet loss by using both the average queue size and the current queue size as the arguments of the packet drop function and was named Active Queue Management in RED to Reduce Packet Loss (AQM-RED-RPL). The third model, the Predictable Active Queue Management to Reduce Sensitivity of RED Parameter (PAQM-RS-RED), was based on the Adaptive RED algorithm [69] with adaptation of the maximum drop probability $p_{\max}$ to retain the average queue size between $Q_{\min}$ and $Q_{\max}$ thresholds. The Revolutionary Active Queue Management Model Utilizing Queue Size Threshold Change (IAQM-TA-QZ) algorithm was presented as the fourth model with dynamic adaptation of $Q_{\min}$ and $Q_{\max}$ thresholds and drop probability to the load condition of traffic. In the fifth model, Nobel Congestion Control Algorithm Using Buffer Occupancy RED (CCA-BO-RED), the dynamic tuning of basic RED parameters based on the rate of occupancy of the queue, as a congestion parameter, was introduced as in [67]. Finally, in the sixth model, Active Queue Management in RED considering Critical Point on Target Queue (AQM-RED-CPTQ), the new parameter called the goal queue q_t, defined as in [64], was introduced. The first and the second models, Application of Dynamic Weight with Distance to Improve the Performance of RED (ADWD-RED-IP) and Active Queue Management in RED to Reduce Packet Loss (AQM-RED-RPL) were also described in [70].

The improved RED-based algorithm, named Quadratic-Linear RED (QLRED), was presented in [71]. In this algorithm the probability drop function was divided into two parts: the quadratic packet drop function in order to achieve a low packet dropping probability at low traffic loads and the linear packet drop function in order to achieve a high packet dropping probability at high traffic loads.

8. 2022

The new AQM algorithm, based on Gentle RED (GRED) [49], Adaptive Gentle RED (AGRED) [13], Effective RED (ERED) [19] and Enhanced Adaptive Gentle RED (Enhanced AGRED) [14] algorithms was considered in [72] and named as Changeable Dynamic Gentle RED (CDGRED). In order to stabilize the average queue size $\hat{q}$ the dynamic changing of two thresholds ($Q_{\max}$ and $2Q_{\max}$) out of three ($Q_{\min}$, $Q_{\max}$ and $2Q_{\max}$) was proposed. Also the comparison of the CDGRED performance with FLRED [39], DGRED [36] and Enhanced AGRED [14] algorithms was conducted.

The new version of classic RED algorithm [3] with two linear packet drop functions for light and heavy traffic loads instead of a single linear drop function in RED [3], named RED-Improved (RED-I), was proposed in [73].

The improved version of Adaptive RED (ARED) algorithm [30] for satellite networks with cubic drop function was presented in [74] and named Improved ARED (I-ARED).

The new AQM algorithm for improving the overall end-to-end delay, throughput, and packet delivery ratio of the massive NB-IoT 5G network while using UDP was introduced in [75] and called Aggressive RED (AgRED). In AgRED algorithm the sigmoid function was used for calculation of drop probabilities for the arriving packets in the initialization phase, with a higher probability of dropping broadcast packets.

In [76] the modification of the RED active queue management algorithm with mixed shape drop function (the linear drop function for low and moderate network traffic loads and the exponential function for computing the packet drop probabilities for high traffic loads) was proposed and named RED-linear exponential (RED-LE).

Another modification of the RED algorithm with combination of a nonlinear (the quadratic drop function to ensure a slow increase of the packet drop probability from 0 to $p_{\max}$ for smaller average queue size when congestion was not too serious) and a linear (to ensure a fast increase of the packet drop probability from $p_{\max}$ to 1 for larger average queue size when congestion was very serious) packet dropping functions was introduced in [77] and named Improved RED (I-RED).

The new version of the RED algorithm with the ascending ridge function as a part of the probability drop function was proposed in [78] and named RED with the ascending ridge function (ARRED). The thresholds $Q_{\min}$ and $Q_{\max}$ were used as parameters of the ascending ridge function.

The novel modification of the RED algorithm, based on fuzzy logic similar to the FCRED (FconRED) [79], GREDFL [80], FERED [20] and FLRED [39] algorithms, and thus named the Fuzzy Comprehensive RED (FCRED), was described in [81] in order to deal with the gap in network monitoring and congestion control at the router buffer by using fuzzy inference process for management of three introduced indicators (average-based queue-related, arrival-related and departure-related) and the drop probability calculation.

The robust version of the RED algorithm, based on ideas from [82, 83], was presented in [84, 85] and called Beta RED (BetaRED) because of using the normalized incomplete beta function as a nonlinear packet drop probability function with dependence on two new parameters μ and σ (the mean and the standard deviation of the beta distribution). The basic idea of BetaRED was to maintain the average queue size $\hat{Q}$ close to a predetermined target queue size q_t and between $Q_{\min}$ and $Q_{\max}$ thresholds. So, the parameter μ was defined as the function of q_t, $Q_{\min}$ and $Q_{\max}$.

Also in [84, 85] two dynamical versions of the BetaRED algorithm were proposed. The first version, Adaptive Beta RED (ABetaRED) algorithm, was based on ideas of the Adaptive RED (ARED) algorithm [30] with the tuning of the thresholds parameters. In the second version, Dynamic Beta RED (DBetaRED) algorithm, the dynamic parameter (virtual target queue length) was introduced to correct deviations between the average queue size $\hat{Q}$ and the actual value of the target queue size q_t.

The Integrated RED (IRED) algorithm was presented in [86] with two new congestion indicators — the integrated arrival factor (based on queue length (current queue combined with a predetermined threshold value divided by the buffer's total capacity) and arrival rate (a weighted average of the current arrival value and the previous arrival value) and the integrated departure factor (based on departure rate and the queue length).

The novel version of the RED algorithm, called Improved RED (IM-RED), with combination of a nonlinear (the quadratic) drop function to deal with light and moderate network traffic loads and a linear drop function for heavy traffic load was introduced in [87]. The proposed drop functions are similar to functions, presented in [77]. The performance evaluation and comparison of IM-RED and IGRED [50] algorithms was conducted in [51].

Similar to I-RED [77] and IM-RED [87] algorithms in [88] the idea of quadratic (for a light or a moderate traffic load) and linear (for a heavy traffic load) probability drop functions was presented, so the new algorithm was named as RED-Quadratic Linear (RED-QL). But unlike the mentioned algorithms, in RED-QL an additional threshold value lying between the minimum $Q_{\min}$ and maximum $Q_{\max}$ thresholds was introduced.

In [89] the modification of Adaptive RED (ARED) [30] and Gentle RED (GRED) [49] algorithms with Markov decision process for adaptation of the queue weight values was presented and named as Markov decision process RED (MDPRED).

In the Linear RED (LRED) algorithm [90] it was proposed to use the adaptive estimated average queue length (the function of the differences in the estimated queue length and the actual queue length) as the congestion indicator and to simplify the calculation of the probability drop function as a linear function of the estimated queue length.

Based on ideas from [16–18, 91, 92] the novel network congestion control algorithm, called the Average Queue Length and Change Rate RED (AC-RED) was introduced in [93] with the average queue length change rate as a new parameter, reflecting the change of network traffic. The proposed AC-RED algorithm was compared with RED [3], ARED [30], URED [94], TRED [26] and AQMRD [17] algorithms.

In [95] the new version of the RED algorithm, the Double Linear RED (DL-RED) was presented with a packet drop function, consisting of two different linear functions (one for light and moderate loads, another for a heavy load).

9. 2023

The modification of the RED algorithm with two nonlinear probability drop functions (the quadratic function for light and moderate traffic loads, the exponential function for a heavy traffic load) was presented in [96] and named as Quadratic Exponential RED (QERED) algorithm. Also the new threshold Q_{target} was introduced ($Q_{\min} < Q_{\text{target}} < Q_{\max}$).

In [97] the Modified Dropping RED (MD-RED) algorithm was proposed. In this algorithm the single linear probability drop function from RED was replaced by two drop functions: the linear for low and moderate traffic loads and the exponential for a high traffic load. Also as in [96], the new threshold Q_{target} was introduced ($Q_{\min} < Q_{\text{target}} < Q_{\max}$).

The modification of the RED algorithm with the exponential probability drop function was considered in [98] with the nonlinearity parameter γ for the adjustment of the bending degree of the exponential function.

In [99] the novel fuzzy-based AQM algorithm based on a computationally efficient precise fuzzy modeling with the Genetic Algorithm (FREDGA) was proposed in order to achieve a balanced and equitable distribution of the resources and avoid bandwidth wasting resulting from unnecessary packet dropping.

The improved version of active queue management algorithm ARED [30] the Switchable ARED (SARED) was introduced in [100] to guarantee the comprehensive QoS performance of a low earth orbit (LEO) satellite network by dynamically updating of RED parameters through the link load information and the predicting the network state of satellite nodes based on obtained link information. The SARED algorithm uses both an S-shaped probability drop function together with a linear function in order to obtain a better stability for the queue size.

The novel version of the classic RED algorithm [3] for the a-priori congestion detection was presented in [101]. The Decision Tree-aided Random Early Detection (DREaD) algorithm proposes to use the machine learning technique (Decision Tree classifier) at the router in order to predict congestion by observing the behavior of the queue at the router (based on the queue characteristics).

10. 2024

Two nonlinear modifications of RED algorithm was presented in [102]: Amended RED (AmRED) and RED-Quadratic Exponential (RED-QE). In AmRED the RED queue threshold range is divided into three sections and three dropping functions (quadratic, linear and exponential) are used. In RED-QE algorithm the RED queue threshold is divided into two sections and two drop functions (quadratic and exponential) are used.

In [103] the Classified Enhanced Random Early Detection (CERED) algorithm for real-time packets and non-real-time packets in Wireless Sensor Networks (WSNs) was proposed. In CERED, the preemption priority was conferred on real-time packets, and the queue management with enhanced initial drop probability $P_{\min>0}$ was implemented only for non-real-time packets. Also in [103] the preemptive priority M/M/1/C vacation queuing model with queue management was presented.

The novel version of RED algorithm — the Amended Dropping Random Early Detection (AD-RED) with two nonlinear packet dropping functions (quadratic plus exponential) as for RED-QE algorithm in [102] was presented in [104].

The Random Early Detection algorithm with static priority scheduling and controlled delay (RED-SP-CoDel) algorithm was introduced in [105]. In this algorithm the RED AQM component was added to the combination of the Controlled Delay (CoDel) [106] AQM mechanism with Static Priority (SP) scheduling for QoS differentiation.

11. Conclusion

The presented bibliographical chronological review of active control algorithms of the RED family is the most complete both in terms of the number of algorithms reviewed (more than two hundred) and in terms of the number of scientific publications analyzed and presented. This review will be useful to researchers in the field of the congestion control.

Active queue management algorithms of the RED family are not something new for the authors of this work, as evidenced by the publications presented below [107–115].

In the future, the authors plan not only to classify the considered algorithms based on the classification criteria presented in [107, 116, 117], but also to review and classify other active queue management algorithms.

Author Contributions: Conceptualization, Zaryadov I.S. and Korolkova A.V.; methodology, Korolkova A.V.; validation, Zaryadov I.S., Korolkova A.V. and Milovanova T.A.; formal analysis, Korolkova A.V.; investigation, Viana C.C. Hilquias.; resources, Zaryadov I.S. and Milovanova T.A.; data curation, Korolkova A.V. and Milovanova T.A.; writing—original draft preparation, Zaryadov I.S.; writing—review and editing, Korolkova A.V.; supervision, Zaryadov I.S.; project administration, Korolkova A.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zaryadov, I. S., Viana, H. C., Korolkova, A. V. & Milovanova, T. A. Chronology of the development of Active Queue Management algorithms of RED family. Part 1: from 1993 up to 2005. *Discrete and Continuous Models and Applied Computational Science* **31,** 305–331. doi:10.22363/2658-4670-2023-31-4-305-331 (2023).
2. Zaryadov, I. S., Viana, H. C., Korolkova, A. V. & Milovanova, T. A. Chronology of the development of Active Queue Management algorithms of RED family. Part 2: from 2006 up to 2015. *Discrete and Continuous Models and Applied Computational Science* **32,** 18–37. doi:10.22363/2658-4670-2024-32-1-18-37 (2024).
3. Floyd, S. & Jacobson, V. Random early detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking* **1,** 397–413. doi:10.1109/90.251892 (1993).
4. Paul, A. K., Kawakami, H., Tachibana, A. & Hasegawa, T. *An AQM based congestion control for eNB RLC in 4G/LTE network* in *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* (IEEE, Vancouver, BC, Canada, 2016), 1–5. doi:10.1109/CCECE.2016.7726792.
5. Syed Masood, M. & Sheik Abdul Khader, P. *Effective queue management using fuzzy logic for congestion control in delay-sensitive applications over mobile Ad hoc networks* in *Emerging Research in Computing, Information, Communication and Applications* (eds Shetty, N. R., Prasad, N. H. & Nalini, N.) (Springer Singapore, Singapore, 2016), 385–395. doi:10.1007/978-981-10-0287-8_36.
6. Zhou, K., Yeung, K. L. & Li, V. O. Nonlinear RED: A simple yet efficient active queue management scheme. *Computer Networks* **50,** 3784–3794. doi:10.1016/j.comnet.2006.04.007 (2006).
7. Domańska, J., Domański, A., Augustyn, D. & Klamka, J. A RED modified weighted moving average for soft real-time application. *International Journal of Applied Mathematics and Computer Science* **24,** 697–707. doi:10.2478/amcs-2014-0051 (2014).
8. Domański, A., Domańska, J. & Czachórski, T. *The impact of the degree of self-similarity on the NLREDwM mechanism with drop from front strategy* in *Computer Networks* (eds Gaj, P., Kwiecień, A. & Stera, P.) **608** (Springer International Publishing, Cham, 2016), 192–203. doi:10.1007/978-3-319-39207-3_17.
9. Lafta, W., Jabbar, S., Kadhim, D. & Ma, G. Heterogeneous network performance improvement Uuing proposed OLRED and OLWRED strategies. *International Journal of Future Computer and Communication* **5,** 199–204. doi:10.18178/ijfcc.2016.5.5.471 (2016).
10. Cisco Systems Inc, C. *Cisco IOS 12.0 Quality of Service* 288 pp. (Cisco Press, USA, 1999).
11. Lafta, W., Jabbar, S., Kadhim, D. & Ma, G. OLWRED: Best Selected Strategy for Data Transmission in Heterogeneous Networks. *International Journal of Computer Applications* **152,** 11–15. doi:10.5120/ijca2016911781 (2016).
12. Zhao, Y.-h., Zheng, X.-f. & Tu, X.-y. Research on the improved way of RED wlgorithm S-RED. *International Journal of u- and e- Service Science and Technology* **9,** 375–384. doi:10.14257/ijunesst.2016.9.2.36 (2 2016).
13. Abdel-jaber, H., Ababneh, J., Thabtah, F., Daoud, A. M. & Baklizi, M. *Performance analysis of the proposed Adaptive Gentle Random Early Detection method under noncongestion and congestion situations* in *Digital Enterprise and Information Systems* (eds Ariwa, E. & El-Qawasmeh, E.) **194** (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011), 592–603. doi:10.1007/978-3-642-22603-8_52.

14. Baklizi, M. & Ababneh, J. Performance evaluation of the proposed Enhanced Adaptive Gentle Random Early Detection algorithm in congestion situations. *International Journal of Current Engineering and Technology* **6,** 1658–1664 (2016).
15. Hanaa, M., Gamal, A. & Samy, E.-D. Active Queue Management for congestion control: performance evaluation, new approach, and comparative study. *International Journal of Computing and Network Technology* **05,** 37–49. doi:10.12785/ijcnt/050201 (2017).
16. Karmeshu, Patel, S. & Bhatnagar, S. *Adaptive mean queue size and its rate of change: queue management with random dropping* 2016. doi:10.48550/arXiv.1602.02241. arXiv: 1602.02241 [cs.NI].
17. Karmeshu, Patel, S. & Bhatnagar, S. Adaptive mean queue size and its rate of change: queue management with random dropping. *Telecommunication Systems* **65,** 281–295. doi:10.1007/s11235-016-0229-4 (2017).
18. Bhatnagar, S., Patel, S. & Karmeshu. A stochastic approximation approach to active queue management. *Telecommunication Systems* **68,** 89–104. doi:10.1007/s11235-017-0377-1 (2018).
19. Abbasov, B. & Korukoglu, S. Effective RED: An algorithm to improve RED's performance by reducing packet loss rate. *Journal of Network and Computer Applications* **32,** 703–709. doi:10.1016/j.jnca.2008.07.001 (2009).
20. Khatari, M. & Samara, G. *Congestion control approach based on Effective Random Early Detection and fuzzy logic* 2017. doi:10.48550/arXiv.1712.04247. arXiv: 1712.04247 [cs.NI].
21. Fgee, E.-B., Smeda, A. & AbouElgaseem, K. MRED: An algorithm to insure high QoS in IP networks. *Journal of Communications* **12,** 200–206. doi:10.12720/jcm.12.4.200-206 (2017).
22. Gyasi-Agyei, A. *Service differentiation in wireless Internet using multiclass RED with drop threshold proportional scheduling* in *Proceedings 10th IEEE International Conference on Networks (ICON 2002). Towards Network Superiority* (IEEE, Singapore, 2002), 175–180. doi:10.1109/ICON.2002.1033307.
23. Alkharasani, A. M., Othman, M., Abdullah, A. & Lun, K. Y. An improved Quality-of-Service performance using RED's Active Queue Management flow control in classifying networks. *IEEE Access* **5,** 24467–24478. doi:10.1109/ACCESS.2017.2767071 (2017).
24. Zhao, Y., Ma, Z., Zheng, X. & Tu, X. An improved algorithm of Nonlinear RED based on membership cloud theory. *Chinese Journal of Electronics* **26,** 537–543. doi:10.1049/cje.2017.03.013 (2017).
25. Paul, A. K., Kawakami, H., Tachibana, A. & Hasegawa, T. Effect of AQM-based RLC buffer management on the eNB scheduling algorithm in LTE network. *Technologies* **5.** doi:10.3390/technologies5030059 (2017).
26. Feng, C.-W., Huang, L.-F., Xu, C. & Chang, Y.-C. Congestion Control Scheme Performance Analysis Based on Nonlinear RED. *IEEE Systems Journal* **11,** 2247–2254. doi:10.1109/JSYST.2014.2375314 (2017).
27. Patel, Z. M. *Queue occupancy estimation technique for adaptive threshold based RED* in *2017 IEEE International Conference on Circuits and Systems (ICCS)* (IEEE, Thiruvananthapuram, India, 2017), 437–440. doi:10.1109/ICCS1.2017.8326038.
28. Aweya, J., Ouellette, M. & Montuno, D. Y. A control theoretic approach to active queue management. *Computer Networks* **36.** Theme issue: Overlay Networks, 203–235. doi:10.1016/S1389-1286(00)00206-1 (2001).
29. Rezaee, A. A. & Pasandideh, F. A fuzzy congestion control protocol based on Active Queue Management in wireless sensor networks with medical applications. *Wireless Personal Communications* **98,** 815–842. doi:10.1007/s11277-017-4896-6 (2018).
30. Floyd, S., Gummadi, R. & Shenker, S. *Adaptive RED: An Algorithm for Increasing the Robustness of RED's Active Queue Management* tech. rep. (AT&T Center for Internet Research at ICSI, 2001).

31. Ahmed, A. & Nasrelden, N. *New congestion control algorithm to improve computer networks performance* in *2018 International Conference on Innovative Trends in Computer Engineering (ITCE)* (IEEE, Aswan, Egypt, 2018), 87–93. doi:10.1109/ITCE.2018.8316605.
32. Mahajan, M. & Singh, T. P. The Modified Gaussian function based RED (MGF-RED) algorithm for congestion avoidance in mobile Ad hoc networks. *International Journal of Computer Applications* **91,** 39–44. doi:10.5120/15889-5112 (2014).
33. Soni, H. & Mishra, P. Reducing packet loss in Active Queue Management. *International Journal of Computer Applications* **81,** 25–28. doi:10.5120/14208-2447 (Nov. 2013).
34. Akshatha, R. & Vedananda, D. E. Implementation of Hybrid Modified RED algorithm for congestion avoidance in MANETS. *International Journal for Research in Applied Science and Engineering Technology* **6,** 2414–2419. doi:10.22214/ijraset.2018.5396 (2018).
35. Sharma, N., Rajput, S. S., Dwivedi, A. K. & Shrimali, M. *P-RED: Probability based Random Early Detection algorithm for queue management in MANET* in *Advances in Computer and Computational Sciences* (eds Bhatia, S. K., Mishra, K. K., Tiwari, S. & Singh, V. K.) **554** (Springer Singapore, Singapore, 2018), 637–643. doi:10.1007/978-981-10-3773-3_62.
36. Baklizi, M., Abdel-Jaber, H., Abualhaj, M., Abdullah, N., Ramadass, S. & Almomani, D. Dynamic stochastic early discovery: A new congestion control technique to improve networks performance. *International Journal of Innovative Computing Information and Control IJICIC* **9,** 1113–1126 (2013).
37. Baklizi, M., Ababneh, J., Abualhaj, M. M., Abdullah, N. & Abdullah, R. Markov-modulated bernoulli dynamic gentle random early detection. *Journal of Theoretical and Applied Information Technology* **96,** 6688–6698 (2018).
38. Chhabra, K., Kshirsagar, M. & Zadgaonkar, A. *An improved RED algorithm with input sensitivity* in *Cyber Security* (eds Bokhari M. U.and Agrawal, N. & Saini, D.) **729** (Springer Singapore, Singapore, 2018), 35–45. doi:10.1007/978-981-10-8536-9_5.
39. Abualhaj, M. M., Abu-Shareha, A. A. & Al-Tahrawi, M. M. FLRED: an efficient fuzzy logic based network congestion control method. *Neural Computing and Applications* **30,** 925–935. doi:10.1007/s00521-016-2730-9 (2018).
40. Kachhad1, K. & Lathigara, A. ModRED : Modified RED an efficient congestion control algorithm for wireless network. *International Research Journal of Engineering and Technology IRJETs* **5,** 1879–1884 (2018).
41. Su, Y., Huang, L. & Feng, C. QRED: A Q-learning-based Active Queue Management scheme. *Journal of Internet Technology* **19,** 1169–1178. doi:10.3966/160792642018081904019 (2018).
42. Cisco Systems Inc, C. *Cisco IOS Quality of Service Solutions Configuration Guide, Release 12.2* tech. rep. (Cisco, 1999).
43. Alhassan, M. S. E. & Hagras, H. *Towards congestion control approach based on Weighted Random Early Detection and type-2 fuzzy logic system* in *2018 10th Computer Science and Electronic Engineering (CEEC)* (IEEE, Colchester, UK, 2018), 71–74. doi:10.1109/CEEC.2018.8674190.
44. Alhassan, M. S. E. & Hagras, H. A congestion control approach based on Weighted Random Early Detection and type-2 fuzzy logic system. *International Journal of Computer Science Trends and Technology IJCST* **8,** 83–94. doi:10.33144/23478578/IJCST-V8I4P14 (2020).
45. Monisha, V. & Ranganayaki, T. *Congestion Avoidance Aware using Modified Weighted Fairness Guaranteed DRED-FDNNPID Congestion Control for MWSN* in *2018 Tenth International Conference on Advanced Computing (ICoAC)* (IEEE, Chennai, India, 2018), 133–137. doi:10.1109/ICoAC44903.2018.8939080.
46. Chen, L. & Cao, J. *Adaptive congestion control of Internet of Things based on Improved RED Algorithm* in *2018 Chinese Automation Congress (CAC)* (IEEE, Xi'an, China, 2018), 295–298. doi:10.1109/CAC.2018.8623124.

47. Hamadneh, N., Obiedat, M., Qawasmeh, A. & Bsoul, M. HRED, An Active Queue Management algorithm for TCP congestion control. *Recent Patents on Computer Science* **12,** 212–217. doi:10.2174/2213275912666181205155828 (2019).
48. Al-Allaf, A. & Jabbar, A. I. A. RED with reconfigurable maximum dropping probability. *International Journal of Computing and Digital Systems* **8,** 61–72. doi:10.12785/ijcds/080107 (2019).
49. Floyd, S. *Recommendation on using the "gentle variant of RED"* tech. rep. (The ICSI Networking ans Security Gropup, 2000).
50. Abdel-Jaber, H., Shehab, A., Barakat, M. & Rashad, M. IGRED: An Improved Gentle Random Early Detection method for management of congested networks. *Journal of Interconnection Networks* **19,** 1950004. doi:10.1142/S021926591950004X (June 2019).
51. Abdel-Jaber, H., Alkhateeb, J. H. & El-Amir, M. *Evaluation of the performance for IM-RED and IGRED algorithms using discrete-time queues* in *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)* (IEEE, Al-Khobar, Saudi Arabia, 2022), 23–28. doi:10.1109/CICN56167.2022.10008318.
52. Baklizi, M. Stabilizing average queue length in Active Queue Management method. *International Journal of Advanced Computer Science and Applications* **10,** 77–83. doi:10.14569/IJACSA.2019.0100310 (2019).
53. Adel, A. A.-S. Enhanced Random Early Detection using responsive congestion indicators. *International Journal of Advanced Computer Science and Applications* **10,** 358–367. doi:10.14569/IJACSA.2019.0100347 (2019).
54. Dash, P., Barpanda, N. & Panda, M. Congestion control in cable network transmission using Novel RED algorithm. *International Journal of Innovative Technology and Exploring Engineering* **8,** 2278–3075. doi:10.35940/ijitee.J9562.0881019 (Aug. 2019).
55. Suwannapong, C. & Khunboa, C. Congestion control in CoAP Observe Group Communication. *Sensors* **19.** doi:10.3390/s19153433 (2019).
56. Adel, A. Controlling delay at the router buffer using Modified Random Early Detection. *International Journal of Computer Networks and Communications* **11,** 63–75. doi:10.5121/ijcnc.2019.11604 (2019).
57. Danladi, S. B. & Ambursa, F. U. *DyRED: An enhanced Random Early Detection based on a new adaptive congestion control* in *2019 15th International Conference on Electronics, Computer and Computation (ICECCO)* (IEEE, Abuja, Nigeria, 2019), 1–5. doi:10.1109/ICECCO48375.2019.9043276.
58. Abdel-Jaber, H. An exponential Active Queue Management method based on Random Early Detection. *Journal of Computer Networks and Communications* **2020,** 63–75. doi:10.1155/2020/8090468 (2020).
59. Adamu, A., Shorgin, V., Melnikov, S. & Gaidamaka, Y. *Flexible Random Early Detection algorithm for queue management in routers* in *Distributed Computer and Communication Networks* (eds Vishnevskiy, V. M., Samouylov, K. E. & Kozyrev, D. V.) **12563** (Springer International Publishing, Cham, 2020), 196–208. doi:10.1007/978-3-030-66471-8_16.
60. Li, S., Xu, Q., Gaber, J., Dou, Z. & Chen, J. Congestion control mechanism based on dual threshold DI-RED for WSNs. *Wireless Personal Communications* **115,** 2171–2195. doi:10.1007/s11277-020-07676-6 (2020).
61. Al-Allaf, A. F. & Jabbar, A. I. A. Reconfigurable Nonlinear GRED algorithm. *International Journal of Computing and Digital Systems* **9,** 1009–1022. doi:10.12785/ijcds/090521 (2020).
62. Xue, L. *The implementation of an improved ARED congestion control algorithm* in *2020 13th International Symposium on Computational Intelligence and Design (ISCID)* (IEEE, Hangzhou, China, 2020), 22–25. doi:10.1109/ISCID51228.2020.00012.

63. Baklizi, M. Weight Queue Dynamic Active Queue Management algorithm. *Symmetry* **12.** doi:10.3390/sym12122077 (2020).
64. Kumhar, D., Kumar, A. & Kewat, A. QRED: an enhancement approach for congestion control in network communications. *International Journal of Information Technology* **13,** 221–227. doi:10.1007/s41870-020-00538-1 (2021).
65. Adamu, A., Surajo, Y. & Jafar, M. T. SARED: A Self-Adaptive Active Queue Management scheme for improving quality of service in network systems. *Computer Science* **22,** 253–267. doi:10.7494/csci.2021.22.2.4020 (Apr. 2021).
66. Suwannapong, C. & Khunboa, C. EnCoCo-RED: Enhanced congestion control mechanism for CoAP observe group communication. *Ad Hoc Networks* **112,** 102377. doi:10.1016/j.adhoc.2020.102377 (2021).
67. Singha, S., Jana, B., Jana, S. & Mandal, N. K. *A novel congestion control clgorithm using buffer occupancy RED* in *Computational Intelligence in Pattern Recognition* (eds Das, A. K., Nayak, J., Naik, B., Dutta, S. & Pelusi, D.) **1349** (Springer Singapore, Singapore, 2022), 519–528. doi:10.1007/978-981-16-2543-5_44.
68. Singha, S., Jana, B. & Mandal, N. K. Active Queue Management in RED considering critical point on target queue. *Journal of Interconnection Networks* **21,** 2150017. doi:10.1142/S0219265921500171 (2021).
69. Feng, W.-C., Kandlur, D. D., Saha, D. & Shin, K. G. *Techniques for Eliminating Packet Loss in Congested TCP/IP Networks* tech. rep. (The University of Michigan, 1997).
70. Singha, S., Jana, B., Mandal, N. K., Jana, S., Bandyopadhyay, S. & Midya, S. *Application of dynamic weight with distance to reduce packet loss in RED based algorithm* in *Advanced Techniques for IoT Applications. EAIT 2021* (eds Mandal, J. K. & De, D.) **292** (Springer Singapore, Singapore, 2022), 530–543. doi:10.1007/978-981-16-4435-1_52.
71. Hassan, S., Oluwatope, A., Ajaegbu, C., Khadijha-Kuburat Adebisi, A. & Olasupo, A. QLREDActive Queue Management Algorithm. *Journal of Computer Science and Its Application* **28,** 95–107. doi:10.4314/jcsia.v28i1.8 (Sept. 2021).
72. Jarrah, A., Alshiab, M. & Shurman, M. High performance Changeable Dynamic Gentle Random Early Detection (CDGRED) for congestion control at router buffer. *International Journal of Grid and High Performance Computing* **14,** 1–14. doi:10.4018/IJGHPC.301585 (Jan. 2022).
73. Hassan, S. O., Ajaegbu, C., Ogunlere, S. O., Kanu, R. U. & Maitanmi, O. S. RED-I: a RED-based algorithm for Internet routers. *Journal of Communications* **17,** 260–266. doi:10.12720/jcm.17.4.260-266 (2022).
74. Bie, Y., Li, Z., Hu, Z. & Chen, J. Queue management algorithm for satellite networks based on traffic prediction. *IEEE Access* **10,** 54313–54324. doi:10.1109/ACCESS.2022.3163519 (2022).
75. Jafri, S. T. A., Ahmed, I. & Ali, S. Queue-buffer optimization based on Aggressive Random Early Detection in massive NB-IoT MANET for 5G applications. *Electronics* **11,** 2955. doi:10.3390/electronics11182955 (2022).
76. Hassan, S. O. RED-LE: a revised algorithm for Active Queue Management. *Journal of Telecommunications and Information Technology* **2,** 91–97. doi:10.26636/jtit.2022.160022 (2022).
77. Hassan, S. O., Rufai, A. U., Ogunlere, S. O., Alao, O. D., Ogundele, L. A., Agbaje, M. O., Adegbenjo, A. A. & Kuyoro, S. O. I-RED: an improved Active Queue Management algorithm. *Journal of Computer Science* **18,** 130–137. doi:10.3844/jcssp.2022.130.137 (2022).
78. Wei, D., Zheng, X., Yan, Z. & Cai, R. *An active queue management algorithm to enhance RED stability* in *2022 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)* (IEEE, Wuhan, China, 2022), 524–529. doi:10.1109/AEMCSE55572.2022.00108.

79. Sun, J., Zukerman, M. & Palaniswami, M. *Stabilizing RED using a Fuzzy Controller* in *2007 IEEE International Conference on Communications* (IEEE, Glasgow, UK, 2007), 266–271. doi:10.1109/ICC.2007.52.
80. Baklizi, M., Abdel-Jaber, H., Adel, A., Abualhaj, M. & Ramadass, S. Fuzzy Logic Controller of Gentle Random Early Detection based on average queue length and delay rate. *International Journal of Fuzzy Systems* **16,** 9–19 (2014).
81. Abu-Shareha, A. A., Al-Kasasbeh, B., Shambour, Q. Y., Abualhaj, M. M. & Al-Khatib, S. N. Fuzzy comprehensive random early detection of router congestion. *Information Technology and Control* **51,** 252–267. doi:10.5755/j01.itc.51.2.30194 (2022).
82. Duran, G., Valero, J., Amigó, J. M., Giménez, A. & Martínez-Bonastre, O. *Stabilizing chaotic behavior of RED* in *2018 IEEE 26th International Conference on Network Protocols (ICNP)* (IEEE, Cambridge, UK, 2018), 241–242. doi:10.1109/ICNP.2018.00033.
83. Amigó, J. M., Duran, G., Giménez, A., Martínez-Bonastre, O. & Valero, J. Generalized TCP-RED dynamical model for Internet congestion control. *Communications in Nonlinear Science and Numerical Simulation* **82,** 105075. doi:10.1016/j.cnsns.2019.105075 (2020).
84. Giménez, A., Murcia, M. A., Amigó, J. M., Martínez-Bonastre, O. & Valero, J. *New RED-type TCP-AQM algorithms based on beta distribution drop functions* 2022. doi:10.48550/arXiv.2201.01105. arXiv: 2201.01105.
85. Giménez, A., Murcia, M. A., Amigó, J. M., Martínez-Bonastre, O. & Valero, J. New RED-type TCP-AQM algorithms based on beta distribution drop functions. *Applied Sciences* **12,** 11176. doi:10.3390/app122111176 (2022).
86. Adel Abu-Shareha, A. Integrated random early detection for congestion control at the router buffer. *Computer Systems Science and Engineering* **40,** 719–734. doi:10.32604/csse.2022.018369 (2022).
87. Hassan, S. O., Rufai, A. U., Agbaje, M. O., Enem, T. A., Ogundele, L. A. & Usman, S. A. Improved random early detection congestion control algorithm for internet routers. *The Indonesian Journal of Electrical Engineering and Computer Science IJEECS* **28,** 384–395. doi:10.11591/ijeecs.v28.i1.pp384-395 (2022).
88. Hassan, S., Nwaocha, V., Rufai, A., Odule, T., Enem, T., Ogundele, L. & Usman, S. Random early detection-quadratic linear: an enhanced active queue management algorithm. *Bulletin of Electrical Engineering and Informatics* **11,** 2262–2272. doi:10.11591/eei.v11i4.3875 (2022).
89. Mahawish, A. A. & Hassan, H. J. Improving RED algorithm congestion control by using the Markov decision process. *Scientific Reports* **12,** 13363. doi:10.1038/s41598-022-17528-x (2022).
90. Abu-Shareha, A., Al-Kasasbeh, B., Shambour, Q. Y., Abualhaj, M. M., Alsharaiah, M. A. & Al-Khatib, S. N. Linear random early detection for congestion control at the router buffer. *Informatica* **46,** 105–114. doi:10.31449/inf.v46i5.3966 (2022).
91. Karmeshu, Patel, S. & Bhatnagar, S. *Adaptive mean queue size and its rate of change: queue management with random dropping* 2016. doi:10.48550/arXiv.1602.02241. arXiv: 1602.02241 [cs.NI].
92. Karmeshu, Patel, S. & Bhatnagar, S. Adaptive mean queue size and its rate of change: queue management with random dropping. *Telecommunication Systems* **65,** 281–295. doi:10.1007/s11235-016-0229-4 (2022).
93. Pan, C., Zhang, S., Zhao, C., Shi, H., Kong, Z. & Cui, X. A novel active queue management algorithm based on average queue length change rate. *IEEE Access* **10,** 75558–75570. doi:10.1109/ACCESS.2022.3189183 (2022).
94. Patel, C. M. *URED: upper threshold RED an efficient congestion control algorithm* in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)* (IEEE, Tiruchengode, India, 2013), 1–5. doi:10.1109/ICCCNT.2013.6726469.

95. Hassan, S., Rufai, A., Ajaegbu, C. & Ayankoya, F. DL-RED: a RED-based algorithm for routers. *International Journal of Computer Applications in Technology* **70,** 244–253. doi:10.1504/IJCAT.2022.130879 (2022).
96. Hassan, S., Rufai, A., Nwaocha, V., Ogunlere, S., Adegbenjo, A., Agbaje, M. & Aniemeka, E. Quadratic exponential random early detection: a new enhanced random early detection-oriented congestion control algorithm for routers. *International Journal of Electrical and Computer Engineering IJECE* **13,** 669–679. doi:10.11591/ijece.v13i1.pp669-679 (Feb. 2023).
97. Hassan, S. & Rufai, A. Modified dropping-random early detection (MD-RED): a modified algorithm for controlling network congestion. *International Journal of Information Technology* **15,** 1499–1508. doi:10.1007/s41870-023-01201-1 (2023).
98. Kato, K., Kato, H., Asahara, H., Ito, D. & Kousaka, T. Effects on random early detection of the packet drop probability function with an adjustable nonlinearity. *Nonlinear Theory and Its Applications IEICE* **14,** 193–206. doi:10.1587/nolta.14.193 (Apr. 2023).
99. Abu-Shareha, A. A., Alsaaidah, A., Alshahrani, A. & Al-Kasasbeh, B. Fuzzy-based active queue management using precise fuzzy modeling and Genetic Algorithm. *Symmetry* **15.** doi:10.3390/sym15091733 (2023).
100. Hou, K., Yang, J., Liu, F. & Zhang, C. *An Active Queue Management Algorithm to Guarantee the QoS of LEO Satellite Network* in *2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS)* (IEEE, Chengdu, China, 2023), 1024–1031. doi:10.1109/ISCTIS58954.2023.10213116.
101. Kesiezie, K. & Murry, L. K. *DREaD: Decision Tree-Aided Random Early Detection - An Intelligent Active Queue Management Technique* in *2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)* (IEEE, Bangalore, India, 2023), 1–6. doi:10.1109/SMARTGENCON60755.2023.10442782.
102. Hassan, S. O., Solanke Olakunle O.and Odule, T. J., Adesina, A. O., Usman, S. A. & Ayinde, S. A. AmRED and RED-QE: redesigning random early detection algorithm. *Telecommunication Systems* **85,** 263–275. doi:10.1007/s11235-023-01082-6 (2024).
103. Zhang, C., Yang, J. & Wang, N. An active queue management for wireless sensor networks with priority scheduling strategy. *Journal of Parallel and Distributed Computing* **187,** 104848. doi:10.1016/j.jpdc.2024.104848 (2024).
104. Hassan, S. O. AD-RED: A new variant of random early detection AQM algorithm. *J. High Speed Netw.* **30,** 53–67. doi:10.3233/JHS-222055 (Jan. 2024).
105. Lhamo, O., Ma, M., Doan, T. V., Scheinert, T., Nguyen, G. T., Reisslein, M. & Fitzek, F. H. RED-SP-CoDel: Random early detection with static priority scheduling and controlled delay AQM in programmable data planes. *Computer Communications* **214,** 149–166. doi:10.1016/j.comcom.2023.11.026 (2024).
106. Nichols, K. & Jacobson, V. Controlling queue delay. *Commun. ACM* **55,** 42–50. doi:10.1145/2209249.2209264 (July 2012).
107. Korolkova, A. V., Kulyabov, D. S. & Tchernoivanov, A. I. On the classification of RED algorithms. Russian. *RUDN Journal of Mathematics Information Sciences and Physics* **3,** 34–46 (2009).
108. Korolkova, A. V. & Zaryadov, I. S. *The mathematical model of the traffic transfer process with a rate adjustable by RED* in *2010 International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* (IEEE, Moscow, Russia, Oct. 2010), 1046–1050. doi:10.1109/ICUMT.2010.5676505.
109. Velieva, T. R., Korolkova, A. V. & Kulyabov, D. S. *Designing installations for verification of the model of active queue management discipline RED in the GNS3* in *The 6th International Congress on Ultra Modern Telecommunications and Control Systems. Saint-Petersburg, Russia. October 6-8, 2014* (IEEE Computer Society, 2015), 570–577. doi:10.1109/ICUMT.2014.7002164.

110. Korolkova, A. V., Kulyabov, D. S. & Sevastianov, L. A. Combinatorial and operator approaches to RED modeling. *Mathematical Modelling and Geometry* **3,** 1–18. doi:10.26456/mmg/2015-331 (2015).
111. Korolkova, A. V. & Zaryadov, I. S. *The mathematical model of the traffic transfer process with a rate adjustable by RED* in *International Congress on Ultra Modern Telecommunications and Control Systems* (IEEE, Moscow, Russia, 2010), 1046–1050. doi:10.1109/ICUMT.2010.5676505.
112. Zaryadov, I. S., Korolkova, A. V., Kulyabov, D. S., Milovanova, T. & Tsurlukov, V. *The survey on Markov-Modulated Arrival Processes and their application to the analysis of active queue management algorithms* in *Distributed Computer and Communication Networks. DCCN 2017. Communications in Computer and Information Science* (eds Vishnevskiy, V. M., Samouylov, K. E. & Kozyrev, D. V.) 417–430 (Springer International Publishing, Cham, 2017). doi:10.1007/978-3-319-66836-9_35.
113. Viana C. C., H., Zaryadov, I. S., Tsurlukov, V. V., Milovanova, T. A., Bogdanova, E. V., Korolkova, A. V. & Kulyabov, D. S. *The general renovation as the active queue management mechanism. Some aspects and results* in *Distributed Computer and Communication Networks. DCCN 2019* (eds Vishnevskiy, V., Samouylov, K. & Kozyrev, D.) 488–502 (Springer, Cham, 2019). doi:10.1007/978-3-030-36625-4_39.
114. Apreutesey, A. M. Y., Korolkova, A. V. & Kulyabov, D. S. *Modeling RED algorithm modifications in the OpenModelica* in *Proceedings of the Selected Papers of the 8th International Conference "Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems" (ITTMM-2019), Moscow, Russia, April 15–19, 2019* (eds Kulyabov, D. S., Samouylov, K. E. & Sevastianov, L. A.) **2407** (CEUR-WS, 2019), 5–14.
115. Viana Carvalho Cravid, H., Zaryadov, I. S. & Milovanova, T. A. Queueing systems with different types of renovation mechanism and thresholds as the mathematical models of active queue management mechanism. *Discrete and Continuous Models and Applied Computational Science* **28,** 305–318. doi:10.22363/2658-4670-2020-28-4-305-318 (2020).
116. Adams, R. Active queue management: a survey. *Communications Surveys & Tutorials IEEE* **15,** 1425–1476. doi:10.1109/SURV.2012.082212.00018 (2013).
117. Abbas, G., Halim, Z. & Abbas, Z. H. Fairness-driven queue management: a survey and taxonomy. *IEEE Communications Surveys & Tutorials* **18,** 324–367. doi:10.1109/COMST.2015.2463121 (2016).

Information about the authors

Zaryadov, Ivan S.—Candidate of Physical and Mathematical Sciences, Asssistant Professor of Department of Probability Theory and Cybersecurity, Institute of Computer Science and Telecommunications, RUDN University (e-mail: zaryadov-is@rudn.ru, ORCID: 0000-0002-7909-6396)

Viana, C.C. Hilquias—Ph.D. student of Department of Probability Theory and Cybersecurity, Institute of Computer Science and Telecommunications, RUDN University (e-mail: hilvianamat1@gmail.com)

Korolkova, Anna V.—Candidate of Physical and Mathematical Sciences, Asssistant Professor of Department of Probability Theory and Cybersecurity, Institute of Computer Science and Telecommunications, RUDN University (e-mail: korolkova-av@rudn.ru, ORCID: 0000-0001-7141-7610)

Milovanova, Tatianna A.—Candidate of Physical and Mathematical Sciences, Asssistant Professor of Department of Probability Theory and Cybersecurity, Institute of Computer Science and Telecommunications, RUDN University (e-mail: milovanova-ta@rudn.ru, ORCID: 0000-0002-9388-9499)

UDC 519.872
DOI: 10.22363/2658-4670-2024-32-2-154–171 EDN: CMFCZB

Хронология развития алгоритмов активного управления очередями семейства RED. Часть 3: 2016–2024

И. С. Зарядов[1,2], К. К. И. Виана[1], А. В. Королькова[1], Т. А. Милованова[1]

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

Аннотация. Работа является третьей частью большого библиографического обзора по алгоритмам семейства RED, представленных в научной печати с 1993 по 2023 год. В данной части будут приведены сведения по алгоритмам, опубликованным с 2016 по 2023 год.

Ключевые слова: активное управление очередями, AQM, RED, управление перегрузками

Discrete & Continuous Models & Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (2) 172–180
http://journals.rudn.ru/miph

Research article

UDC 519.63, 539.6, 539.1
PACS 02.60.–x 02.60.Cb 02.70.Bf 03.65.–w 34.20.–b 37.10.Gh
DOI: 10.22363/2658-4670-2024-32-2-172–180 EDN: CRIVYI

Solution of a two-dimensional time-dependent Schrödinger equation describing two interacting atoms in an optical trap

I. S. Ishmukhamedov[1,2], A. S. Ishmukhamedov[1,2], Zh. E. Jalankuzov[1,2]

[1] *Institute of Nuclear Physics, Almaty, 050032, Kazakhstan*
[2] *Al-Farabi Kazakh National University, Almaty, 050040, Kazakhstann*

(received: August 25, 2023; revised: December 20, 2023; accepted: January 10, 2024)

Abstract. We introduce a numerical method to solve the two-dimensional time-dependent Schrödinger equation, which characterizes a system of two atoms with a finite-range interaction potential confined within a harmonic oscillator trap. We choose a Gaussian-shaped potential for the interaction potential. Such a system has been previously studied analytically, except that a zero-range interaction potential was used instead. We observe a strong agreement between the results for the two types of interactions. Also, we investigate the one-dimensional time-dependent Schrödinger equation for the relative motion and compute the ground state energy level as a function of the coupling strength.

Key words and phrases: split operator method, finite differences, time-dependent Schrödinger equation, quantum harmonic oscillator, Gaussian interaction potential, zero-range interaction, pseudopotential, cold atoms, optical trap

Citation: Ishmukhamedov I. S., Ishmukhamedov A. S., Jalankuzov Z. E., Solution of a two-dimensional time-dependent Schrödinger equation describing two interacting atoms in an optical trap. *Discrete and Continuous Models and Applied Computational Science* **32** (2), 172–180. doi: 10.22363/2658-4670-2024-32-2-172–180. edn: CRIVYI (2024).

1. Introduction

The physics of ultracold atoms has become one of the most intriguing fields of research. Due to the ability to precisely control the interaction between particles, the dimensions of the system, and its quantum state, researchers can explore a diverse range of phenomena with wide-ranging applications [1]. It is now even possible to experimentally realize systems with only a few atoms, or even just two atoms [2, 3]. This opens up a pathway for experimental verification of few-body models, highlighting the significant importance of solving the Schrödinger equation that describes such systems [4–12].

In this study, we present the solution of the two-dimensional time-dependent Schrödinger equation that describes two interacting atoms subject to the external harmonic oscillator confinement. We assume the bosonic symmetry for the particles. We choose the short-range Gaussian interaction potential to describe the interaction between the particles. Similar system, only for the zero-range interaction potential, or pseudopotential, has been studied previously [13–15]. It is a common practice to represent the interaction potential in terms of the coupling strength or scattering length. In this regard, we adjust the parameters of the Gaussian interaction potential to match the ground state energy level of the pseudopotential. This allows us to effectively analyze the system in the context of coupling strength.

Because of the harmonic trap potential, the system can be separated into relative and center-of-mass motions. First, we compute the ground state energy for the relative motion as a function of the coupling strength, Gaussian interaction depth and the correlation between these two parameters. Next, we compare the evolution of the wave function with the solutions derived from the pseudopotential, both for the relative motion and when including the center-of-mass motion. Our comparison reveals a strong agreement with the zero-range interaction potential in both of these cases.

The presented method has been employed for anharmonic trap potentials and various scenarios in the works [16–20]. In our present study, however, we illustrate that fitting the energy for the short-range Gaussian interaction with that of the zero-range potential yields an almost exact match of both wave functions within the designated range of the Gaussian interaction. Notably, in the vicinity of the origin, where the range is less than that of the Gaussian interaction, discernible distinctions between the two interaction types become evident. This contrast is even more pronounced in the two-dimensional case within the same region.

The remainder of this paper is organized as follows. Section 2 describes the Hamiltonian of two atoms in the harmonic oscillator potential. The interaction between the particles is modeled using a Gaussian-shaped function with a fixed interaction width. We also consider the system with zero-range interaction, for which the analytical solution is known. Section 3 presents the ground state energy results as a function of the interaction parameters. We then proceed to study the dynamics of the wave function for both relative motion and the total Hamiltonian. Section 4 provides a summary of the results obtained.

2. Model Hamiltonian and the method

2.1. Gaussian interaction

The Hamiltonian describing two interacting atoms in the harmonic trap reads:

$$H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x_1^2} - \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x_2^2} + V(x_1) + V(x_2) + V_{\text{int}}(x_1 - x_2),$$

where the $V(x)$ stands for the trapping potential:

$$V(x) = \frac{1}{2}m\omega x^2,$$

and V_{int} represents the interaction potential, which we have selected in the Gaussian form:

$$V_{\text{int}}(x_1 - x_2) = -V_0 \exp\left\{-\frac{(x_1 - x_2)^2}{2r_0^2}\right\}.$$

In the following, we use oscillator units, wherein length is given by $\ell = \sqrt{\frac{\hbar}{m\omega}}$, energy by $\hbar\omega$, and time by ω^{-1} units. We set the interaction range of (2) to a fixed value of $r_0 = 0.1$. This value provides a good approximation for the contact interaction between the atoms [11, 12, 20, 21].

Next, it is convenient to switch to the relative and center-of-mass motion coordinates given by:

$$\begin{cases} r = \dfrac{x_1 - x_2}{\sqrt{2}} \\ R = \dfrac{x_1 + x_2}{\sqrt{2}} \end{cases}, \tag{3}$$

where the factor $\sqrt{2}$ ensures that the kinetic operators and trapping potentials have equal coefficients. Thus, the Hamiltonian in these variables is represented as:

$$H(r, R) = H_r + H_R, \tag{4}$$

where

$$H_r = -\frac{1}{2}\frac{\partial^2}{\partial r^2} + \frac{1}{2}r^2 + V_{\text{int}}(r),$$

and

$$H_R = -\frac{1}{2}\frac{\partial^2}{\partial R^2} + \frac{1}{2}R^2.$$

The dynamics of the system are governed by the time-dependent Schrödinger equation:

$$i\frac{\partial \Psi(r, R, t)}{\partial t} = H(r, R)\Psi(r, R, t).$$

Due to the separability of the Hamiltonian, the evolution operator can be factorized. Therefore, the dynamics of the wave function can be expressed as:

$$\Psi(r, R, t + \Delta t) = \exp\{-i\Delta t H_r\} \exp\{-i\Delta t H_R\}\, \Psi(r, R, t). \tag{6}$$

The operators on the right-hand side of (6) are approximated using the Crank–Nicolson formula:

$$\exp\left(-i\Delta t\hat{A}\right) = \left(1 + \frac{i}{2}\Delta t\hat{A}\right)^{-1}\left(1 - \frac{i}{2}\Delta t\hat{A}\right), \tag{7}$$

where $\hat{A}$ represents the operators in (6). The equation (7) is accurate up to third order in terms of $\mathcal{O}(\Delta t^3)$. The partial derivatives are approximated using finite differences of sixth-order accuracy. The system of linear equations thus obtained is of a band diagonal type and is solved using the sweep method.

We employ the imaginary-time propagation method to compute the stationary states, i.e., $\Psi(r, R, t = 0)$, and the bound energy levels.

2.2. Zero-range interaction

For the purpose of comparison, we also consider an analytical model for two bosons with zero-range interaction. In this approach, the total Hamiltonian (in oscillator units) is given by:

$$H = -\frac{1}{2}\frac{\partial^2}{\partial x_1^2} - \frac{1}{2}\frac{\partial^2}{\partial x_2^2} + \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + \sqrt{2}g\delta(x_1 - x_2), \tag{8}$$

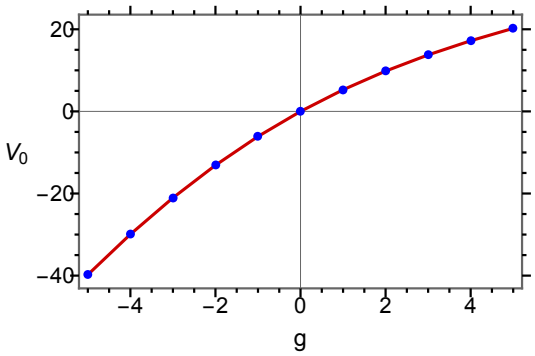


Figure 1. The relationship between the depth of the Gaussian interaction and the coupling strength g of the pseudopotential (data points with a guiding line)

where $\delta(x_1 - x_2)$ is the Dirac Delta function, and the factor $\sqrt{2}$ is used for convenience. Applying the transformation (3), the Hamiltonian decouples into the form given by (4), where the relative motion part is expressed as:

$$H_r = -\frac{1}{2}\frac{\partial^2}{\partial r^2} + \frac{1}{2}r^2 + g\delta(r). \tag{9}$$

The solution to (9) is established through the work of Busch et al. [13, 14]. Specifically, the energy levels are determined as the roots of the transcendental equation:

$$-g = 2\frac{\Gamma\left(-\frac{E_r}{2} + \frac{3}{4}\right)}{\Gamma\left(-\frac{E_r}{2} + \frac{1}{4}\right)}, \tag{10}$$

and the wave functions are given as:

$$\psi_r(r) = \mathcal{N}e^{-r^2/2}U\left(\frac{1}{4} - \frac{E_r}{2}, \frac{1}{2}, r^2\right).$$

Here $\mathcal{N}$ is the normalization constant and $U(a, b, z)$ are the Kummer functions. The solutions of the center-of-mass motion Hamiltonian H_R are the well-known solutions of the harmonic oscillator.

The time-dependent solution for the zero-range interaction of equation (5) can then be expressed as follows:

$$\Psi(r, R, t) = e^{-itE}\Psi(r, R, t = 0), \tag{11}$$

where E is the eigenenergy of the total Hamiltonian (8).

3. Results

Further on, we will focus solely on the ground state solutions. To represent the results of solving (5) for the Gaussian interaction, we adjust the depth of $V_{\text{int}}(r)$ such that the ground state energy level matches that of the pseudopotential in (10). This way we can directly compare the outcomes obtained for the Gaussian interaction potential with those of the zero-range interaction potential. The result of this adjustment is depicted in Fig. 1. From Fig. 1, it is evident that negative values of g correspond to attractive interactions between the particles, while positive values indicate repulsion.

Figure 2 illustrates the dependence of the ground state energy on the parameters for both the Gaussian interaction potential (Fig. 2(a)) and the zero-range interaction potential (given by (10), Fig. 2(b)).

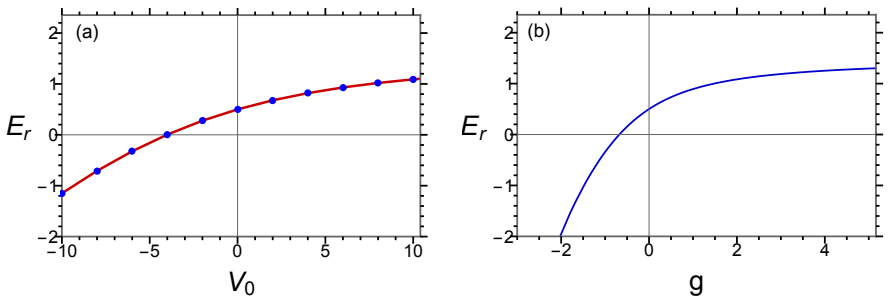


Figure 2. (a) Ground state energy for the relative motion as a function of the Gaussian interaction depth V_0 (data points with a guiding line). (b) Ground state energy for the relative motion as a function of the coupling strength g

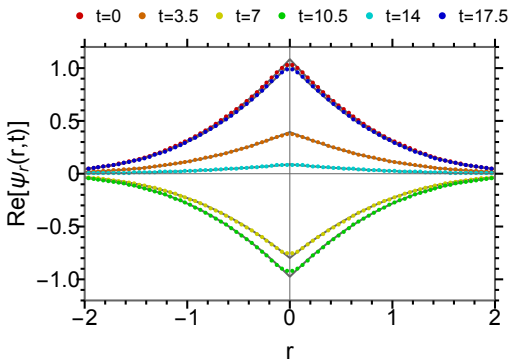


Figure 3. The real part of the wave function for the relative motion at various time points, calculated numerically (data points) and analytically (gray lines), for the coupling strength $g = -1$

3.1. Relative motion

First, we consider the Hamiltonian of the relative motion and compare the dynamics of the real part of the wave function (data points) with that of the zero-range interaction (gray lines), as shown in Fig. 3. We focus on the case where the coupling strength is $g = -1$. In Fig. 3, one can observe that the wave functions differ mainly in the region close to the origin, approximately $r \sim 0.1$, which corresponds to the length of the Gaussian interaction potential. Apart from that region, both solutions are in excellent agreement.

3.2. Relative and center-of-mass motions

Here, we incorporate the center-of-mass motion component of the Hamiltonian (4) into our analysis. We numerically calculate the dynamics of the wave function $\Psi(r, R, t)$ for the Gaussian interaction potential using (6) and for the pseudopotential using (11). Our focus is on the case where the coupling strength is $g = -1$. The results are presented in Fig. 4. In this figure, differences become noticeable around the region close to the origin, once again attributed to the width of the Gaussian interaction $r_0 = 0.1$. However, the discrepancy between the two types of interaction is slightly more pronounced.

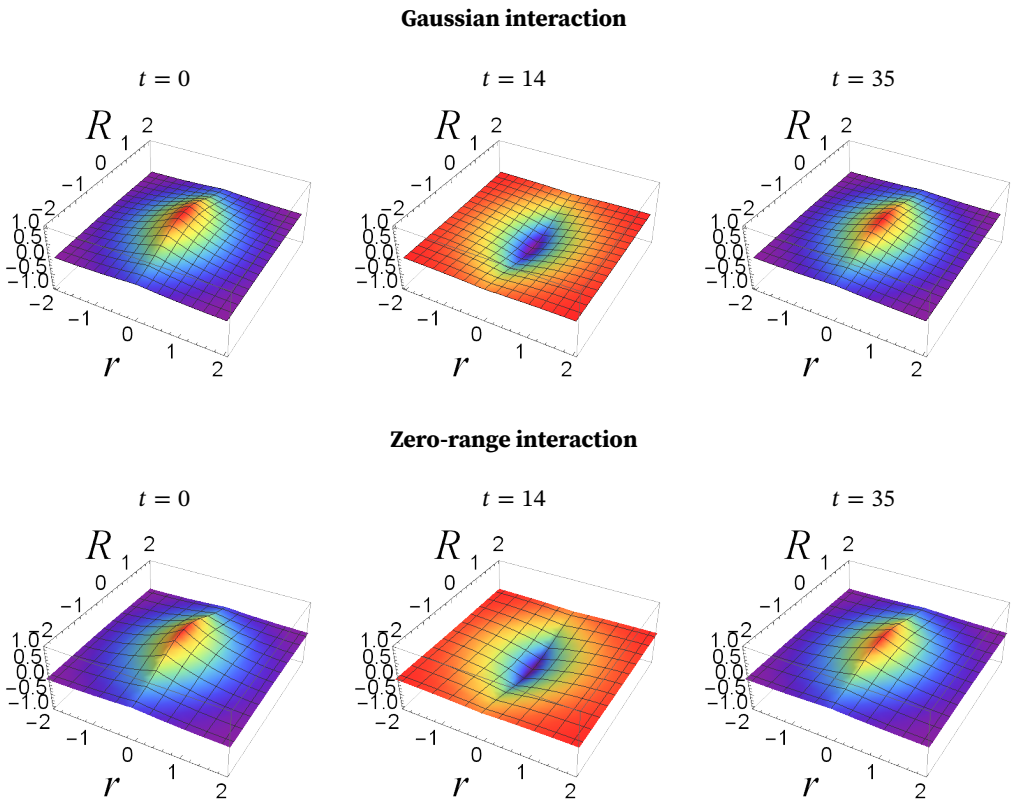


Figure 4. Real part of the wave function, $\mathrm{Re}\,[\Psi(r,R,t)]$, is shown at different time moments for both the Gaussian interaction (upper graphs) and the zero-range interaction (lower graphs), for the coupling strength $g=-1$

4. Summary

We studied the solutions of the two-dimensional Schrödinger equation for Gaussian and zero-range interaction potentials. The trap potential has the form of a harmonic oscillator, which results in the separability of the relative and center-of-mass motions. The solution for the Gaussian interaction potential was numerically obtained using the Crank-Nicolson approximation for the evolution operator and sixth-order finite differences for the spatial derivatives. The solution for the zero-range interaction is known analytically [13, 14].

The comparison between the two types of interaction potentials for the relative motion revealed that the discrepancy manifests around the region where $r \sim 0.1$, corresponding to the width of the Gaussian interaction potential. In the case where the center-of-mass motion is included, the difference becomes even more noticeable.

The provided method can be extended to more complex interaction and trap potentials [16, 17], as well as to systems involving more particles and/or higher dimensions [20].

Author Contributions: Conceptualization, I.S. Ishmukhamedov; methodology, I.S. Ishmukhamedov; software, I.S. Ishmukhamedov, A.S. Ishmukhamedov and Zh.E. Jalankuzov; validation, I.S. Ishmukhamedov, A.S. Ishmukhamedov and Zh.E. Jalankuzov;

formal analysis, I.S. Ishmukhamedov, A.S. Ishmukhamedov and Zh.E. Jalankuzov; investigation, I.S. Ishmukhamedov, A.S. Ishmukhamedov and Zh.E. Jalankuzov; resources, I.S. Ishmukhamedov, A.S. Ishmukhamedov and Zh.E. Jalankuzov; data curation, I.S. Ishmukhamedov, A.S. Ishmukhamedov and Zh.E. Jalankuzov; writing—original draft preparation, I.S. Ishmukhamedov, A.S. Ishmukhamedov and Zh.E. Jalankuzov; writing—review and editing, I.S. Ishmukhamedov, A.S. Ishmukhamedov and Zh.E. Jalankuzov; visualization, I.S. Ishmukhamedov, A.S. Ishmukhamedov and Zh.E. Jalankuzov; supervision, I.S. Ishmukhamedov; project administration, I.S. Ishmukhamedov; funding acquisition, I.S. Ishmukhamedov. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All data is available upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Funding: This research has been funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP13067639).

References

1. Bloch, I., Dalibard, J. & Zwerger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80,** 885 (2008).
2. Serwane, F., Zürn, G., Lompe, T., Ottenstein, T. B., Wenz, A. N. & Jochim, S. Deterministic preparation of a tunable few-fermion system. *Science* **332,** 336 (2011).
3. Zürn, G., Serwane, F., Lompe, T., Wenz, A. N., Ries, M. G., Bohn, J. E. & Jochim, S. Fermionization of Two Distinguishable Fermions. *Phys. Rev. Lett.* **108,** 075303 (2012).
4. Melezhik, V. S., Kim, J. I. & Schmelcher, P. Wave-packet dynamical analysis of ultracold scattering in cylindrical waveguides. *Phys. Rev. A* **76,** 053611 (2007).
5. Melezhik, V. S. & Schmelcher, P. Quantum dynamics of resonant molecule formation in waveguides. *New J. Phys.* **11,** 073031 (2009).
6. Kościk, P. On the Exponential Decay of Strongly Interacting Cold Atoms from a Double-Well Potential. *Few-Body Syst.* **64,** 11 (2023).
7. Kościk, P. & Sowiński, T. Universality of Internal Correlations of Strongly Interacting p-Wave Fermions in One-Dimensional Geometry. *Phys. Rev. Lett.* **130,** 253401 (2023).
8. Dobrzyniecki, J. & Sowiński, T. Two Rydberg-dressed atoms escaping from an open well. *Phys. Rev. A* **103,** 013304 (2021).
9. Bougas, G., Mistakidis, S. I., Giannakeas, P. & Schmelcher, P. Dynamical excitation processes and correlations of three-body two-dimensional mixtures. *Phys. Rev. A* **106,** 043323 (2022).
10. Bougas, G., Mistakidis, S. I., Giannakeas, P. & Schmelcher, P. Few-body correlations in two-dimensional Bose and Fermi ultracold mixtures. *New J. Phys.* **23,** 093022 (2021).
11. Gharashi, S. E. & Blume, D. Tunneling dynamics of two interacting one-dimensional particles. *Phys. Rev. A.* **92,** 033629 (2015).
12. Kestner, J. P. & Duan, L. M. Anharmonicity-induced resonances for ultracold atoms and their detection. *New J. Phys.* **12,** 053016 (2010).
13. Busch, T., Englert, B.-G., Rzażewski, K. & Wilkens, M. Two Cold Atoms in a Harmonic Trap. *Found. Phys.* **28,** 549 (1998).
14. Murphy, D. S., McCann, J. F., Goold, J. & Busch, T. Boson pairs in a one-dimensional split trap. *Phys. Rev. A* **76,** 053616 (2007).
15. Ishmukhamedov, I. S., Aznabayev, D. T. & Zhaugasheva, S. A. Two-body atomic system in a one-dimensional anharmonic trap: The energy spectrum. *Phys. Part. Nucl. Lett.* **12,** 680 (2015).

16. Ishmukhamedov, I. S. & Melezhik, V. S. Tunneling of two bosonic atoms from a one-dimensional anharmonic trap. *Phys. Rev. A.* **95,** 062701 (2017).
17. Ishmukhamedov, I. S. & Ishmukhamedov, A. S. Tunneling of two interacting atoms from excited states. *Physica E* **109,** 24 (2019).
18. Ishmukhamedov, I. S. Quench dynamics of two interacting atoms in a one-dimensional anharmonic trap. *Physica E* **142,** 115228 (2022).
19. Ishmukhamedov, I. S., Ishmukhamedov, A. S. & Melezhik, V. S. Numerical Solution of the Time Dependent 3D Schrödinger Equation Describing Tunneling of Atoms from Anharmonic Traps. *EPJ Web Conf.* **173,** 03011 (2018).
20. Ishmukhamedov, I. S., Ishmukhamedov, A. S., Jalankuzov, Z. E. & Ismailov, D. V. Molecular excited state in the interaction quench dynamics of two different atoms in a two-dimensional anisotropic trap. *Eur. Phys. J. Plus* **139,** 53 (2024).
21. Ishmukhamedov, I. S., Valiolda, D. S. & Zhaugasheva, S. A. Description of ultracold atoms in a one-dimensional geometry of a harmonic trap with a realistic interaction. *Phys. Part. Nuclei Lett.* **11,** 238 (2014).

Information about the authors

Ishmukhamedov, I. S.—Candidate of Physical and Mathematical Sciences, Researcher at Institute of Nuclear Physics (e-mail: i.ishmukhamedov@mail.ru, ORCID: 0000-0002-7903-3432)

Ishmukhamedov, A. S.—Researcher at Institute of Nuclear Physics (e-mail: altaymedoed@gmail.com, ORCID: 0000-0001-5248-3022)

Jalankuzov, Zh. E.—Researcher at Institute of Nuclear Physics (e-mail: jalankuzov.zhanibek@gmail.com, ORCID: 0009-0003-1962-8834)

UDC 519.63, 539.6, 539.1
PACS 02.60.–x 02.60.Cb 02.70.Bf 03.65.–w 34.20.–b 37.10.Gh
DOI: 10.22363/2658-4670-2024-32-2-172–180 EDN: CRIVYI

Решение двумерного нестационарного уравнения Шрёдингера, описывающего два взаимодействующих атома в оптической ловушке

И. С. Ишмухамедов[1,2], А. С. Ишмухамедов[1,2], Ж. Е. Джаланкузов[1,2]

[1] *Институт ядерной физики, Алматы, 050032, Казахстан*
[2] *Казахский национальный университет им. аль-Фараби, Алматы, 050040, Казахстан*

Аннотация. Мы представляем численный метод решения двумерного нестационарного уравнения Шредингера, которое характеризует систему двух атомов с потенциалом взаимодействия конечного радиуса действия, заключенную в ловушку гармонического осциллятора. В качестве потенциала взаимодействия мы выбираем гауссовский потенциал. Такая система ранее изучалась аналитически, с той лишь разницей, что вместо нее использовался нулевой потенциал взаимодействия. Мы наблюдаем хорошее согласие между результатами для двух типов взаимодействий. Кроме того, мы исследуем одномерное нестационарное уравнение Шредингера для относительного движения и вычисляем уровень энергии основного состояния в зависимости от константы связи.

Ключевые слова: метод расщепления, конечные разности, нестационарное уравнение Шрёдингера, квантовый гармонический осциллятор, гауссов потенциал взаимодействия, взаимодействие нулевого радиуса, псевдопотенциал, холодные атомы, оптическая ловушка

Discrete & Continuous Models
& Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (2) 181–201
http://journals.rudn.ru/miph

Review

UDC 004.2, 004.7
PACS 07.05.Tp
DOI: 10.22363/2658-4670-2024-32-2-181–201 EDN: CFLJNF

The recent progress in terahertz channel characterization and system design

Alexander S. Shurakov[1,2], Evgeny V. Mokrov[3], Anatoliy N. Prikhodko[1,2,4], Margarita I. Ershova[1], Vyacheslav O. Begishev[3], Abdukodir A. Khakimov[3], Yevgeny A. Koucheryavy[3], Gregory N. Gol'tsman[1,2,4]

[1] *Moscow Pedagogical State University, 1 M. Pirogovskaya St, bldg. 1, Moscow, 119991, Russian Federation*
[2] *HSE University, 20 Myasnitskaya St, Moscow, 101000, Russian Federation*
[3] *RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation*
[4] *Russian Quantum Center, Skolkovo, 143025, Russian Federation*

(received: March 25, 2024; revised: April 10, 2024; accepted: April 15, 2024)

Abstract. As the standardization of 5G New Radio (NR) systems operating in micro- and millimeter-wave frequency bands is over, scientific and industrial communities have begun to address the question of what 6G communications systems might or should be. While technological specifics are still in their early development phase, there is a common agreement that these systems will utilize the lower part of the terahertz band, namely, 100–300 GHz. This band poses a number of specific challenges for system designers, including the effects related to channel characteristics and the conceptually new requirements for electronics. This paper aims to report the current state-of-the-art channel characterization and communications system design. With respect to the former, we consider dynamic human body blockages and micromobility impairments. For the latter, we mainly concentrate on the physical layer devices for direct conversion schemes and the design of the so-called reconfigurable intelligent surfaces that will potentially serve as a cost-efficient blockage mitigation technique.

Key words and phrases: terahertz band, 6G, dynamic blockage, micromobility, experiment, coherent communication, direct conversion, reconfigurable intelligent surface

Citation: Shurakov A. S., Mokrov E. V., Prikhodko A. N., Ershova M. I., Begishev V. O., Khakimov A. A., Koucheryavy Y. A., Gol'tsman G. N., The recent progress in terahertz channel characterization and system design. *Discrete and Continuous Models and Applied Computational Science* **32** (2), 181–201. doi: 10.22363/2658-4670-2024-32-2-181–201. edn: CFLJNF (2024).

1. Introduction

Expanding the available bandwidth is the main tool for increasing rates at the access interface in cellular systems. This trend has been in place since 3G systems with 3G UMTS, 4G LTE, and 5G NR utilizing 2 MHz, 20 MHz, and 50 MHz wide basic channels, respectively, thus offering higher bandwidth. With 5G New Radio systems operating in the millimeter wave (mmWave, 24–100 GHz) band, the next logical step is the sub-terahertz/terahertz (sub-THz/THz, 100–300 GHz) band and even higher, where enormous yet unregulated bandwidth is available. Such bandwidth may result in data rates potentially supporting a plethora of upcoming applications, including virtual/augmented reality (VR/AR), holographic communications, etc.

The use of the THz band poses several challenges to systems designers. The first set is related to still unsolved propagation effects that were already evident in the mmWave band, including dynamic blockage by human bodies and the micromobility phenomenon. The former makes mmWave/THz systems unusable under the conditions they were invented for large dense crowds generating enormous traffic demands. The solution to this problem requires fast and accurate blockage detection mechanisms and methods to decrease the frequency of blockage events. Micromobility is closely related to the utilization of directional antennas and is expected to become more relevant at THz frequencies, where the directivity of antenna radiation patterns will further increase.

The development of large-scale antenna devices for the front ends of ultra-directional THz transmitters and receivers, as well as the smart 6G antenna environment, must comply with the criteria of cost and power efficiency. Moreover, supporting extremely large bandwidths is another challenge. This solution requires the use of advanced circuits for THz signal generation and processing that are naturally compatible with ultrafast and spectrally effective communication schemes. This, in turn, can be accomplished via partial replacement of intrinsically limited Si-based electronics with more capable A3B5 electronic and radiophotonic devices.

This paper aims to report the recent progress in THz channel characterization and system design. In the former topic, we concentrate on two effects currently affecting the public deployment of mmWave/THz communication systems: dynamic human body blockage and micromobility. In the context of system design, we consider state-of-the-art coherent communications with direct conversion and reconfigurable intelligent surfaces (RISs) for reflection-aided channels. The latter can be utilized as a simple and efficient block avoidance technique.

The remainder of this paper is organized as follows. We begin with the effect of dynamic human body blockage in Section 2. In Section 3, we discuss recent progress in micromobility characterization. Section 4 describes the design of state-of-the-art THz communication systems. Finally, in Section 4.2 we consider a potential blockage avoidance solution for 5G and 6G systems – RISs. Finally, the conclusions are presented in the last section.

2. Dynamic blockage

2.1. Motivation and impact

Dynamic blockage of the propagation path by human bodies is one of the critical factors affecting the performance of THz communications systems. This property was first observed in the mmWave domain, where it was demonstrated to result in 15–25 dB attenuation [1] depending on the propagation conditions and environmental characteristics. The eventual impact of this phenomenon is a drastic drop in the received signal strength resulting in either the downgrading of the utilized modulation and coding scheme (MCS) or outage conditions. Due to an even Shorter wavelength attenuation in the THz band is expected to be even more drastic.

Several methods have been proposed for human body blockage avoidance. First, because of the inherent multi-path propagation at mmWave/THz frequencies, a non-line-of-sight (nLoS) path can be utilized for continuing service in the case of blockage. When no outage is experienced bandwidth reservation techniques may still support the active session as proposed in [2, 3]. However, as indicated in the 3GPP TR 38.901 standard, even at mmWave frequencies these paths generally generate less than 10% of the LoS path power. When blockage leads to an outage, multi-connectivity functionality standardized by 3GPP can be utilized [4, 5]. However, this approach requires dense deployment of mmWave/THz base stations (BS), which may not be available during the first deployment phases. In

addition, even when both approaches are utilized together including the implementation of multi-connectivity between mmWave and THz radio access technologies (RAT), there is still a non-negligible probability that a session initially accepted for service will eventually be lost as a result of blockage. Finally, an alternative method is to embed blockage avoidance techniques in the physical layer by utilizing beam focusing [6] and/or non-diffractive beams [7]. However, these studies are still in their infancy.

Blockage detection techniques are required to enforce blockage avoidance. These techniques can be classified as proactive and reactive techniques [8]. However, for both one needs to understand not only the duration of the blockage periods but also the duration of the signal fall and rise times. The signal fall time is of critical importance as it is not allowed to benchmark the development test but can serve as an indicator of the type of impairment allowing to distinguish blockage from other events including fast fading and/or micromobility discussed in Section 3.

2.2. Related work

In this section, we outline related studies that characterize the impact of dynamic human body blockages in the THz band. We start by briefly reviewing the results of the LoS path propagation and blockage.

First, we highlight a notable deficiency in the current body of research on propagation phenomena beyond the 100 GHz. Studies on the attenuation and time dynamics of the blockage process remain scarce. For instance, [9] provides an overview of propagation loss measurements at 140 GHz for various building materials: however, the characteristics of the blockage process have not been reported. The authors of [10, 11] reported attenuation values for 300 GHz for vehicular blockage in different configurations, but these results cannot be extrapolated to human body blockage owing to the different materials and geometries.

Numerous authors have investigated the transparency characteristics of materials, as evidenced in [9, 12, 13]. For example, [14] explored the attenuation induced by various materials in the sub-THz range, whereas [12] reported measurements at four different frequencies, with two considered to belong to the sub-THz range. The study in [15] presented an analytical model based on the average blockage probability estimated empirically as follows a function of the relative positions of the receiver (Rx) and transmitter (Tx), average Tx/Rx heights, antenna parameters, and room size.

In general, despite the existing body of data regarding the penetration of sub-THz signals through diverse materials and measurement campaigns related to vehicular communications, a comprehensive characterization of the human body attenuation process within the sub-THz frequency spectrum is lacking. This is especially concerning for the signal fall and rise times.

2.3. Recent results

The first step towards this direction was taken in [8], where the authors investigated the signal blockage by the human body at a frequency of 156 GHz in the indoor environment, as shown in Fig. 2a. An example of the signal behavior under blockage is shown in Fig. 1, where it is marked by a ”direct trace”. The authors focused on blockage characteristics, including average signal attenuation and timing metrics such as signal fall and rise times, and blockage duration. For transmission over a distance of 3–7 m, the average attenuation was found to range from 8 to 15 dB depending on the LoS height and the distance between the Tx and Rx. The blockage duration varied from 5 to 10% for different Tx-Rx distances (with corresponding nominal values of 360–390 ms), while the rise and fall times of the signal gradually increased from 60 to 100 ms as the Tx-Rx distance increased and remained unchanged for different LoS altitudes.

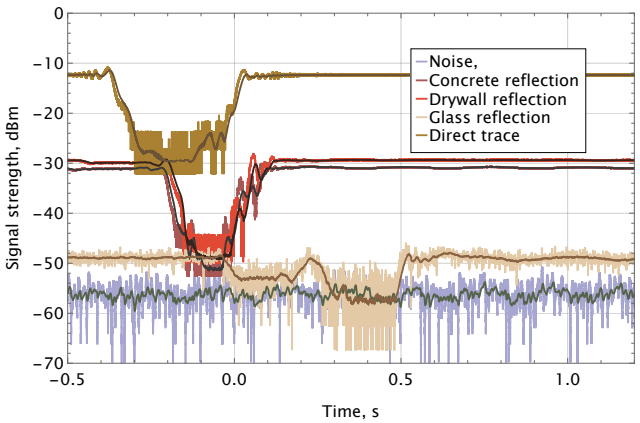
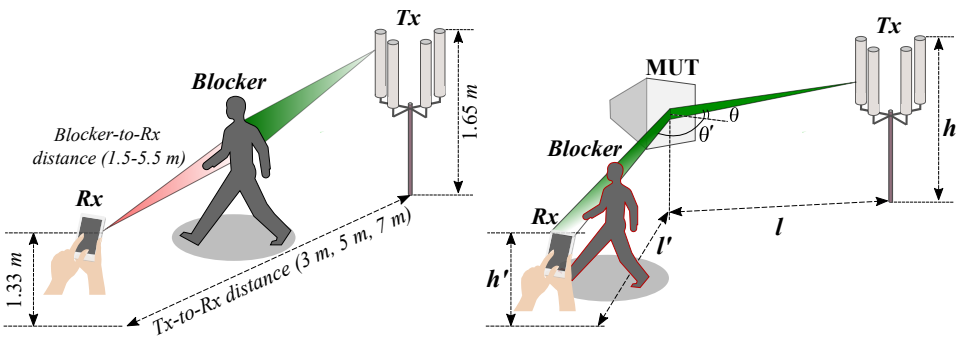


Figure 1. Signal behavior under blockage for straight and reflected propagation



(a) LoS propagation

(b) Reflected propagation

Figure 2. Illustrating considerations in LoS and reflective propagation

In [16], the authors investigated the effect of dynamic blockage by the human body on the signal strength in a highly directive channel in the sub-THz frequency range under typical indoor conditions with reflection from different materials, as shown in Fig. 2b. An illustration of the blocked signal under the reflected propagation of different materials is shown in Fig. 1. Through an extensive measurement campaign at the 156 GHz carrier frequency, reflection and blocking losses along various nLoS paths were reported, as well as the corresponding blockage duration, fall, and rise times. Empirical studies were validated using material-dependent 156 GHz radio propagation models for 24 different measurement configurations.

Upon visual examination of Fig. 1, it is clear that signal levels received from paths involving reflection in unobstructed conditions are notably lower than those of direct line-of-sight paths. This discrepancy is primarily attributed to the losses incurred during reflection. Additionally, distinct variations were observed in both obstruction patterns and signal strengths both preceding and following the occurrence of an obstruction event. It is also visible that the plots representing the incidence of radio signals are closely clustered together, almost overlapping, particularly in the case of concrete and drywall. Any slight disparities observed, especially in the case of glass, can likely be

Table 1

Mean signal attenuation, fall, and rise times

LoS propagation			
Tx-RX distance, *m*	Attenuation, dB	Rise time, ms	Fall time, ms
3 m	13.87	59.26	62.39
5 m	12.16	85.73	79.17
7 m	7.29	102.51	95.57
Reflected propagation, 3 m Tx-Rx distance			
Type of material	Attenuation, dB	Rise time, ms	Fall time, ms
Concrete	11.59	51.615	67.475
Drywall	11.15	46.231	53.186
Glass	9.76	44.081	39.338

ascribed to external variables such as the dynamic range of the Rx and the consistency of the motion trajectories of the obstructions.

Table 1 summarizes the results in [8, 16] for signal attenuation, fall, and rise times. The presented results show notice that the presence of reflection does not quantitatively change the mean signal attenuation as it remains in the range of 10–13 dB. For LoS propagation conditions the mean signal fall time varied between 60–90 ms depending on the propagation distance. The type of reflecting material affects these values. Specifically, for glass, it becomes as small as 39 ms. Even for 5G mmWave NR technology, the time gap is at least a few frames making it possible to utilize reactive blockage detection tests for multiconnectivity operations.

3. Micromobility

3.1. Motivation and impact

Modern 5G NR and future 6G cellular systems operate using antenna arrays. The use of directional antenna radiation patterns requires a beam-tracking procedure. For example, in 5G NR systems, beam tracking is performed regularly every 10–320 ms, where the actual interval is chosen by the network operator. Because of the much smaller wavelength in the 100–3000 GHz band, the half-power beamwidth (HPBW) of antenna arrays in 6G systems is substantially higher than that of 5G NR systems. This increase leads to extreme overhead in terms of (i) the time needed to perform beam tracking, (ii) frequency resources, and (iii) energy.

In terms of power, it has been shown that although the power consumption induced by beam tracking is smaller than that caused by the actual data transmission, it still adds significant overheads, especially for low data rate applications having intermittent connectivity such as web browsing [17]. The time required to find the best antenna configuration is directly proportional to the number of antenna array elements on the BS and UE sides. That is when hierarchical beam searching is utilized, as $(N_B + N_U)\delta$, where δ is the array switching time ($2 - 10\mu$s for modern arrays), N_B and N_U are the numbers of antenna configurations at the BS and UE sides, respectively. In terms of frequency resources, the overhead also scales proportionally with the number of Synchronization Signal Blocks

(SSB) per frame

$$R_O = \frac{N_{pss} + N_{sss} + N_{pbch} + N_{ssb}}{N_{rb} + N_{sc} + N_{sl} + N_{sym}},$$

where N_{sym} is the number of symbols in a slot, N_{sl} is the number of slots per frame, N_{sc} is the number of sub-carriers, N_{rb} is the total number of resource blocks (RB), N_{ssb} is the number of SSB per frame, N_{pbch} is the number of primary broadcast channel resource elements (RE), N_{sss} is the number of secondary synchronization signal REs, N_{pss} is the number of primary synchronization signal REs. Since the HPBW of the antenna array can be approximated by $102°/N$ [18] and assuming sectoral coverage of a single BS antenna, N_{ssb} can be approximated as $N_{ssb} = 120°(N_B + N_U)^2/102° = 1.18(N_B + N_U)^2$ implying quadratic dependence.

These overheads imply that the interval between the beam-tracking time instants must be minimized. In current 5G NR systems regular beam tracking is utilized with the interval between synchronization time instants set at the BS and UE and between 10 ms and 320 ms. Moving forward towards 6G systems such timings may become infeasible owing to the considered overheads, and thus, there is a need to optimize the time interval between beam tracking time instants.

3.2. Related work

The effect of micromobility is inherent to systems with directional antennas. Although 5G NR systems already utilize directional antenna radiation patterns the HPBW of the antenna arrays is rather large for micromobility to produce a noticeable impact. Nevertheless, it has already been experimentally characterized for 5G NR systems in [19], where the authors demonstrated that it may lead to degradation of the received signal strength.

As we move towards the THz band, the dimensions of the antenna arrays grow, leading to a higher HPBW directivity. The seminal work on, where the impact of user micromobility on link performance was characterized in [20]. The authors utilized inertial sensors embedded in a smartphone including a gyroscope and accelerometer to characterize the stochastic trajectories of the imaginary boresight of the beam. Shows the basic trade-off between outage time and spectral efficiency for different HPBW sizes. Furthermore in [21], the same authors utilized simple micromobility models based on diffusion processes to assess the performance of on-demand and regular beam-tracking procedures in the THz band. Their results illustrate that the latter, which is currently utilized for 5G NR systems, may not be the best option when the number of antenna elements reaches a level that is typical for 6G THz systems.

The authors in [22] carried out a micromobility emulation, where the boresight of the beam was represented by a laser pointer firmly connected to the smartphone. They considered four types of applications including video watching, VR watching, racing games, and phone calling, and demonstrated that the different types of applications are characterized by different micromobility properties. They also showed that some applications such as video watching and phone calling may allow for large times between beam tracking time instants. In [22], based on the statistical data obtained in [23], three models were formulated and parameterized: (i) two-dimensional Markov model, (ii) decomposed over Oy and Ox axes one-dimensional Markov model, and (iii) decomposed over Oy and Ox axis Brownian motion model. The models were compared using the time to outage after the beam-tracking time instant. The results demonstrated that the two-dimensional Markov model produced the best approximation of the considered metric but was characterized by the highest complexity as the number of states reached thousands. On the other hand, decomposed models are suitable as first-order approximations for those applications, where the user actions are not explicitly controlled by the application, for example, all applications excluding racing games.

Once the effect of micromobility has been characterized, studies assessing its impact on user- and system-level performance, as well as proposing solutions for remote application detection, have started to appear. First, in [24], the authors characterized the impact of micromobility on the outage performance of 6G THz systems with micromobility functionality. They demonstrated to achieve a meaningful compromise between outage and spectral efficiency the beam searching time must be decreased by at least ten times. This can be achieved by utilizing a faster switching time of the antenna configuration. Furthermore, in [5], a joint deployment of mmWave and THz 6G communication systems was considered. By accounting for both channel- and resource-dependent factors, the authors demonstrated that a system with intermittent connectivity induced by micromobility must rely on fully reliable backup options such as 4G LTE.

With respect to the applications classification for maximizing the time between beam-tracking time instants, the authors in [25] advocated the use of inertial sensors including gyroscopes and accelerometers to decrease the search space for antenna configurations on both the BS and UE sides. However, as shown by the authors, the inaccuracy of modern embedded sensors is insufficient for achieving sufficient gains in terms of the time spent searching for beams. An alternative approach based on the boresight of the antenna main lobe detection was proposed in [26]. Although nearly perfect application detection can be achieved which may result in optimal beam tracking intervals, the considered metric (boresight location) is not immediately available at the BS and UE sides.

Despite significant efforts to date, there are neither measurements nor methodologies that characterize the impact of micromobility on the received signal strength immediately after the beam tracking time instant.

3.3. Measurements methodology

Owing to the lack of miniaturized equipment operating in the sub-THz frequency band, to carry out the micromobility measurement campaign we applied the following multi-stage methodology: (i) beam center mobility emulation for different applications, (ii) Markov modeling of the time-series, (iii) generation of control traces for real measurements, and (iv) actual micromobility measurements in the sub-THz band.

Beam center micromobility emulation. In the first stage, we emulated the micromobility of the beam center using a laser pointer firmly fixed to the UE. A user utilizes UE to run applications including (i) racing games, (ii) phone calls, (iii) VR watching, and (iv) video watching. A large white screen is used to track the moving spot produced by the laser. A high-resolution camera operating at 120 fps was used to capture the time-dependent position of the laser spot. Standard signal processing techniques were utilized to convert the stop location to grid coordinates. Overall, we performed 10 independent measurements for four considered applications. The duration of each measurement was set to 120 seconds. Additional details of the experimental setup are provided in [23].

Markov modeling of time-series. Once the traces were collected, we developed models for the time series. To explicitly capture both distributional and correlational properties, we utilized a two-dimensional Markov chain framework. Accordingly, the entire range of the x and y coordinates was divided into $N \times N$ cells. The transition probabilities were then calculated to produce the Markov model. The value of N can be tuned based on the required model accuracy. Based on the analysis conducted in [23], the value of $N = 100$ was chosen. The accuracy of the model was assessed in [22], where the reported results demonstrated that it performs better than simplified models including decomposed Markov and Brownian motion models.

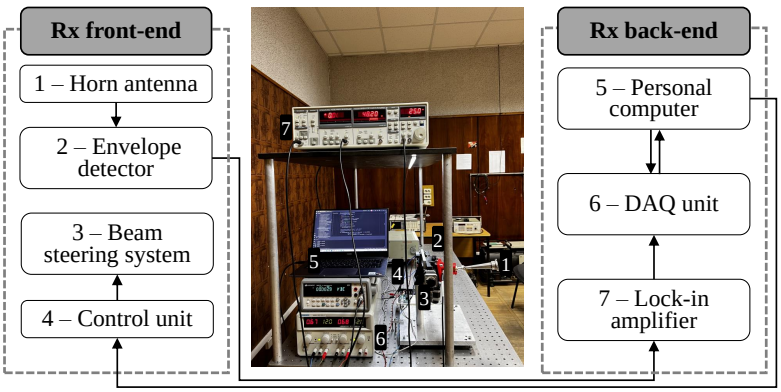
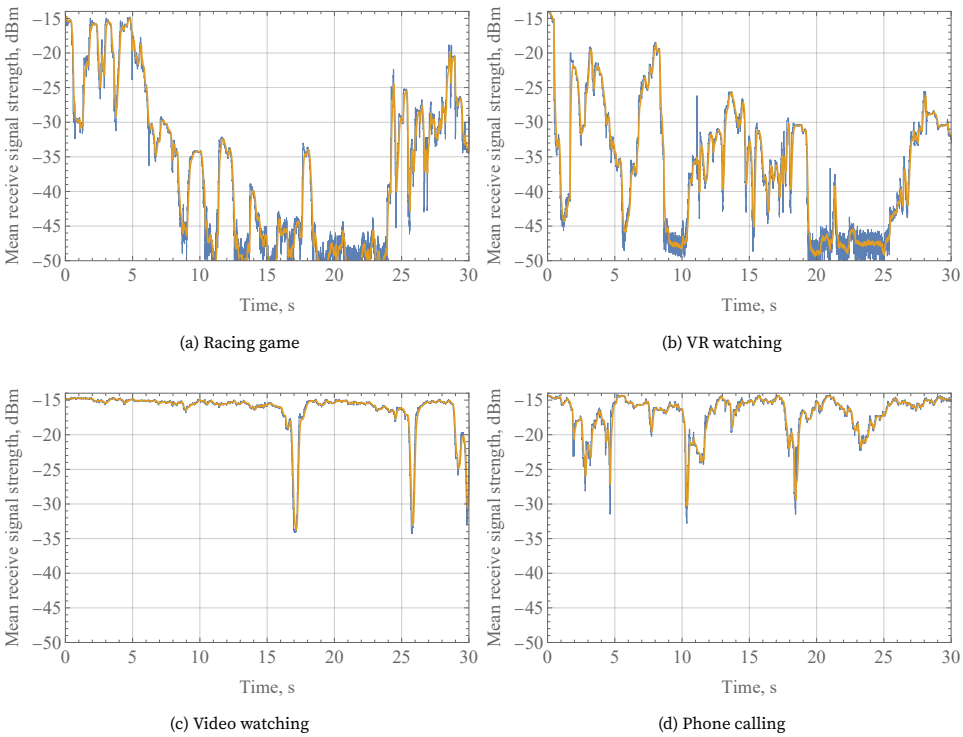


Figure 3. Schematic diagram and photograph of a steerable Rx

Control traces generation. In real measurements of the received signal strength affected by the micromobility patterns we utilized high-precision goniometers to control the direction of the antenna. We utilized the specified models to generate control traces consisting of 0s and 1s indicating positive and negative, respectively, steps for the vertical and horizontal planes. Note that the maximum time resolution of the goniometers was 150 ms/° which was slower than the micromobility speed of the racing game and VR-watching applications. To this end, we utilized the time expansion technique to align the maximum speed of the applications with that of the goniometers. The reverse procedure was implemented to restore the actual time after the measurements were performed.

Micromobility measurement campaign. The experimental setup consisted of two key components: a static transmitting module (Tx) and a mobile receiving module (Rx). The first is based on a solid-state source comprising a microwave synthesizer, followed by a sub-THz frequency multiplication chain (FMC). The Tx provides 44 mW of output power in the form of a modulated constant waveform with a modulation frequency of 25 kHz at a carrier frequency of 156 GHz. The voltage-controllable attenuator terminating the FMC provides strict amplitude control of the Tx-generated signal and performs the function of a modulator. It was also synchronized with a lock-in amplifier used at the back end of the Rx (Fig. 3). The Rx front end included a waveguide envelope detector based on a Schottky diode with 100 V/W of responsivity. The detector is equipped with a high-directivity horn antenna identical to that of Tx and is integrated with a beam steering system consisting of two motorized single-axis goniometers. The control traces used to drive the goniometers along the predefined angular trajectories have the form of matrices containing four columns with 0s and 1s (the generation procedures are provided earlier in the text). The values from the first and third columns are responsible for changing the direction of the instantaneous angular step of each goniometer (1—positive, 0—negative), the second and fourth, in turn, – for the corresponding magnitudes of the angular steps (1—24.8 arc seconds, 0—0 arc seconds). From a personal computer, the traces are submitted to control and data acquisition (DAQ) units synchronizing the transmission and directive reception of the sub-THz signal. Rx was placed 400 cm away from Tx such that their optical axes were co-aligned at the beginning of each measurement. The recorded time series of the received signal strength is statistically processed.

Time-series. The selected time-series of signal strength impaired by the micromobility of applications is shown in Fig. 4. The visual analysis of the observed data shows that one application stands out from the rest – video watching is characterized by extremely steady behavior with signal strength remaining close to the maximum even after significant time after beam alignment. On some



(a) Racing game

(b) VR watching

(c) Video watching

(d) Phone calling

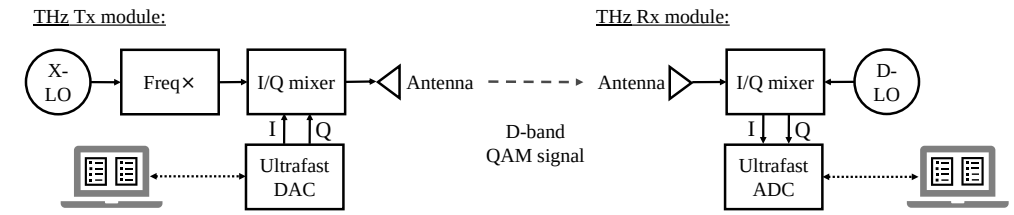
Figure 4. Selected trajectories of micromobility patterns

traces, occasional falls are observed but the signal strength quickly recovers after them. This is in close agreement with the beam center behavior reported in [23], where the center of the beam rarely fluctuated.

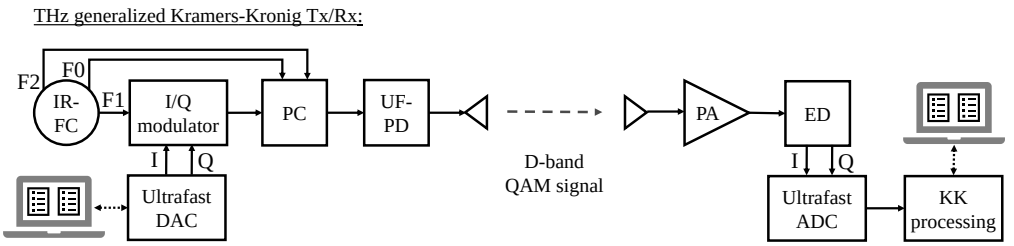
Visual analysis of the rest of the time series allows identification of the phone application as having lower received signal strength fluctuations as compared to VR watching and racing games. The maximum deviations observed for this application are generally limited to 15 dBm with an occasional drop of 30 dBm. The two remaining applications are characterized by visually similar behaviors with rapid variations. Nevertheless, as one may observe there are still large intervals for both VR watching and racing game applications, where the received signal strength remains relatively constant. This implies that the resources utilized for beam tracking may still be saved, even for these applications.

4. Communication system design

In this section, we focus on physical layer devices within coherent THz communication schemes with far-beyond-gigabit data rates. Classic microwave designs of direct conversion transmitters/receivers are briefly discussed together with their modern and actively developing radiophotonic alternatives. Path loss management via smart antenna devices embedded in reflection-aided ultra directional THz channels was further assessed in terms of radiophysics and engineering challenges.



(a) Classic direct conversion with QAM modulation



(b) Radiophotonic conversion with QAM modulation

Figure 5. Coherent THz communication schemes

4.1. Physical layer and direct conversion schemes

It is conventionally stated that the higher the carrier frequency in the communication channel, the larger the channel capacity. Considering the evolution of wireless networks, this statement can be recognized as a gradual migration of their operating frequencies from the L to D band over the last several decades. As expected, the technological background of the predecessor is not always easily adopted by the successor. This situation has several reasons that eventually originate from the available hardware fabrication and operation technologies. Currently, state-of-the-art wireless systems are being used in the THz technology. They are expected to occupy the D band, which is the lower part of the THz spectrum and spans from 110 to 170 GHz. For coherent communications, it should potentially enable baud rates of up to 55–85 GHz at the air interface. However, such a performance is only possible when wireless transceivers are compatible with the amplitude and phase manipulation methods at either i) the carrier frequency itself in the case of classic microwave modulation schemes or ii) higher-than-carrier frequencies if radiophotonic schemes are implemented [27, 28].

Quadrature amplitude modulation (QAM) schemes with 16–256 levels are commonly used in 4G/LTE and 5G systems. QAM schemes with up to 1024 modulation levels are expected to appear in 6G THz channels in the near future. Fig. 5 shows simplified examples of conventional and radiophotonic hardware implementations of coherent QAM-based THz communications. The fundamental difference between them is related to the manner in which a D-band QAM signal is generated and processed.

As shown in Fig. 5(a), the conventional method is to build a direct conversion transmitter/receiver channel with narrow-beam antenna front-ends. It utilizes four key analog elements, including X- and D-band local oscillators (X-LO and D-LO), an X-to-D-band frequency multiplication and amplification chain (Freq×), and two D-band in-phase/quadrature (I/Q) mixers, as well as two key digital elements, including ultrafast digital-to-analog and analog-to-digital converters (DAC and ADC). The robust

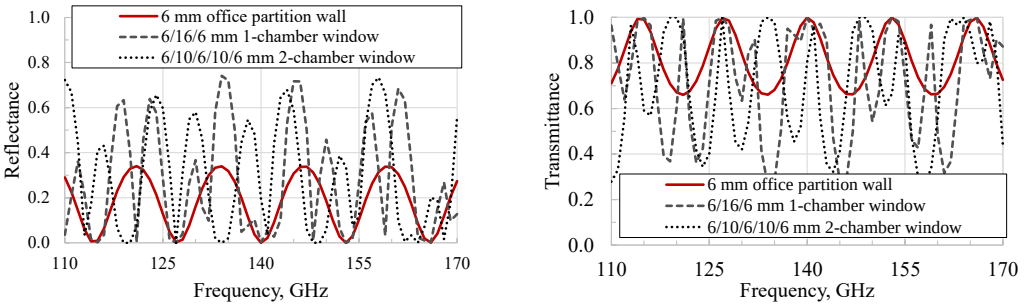
technology of X-LO acting as a fundamental frequency source is yttrium iron garnet (YIG) voltage-controlled oscillators (VCOs). When frequency- or phase-locked, 8–12 GHz YIG VCOs ensure high spectral purity with a phase noise close to the thermal limit of approximately −180 dBc/Hz above a corner frequency offset from the carrier and the output power beyond 13 dBm [29]. To elevate the generated X-band frequency to 140–150 GHz and an output power of 20–26 dBm, it is efficient to use Freq× with an A3B5-based frequency multiplier and power amplifier (PA) circuits [30] which can be tuned and cascaded to meet the desired specifications. Coding and decoding of D-band carrier signal within communication schemes with I/Q modulation ensure data rates beyond 50 Gbit/s in short distance links for Monolithic Microwave Integrated Circuits (MMICs) based on Fin Field Effect Transistor (FinFET) Complementary Metal Oxide Semiconductor (CMOS), SiGe Bipolar CMOS (BiCMOS), and SiGe Heterojunction Bipolar Transistor (HBT) platforms [31–33]. Further increases in the communication range and data transfer rates are possible if the signal strength and baud rates are increased beyond 1 W and 30 GHz levels. This, in turn, can be accomplished by replacing intrinsically limited Si-based electronics with more capable A3B5 devices [34].

The radiophonic generation and processing of THz carrier signals are also considered promising. Referring to Fig. 5(b), the I/Q modulation of the infrared frequency comb (IR-FC) components can be used, followed by their summation and downconversion with the aid of a fiber-optic power combiner (PC) and ultrafast photodiode (UF-PD). This approach enables the simple implementation of QAM modulation, which is capable of providing a high data rate and high spectral efficiency. The coherent reception of data signals at THz frequencies is possible within a simplified implementation scheme that relies on the use of an envelope detector (ED) and subsequent digital signal processing. This scheme enables the restoration of the phase of the original signal from the envelope measurements and is a variation of the Kramers-Kronig (KK) receiver. Its application made it possible to achieve the highest data transfer rate, namely, 115 Gbit/s for a 110 m channel [35], demonstrated for wireless D-band communications. The choice of ED technology in the KK receiver scheme affects the resource intensity of the transmitter/receiver fabrication and operation. In addition, it affects the complexity of generalizing the algorithm for digital processing at Rx back-end which ensures high-speed data transfer, with a low error rate.

We propose combining the features of electronic and photonic integrated circuits (ePICs) to facilitate the creation of an ultrawideband I/Q mixer that is compatible with the GaAs-on-Si platform. Our GaAs-on-Si I/Q mixer utilizes a multistage Y-junction-based power distribution network (PDN) to couple LO and carrier signals with I/Q mixing elements. The PDN was implemented in a 500 μm thick high-resistivity Si by creating a perforated dielectric structure exhibiting permittivity anisotropy and pronounced single-mode propagation paths [36]. Propagating THz was delivered to two diode-loaded tapered slot lines (TSLs) implemented on top of the single-mode waveguides of the PDN. To improve the coupling, the waveguides were terminated using a pair of distributed Brag reflectors. The I/Q control and readout signals are delivered through 2 50-Ohm slot lines connected to the rear ends of the TSLs. The diodes rely on the n+/n/etch-stopper/handle GaAs heterostructure, enabling face-down attachment to the Si wafer with further removal of the handle and etch-stopper layers before structuring the n/n+ bilayer [37]. To reduce the parasitic capacitances in the diode intrinsic circuitry, a gamma-shaped anode-suspended bridge can be implemented [38].

The expected performance and technological feasibility of the proposed ePIC for a D-band I/Q mixer were assessed in [39]. The performance simulations use numerical S-parameter analyses similar to those described in detail elsewhere [40].

The developed PDN of the mixer ePIC utilizes a constant waveguide cross-section of $0.29\lambda_0 \times 0.23\lambda_0$ along all the single-mode propagation paths. In turn, the circular segments of the integrated Y-junctions have radii of curvature of $0.69\lambda_0$ with λ_0 presenting a free-space wavelength of electromagnetic radiation. To implement a 90-degree lag in the phase required for quadrature up-



(a) Reflection spectra **(b)** Transmission spectra

Figure 6. Properties of inner and outer glass elements of the office

and down-conversions, two straight waveguide sections of different optical lengths are incorporated into the Y-junction adjacent to the LO input. Such a geometry in the first-run design ensures 7–8 dB of insertion losses for the LO and carrier signals including 1.5 dB intrinsically attributed to the transitions from PDN single-mode waveguides to diode-loaded TSLs. The corresponding reflection losses reach 28 dB whereas the isolation exceeds 13–16 dB. For carrier frequencies between 135–145 GHz, a phase difference of 89 ± 5 deg and an amplitude imbalance close to 1 dB were maintained between the I and Q mixing elements. In terms of packaging, the mixer is compatible with the WR6.5 waveguide at the UG387 flange as an input-output LO/carrier signal interface. A coaxial SMA/SMP interface was chosen for the control and readout of the I and Q states. On-chip integration with V/Q-band GaAs Schottky diode frequency multipliers [41] was also possible. Moreover, given the technological robustness of integrated Si photonic crystals of up to a few terahertz [40, 42], the feasibility of the design for D-band operation is evident. Prototyping and a detailed inspection of the forecasted mixer performance at 110-170 GHz are among our future plans.

4.2. Reflection-aided channels

Given the severe attenuation by high-humidity outdoor environments and the limitation of available carrier power, D-band 6G systems have been widely proposed for indoor ultra directional data links. However, such deployments are characterized by occasional distortions of the wireless signal upon interaction with moving users or static furniture-filling premises. Knowing the radiophysical properties of construction materials, one can find a sweet spot for placing a pencil-beam transmitter to meet the desired needs of single- and multi-user scenarios. For instance, as shown in Fig. 6, the same inner and outer flat glass elements typical of modern workspaces can effectively behave under the deployment of either LoS or reflection-aided links if the carrier frequency is shifted by 2–5% only.

Using the THz reflectometry together with characteristic matrix (CM) analysis of layered dielectric media [43], we extracted the complex permittivities of widely used construction materials, including flat glass, drywall, and aerated concrete [44]. We experimentally assessed the accuracy of the Fresnel equations (FE) and found that pronounced self-interference may potentially lead to an increase in calculation error up to 14.4–29.1%. The accuracy of FE was quantified using the normalized mean squared deviation (NMSD) of the calculated reflectances from their frequency profiles forecasted by permittivity-based CM analysis. A comparison was performed for s- and p-polarized signals in the frequency range of 130–160 GHz with the angle of incidence approaching Brewster's angle. As

Table 2

Material-dependent accuracy of reflectance calculations via FE

Type of material	Permittivity	NMSD(s)	NMSD(p)
Glass	$7.23+0.22i$	0.012	0.065
Drywall	$2.63+0.026i$	0.144	0.291
Concrete	$1.9+0.017i$	0.002	0.003

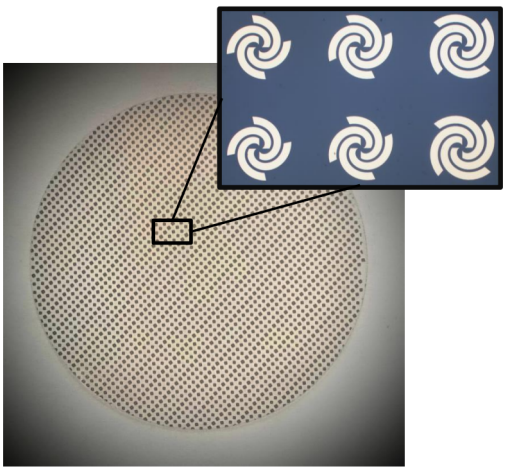


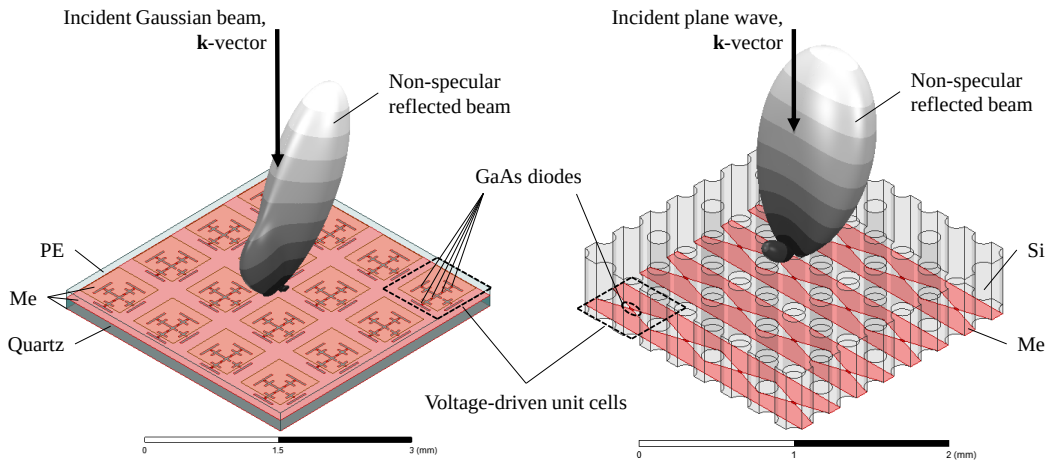
Figure 7. A 1k+ pixel static RIS for D-band operation

can be observed from Tab. 2, the accuracy is insufficient for precise channel modeling at the system level, and hybrid propagation models [45] must account for interference effects when scattering off furniture is considered.

As discussed earlier, the D-band wireless data links of 6G networks are vulnerable to dynamic signal blockage and user micromobility. This feature has an intrinsic nature and is attributed to the extreme directionality of the antenna elements at both the BS and UE operated at carrier frequencies of approximately 150 GHz. Efficient pencil-beam propagation management must be relied upon to eliminate signal fading and potential connection failure. Thus, a smart antenna environment is proposed as a promising solution to meet the demands of ultrawideband and stable communications in the D-band or beyond.

RIS-aided D-band communication technologies are expected to be deployed in three logical steps, starting with the adoption of static RISs by the 5/6G architectures with their further replacement by semi-static and dynamically controlled RIS modules. Although continuous beam steering with a wide angular field of view is attractive, practical systems are also seriously judged by their cost and operation efficiencies. Thus, technologically robust multi-state fast-beam steerers are of particular interest to commercial system developers.

Fig. 7 shows an image of a static RIS designed for operation at 150 GHz. It relies on a 140 μm thick quartz wafer with a two-sided metallization system. The rear side metallization was implemented in the form of a wafer-sized circular screen that acted as a ground plane. The front side is covered by a spiral-based periodic array structure with uniform amplitude and spacing, which utilizes more



(a) 2-bit RIS: electronic 16px cell

(b) 1-bit RIS: photonic-electronic 16px cell

(c) 2-bit RIS: embedded tuning diodes

(d) 1-bit RIS: embedded tuning diodes

Figure 8. Multi-state dynamic RISs for D-band operation

than 1000 (1k+) identical elements. The array phase increment of 120 deg results in a deflection angle of 50 deg as compared to the case of specular reflections. This design ensures operation at angles of incidence within ±75 deg supporting both linear and circular polarizations. The reflection losses were below 1 dB and the cross-polarization level was −24 dB. This performance is in good agreement with the simulation results. The simulation sequence relies on numerical Floquet port analysis and analytical beam profile construction via the array factor method. The fabricated RIS sample was successfully used for the non-distortive repetition of a 6 deg wide Gaussian beam, as reported in detail elsewhere [46]. The developed passive platform is well-suited for further integration with lumped tuning elements. These elements, in turn, need to ensure system reconfiguration speed on a microsecond scale when dynamically controlled RISs are considered. This can be achieved by using high-speed transistors or diode electronics, which greatly outperform low-speed devices based on microelectromechanical systems, liquid crystals, and phase change materials [47]. However, the latter demonstrates acceptable speed and repeatability upon state cycling [48]. This makes them attractive candidates for power-efficient semi-static RIS modules.

In the THz research community, GaAs planar Schottky diodes (PSDs) are well recognized for their excellent performance as frequency multipliers and mixers in practical high-resolution spectral instruments [41]. Fig. 8 provides examples of their use as tuning elements in multi-state dynamic RISs designed for D-band operation. The design shown in Fig. 8(a) was inspired by the award-winning 2-bit RIS, supporting the linear and circular polarizations reported in [49]. The 2 major

modifications are introduced by us to make it technologically robust and well-suited for MMIC implementation [50]. First, the air gap under the patch-carrying upper substrate was eliminated. We also replaced the substrate with a lower-permittivity material, namely, 100 μm thick polyethylene (PE) film, to maintain proper electrical spacing to the metallic (Me) slotted screen with built-in diode elements. These elements were originally presented by positive-intrinsic-negative (PIN) diodes exhibiting performance degradation in the D-band. Thus, we replaced them with PSDs, whose junction transport properties were incorporated into the parameterized unit cell hybrid model through a voltage-dependent admittance, $\hat{Y}(V_b)$:

$$\hat{Y}(V_b) = \frac{d}{dV_b}\left(I_s \exp\left(\frac{V_b}{\eta V_t}\right) - I_s\right) + i\omega\frac{\pi d_j^2}{4}\sqrt{\frac{q\varepsilon N_d}{2(\psi_{bi} - V_b)}}.$$

Physical quantities attributed to the typical diode layered structure and operating regimes are tabulated along with the description of an auxiliary ABCD-matrix math in [51]. Regarding the RIS structure, the Me screen was implemented on top of a back-metalized 140 μm thick quartz wafer. Our first-run prototyping with beam pattern measurement at 150 GHz confirmed a simulated reflection loss of 2 dB. Moreover, we measured a far-field beamwidth of 2.5 deg for the developed 16-pixel (16px) passive platform illuminated by a 6 deg wide Gaussian beam. Although the design is intrinsically compatible with the GaAs-on-quartz platform [37] of PSDs, we found that the fabrication tolerances were quite tight in terms of the radiophysical properties of the integrated RIS structure. Thus, we have revised this by Fig. 8(b).

The revised RIS has 1-bit elements represented by GaAs quarter-wave photonic crystal cells integrated with PSD-loaded tapered slot lines. The elegance of this solution is related to the use of a GaAs wafer-thick RIS structure without mechanical lapping or substrate transfer onto a low-permittivity wafer upon fabrication. Using the developed structure as part of a reconfigurable metasurface, one can implement reflective and refractive angular scanning within 60° with signal attenuation on a decibel scale.

Our simulations were strengthened by the MMIC fabrication and RIS operation developments. Scanning electron microscopy (SEM) images of the fabricated nonlinear tuning elements for the 1- and 2-bit RIS designs discussed above are shown in Fig. 8(c)–(d). They rely on Schottky contacts with at the diameter $d_j = 3\,\mu$m implemented at the interface of Ti/Pt/Au and n-GaAs, which are doped by Si to the dopant level $N_d = 4 \times 10^{-17}$ cm^3. The developed fabrication process provided an ideality factor $\eta \approx 1.5$, saturation current $I_s \approx 10$ pA, and series resistance $R_s \approx 15$ Ohm. Only two direct current (DC) control lines per row are required to implement planar angular scanning in both the proposed designs. Moreover, in the case of a 1-bit RIS circuit, it is sufficient to use a dual-channel digital-to-analog converter (DAC), controlling odd and even pairs of rows connected in parallel. For non-specular reflective beam steering, a DC bias voltage of 1 V must be delivered to them out-of-phase. In-phase biasing or no bias ensures specular reflection. The fabricated 64-element MMIC of a 140–150 GHz 1-bit RIS demonstrates decent stability when operated directly by a field-programmable gate array (FPGA) board equipped with a digital user interface and 2 50 MHz DAC channels. The normalized Allan variances were equal to 1×10^{-8} and 4×10^{-9} at integration times of 20 ms and 1 s, respectively. The consumed DC power is 13.7 mW, which corresponds to 219 mW when the number of RIS elements is increased to 1024. Such performance is on par with the power efficiency of D-band integrated transceivers relying on the SiGe BiCMOS platform [32]. It is also worth noting that the design bears structural similarity to the varactor-embedded reconfigurable metasurface recently proposed for X-band reflection-aided channels [52]. Thus, continuous beam steering can be implemented after minor tunings.

5. Conclusions

The THz frequency band is commonly considered an enabler for 6G communication systems. However, in addition to promises to deliver even more bandwidth to the air interface compared to mmWave 5G, it also brings novel challenges related to system design.

This paper aims to report on the current state-of-the-art channel characterization and communication system design. With respect to the former, we summarized recent results related to characterizing the dynamic human blockage and micromobility dynamics in LoS and nLoS THz channels. Then, we highlighted the challenges and discussed progress in electronics development, including the development of RIS intended for blockage avoidance in reflection-aided THz links.

Author Contributions: Conceptualization, G.G. and A.S.; methodology, A.P.; software, E.M.; validation, V.B., A.K. and M.E.; formal analysis, V.B.; investigation, A.K.; resources, A.P.; data curation, E.P.; writing—original draft preparation, A.S.; writing—review and editing, A.S.; visualization, E.M.; supervision, E.K.; project administration, G.G.; funding acquisition, E.K. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Funding: This paper has been supported by the Russian Science Foundation, projects no. 22-79-10128 (Sections 1–2) and no. 22-79-10279 (Section 4). The studies reported in Section 3 has been conducted as a part of strategic project "Digital Transformation: Technologies, Effectiveness, Efficiency" of Higher School of Economics development programme granted by Ministry of science and higher education of Russia "Priority-2030" grant as a part of "Science and Universities" national project. Support from the Basic Research Program of the National Research University Higher School of Economics is also gratefully acknowledged.

References

1. Park, J.-J., Lee, J., Liang, J., Kim, K.-W., Lee, K.-C. & Kim, M.-D. *Millimeter wave vehicular blockage characteristics based on 28 GHz measurements* in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)* (2017), 1–5.
2. Begishev, V., Sopin, E., Moltchanov, D., Kovalchukov, R., Samuylov, A., Andreev, S., Koucheryavy, Y. & Samouylov, K. Joint use of guard capacity and multiconnectivity for improved session continuity in millimeter-wave 5G NR systems. *IEEE Transactions on Vehicular Technology* **70,** 2657–2672 (2021).
3. Kovalchukov, R., Moltchanov, D., Begishev, V., Samuylov, A., Andreev, S., Koucheryavy, Y. & Samouylov, K. *Improved session continuity in 5G NR with joint use of multi-connectivity and guard bandwidth* in *2018 IEEE Global Communications Conference (GLOBECOM)* (2018), 1–7.
4. Begishev, V., Sopin, E., Moltchanov, D., Pirmagomedov, R., Samuylov, A., Andreev, S., Koucheryavy, Y. & Samouylov, K. Performance analysis of multi-band microwave and millimeter-wave operation in 5G NR systems. *IEEE Transactions on Wireless Communications* **20,** 3475–3490 (2021).
5. Sopin, E., Moltchanov, D., Daraseliya, A., Koucheryavy, Y. & Gaidamaka, Y. User association and multi-connectivity strategies in joint terahertz and millimeter wave 6G systems. *IEEE Transactions on Vehicular Technology* **71,** 12765–12781 (2022).
6. Zhang, H., Shlezinger, N., Guidi, F., Dardari, D., Imani, M. F. & Eldar, Y. C. Beam focusing for near-field multiuser MIMO communications. *IEEE Transactions on Wireless Communications* **21,** 7476–7490 (2022).
7. Durnin, J., Miceli, J. & Eberly, J. H. Comparison of Bessel and Gaussian beams. *Optics letters* **13,** 79–80 (1988).

8. Shurakov, A., Moltchanov, D., Prikhodko, A., Khakimov, A., Mokrov, E., Begishev, V., Belikov, I., Koucheryavy, Y. & Gol'tsman, G. Empirical blockage characterization and detection in indoor sub-THz communications. *Computer Communications* **201,** 48–58. doi:10.1016/j.comcom.2023.01.017 (2023).
9. Xing, Y. & Rappaport, T. S. *Propagation measurement system and approach at 140 GHz-moving to 6G and above 100 GHz* in *2018 IEEE global communications Conference (GLOBECOM)* (2018), 1–6.
10. Eckhardt, J. M., Petrov, V., Moltchanov, D., Koucheryavy, Y. & Kürner, T. Channel Measurements and Modeling for Low-Terahertz Band Vehicular Communications. *IEEE Journal on Selected Areas in Communications* **39,** 1590–1603 (2021).
11. Petrov, V., Eckhardt, J. M., Moltchanov, D., Koucheryavy, Y. & Kurner, T. *Measurements of reflection and penetration losses in low terahertz band vehicular communications* in *2020 14th European Conference on Antennas and Propagation (EuCAP)* (2020), 1–5.
12. Du, K., Ozdemir, O., Erden, F. & Guvenc, I. Sub-Terahertz and mmWave Penetration Loss Measurements for Indoor Environments. *arXiv preprint arXiv:2103.02745* (2021).
13. Kokkoniemi, J., Lehtomäki, J., Petrov, V., Moltchanov, D. & Juntti, M. *Frequency domain penetration loss in the terahertz band* in *2016 Global Symposium on Millimeter Waves (GSMM) & ESA Workshop on Millimetre-Wave Technology and Applications* (2016), 1–4.
14. Xing, Y. & Rappaport, T. S. *Propagation measurement system and approach at 140 GHz-moving to 6G and above 100 GHz* in *2018 IEEE global communications Conference (GLOBECOM)* (2018), 1–6.
15. Bilgin, B. A., Ramezani, H. & Akan, O. B. *Human blockage model for indoor terahertz band communication* in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)* (2019), 1–6.
16. Shurakov, A., Rozhkova, P., Khakimov, A., Mokrov, E., Prikhodko, A., Begishev, V., Koucheryavy, Y., Komarov, M. & Gol'tsman, G. Dynamic Blockage in Indoor Reflection-Aided Sub-Terahertz Wireless Communications. *IEEE Access* **11,** 134677–134689. doi:10.1109/ACCESS.2023.3337050 (2023).
17. Ostrikova, D., Beschastnyi, V., Moltchanov, D., Gaidamaka, Y., Koucheryavy, Y. & Samouylov, K. System-level analysis of energy and performance trade-offs in mmWave 5G NR systems. *IEEE Transactions on Wireless Communications* (2023).
18. Balanis, C. Antenna theory: analysis and design. *Microstrip Antennas, John Wiley & Sons* (2005).
19. Ichkov, A., Gehring, I., Mähönen, P. & Simić, L. *Millimeter-wave beam misalignment effects of small-and large-scale user mobility based on urban measurements* in *Proceedings of the 5th ACM Workshop on Millimeter-Wave and Terahertz Networks and Sensing Systems* (2021), 13–18.
20. Petrov, V., Moltchanov, D., Koucheryavy, Y. & Jornet, J. M. *The effect of small-scale mobility on terahertz band communications* in *Proceedings of the 5th ACM International Conference on Nanoscale Computing and Communication* (2018), 1–2.
21. Petrov, V., Moltchanov, D., Koucheryavy, Y. & Jornet, J. M. Capacity and outage of terahertz communications with user micro-mobility and beam misalignment. *IEEE Transactions on Vehicular Technology* **69,** 6822–6827 (2020).
22. Stepanov, N., Turlikov, A., Begishev, V., Koucheryavy, Y. & Moltchanov, D. *Accuracy assessment of user micromobility models for thz cellular systems* in *Proceedings of the 5th ACM Workshop on Millimeter-Wave and Terahertz Networks and Sensing Systems* (2021), 37–42.
23. Stepanov, N., Moltchanov, D., Begishev, V., Turlikov, A. & Koucheryavy, Y. Statistical analysis and modeling of user micromobility for THz cellular communications. *IEEE Transactions on Vehicular Technology* **71,** 725–738 (2021).
24. Moltchanov, D., Gaidamaka, Y., Ostrikova, D., Beschastnyi, V., Koucheryavy, Y. & Samouylov, K. Ergodic outage and capacity of terahertz systems under micromobility and blockage impairments. *IEEE Transactions on Wireless Communications* **21,** 3024–3039 (2021).

25. Dugaeva, S., Begishev, V., Mokrov, E. & Samouylov, K. *Using Motion Sensors For Improved Beam Tracking in THz Communications with User Micromobility* in *2022 International Conference on Modern Network Technologies (MoNeTec)* (2022), 1–8.
26. Dugaeva, S., Begishev, V. & Stepanov, N. *Utilization of Machine Learning Algorithms to Identify User Applications* in *International Conference on Distributed Computer and Communication Networks* (2023), 410–422.
27. Du Preez, J., Sinha, S. & Sengupta, K. SiGe and CMOS technology for state-of-the-art millimeter-wave transceivers. *IEEE Access* (2023).
28. Jia, S., Lo, M.-C., Zhang, L., Ozolins, O., Udalcovs, A., Kong, D., Pang, X., Guzman, R., Yu, X., Xiao, S., *et al.* Integrated dual-laser photonic chip for high-purity carrier generation enabling ultrafast terahertz wireless communications. *Nature communications* **13,** 1388 (2022).
29. Gupta, A. S., Howe, D. A., Nelson, C., Hati, A., Walls, F. L. & Nava, J. F. High spectral purity microwave oscillator: Design using conventional air-dielectric cavity. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency control* **51,** 1225–1231 (2004).
30. Makhlouf, S., Cojocari, O., Hofmann, M., Nagatsuma, T., Preu, S., Weimann, N., Hübers, H.-W. & Stöhr, A. Terahertz sources and receivers: From the past to the future. *IEEE Journal of Microwaves* **3,** 894–912 (2023).
31. Brown, T. W., Dogiamis, G. C., Yeh, Y.-S., Correas-Serrano, D., Rane, T. S., Ravikumar, S., Chou, J. C., Neeli, V. B., Koo, J., Marulanda, M., *et al.* A 50-Gb/s 134-GHz 16-QAM 3-m dielectric waveguide transceiver system implemented in 22-nm FinFET CMOS. *IEEE Solid-State Circuits Letters* **4,** 206–209 (2021).
32. Carpenter, S., Zirath, H., He, Z. S. & Bao, M. A fully integrated D-band direct-conversion I/Q transmitter and receiver chipset in SiGe BiCMOS technology. *Journal of Communications and Networks* **23,** 73–82 (2021).
33. Rodríguez-Vázquez, P., Grzyb, J., Heinemann, B. & Pfeiffer, U. R. A 16-QAM 100-Gb/s 1-M wireless link with an EVM of 17% at 230 GHz in an SiGe technology. *IEEE Microwave and Wireless Components Letters* **29,** 297–299 (2019).
34. Gustavsson, U., Frenger, P., Fager, C., Eriksson, T., Zirath, H., Dielacher, F., Studer, C., Pärssinen, A., Correia, R., Matos, J. N., *et al.* Implementation challenges and opportunities in beyond-5G and 6G communication. *IEEE Journal of Microwaves* **1,** 86–100 (2021).
35. Harter, T., Füllner, C., Kemal, J. N., Ummethala, S., Steinmann, J. L., Brosi, M., Hesler, J. L., Bründermann, E., Müller, A.-S., Freude, W., *et al.* Generalized Kramers–Kronig receiver for coherent terahertz communications. *Nature Photonics* **14,** 601–606 (2020).
36. Shurakov, A., Prikhodko, A., Belikov, I. & Gol'tsman, G. *Terahertz Hot Electron Bolometer Coherent and Direct Detectors Utilizing Si Waveguiding Structures* in *2022 IEEE 8th All-Russian Microwave Conference (RMC)* (2022), 19–22.
37. Alijabbari, N., Bauwens, M. F. & Weikle, R. M. Design and characterization of integrated submillimeter-wave quasi-vertical Schottky diodes. *IEEE Transactions on Terahertz Science and Technology* **5,** 73–80 (2014).
38. Shurakov, A., Mikhailov, D., Belikov, I., Kaurova, N., Zilberley, T., Prikhodko, A., Voronov, B., Vasil'evskii, I. & Goltsman, G. *Planar Schottky diode with a Γ-shaped anode suspended bridge* in *Journal of Physics: Conference Series* **1695** (2020), 012154.
39. Shurakov, A., Belikov, I., Prikhodko, A. & Gol'tsman, G. *Photonic-Electronic IC of a Subterahertz GaAs-on-Si IQ mixer* in *2024 49th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)* (2024 (to be published)).
40. Shurakov, A., Belikov, I., Prikhodko, A., Ershova, M. & Goltsman, G. Superconducting Electronic–Photonic Platform for HEB-Based Terahertz Spectrometers. *Applied Sciences* **13,** 5892 (2023).

41. Mehdi, I., Siles, J. V., Lee, C. & Schlecht, E. THz diode technology: Status, prospects, and applications. *Proceedings of the IEEE* **105,** 990–1007 (2017).
42. Koala, R. A., Fujita, M. & Nagatsuma, T. Nanophotonics-inspired all-silicon waveguide platforms for terahertz integrated systems. *Nanophotonics* **11,** 1741–1759 (2022).
43. Born, M. & Wolf, E. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Elsevier, 2013).
44. Shurakov, A., Rozhkova, P., Khakimov, A., Mokrov, E., Prikhodko, A., Begishev, V., Koucheryavy, Y., Komarov, M. & Gol'tsman, G. Dynamic Blockage in Indoor Reflection-Aided Sub-Terahertz Wireless Communications. *IEEE Access* **11,** 134677–134689 (2023).
45. Gargari, A. A., Polese, M. & Zorzi, M. *Full-stack comparison of channel models for networks above 100 GHz in an indoor scenario* in *Proceedings of the 5th ACM Workshop on Millimeter-Wave and Terahertz Networks and Sensing Systems* (2021), 43–48.
46. Razakova, A., Prikhodko, A., Lvov, A., Shurakov, A. & Gol'tsman, G. Subterahertz circularly polarized 1k-pixel reflective surface for 6G applications. *St. Petersburg Polytechnic University Journal - Physics and Mathematics* (2024, to be published).
47. Fu, X., Yang, F., Liu, C., Wu, X. & Cui, T. J. Terahertz beam steering technologies: from phased arrays to field-programmable metasurfaces. *Advanced optical materials* **8,** 1900628 (2020).
48. Kozyukhin, S. A., Lazarenko, P. I., Popov, A. I. & Eremenko, I. L. Phase change memory materials and their applications. *Russ. Chem. Rev* **91,** 9 (2022).
49. Dai, L., Wang, B., Wang, M., Yang, X., Tan, J., Bi, S., Xu, S., Yang, F., Chen, Z., Di Renzo, M., *et al.* Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and experimental results. *IEEE access* **8,** 45913–45923 (2020).
50. Shurakov, A., Prikhodko, A., Belikov, I., Razakova, A. & Gol'tsman, G. *Integrated Circuit of an Intelligent Reflecting Surface for sub-THz Wireless Communication* in *2023 IEEE 18th International Conference on Nano/Micro Engineered and Molecular Systems (NEMS)* (2023), 183–187.
51. Prikhodko, A., Yaropolov, T., Shurakov, A. & Gol'tsman, G. *Unit cell model of a terahertz intelligent reflecting surface with Schottky microcontacts* in *35th European Modeling and Simulation Symposium (EMSS 2023)* (2023), 019.
52. Ratni, B., de Lustrac, A., Piau, G.-P. & Burokur, S. N. Active metasurface for reconfigurable reflectors. *Applied Physics A* **124,** 1–8 (2018).

Information about the authors

Shurakov, Alexander S.—PhD, Associate Professor, Department of General and Experimental Physics, Moscow Pedagogical State University (e-mail: alexander@rplab.ru, phone: +7 (499) 766-48-87, ORCID: 0000-0002-4671-7731, ResearcherID: E-4118-2014, Scopus Author ID: 55266061300)

Mokrov, Evgeny V.—Candidate of Physical and Mathematical Sciences, Senior lecturer of Department of Probability Theory and Cyber Security of Peoples' Friendship University of Russia (RUDN University) (e-mail: mokrov-ev@rudn.ru, phone: +7 (495) 952-28-23, ORCID: 0000-0003-3290-4541, ResearcherID: AAK-6348-2021, Scopus Author ID: 56512031300)

Prikhodko, Anatoliy N.—Junior Researcher, Specialized Department of Quantum Optics and Telecommunications of Scontel CJSC, HSE University (e-mail: anprihodko@hse.ru, phone: +7 (495) 772-95-90, ORCID: 0000-0002-4859-8975, ResearcherID: ADC-0507-2022, Scopus Author ID: 57207500541)

Ershova, Margarita I.—Junior Researcher, Laboratory of quantum detectors, Moscow Pedagogical State University (e-mail: mi.ershova@mpgu.su, phone: +7 (499) 766-48-87, ORCID: 0009-0009-6785-4389, ResearcherID: JNB-5214-2023, Scopus Author ID: 58298409900)

Begishev, Vyacheslav O.—Candidate of Physical and Mathematical Sciences, Assistant professor of the Department of Probability Theory and Cyber Security of Peoples' Friendship University of Russia (RUDN University) (e-mail: begishev-vo@rudn.ru, phone: +7 (495) 952-28-23, ORCID: 0000-0002-7232-4157, ResearcherID: AAF-6491-2019, Scopus Author ID: 56562837400)

Khakimov, Abdukodir A.—Candidate of Technical Sciences, Junior researcher of the Department of Probability Theory and Cyber Security of Peoples' Friendship University of Russia (RUDN University) (e-mail: khakimov-aa@rudn.ru, phone: +7 (495) 952-28-23, ORCID: 0000-0003-2362-3270, ResearcherID: AAD-1134-2019, Scopus Author ID: 57194233776)

Koucheryavy, Yevgeny A.—Doctor of Technical Sciences, Professor of the Department of Probability Theory and Cyber Security of Peoples' Friendship University of Russia (RUDN University) (e-mail: kucheryavyy-ea@rudn.ru, phone: +7 (495) 952-28-23, ORCID: 0000-0003-3976-2971, ResearcherID: D-7976-2018, Scopus Author ID: 6507253900)

Gol'tsman, Gregory N.—Doctor of Physical and Mathematical Sciences, Head of Department, Department of General and Experimental Physics, Moscow Pedagogical State University (e-mail: goltsman@rplab.ru, phone: +7 (499) 766-48-87, ORCID: 0000-0002-1960-9161, ResearcherID: A-4189-2014, Scopus Author ID: 7006771637)

UDC 004.2, 004.7
PACS 07.05.Tp
DOI: 10.22363/2658-4670-2024-32-2-181–201 EDN: CFLJNF

Текущеее состояние исследований в области моделирования каналов и разработки систем терагерцового диапазона

А. С. Шураков[1,2], Е. В. Мокров[3], А. Н. Приходько[1,2,4], М. И. Ершова[1], В. О. Бегишев[3], А. А. Хакимов[3], Е. А. Кучерявый[3], Г. Н. Гольцман[1,2,4]

[1] *Московский педагогический государственный университет, ул. Малая Пироговская, д. 1, стр. 1, Москва, 119991, Российская Федерация*
[2] *Национальный исследовательский университет «Высшая школа экономики», ул. Мясницкая улица, д. 20, Москва, 101000, Российская Федерация*
[3] *Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация*
[4] *Российский квантовый центр, Сколково, 143025, Российская Федерация*

Аннотация. На текущий момент стандартизация систем 5G Новое Радио, работающих в диапазоне частот микро- и миллиметровых волн, завершена. Исследовательское сообщество, стандартизищирующие организации и производители оборудования начинают решать вопрос о том, какой может или должна быть система связи 6G. Хотя технологические особенности все еще находятся на ранней стадии разработки, существует общее мнение, что эти системы будут использовать нижнюю часть терагерцового диапазона, 100–300 ГГц. Этот диапазон создает ряд специфических проблем для разработчиков систем, включая эффекты, связанные с характеристиками канала, а также концептуально новые требования к электронике. Целью данной статьи является обзор текущего состояния исследований в области определения характеристик каналов и проектирования систем связи. В отношении первых мы будем рассматривать динамическую блокировку путей распространения сигнала телом человека и эффект микромобильность. С точки зрения системного уровня, мы концентрируемся на устройствах для схем прямого преобразования и разработке так называемых реконфигурируемых интеллектуальных поверхностей, которые потенциально могут служить экономичным методом устранения блокировок.

Ключевые слова: терагерцовый диапазон, 6G, блокировка, микромобильность, эксперимент, когерентная связь, прямое преобразование, реконфигурируемя интеллектуальная поверхность

Discrete & Continuous Models
& Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (2) 202–212
http://journals.rudn.ru/miph

Research article

UDC 519.21
PACS 52.25.Fi
DOI: 10.22363/2658-4670-2024-32-2-202–212 EDN: CRLKAJ

Clenshaw algorithm in the interpolation problem by the Chebyshev collocation method

Konstantin P. Lovetskiy, Anastasiia A. Tiutiunnik,
Felix Jose do Nascimento Vicente, Celmilton Teixeira Boa Morte

RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation

(received: October 14, 2023; revised: November 12, 2023; accepted: November 18, 2023)

Abstract. The article describes a method for calculating interpolation coefficients of expansion using Chebyshev polynomials. The method is valid when the desired function is bounded and has a finite number of maxima and minima in a finite domain of interpolation. The essence of the method is that the interpolated desired function can be represented as an expansion in Chebyshev polynomials; then the expansion coefficients are determined using the collocation method by reducing the problem to solving a well-conditioned system of linear algebraic equations for the required coefficients. Using the well-known useful properties of Chebyshev polynomials can significantly simplify the solution of the problem of function interpolation. A technique using the Clenshaw algorithm for summing the series and determining the expansion coefficients of the interpolated function, based on the discrete orthogonality of Chebyshev polynomials of the 1st kind, is outlined.

Key words and phrases: interpolation of functions by the Chebyshev collocation method, Clenshaw algorithm for accelerating calculations

Citation: Lovetskiy K. P., Tiutiunnik A. A., do Nascimento Vicente F. J., Boa Morte C. T., Clenshaw algorithm in the interpolation problem by the Chebyshev collocation method. *Discrete and Continuous Models and Applied Computational Science* **32** (2), 202–212. doi: 10.22363/2658-4670-2024-32-2-202–212. edn: CRLKAJ (2024).

1. Introduction

The construction of efficient numerical methods for solving differential and integral equations is an important element in solving applied problems in various fields, such as aerospace engineering, meteorology, physical oceanography, mechanical engineering, and nuclear energy. Taking this into account, we will consider and analyze the efficiency of some spectral algorithms for function interpolation, which are often used when solving equations of mathematical physics.

Spectral methods are a class of methods used in applied mathematics for the numerical solution of certain differential and integral equations, sometimes using the fast Fourier transform [1–4]. The idea is to present the desired solution $u(x)$ in the form of a finite sum of "basis functions" $\varphi_n(x)$ with subsequent choice of the coefficients in the sum that satisfy the specified equations.

$$u(x) \approx u_N(x) = \sum_{n=0}^{N} a_n \varphi_n(x). \tag{1}$$

Substituting this expression into equation

$$Lu(x) = f(x),$$

where L is the operator of the differential or integral equation, results in the appearance of the so called residual function

$$R(x; a_0, a_1, ..., a_N) = Lu_N(x) - f(x). \tag{2}$$

The residual function $R(x; a_0, a_1, ..., a_N)$ is identically equal to zero, when $u_N(x)$ is an exact solution. Therefore, the main goal of an algorithm for solving the problem studied is to minimize the residual function by choosing the appropriate spectral coefficients a_n, $n = 0, 1, ..., N$.

In the Galerkin–Petrov method, the solution is expanded in one basis, the coordinate functions, and the orthogonality of the residual is required to another basis, the projection functions.

The choice of trial basis functions $\varphi_n(x)$ in Eq. (1) and the testing functions, the basis for minimizing the residual (2), is the key feature that distinguishes spectral methods from the finite difference and finite element methods. In the latter two methods, the trial/testing functions are local functions with finite carriers. On the contrary, spectral methods use globally smooth functions as trial/testing functions. The simplest basis functions are power functions — the monomials $\varphi_n(x) = x^n$ such that

$$u_N(x) = a_0 + a_1 x + a_2 x^2 + ... + a_N x^N.$$

The most frequently used trial/testing functions are trigonometric functions or orthogonal polynomials (usually, eigenfunctions of singular Sturm–Liouville problems), which comprise

Fourier spectral method	$\varphi_k(x) = e^{ikx}$,
Chebyshev spectral method	$\varphi_k(x) = T_k(x)$,
Legendre spectral method	$\varphi_k(x) = L_k(x)$,
Laguerre spectral method	$\varphi_k(x) = \boldsymbol{L}_k(x)$,
Hermit spectral method	$\varphi_k(x) = H_k(x)$.

Here $T_k(x)$, $L_k(x)$, $\boldsymbol{L}_k(x)$ and $H_k(x)$ are Chebyshev, Legendre, Laguerre and Hermit polynomials of the k-th power, respectively.

It is exactly the choice of the testing functions for calculating the residual that determines the name of the methods for solving equations:

- **Galerkin method**. The main feature of the method is the coincidence of the trial basis and the testing one.
- **Petrov–Galerkin method**. The trial basis and the testing one are different.
- **Collocation method**. For the grid points chosen in advance (collocation) from the search domain it is required that the residual function is zero.

2. Variants of the colocation method for interpolation of functions

Let us study the collocation method for solving the interpolation problem, based on the representation of the approximating function in the form of a finite sum (1) of its expansion in the Chebyshev polynomials of the first kind. In this case, the Chebyshev functions of the first kind $\varphi_n(x) = T_n(x)$ are considered as the orthogonal basis.

The approach to solving the problem of approximating $f(x)$ based on the collocation method consists in choosing not only the finite-dimensional space of possible solutions (usually, polynomials up to a certain power), but also the number and position of grid points in the solution's search domain (called collocation points). Then it is necessary to choose such coefficients a_n, $n = 0, 1, ..., N$ of the expansion of the solution in the basis polynomials, which provide (2) to be exactly satisfied at the collocation points.

$$R(x; a_0, a_1, ..., a_N) \equiv \sum_{n=0}^{N} a_n T_n(x_i) - f(x_i) = 0, \quad i = 0, ..., N. \tag{3}$$

The expansion coefficients a_n, $n = 0, 1, ..., N$ are conventionally found from the solution of system (3), for the unambiguous solvability of which it is necessary that the Chebyshev matrix determinant should be nonzero: $\det[T_j(x_k)] \neq 0$, $j, k = 0, ..., n$. The choice of different grid points not coincident with each other guarantees the nondegeneracy of the determinant and, thus, uniqueness of the solution of (3) upon such a choice [3].

In the matrix form, the system of equations (3) can be written as

$$\boldsymbol{T}\boldsymbol{a} = \boldsymbol{f},$$

where the elements $T_j(x_k)$ of the k-th row of matrix $\boldsymbol{T}$ are Chebyshev polynomials of the first kind of the j-th degree, $\boldsymbol{a}^T = (a_0, a_1, ..., a_N)$ is the vector of interpolation parameters, $\boldsymbol{f}^T = (f_0, f_1, ..., f_N)^T$ is the vector of values of the interpolated function at the grid points $(x_0, x_1, ..., x_N)$.

Using the properties of the discrete orthogonality of the polynomials $T_n(x)$ on the Chebyshev–Lobatto grid allows constructing an efficient algorithm for finding the interpolation coefficients. Let us transform the system of residual equations (3) so that the SLAE matrix would be almost orthogonal. For this purpose, multiply the first and the last equation in (3) by the factor $1/\sqrt{2}$ to obtain an equivalent "modified" system with the new matrix $\tilde{\boldsymbol{T}}$ (instead of $\boldsymbol{T}$) and vector $\tilde{\boldsymbol{f}}$ instead of $\boldsymbol{f}$. The new system's advantage is that its matrix is "almost orthogonal", and multiplying it from the left by the transposed matrix $\tilde{\boldsymbol{T}}^T$ we obtain the diagonal matrix:

$$\tilde{\boldsymbol{T}}^T\tilde{\boldsymbol{T}} = \begin{bmatrix} n & 0 & 0 & ... & 0 \\ 0 & \frac{n}{2} & 0 & ... & 0 \\ 0 & 0 & \frac{n}{2} & ... & 0 \\ ... & ... & ... & \ddots & ... \\ 0 & 0 & 0 & ... & n \end{bmatrix}$$

The multiplication of the modified system (3) from the left by the transposed matrix $\tilde{\boldsymbol{T}}^T$ leads to a simple matrix equation with a diagonal matrix that determines the desired expansion coefficients

$$\tilde{\boldsymbol{T}}^T\tilde{\boldsymbol{T}}\boldsymbol{a} = \tilde{\boldsymbol{T}}^T\tilde{\boldsymbol{f}}, \tag{4}$$

where $\tilde{\boldsymbol{f}} = (\boldsymbol{f}_0/\sqrt{2}, \boldsymbol{f}_1, ..., \boldsymbol{f}_{n-1}, \boldsymbol{f}_n/\sqrt{2})$. In the right-hand side of Eq. (4), vector $\boldsymbol{g} = \tilde{\boldsymbol{T}}^T\boldsymbol{f}$ is obtained. In this notation, the coefficients of function $f(x)$ expansion in Chebyshev polynomials of the first kind are easily written explicitly as

$$\begin{cases} a_0 = g_0/n, \\ a_1 = 2g_1/n, \\ a_2 = 2g_2/n, \\ ... \\ a_n = g_n/n. \end{cases} \tag{5}$$

Thus, relations (5) unambiguously determine the expansion coefficients of the approximating polynomial $u_N(x) = \sum_{n=0}^{N} a_n T_n(x)$.

The described approach to the solution of interpolation problem allows stable solution of both the problem of reconstructing the approximating polynomial expansion coefficients and the problem of calculating the interpolant values at an arbitrary point in the domain of definition of the desired function. However, the speed of executing these operations still leaves much to be desired, even though the use of the Gauss–Lobatto grid eliminates the need to solve a system of linear algebraic equations (3) with completely filled matrix. The problem turned out to reduce to multiplying the matrix $\tilde{\boldsymbol{T}}^T$ by the vector $\tilde{\boldsymbol{f}}$ and dividing the components of the resulting vector by the appropriate elements of the diagonal matrix $\tilde{\boldsymbol{T}}^T\tilde{\boldsymbol{T}}$.

The use of various modifications of the Clenshaw algorithm can significantly speed up the solution of the interpolation problem.

3. Clenshaw algorithm – increasing the efficiency of calculating the Chebyshev series at an arbitrary point in the approximation interval

Having the coefficients of the polynomial expansion of the desired function makes it possible to calculate the values of the interpolating polynomial at arbitrary points of the approximation interval $x \in [-1, 1]$ directly as

$$u_N(x) = a_0 T_0(x) + a_1 T_1(x) + a_2 T_2(x) + \ldots + a_N T_N(x). \tag{6}$$

However, calculating the sum using Eq. (6) is not optimal [5]. Efficient and stable summation of this series is possible based on the Clenshaw algorithm using the recurrent three-term relation.

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x).$$

This approach allows calculating the value of the next polynomial in the summable series from the values of a pair of previous polynomials using simple multiplication and addition operations. First, it is just necessary to calculate

$$T_0(x) = 1, \quad T_1(x) = x.$$

and to launch the iteration process [5]. Detailed information on the algorithm and stability of the summation process is presented in the paper by Fox and Parker [6].

Clenshaw's algorithm generalizes not only to various types of Chebyshev polynomials; it applies to any class of functions that can be defined by a three-term recurrence relation. Clenshaw's algorithm calculates the weighted sum of a finite series of functions $\varphi_k(x)$:

$$S(x) = \sum_{k=0}^{n} a_k \varphi_k(x),$$

where $\varphi_k(x)$, $k = 0, 1, \ldots, n$ is a sequence of functions satisfying three-term relations

$$\varphi_{k+1}(x) = \alpha_k(x)\varphi_k(x) + \beta_k(x)\varphi_{k-1}(x),$$

with pre-known coefficients $\alpha_k(x)$ and $\beta_k(x)$.

The algorithm is most efficient when $\varphi_k(x)$ are functions hard to calculate directly, whereas the calculation of coefficients $\alpha_k(x)$ and $\beta_k(x)$ is relatively simple. In the most widespread applications, $\alpha_k(x)$ is independent of k and $\beta_k(x)$ is a constant depending on neither x, nor k. In our case (6), $\alpha_k(x) = 2x$ and $\beta_k(x) = -1$.

To execute the series summation for a given sequence of coefficients $\alpha_0, \alpha_1, \ldots, \alpha_n$, it is first necessary to calculate the values $b_k(x)$ of the auxiliary sequence using the "inverse" recurrence relation

$$\begin{aligned} b_{n+1}(x) &= b_{n+2}(x) = 0, \\ b_k(x) &= a_k + \alpha_k(x) b_{n+1}(x) + \beta_{k+1}(x) b_{n+2}(x). \end{aligned} \tag{7}$$

It is important to note that to construct the sequence $b_k(x)$, $k = n, ..., 1$ no calculation (and knowledge) of $\varphi_k(x)$ values is necessary. After determining the coefficients $b_2(x)$ and $b_1(x)$ only two simplest values of $\varphi_0(x)$ and $\varphi_1(x)$ are enough to obtain the desired sum

$$S(x) = a_0\varphi_0(x) + \varphi_1(x) b_1(x) + \beta_1(x)\varphi_0(x) b_2(x).$$

Let us consider in more detail the Clenshaw summation method for fast and stable calculation of the sum of series (6) — the recursive method for calculating a linear combination of Chebyshev polynomials.

$$u_n(x) = a_0 + a_1 T_1(x) + a_2 T_2(x) + ... + a_n T_n(x).$$

Let us take into account that the coefficients in the recurrence relation for the Chebyshev polynomials are $\alpha(x) = 2x$ and $\beta(x) = -1$ and the initial polynomials are $T_0(x) = 1$, $T_1(x) = x$. Then the "reverse" recurrent sequence for calculating coefficients $b_k(x)$ has the form

$$b_k(x) = a_k + 2x b_{k+1}(x) - b_{k+2}(x), \quad k = n, n-1, ..., 1,$$

with zero "initial" coefficients

$$b_{n+1}(x) = b_{n+2}(x) = 0.$$

Now we calculate the value of the desired sum using the "reverse" recurrence relation

$$\begin{aligned} u_n(x) = a_0 + \sum_{k=1}^{n} a_k T_k(x) = \sum_{k=1}^{n} (b_k - 2x b_{k+1} + b_{k+2}) T_k(x) = \\ = a_0 + b_1 x + b_2(2x T_1(x) - T_0(x)) - 2x b_2 T_1(x) + \sum_{k=3}^{n} b_k (T_k(x) - 2x T_{k-1}(x) + T_{k-2}(x)). \end{aligned}$$

The last term turns into zero and the ultimate value of the sum of series $u_n(x)$ is determined by the formula

$$u_n(x) = a_0 + x b_1(x) - b_2(x). \tag{8}$$

Example 1. A program in pseudocode implementing the summation according to the Clenshaw algorithm.

The simplest (using no Clenshaw algorithm) program for calculating the sum of series (6) at a fixed point $x \in [-1, 1]$ can be presented directly as

```
Sum := a[0] + x * a[1]
t := x; t1 := 1; c := 2 * x
for k := 2 to n do
begin
  t2 := t1; t1 := t;
  t := c * t1 - t2;
  Sum := Sum + t * a[k]
end
```

On completion of the program operation, the variable Sum contains the desired $S(x)$ value. The program uses $2n + 1$ multiplications and $2n - 1$ additions.

Following Clenshaw and using the auxiliary "reverse" recurrence formula (7), we transform the calculation program so that the number of multiplications reduces (practically by two times) to $n + 2$ and the number of additions remains to be $2n + 1$.

The subroutine pseudocode is

```
Sum := a[n]; b1 := 0; c := 2 * x;
for k := n - 1 downto 0 do
begin
  b2 := b1; b1 := Sum;
  Sum := a[k] + c * b1 - b2
end;
Sum := 0.5 * (Sum - b2)
```

Instead of the above "beautiful" version of the pseudocode, it is possible to use an alternative version with even fewer operations, namely, $n + 1$ multiplications and $2n$ additions

```
Sum := a[n]; b1 := 0; c := 2 * x;
for k := n - 1 downto 1 do
begin
  b2 := b1; b1 := Sum;
  Sum := a[k] + c * b1 - b2
end;
Sum := a[0] + x * b1 - b2
```

4. Expanding the scope of application of the Clenshaw method (calculation of interpolant expansion coefficients)

The calculation of expansion coefficients $\{c_0, c_1, ..., c_n\}$ reduces to a solution of the system (4) of linear algebraic equations. The most laborious stage is the operation of multiplying the vectors of transposed Chebyshev matrix $\tilde{\boldsymbol{T}}^T$ by vector $\tilde{\boldsymbol{f}} = (\boldsymbol{f}_0/\sqrt{2}, \boldsymbol{f}_1, ..., \boldsymbol{f}_{n-1}, \boldsymbol{f}_n/\sqrt{2})^T$. It turns out that the structure of the Chebyshev matrix $\boldsymbol{T}$ allows efficient use of the Clenshaw algorithm for substantial reduction of the number of operations compared to the conventional algorithm of multiplying a matrix by a vector.

Let us present the modification of the Clenshaw algorithm simplifying the procedure of multiplying the transposed Chebyshev matrix by the right-hand side vector $\boldsymbol{T}^T\boldsymbol{f}$ when using the Gauss–Lobatto grid.

$$\boldsymbol{T}^T\boldsymbol{f} = \begin{bmatrix} T_{0,0} & T_{0,1} & T_{0,2} & ... & T_{0,n} \\ T_{1,0} & T_{1,1} & T_{1,2} & ... & T_{1,n} \\ T_{2,0} & T_{2,1} & T_{2,2} & ... & T_{2,n} \\ ... & ... & ... & \ddots & ... \\ T_{n,0} & T_{n,1} & T_{n,2} & ... & T_{n,n} \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ ... \\ f_n \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ ... \\ c_n \end{bmatrix}$$

To get the value of a particular j-th component of the vector of coefficients, it is necessary to multiply the j-th row of matrix $\boldsymbol{T}^T$ by the vector $\boldsymbol{f}$.

$$c_j(x) = \sum_{k=0}^{n} f_k T_j(x_k) = f_0 T_j(x_0) + f_1 T_j(x_1) + f_2 T_j(x_2) + ... + f_n T_j(x_n).$$

To simplify the calculation of the desired sum, the product of the row of Chebyshev transposed matrix by the vector on the right-hand side, we use the representation of a Chebyshev polynomial of the first kind of the j-th order in the trigonometric form:

$$T_j(\cos\theta) = \cos(j\theta)$$

$$x_k = \cos(k\pi/n), \quad k = 0, 1, ..., n.$$

Each j-th row of the transposed Chebyshev matrix contains the values of a Chebyshev polynomial of the j-th order at the Gauss–Lobatto grid points $x_k = \cos(k\pi/n)$, $k = 0, 1, ..., n$.

Let us execute the change of variables

$$y_j = j\pi/n, \quad j = 0, 1, ..., n.$$

$$T_j(x_k) = \cos(ky_j).$$

This change allows using the identity $T_j(x_k) = \cos(j(k\pi/n)) = \cos(k(j\pi/n)) = T_k(x_j)$ for a transition to the conventional Clenshaw scheme of calculating the product of the transposed (symmetric) Chebyshev matrix by the vector of the right-hand side. That is, the product of coefficients f_k,$k = 0, 1, ..., n$ by the corresponding values of the polynomial of the j-th order $T_j(x_k)$ at the grid points x_k, $k = 0, 1, ..., n$ for calculating the interpolant expansion coefficient $c_j(x)$ can be replaced with a product of the same coefficients by the corresponding values of the polynomials of the k-th order $T_k(x_j)$ at one point x_j of the grid with the number corresponding to that of the calculated interpolant expansion coefficient.

$$c_j = \sum_{k=0}^{n} f_k T_j(x_k) = f_0 T_j(0 \cdot y_i) + f_1 T_j(1 \cdot y_i) + f_2 T_j(2 \cdot y_i) + ... + f_n T_j(n \cdot y_i).$$

Denoting

$$\theta = j\pi/n,$$

we arrive at a formula for calculating the j-th expansion coefficient of the desired interpolant in the reduced form:

$$c_j = \sum_{k=0}^{n} f_k T_j(x_k) = \sum_{k=0}^{n} f_k \cos(k\theta), \quad j = 0, 1, ..., n.$$

To calculate each expansion coefficient c_j, $j = 0, 1, ..., n$ using the Clenshaw scheme, two approaches are possible.

In the first of them, it is possible to use the three-term recurrence relation expressing cosines of multiple angles $\cos nx$ through two cosines of preceding multiplicities $\cos((n-1)x)$, $cos((n-2)x)$ and the value of $\cos(x)$:

$$c_j = \sum_{k=0}^{n} f_k \cos(k\theta), \quad \theta = \frac{j\pi}{n}, \quad j = 0, 1, ..., n.$$

The second approach leads to using the variant of Clenshaw scheme (7)–(8) already studied above.

$$c_j = \sum_{k=0}^{n} f_k T_j(x_k), \quad j = 0, 1, ..., n.$$

It is known that the coefficients of the three-term recurrence relation for the Chebyshev polynomials of the first kind are

$$\alpha(n,x) = \alpha(x) = 2x, \quad \beta(n,x) = \beta(x) = -1,$$

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x)$$

and the polynomials of the 1-st and the 2-nd order have the form

$$T_0(x) = 1, \quad T_1(x) = x.$$

Having summed over the specified coefficients $c_0, c_1, c_2, ..., c_n$, we calculate the values $b_k(x)$ using the "reverse" recurrence formula:

$$b_{n+1}(x) = b_{n+2}(x) = 0,$$

$$b_k(x) = c_k + 2xb_{k+1}(x) - b_{k+2}(x), \quad k = n, n-1, ..., 1.$$

Then

$$c_k = b_k - 2xb_{k+1}(x) + b_{k+2}(x), \quad k = n, n-1, ..., 1.$$

Now we substitute these coefficients into the series sum formula

$$p_n(x) = \sum_{k=0}^{n} c_k T_k(x) = c_0 + b_1 x - b_2.$$

To get rid of using the value of argument in the ultimate formula for the sum of the truncated Chebyshev series, it is possible to continue the loop process by one additional step and to calculate the zeroth coefficient of the reverse sequence by the formula

$$b_0 = c_0 + 2xb_1 - b_2,$$

which allows calculating the product $b_1 x = \dfrac{b_0 - c_0 + b_2}{2}$,

Substituting the calculated value of b_1 x into the sum expression, we get an alternative variant of the formula for the sum of products of indexed coefficients by the Chebyshev polynomials of the first kind of the appropriate degree:

$$p_n(x) = c_0 + b_1 x - b_2 = c_0 + \frac{b_0 - c_0 + b_2}{2} - b_2 = \frac{1}{2}[c_0 + b_0 - b_2].$$

The ultimate value of the desired sum depends only on the coefficient c_0 of the series sought for and two coefficients b_0, b_2 obtained as a result of running the reverse recurrence sequence.

5. Conclusions

Clenshaw described an algorithm that allows calculating the final sum of the Fourier series in terms of sines and cosines. Clenshaw's algorithm is a recursive method for summing a linear combination of Chebyshev polynomials [5, 6]. The method was published by Charles William Clenshaw in 1955. This is a generalization of Horner's method for summing a linear combination of monomials. Although this method is named after William George Horner, it has been known for a long time – Horner himself attributed it to Joseph–Louis Lagrange. But the method was described and used many hundreds of years ago by Chinese and Persian mathematicians. After the advent of computers, this algorithm became fundamental for efficient calculations with polynomials.

The use of many well-known useful properties of Chebyshev polynomials can significantly improve the software implementation of the function interpolation algorithm based on the Clenshaw method. In future, the authors intend to use the outlined interpolation technique for a stable implementation of algorithms for calculating definite integrals, derivatives of functions using matrices of spectral Chebyshev differentiation and finding antiderivative functions using integration matrices.

For example, to calculate definite integrals, it may be useful to calculate the sums of modified series of Fourier type — only even or only odd terms of the series.

With the rapid development of modern technology, many types of interpolation methods have been proposed, including piecewise constant, linear, polynomial and spline interpolation methods [2, 7, 8]. Among them, interpolation based on Chebyshev polynomials is of great interest, which has been shown to be one of the important methods in the literature [9–15], since this type of interpolation polynomials eliminates the problem of the Runge phenomenon [16, 17].

Function approximation based on Chebyshev polynomial interpolation and discrete cosine transform is discussed in many papers [1, 2, 18–20]. In these methods, the points of a non-uniform grid corresponding to the roots or extremals of the Chebyshev polynomials are first obtained, and then the approximation coefficients are calculated at these points using collocation methods. The results showed that the use of Chebyshev polynomials provides almost optimal accuracy for solving problems of interpolation, differentiation, and integration of smooth functions.

Author Contributions: Konstantin P. Lovetskiy—Conceptualization, investigation, writing—original draft preparation, supervision. Anastasiia A. Tiutiunnik—methodology, writing—review and editing, project administration, funding acquisition. Felix Jose do Nascimento Vicente, Celmilton Teixeira Boa Morte—software, validation. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data sharing is not applicable.

Acknowledgments: The authors of the article are grateful to Prof. Sevastianov L. A. for help with the work.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Funding: This research was funded by the RUDN University Scientific Projects Grant System, project No. 021934-0-000 (Konstantin P. Lovetskiy, Anastasiia A. Tiutiunnik).

References

1. Boyd, J. P. *Chebyshev and Fourier Spectral Methods: Second Revised Edition. Dover Books on Mathematics* (Courier Corporation, 2013).
2. Fornberg, B. *A practical guide to pseudospectral methods* doi:10 . 1017 / cbo9780511626357 (Cambridge University Press, 1996).
3. Mason, J. C. & Handscomb, D. C. *Chebyshev Polynomials* in *Chebyshev Polynomials* (Chapman and Hall/CRC Press, 2002).
4. Orszag, S. A. Comparison of Pseudospectral and Spectral Approximation. *Studies in Applied Mathematics* **51,** 253–259. doi:10.1002/sapm1972513253 (1972).
5. Clenshaw, C. W. A note on the summation of Chebyshev series. *Mathematics of Computation* **9,** 118–120. doi:10.1090/S0025-5718-1955-0071856-0 (1955).
6. Fox, L. & Parker, I. B. *Chebyshev polynomials in numerical analysis* (Oxford, 1968).
7. Shen, Z. & Serkh, K. *Is polynomial interpolation in the monomial basis unstable?* 2023. doi:10.48550/arXiv.2212.10519.
8. Zhang, X. & Boyd, J. P. *Asymptotic Coefficients and Errors for Chebyshev Polynomial Approximations with Weak Endpoint Singularities: Effects of Different Bases* 2021. doi:10.48550/arXiv.2103.11841.

9. Lovetskiy, K. P., Sevastianov, L. A. & Nikolaev, N. E. Regularized Computation of Oscillatory Integrals with Stationary Points. *Procedia Computer Science* **108,** 998–1007. doi:10.1016/j.procs.2017.05.028 (2017).
10. Lovetskiy, K. P., Sevastianov, L. A., Kulyabov, D. S. & Nikolaev, N. E. Regularized computation of oscillatory integrals with stationary points. *Journal of Computational Science* **26,** 22–27. doi:10.1016/j.jocs.2018.03.001 (2018).
11. Lovetskiy, K. P., Kulyabov, D. S. & Hissein, A. W. Multistage pseudo-spectral method (method of collocations) for the approximate solution of an ordinary differential equation of the first order. *Discrete and Continuous Models and Applied Computational Science* **30,** 127–138. doi:10.22363/2658-4670-2022-30-2-127-138 (2022).
12. Lovetskiy, K. P., Sevastianov, L. A., Hnatič, M. & Kulyabov, D. S. Numerical Integration of Highly Oscillatory Functions with and without Stationary Points. *Mathematics* **12,** 307. doi:10.3390/math12020307 (2024).
13. Sevastianov, L. A., Lovetskiy, K. P. & Kulyabov, D. S. *An Effective Stable Numerical Method for Integrating Highly Oscillating Functions with a Linear Phase* in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **12138 LNCS** (2020), 29–43. doi:10.1007/978-3-030-50417-5_3.
14. Sevastianov, L. A., Lovetskiy, K. P. & Kulyabov, D. S. *Numerical integrating of highly oscillating functions: effective stable algorithms in case of linear phase* 2021. doi:10.48550/arXiv.2104.03653.
15. Sevastianov, L. A., Lovetskiy, K. P. & Kulyabov, D. S. A new approach to the formation of systems of linear algebraic equations for solving ordinary differential equations by the collocation method. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics* **23,** 36–47. doi:10.18500/1816-9791-2023-23-1-36-47 (2023).
16. Berrut, J. & Trefethen, L. N. Barycentric Lagrange Interpolation. *SIAM Review* **46,** 501–517. doi:10.1137/S0036144502417715 (2004).
17. Epperson, J. F. On the Runge Example. *The American Mathematical Monthly* **94,** 329. doi:10.2307/2323093 (1987).
18. Amiraslani, A., Corless, R. M. & Gunasingam, M. Differentiation matrices for univariate polynomials. *Numerical Algorithms* **83,** 1–31. doi:10.1007/s11075-019-00668-z (2020).
19. Wang, Z. Interpolation using type i discrete cosine transform. *Electronics Letters* **26,** 1170. doi:10.1049/el:19900757 (1990).
20. Wang, Z. Interpolation using the discrete cosine transform: reconsideration. *Electronics Letters* **29,** 198. doi:10.1049/el:19930133 (1993).

Information about the authors

Lovetskiy, Konstantin P.—Candidate of Sciences in Physics and Mathematics, Associate Professor of Department of Computational Mathematics and Artificial Intelligence of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University) (e-mail: lovetskiy-kp@rudn.ru, phone: +7 (495) 952-25-72, ORCID: 0000-0002-3645-1060)

Tiutiunnik, Anastasiia A.—Candidate of Sciences in Physics and Mathematics, Associate Professor of Department of Computational Mathematics and Artificial Intelligence of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University) (e-mail: tyutyunnik-aa@rudn.ru, phone: +7 (495) 955-07-83, ORCID: 0000-0002-4643-327X)

do Nascimento, Vicente Felix Jose—student of Department of Computational Mathematics and Artificial Intelligence of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University) (e-mail: 1032199092@rudn.ru)

Teixeira Boa, Morte Celmilton—student of Department of Computational Mathematics and Artificial Intelligence of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University), (e-mail: 1032199094@rudn.ru)

UDC 519.21
PACS 52.25.Fi
DOI: 10.22363/2658-4670-2024-32-2-202–212 EDN: CRLKAJ

Алгоритм Кленшоу в задаче интерполяции методом Чебышевской коллокации

К. П. Ловецкий, А. А. Тютюнник, Ду Нашсименту Висенте Феликс Жозе, Тейшейра Боа Морте Селмилтон

Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация

Аннотация. В статье описан метод вычисления интерполяционных коэффициентов разложения по полиномам Чебышева. Метод справедлив, когда искомое функция ограничена и имеет конечное число максимумов и минимумов в конечной области интерполирования. Суть метода состоит в том, что интерполируемая искомая функция может быть представлена в виде разложения по полиномам Чебышева; затем коэффициенты разложения определяются по методу коллокаций сведением задачи к решению хорошо обусловленной системы линейных алгебраических уравнений относительно искомых коэффициентов. Использование известных полезных свойств полиномов Чебышева позволяет значительно упростить решение задачи интерполяции функций. Изложена методика использования алгоритма Кленшоу для суммирования рядов и определения коэффициентов разложения интерполируемой функции, основанная на дискретной ортогональности полиномов Чебышева 1-го рода.

Ключевые слова: интерполяция функций методом Чебышевской коллокации, алгоритм Кленшоу ускорения вычислений

Discrete & Continuous Models
& Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (2) 213–221
http://journals.rudn.ru/miph

Research article

UDC 004.8
DOI: 10.22363/2658-4670-2024-32-2-213–221
EDN: CPUADE

Statistical causality analysis

Alexander A. Grusho[1,2], Nikolai A. Grusho[1], Michael I. Zabezhailo[1],
Konstantin E. Samouylov[2], Elena E. Timonina[1,2]

[1] *Federal Research Center "Computer Sciences and Control" of the Russian Academy of Sciences, 44 Vavilova St, bldg. 2, Moscow, 119133, Russian Federation*
[2] *RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation*

(received: April 14, 2024; revised: April 25, 2024; accepted: April 30, 2024)

Abstract. The problem of identifying deterministic cause-and-effect relationships, initially hidden in accumulated empirical data, is discussed. Statistical methods were used to identify such relationships. A simple mathematical model of cause-and-effect relationships is proposed, in the framework of which several models of causal dependencies in data are described – for the simplest relationship between cause and effect, for many effects of one cause, as well as for chains of cause-and-effect relationships (so-called transitive causes). Estimates are formulated that allow using the de Moivre–Laplace theorem to determine the parameters of causal dependencies linking events in a polynomial scheme trials. The statements about the unambiguous identification of cause-and-effect dependencies that are reconstructed from accumulated data are proved. The possibilities of using such data analysis schemes in medical diagnostics and cybersecurity tasks are discussed.

Key words and phrases: finite classification task, cause-and-effect relationships, machine learning

Citation: Grusho A. A., Grusho N. A., Zabezhailo M. I., Samouylov K. E., Timonina E. E., Statistical causality analysis. *Discrete and Continuous Models and Applied Computational Science* **32** (2), 213–221. doi: 10.22363/2658-4670-2024-32-2-213–221. edn: CPUADE (2024).

1. Introduction

The simplest idea of a causal relationship is given by the functional dependence $y = f(x)$, in which x is the cause of the appearance of the consequence y, if their dependence is described by the functional relationship f. In this simplest example, the determinism of the appearance of an effect is expressed when a cause appears and a known f. However, the mechanism of generating an effect from a cause is not always known, but this does not reduce the importance of causality in solving practical problems. Therefore, the task arises of searching for cause-and-effect relationships using statistical methods in some limited conditions. There are many probabilistic models and ways to restore\reconstruct cause-and-effect relationships with some confidence [1]. To a large extent, these models and methods are associated with various practical applications of cause-and-effect relationships [2]. Let's look at some of them.

Judea Pearl's causal inference model [3, 4] can be expressed in its simplest form by the ratio $y = f(x, u)$, where, in addition to the initial definition of causality, an argument u of randomness is added. Then all calculations related to the causal effect are better expressed in terms of mathematical statistics. This is due to the fact that, as a rule, nothing is known about the distribution of the random component and statistical estimates are used in practice.

This model, among other things, allows us to formulate and solve problems of the influence of therapeutic effects of drugs on patients with a certain disease. Here, u plays the role of individual patient characteristics that affect the effectiveness of treatment with the drug in question.

Another interpretation of this model is that the degree of lesion in a certain diagnosis depends not only on the underlying cause of the disease, but also on the individual characteristics of a particular patient. These features form a random component of the depth of the patient's lesion [5].

Let's turn to the practical procedure of making a diagnosis in a medical study. As a rule, there are a large number of parameters that the doctor considers when making a diagnosis. Each value of these parameters defines a property that can carry information about a possible diagnosis, but may not affect the diagnosis at all. Based on experience, the doctor identifies properties that have a causal relationship with a possible diagnosis. Then, in the collected analyses and other auxiliary data, we are talking about the consequences of the cause, which is determined by the property of diagnosis (disease) and manifests itself in the consequences observed by the doctor.

In this paper, we consider a model of deterministic causation in the presence of a significant number of properties that are not related to the causal effect of some properties on others. Despite the simplified nature of the model, it manages to describe an important case of the appearance of some properties that can enhance or weaken the degree of damage when exposed to the underlying cause (confounding [6, 7]). This case is described in the article with the help of a complex cause, when several properties form a cause, as a result of which a consequence property appears.

Another class of applications of the models of the appearance of deterministic cause-and-effect relationship considered in the article against the background of a sequence with other properties that are not directly related to the desired causality is the search for traces of an insider in the observed sequences of data monitoring user actions in banking applications [8, 9]. In the considered task, several possible scenarios of insider actions were analyzed, the properties of which could manifest themselves in the monitoring data. These properties were tracked in Big Data monitoring, which made it possible to successfully solve operational information security tasks. The desired properties appear in the monitoring data as inclusions of effect properties, where the cause is not the observed actions of an insider. This scheme is transferred to interspersed anomalies arising from causes caused by failures of software and hardware systems.

The problem of trust in the results of computer data analysis in diagnostic tasks and critical calculations is important. The scheme proposed in the article can be used to increase confidence in the results of complex computer calculations. For example, we highlight a number of input and output parameters that consistently belong to input and output data clusters. If consistency is maintained in the next iteration of calculations, then this fact increases confidence in the results of calculations. If there are contradictions in the consistency of input and output data, then it is necessary to look for the reasons\causes for such a mismatch. At the same time, the remaining input and output data do not participate in such control and are considered random noise [10]. The explicability of the formed conclusions is most often realized with the help of causal relationships identified in the analyzed data [11].

2. A simple mathematical model of cause-and-effect relationships

Let $A = \{x_1, ..., x_n\}$ be the set of observable properties of the data. Let's consider the simplest case, when the data is a random sequence of properties. Randomness is introduced using a polynomial scheme of M trials with known probabilities $p(x_k) = p_k$ where $k = 1, ..., n$. After the implementation of the polynomial scheme trials, deterministic appearances of consequences are additionally embedded in the sequence, the causes of which appeared randomly in the sequence under consideration, and

the consequences increase the total length of the sequence M by an amount equal to the number of appearances of causes. The constructed model will be called the information space.

First, let's consider the following schemes of cause-and-effect relationships.

1) The scheme "$a \rightarrow b$" is determined when property a arises, then property b arises deterministically.
2) The scheme "$a \rightarrow b \,\&\, a \rightarrow c$" occurs when property a has several consequences of b and c.
3) The scheme "$a \rightarrow b \,\&\, b \rightarrow c$" is defined as a sequence of causal relationships: a entails the appearance of b and b entails the appearance of c. This scheme is also called transitive cause a for effect c.

In scheme 1), the effect should appear after a fixed t steps from the cause. *At* $t = 1$, the effect immediately follows the cause. In scheme 2) each consequence arises with its own interval t_1 or t_2.

In the future, we will change the models in accordance with approximations to real conditions.

3. Statistical identification of cause-and-effect relationships in a simple model

Under the conditions of the constructed model, it is possible to statistically determine cause-and-effect relationships. For the original polynomial scheme trials

$$p_{ij} = p(x_i, x_j) = p(x_i)p(x_j).$$

If (x_i, x_j) are connected by a causal relationship, then after t steps after x_i the next property x_j is embedded in the polynomial trial sequence.

Suppose that M is large enough so that the relative frequencies of properties in a polynomial scheme trials with a sufficient degree of confidence uniquely identify the various unequal probabilities of this scheme. Let's denote $\nu(i)$ the frequency of occurrence x_i in a polynomial sequence of outcomes of length M. If (x_i, x_j) are causally related, then the length of the sequence M^* increases from M by $\nu(i)$, and the frequency of x_j occurrence increases and will be equal to $\nu^*(j) = \nu(i) + \nu(j)$. Then

$$\frac{\nu^*(j)}{M} = \frac{\nu(i)}{M} + \frac{\nu(j)}{M} > p(x_j).$$

It follows first of all that the property x_j is a consequence of some cause. If there is only one cause x_i in the observed sequence, then

$$\frac{M^* - M}{M} = \frac{\nu(i)}{M},$$

which uniquely determines the cause x_i in the event that the probability value $p(x_i)$ differs from any other probabilities of the polynomial scheme trials. If there are several such probabilities, then each occurrence of a cause in a random sequence deterministically entails the appearance of x_j exactly after t steps. The random repetition of this distance from other properties to x_j, having a probability equal to $p(x_j)$, decreases exponentially, which makes such a possibility unlikely with a sufficiently large occurrence of $\nu(i)$. In particular, the requirement of a priori knowledge of t is used to simplify the scheme. It is sufficient to assume the existence of such a t because, when potential variants of the cause are identified, the statistics of the distances between (x_i, x_j) will dominate and uniquely determine t. Thus, the following theorem is proved.

Theorem 1. *With a well-known polynomial scheme trials of length M and assuming the existence of one causal relationship that generates an effect always through the same number of sequence steps, such a causal relationship is uniquely identified.*

It is possible that with known probabilities of a polynomial scheme trials, its true length M is unknown. Let's consider this case. Let's calculate the frequency of occurrence of all outcomes of the observed sequence of properties $\{\nu(k), k = 1, ..., n\}$. According to the de Moivre–Laplace theorem [12] for each $\{\nu(k), k = 1, ..., n\}$ for large M^* we have the following estimates

$$M_k = \frac{\nu(k)}{p(x_k)} = M + o(\ln M\sqrt{M}).$$

Then

$$\frac{\nu(i) + \nu(j)}{p(x_j)} = M + \frac{Mp(x_i)}{p(x_j)} + o(\ln M\sqrt{M}) > M_k,$$

which is the most of all M_k for $k \neq j$. Thus, in this case, the consequence is identified. Using a fixed distance between (x_i, x_j), as shown above, we establish the cause x_i.

Let's consider the simplest case of Scheme 2). In this case x_i, the cause has two effects x_j and x_k, which appear at distances t_1 and t_2 from the cause. If all the probabilities of the polynomial scheme trials are known and the length of the polynomial sequence is M, then the length of the sequence with included consequences is $M^* = M + 2\nu(i)$. From here we find $\nu(i)$ and for sufficiently large M we find $p(x_i)$. If there are no other similar probabilities, then we find the cause. If there are more probabilities $p(x_i)$, then the frequencies $\nu^*(j) = \nu(i) + \nu(j)$ and $\nu^*(k) = \nu(i) + \nu(k)$ allow us to reconstruct the effects and by the distances t_1 and t_2 we find the true cause. Here it is also sufficient to allow constant distances between cause and effect x_i. The values t_1 and t_2 are easily calculated from the repeatability of each cause-effect pair.

If M is unknown, then the method proposed for Scheme 1) can be used to identify each consequence.

Scheme 3) is equivalent to the two simplest Schemes 1). In this case, there are two causes x_i and x_v, each of which has a corresponding effect x_j or x_k with fixed distances t_1 and t_2 to the causes. For a known M, the real length $M^* = M + \nu(i) + \nu(v)$. For an increased frequency of $\nu^*(j) = \nu(i) + \nu(j)$ and $\nu^*(k) = \nu(v) + \nu(k)$ we are recovered effects in the investigation. We find the causes by the distances. Here it is also possible to abandon a priori knowledge of the values t_1 and t_2, but only assume the constancy of these values. In the case of a transitive cause differs in that if (x_i, x_j, x_k) are connected by a transitive causal relationship, then $\nu^*(k) = \nu(i) + \nu(j) + \nu(k)$.

If the parameter M is unknown, the effects of various causes can be determined using the method for Scheme 1). Then, using the distances t_1 and t_2, we find both causes.

Let's consider a case where there are two causes with the same effect. Let the effect x_j be a common consequence of each of the two causes x_i and x_k. Then, using the method described above, we find the consequence x_j and the frequency of its occurrence $\nu^*(j) = \nu(i) + \nu(j) + \nu(k)$. If the distances t_1 and t_2 are known, then it is easy to find the causes. From each encountered x_j, we select two elements at distances t_1 and t_2. The set of selected elements will also include (x_i, x_k) and random elements with polynomial probabilities. Let's calculate the frequency of occurrence of each of the encountered elements. It will be shown further that the two highest frequencies corresponding to the causes and the causes themselves are uniquely distinguished from them.

We will carry out the proof for the most difficult case, when M is unknown and only the constraints $t_1 \leq t$ and $t_2 \leq t$ are known. From each encountered x_j, we select all the elements on the left at a distance t. The random component of occurrence x_j is equal to $\nu(j)$. By the de Moivre–Laplace theorem

$$\nu(j) = Mp(x_j) + o(\ln M\sqrt{M}),$$

where the parameter M is unknown, but

$$M \geq \frac{M^*}{3}.$$

An error in determining the causes may occur due to the accidental coincidence of conditions of the same distance from x_j a large number of randomly appearing properties. The largest number of identical random elements does not exceed $\nu(j)$. Such a random sequence can be located at a distance from each x_j at a distance t^* with a probability of $1/t$. If this is a property x_f, then the probability of a random occurrence of a long sequence of length

$$M^{**} = \frac{\min\limits_{i,j,k}\{p(x_i), p(x_j), p(x_k)\}M}{2},$$

is estimated by the value

$$\left(\frac{p(x_f)}{t}\right)^{M^{**}},$$

which is significantly less than the probability estimates in the de Moivre–Laplace theorem. Given that t and the number of properties are limited, we obtain the following theorem.

Theorem 2. *With a well-known polynomial scheme trials of length M and assuming the existence of two causal relationships that generate a common consequence always through the same number of steps in the resulting sequence, such a causal relationship is uniquely identified. Moreover, it can be assumed that only M^* is known, which includes the length of the random part of the sequence and the number of all the consequences of the desired causes.*

4. Causal relationships in several information spaces

Let's consider a finite set of information spaces (IS) [13], which we will denote as $\mathrm{IS}^* = \{\mathrm{IS}_1, ..., \mathrm{IS}_k\}$. Each IS has its own set of properties A_i. Randomness is determined using polynomial scheme from M trials and with known probabilities, but the probabilities of polynomial schemes trials for different IS_i are different. If a belongs to IS_1, and b belongs to IS_2 and a is the cause of the deterministic appearance of consequence b, then if there is a connection from IS_1 to IS_2 that allows initiating the consequence, property b appears in the sequence of properties of IS_2 at the same moment (the sequences in all IS are synchronized), and from this moment the remaining part of the sequence in IS_2 is shifted by 1. These shifts do not interfere with finding synchronous pairs in IS_1 and IS_2, since the number of shifts each time is equal to the number of occurrences in IS_1 of the cause candidate minus one.

If cause a generates consequences in several different IS, then the scheme of the appearance of consequences a is such as described above.

For simplicity, let's first consider the case of two spaces IS_1 and IS_2. The task is to statistically identify cause-and-effect relationships. Let $A_1 = \{x_1, ..., x_n\}$ – are properties in IS_1, $A_2 = \{y_1, ..., y_m\}$ – are properties in IS_2.

The probabilities of random synchronous pairs are equal to $p(x_i, y_j) = p(x_i)p(y_j)$ If x_i is the cause of occurrence y_j, then the probability of occurrence of such a pair is equal to $p(x_i)$.

Let's denote the frequencies of the properties in IS_1 by ν, and the frequencies in IS_2 by μ. If x_i is the cause of occurrence y_j, then the resulting frequency of the property

$$\mu^*(y_j) = \nu(x_i) + \mu(y_j).$$

Hence, as before, we conclude that y_j is a consequence of the cause of IS_1. The restoration of the cause is reduced to calculating the statistics of properties from IS_1, which are located in IS_1 at the moments when properties y_j appear in the sequence of IS_2.

By the de Moivre–Laplace theorem

$$\mu(y_j) = Mp(y_j) + o(\ln M\sqrt{M}),$$
$$\nu(x_i) = Mp(x_i) + o(\ln M\sqrt{M}).$$

Let $\hat{\nu}(x_k)$ be the frequency of occurrence of properties x_k from IS_1, which are located in IS_1 at the moments when properties y_j appear in the sequence of IS_2. If the inequality

$$p(x_k)p(y_j) < p(x_i),$$

holds for any $k \neq i$, then the estimate

$$\hat{\nu}(x_k) = Mp(y_j)p(x_k) + o(\ln M\sqrt{M}) < \nu(i),$$

is valid for any $k \neq i$ and completely determines the cause for sufficiently large M.

Theorem 3. *With known polynomial scheme trials of length M in* IS_1 *and* IS_2 *and under the assumption*

$$p(x_k)p(y_j) < p(x_i), \quad k \neq i,$$

if there is one connection by which a cause belonging to IS_1*, through this connection generates an effect in* IS_2 *always simultaneously with the appearance of the cause, such a causal relationship is uniquely identified for sufficiently large M.*

If a given property in IS_1 has several consequences located in different IS, then for large M these cause-and-effect relationships are determined in the same way as it is done for two IS.

Similarly, an algorithm is built to identify a transitive causal relationship in Scheme 3), when causes and effects are located in different IS.

5. Complex causes

In practice, there is often a situation where the cause is complex. Let (x_i, y_j, z_k) be properties that form causal relationships as follows. A property x_i from IS_1 is not the cause of an effect z_k from IS_3, and a property y_j from IS_2 is not the cause of an effect z_k from IS_3. However, if, with the simultaneous appearance of x_i,y_j in IS_1 and IS_2 and the presence of their connection with IS_3, an effect z_k appears in IS_3, then the cause constructed in this way is called complex.

The task is to identify the only complex causal relationship in IS^*.

If $\mathrm{IS}_1 \times \mathrm{IS}_2$ is known, then the number of occurrences of the cause for the effect z_k is distributed according to a polynomial law with probabilities $p(x_i)p(y_j)$.

If we denote the frequency of occurrence of a pair (x_i, y_j) in the probability space $\mathrm{IS}_1 \times \mathrm{IS}_2$ as $\nu(x_i, y_j)$, then as a result of the occurrence of the cause for the effect z_k, the statistics $\nu(z_k)$ will increase by $\nu(x_i, y_j)$. From which it follows that z_k is a consequence of any cause, it is established as before. If $\mathrm{IS}_1 \times \mathrm{IS}_2$ is known, then the cause is determined in the same way as for one information space. If $\mathrm{IS}_1 \times \mathrm{IS}_2$ is unknown, then it is necessary to identify the spaces in which the parts of the cause are located.

To identify spaces in which parts of the cause are present, you can use the procedure for disconnecting links between information spaces. At the first stage, it is necessary to create a sequence of properties in each IS according to a polynomial scheme trials of length M. At the second stage, for each individual IS that may have a connection with IS_3, a procedure is used to turn on the connection between them and a frequency of z_k change is considered. In the absence of a frequency change, the connection of the next IS with the IS_3 is disabled and the same experiment is performed with the next IS.

If all IS that could have a connection with IS_3 did not lead to a change in the frequency of the property, then the pairs of included connections are also sorted out. In the case under consideration, the $IS_1 \times IS_2$ pair will give the desired frequency of z_k change. This means that there is a complex reason for this pair. There are a couple of implementations of synchronous polynomial schemes in the $IS_1 \times IS_2$ space. Consequently, as noted above, they generate a new polynomial scheme $IS_4 = IS_1 \times IS_2$ of length M with probabilities $p(x_i, y_j) = p(x_i)p(y_j)$. Using the methods described earlier, there is a complex cause.

6. Conclusion

The paper considers the simplest case of identifying deterministic cause-and-effect relationships in the presence of random properties that do not carry information about causes and effects.

To identify cause-and-effect relationships, information is used on the occurrence of causes and the time of occurrence of consequences after the appearance of causes. First, the consequences of some unknown causes are revealed, and then additional information allows you to identify the causes of the found consequences themselves.

This model roughly corresponds to the search for additional information in medical diagnostics and information security. The analysis of the above model allows you to find new ways to obtain additional information in search tasks.

In the future, it is supposed to investigate the model in conditions of striving for infinity of parameters M and n. Another area of further research is to increase the size of causes and effects to sets containing several properties.

Author Contributions: Conceptualization, A. Grusho and M. Zabezhailo; methodology, K. Samouylov; validation, A. Grusho, M. Zabezhailo and N. Grusho; formal analysis, A. Grusho and E. Timonina; investigation, N. Grusho; data curation, A. Grusho and N. Grusho; writing—original draft preparation, E. Timonina; writing—review and editing, K. Samouylov. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, X., Hu, W. & Yang, F. Detection of Cause-Effect Relations Based on Information Granulation and Transfer Entropy. *Entropy* **24,** 212. doi:10.3390/e24020212 (2022).
2. Reimer, J., Wang, Y., Laridi, S., Urdich, J., Wilmsmeier, S. & Palmer, G. Identifying cause-and-effect relationships of manufacturing errors using sequence-to-sequence learning. *Scientific Reports* **12,** 22332. doi:10.1038/s41598-022-26534-y (2022).
3. Pearl, J. *Causal Inference* in *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008* (eds Guyon, I., Janzing, D. & Schölkopf, B.) **6** (PMLR, Whistler, Canada, Dec. 2010), 39–58.
4. Pearl, J. *The mathematics of causal inference* in *Joint Statistical Meetings Proceedings. ASA* (2013), 2515–2529.
5. Yao, L., Chu, Z., Li, S., Li, Y., Gao, J. & Zhang, A. A Survey on Causal Inference. *ACM Transactions on Knowledge Discovery from Data* **15,** 1–46. doi:10.1145/3444944 (2021).
6. Höfler, M. Causal inference based on counterfactuals. *BMC Medical Research Methodology* **5,** 1–28. doi:10.1186/1471-2288-5-28 (2005).

7. Richens, J. G., Lee, C. M. & Johri, S. Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications* **11,** 3923. doi:10.1038/s41467-020-17419-7 (2020).
8. Grusho, A. A., Grusho, N. A. & Timonina, E. E. Root Cause Anomaly Localization [Lokalizatsiya iskhodnoy prichiny anomalii]. *Information Security Problems. Computer Systems.* in Russian, 9–16 (2020).
9. Smirnov, D. V. Methodology of problem-oriented Big Data analysis in limited time mode [Metodika problemno-orientirovannogo analiza Big Data v rezhime ogranichennogo vremeni]. *International Journal of Open Information Technologies* **9.** in Russian, 88–94 (2021).
10. Grusho, A., Grusho, N., Zabezhailo, M. & Timonina, E. *Evaluation of Trust in Computer-Computed Results* in *Distributed Computer and Communication Network* (eds Vishnevskiy, V. M., Samouylov, K. E. & Kozyrev, D. V.) (Springer International Publishing, Cham, 2022), 420–432. doi:10.1007/978-3-030-97110-6_33.
11. Guyatt, G. *et al.* Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA* **268,** 2420–2425. doi:10.1001/jama.1992.03490170092032 (Nov. 1992).
12. Shiryaev, A. N. *Probability [Veroyatnost']* in Russian. 521 pp. (MTsNMO, Moscow, 2004).
13. Grusho, A., Grusho, N. & Timonina, E. *Method of Several Information Spaces for Identification of Anomalies* in *Intelligent Distributed Computing XIII* (eds Kotenko, I., Badica, C., Desnitsky, V., El Baz, D. & Ivanovic, M.) (Springer International Publishing, Cham, 2020), 515–520. doi:10.1007/978-3-030-32258-8_60.

Information about the authors

Grusho, Alexander A.—Principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; professor of Department of Probability Theory and Cyber Security of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University), (e-mail: grusho@yandex.ru, ORCID: 0000-0003-4400-2158)

Grusho, Nikolai A.—Candidate of Physical and Mathematical Sciences, Senior scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (e-mail: info@itake.ru, ORCID: 0000-0002-5005-2744)

Zabezhailo, Michael I.—Professor, Doctor of Physical and Mathematical Sciences, Principal scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (e-mail: m.zabezhailo@yandex.ru, ORCID: 0000-0002-5067-5909)

Samouylov, Konstantin E.—Professor, Doctor of Technical Sciences, Head of the Department of Probability Theory and Cyber Security of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University) (e-mail: samuylov-ke@rudn.ru, ORCID: 0000-0002-6368-9680)

Timonina, Elena E.—Professor, Doctor of Technical Sciences, Leading scientist, Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; professor of Department of Probability Theory and Cyber Security of Peoples' Friendship University of Russia named after Patrice Lumumba (RUDN University) (e-mail: eltimon@yandex.ru, ORCID: 0000-0002-6493-3622)

UDC 004.8
DOI: 10.22363/2658-4670-2024-32-2-213–221 EDN: CPUADE

Статистический анализ причинно-следственных связей

А. А. Грушо[1, 2], Н. А. Грушо[1], М. И. Забежайло[1], К. Е. Самуйлов[2], Е. Е. Тимонина[1, 2]

[1] *Федеральный исследовательский центр «Информатика и управление» РАН, ул. Вавилова, д. 44, стр. 2, Москва, 119133, Российская Федерация*
[2] *Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация*

Аннотация. Рассмотрена проблема выявления детерминированных причинно-следственных связей, изначально скрытых в накопленных эмпирических данных. Для выявления таких связей использовались статистические методы. Предложена простая математическая модель причинно-следственных отношений, в рамках которой описано несколько моделей причинно-следственных связей в данных – для простейших отношений между причиной и следствием, для многих следствий одной причины, а также для цепей причинно-следственных связей (так называемых транзитивных причин). Сформулированы оценки, позволяющие с помощью теоремы Муавра–Лапласа определить параметры модели, связывающие события в испытаниях полиномиальной схемы. Обоснованы утверждения об однозначной идентификации причинно-следственных связей, которые восстанавливаются по накопленным данным. Обсуждаются возможности использования таких схем анализа данных в задачах медицинской диагностики и кибербезопасности.

Ключевые слова: задача конечной классификации, причинно-следственные связи, машинное обучение

Discrete & Continuous Models & Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (2) 222–233
http://journals.rudn.ru/miph

Research article

UDC 519.872, 519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-2-222–233 EDN: CDJVIL

Well-posedness of the microwave heating problem

Baljinnyam Tsangia

Mongolian University of Science and Technology, Ulaanbaatar, Mongolia

(received: May 1, 2024; revised: May 10, 2024; accepted: May 15, 2024)

Abstract. A number of initial boundary-value problems of classical mathematical physics is generally represented in the linear operator equation and its well-posedness and causality in a Hilbert space setting was established. If a problem has a unique solution and the solution continuously depends on given data, then the problem is called well-posed. The independence of the future behavior of a solution until a certain time indicates the causality of the solution. In this article, we established the well-posedness and causality of the solution of the evolutionary problems with a perturbation, which is defined by a quadratic form. As an example, we considered the coupled system of the heat and Maxwell's equations (the microwave heating problem).

Key words and phrases: evolutionary problems, nonlinear perturbation, Lipschitz continuous, quadratic form, coupled problems

Citation: Tsangia B., Well-posedness of the microwave heating problem. *Discrete and Continuous Models and Applied Computational Science* **32** (2), 222–233. doi: 10.22363/2658-4670-2024-32-2-222–233. edn: CDJVIL (2024).

1. Introduction

Here we consider a non-linear, coupled system in thermoelectricity. Thermoelectric effects are viewed as the result of the mutual interference of heat flow and electric flow in a system. The interaction of thermal and electric processes is modeled by the heat equation

$$\rho C_\rho \partial_0 \vartheta + \mathrm{div} q = Q$$

and Maxwell's equations

$$-\mathrm{curl} H + J + \partial_0 D = J_1$$
$$\mathrm{curl} E + \partial_0 B = 0.$$

Here q is the thermal current flux, ρ is the volumetric mass density, C_ρ is the specific heat density, ϑ is the absolute temperature, J is the electric current flux, E, H are the electric and magnetic fields, respectively, D is the displacement current, B is the magnetic induction and J_1 is the given electric source. Q describes the production of internal energy by various mechanisms, such as the Joule heating, radioactive decay, etc. In our system the Joule heating $Q = \langle E \mid J \rangle$ produces the internal energy. This term governs the non-linearity in the system and, moreover, it couples the heat and Maxwell's equations. The system of these equations has to be supplemented by so-called constitutive equations, which describe the material's properties and effects. As constitutive equations, we deal with the following thermoelectric material relations

$$J = \sigma E$$

$$\begin{aligned} q &= -\lambda \mathrm{grad}\vartheta \\ D &= \varepsilon E \\ B &= \mu H. \end{aligned}$$

Here σ is the electric conductivity, λ is the thermal conductivity, ε is the electric permittivity, μ is the magnetic permeability. The coupled system of the heat and Maxwell's equations with these constitutive equations becomes the microwave heating problem. The microwave heating problem has wide industrial applications and it has been studied theoretically and numerically in various situations (see e.g. [1–3] and the references therein). We study the coupled systems in the three-dimensional case. Moreover, we consider this system with the physical coefficients defined as 3-by-3−matrix-valued functions depending on the spatial variables only. We assume (homogeneous) Dirichlet boundary conditions for ϑ, (homogeneous) electric boundary conditions for E and non-vanishing initial values. We say that a problem is *well-posed* if the problem has a unique solution and the solution continuously depends on the given data. The independence of the future behavior of a solution until a certain time indicates the *causality* of the solution. In our solution theory the well-posedness and causality of a given problem are discussed.

The idea of tackling well-posedness and causality of the problem just discussed is to frame the above system in the theory of evolutionary equations: In [4, 5] it has been found that a number of initial boundary-value problems of classical mathematical physics is represented by the following general form

$$\left(\partial_0 M\left(\partial_0^{-1}\right) + A\right) u = F. \tag{3}$$

Here ∂_0 is the (continuously invertible) derivative with respect to time in a suitable weighted Hilbert space, A is a skew-selfadjoint operator in a suitable Hilbert space; the mapping $(z \mapsto M(z))$ is bounded operator valued and holomorphic in an open ball $B_{\mathbb{C}}(r, r)$ with some positive radius r centered at r. The operator $M\left(\partial_0^{-1}\right)$ is interpreted in the sense of a function calculus by establishing ∂_0 as a normal operator in a suitable Hilbert space. The solution theory associated to (3) was established in [4, 5] and many diverse problems were studied there. For applications, we focus on a particular case of $M\left(\partial_0^{-1}\right)$, namely,

$$M\left(\partial_0^{-1}\right) = M_0 + \partial_0^{-1} M_1.$$

Here M_0 is a selfadjoint, bounded, linear operator with $M_0|_{N(M_0)} \geqslant c_0 > 0$, M_1 is a bounded, linear operator satisfying $\Re M_1|_{R(M_0)} \geqslant c_1 > 0$. In the next section we establish the solution theory of the following problem

$$\left(\partial_0 M_0 + M_1 + A\right) u + \tilde{F}(u) = F, \tag{4}$$

which covers the aforementioned non-linear coupled system. Here $\tilde{F}$ is a quadratic form. The non-linear problem (1) yields a fixed point problem. In our approach the well-posedness of (4) is based on the strict positive definiteness of the operators $\Re\left(\partial_0 M_0 + M_1 + A\right)$ and $\Re\left(\partial_0 M_0 + M_1 + A\right)^*$ and a Lipschitz continuous approximation of $\tilde{F}$. Due to the strict positive definiteness result, the inverse operator $\left(\partial_0 M_0 + M_1 + A\right)^{-1}$ becomes Lipschitz continuous in a suitable Hilbert space. Thus, (4) amounts to be an evolutionary problem in the sense of (3) with a Lipschitz continuous perturbation, which is eventually solved by the contraction mapping principle. As an application we shall consider the microwave heating problem in the third section.

2. Solution theory

We start by establishing time differentiation ∂_0 as a normal operator. It is initially considered on $\overset{\circ}{C}_\infty(\mathbb{R})$, which is the set of infinitely often differentiable, complex-valued functions defined on the real line $\mathbb{R}$ having compact support. Hence ∂_0 is a densely defined, closed linear operator on $L^2(\mathbb{R})$, moreover, it is an essentially skew-selfadjoint operator on $L^2(\mathbb{R})$. We define the following weighted L^2-space

$$H_{\nu,0}(\mathbb{R}) := L^2(\mathbb{R}, \exp(-2\nu x)\,dx) := \left\{\varphi \in L^1_{loc}(\mathbb{R}) \mid \exp(-\nu m_0)\,\varphi \in L^2(\mathbb{R})\right\}$$

equipped with the norm

$$|\varphi|_{\nu,0} := \sqrt{\int_{\mathbb{R}} |\varphi(x)|^2 \exp(-2\nu x)\,dx},\ \varphi \in H_{\nu,0}(\mathbb{R}).$$

Here m_0 is the closure of the following operator

$$\begin{aligned} \overset{\circ}{C}_\infty(\mathbb{R}) \subseteq L^2(\mathbb{R}) &\to L^2(\mathbb{R}) \\ \varphi &\mapsto (x \mapsto x\varphi(x)). \end{aligned}$$

The operator m_0 is called the multiplication by argument operator and it is densely defined, Hermitian and moreover, it is self-adjoint. Let $\nu \in \mathbb{R}$. We also define the operator $\exp(-\nu m_0)$ such that $\exp(-\nu m_0)\,\varphi := (x \mapsto \exp(-\nu x)\,\varphi(x))$ for $\varphi \in L^1_{loc}(\mathbb{R})$. Note that $\exp(-\nu m_0)\left[\overset{\circ}{C}_\infty(\mathbb{R})\right] = \overset{\circ}{C}_\infty(\mathbb{R})$. Due to the density of $\overset{\circ}{C}_\infty(\mathbb{R})$ in both the spaces $L^2(\mathbb{R})$ and $H_{\nu,0}(\mathbb{R})$, $\exp(-\nu m_0)$ can be extended to a unitary operator from $L^2(\mathbb{R})$ onto $H_{\nu,0}(\mathbb{R})$ and the unitary extension is denoted again by $\exp(-\nu m_0)$. The inverse of $\exp(-\nu m_0)$ is

$$\exp(\nu m_0) : L^2(\mathbb{R}) \to H_{\nu,0}(\mathbb{R}).$$

Note that we may utilize the notation $H_{0,0}(\mathbb{R})$ for the space $L^2(\mathbb{R})$ with the inner product $\langle \cdot \mid \cdot \rangle_{0,0}$ and the norm $|\cdot|_{0,0}$. The following operator

$$\partial_\nu := \exp(\nu m_0)\,\partial_0 \exp(-\nu m_0)$$

is unitarily equivalent to ∂_0 on $H_{\nu,0}(\mathbb{R})$. The operator $\partial_\nu + \nu$ is the time derivative on $H_{\nu,0}(\mathbb{R})$ and we denote it again by ∂_0. Moreover, for all $\nu \in \mathbb{R}_{>0}$, $\partial_0 : D(\partial_0) \subseteq H_{\nu,0}(\mathbb{R}) \to H_{\nu,0}(\mathbb{R})$ is continuously invertible on $H_{\nu,0}(\mathbb{R})$, that is

$$\left\|\partial_0^{-1}\right\|_{L(H_{\nu,0}(\mathbb{R}))} \leq \frac{1}{\nu}$$

and a normal operator for all $\nu \in \mathbb{R}\setminus\{0\}$ on $H_{\nu,0}(\mathbb{R})$. Furthermore, $\partial_0^{-1} : H_{\nu,0}(\mathbb{R}) \to H_{\nu,0}(\mathbb{R})$ is a normal operator (see e.g. [6, Theorem 5.42]) and there is the Sobolev chain

$$H_{\nu,k+1}(\partial_0) \hookrightarrow H_{\nu,k}(\partial_0),\ k \in \mathbb{N}$$

with respect to ∂_0, where $H_{\nu,k}(\partial_0) := D\left(\partial_0^k\right)$ is the Hilbert space with the norm

$$|\cdot|_{\nu,k} = \left|\partial_0^k \cdot\right|_{\nu,0}$$

for each $k \in \mathbb{N}$. Furthermore, we have

$$H_{\nu,-k}(\partial_0) \hookrightarrow H_{\nu,-k-1}(\partial_0),\ k \in \mathbb{N},$$

where $H_{\nu,-k}(\partial_0)$ are completions of $H_{\nu,0}(\mathbb{R})$ for all $k \in \mathbb{N}$ with the norms $|\cdot|_{\nu,-k} := \left|\partial_0^{-k}\cdot\right|_{\nu,0}$. Note that we can unitarily extend the following operator

$$\begin{aligned} H_{\nu,0}(\mathbb{R}) \subseteq H_{\nu,-1}(\partial_0) &\to H_{\nu,0}(\mathbb{R}) \\ \varphi &\mapsto \partial_0^{-1}\varphi. \end{aligned}$$

We denote its extension again by ∂_0^{-1}. This motivates the unitary extension of ∂_0 from $H_{\nu,0}(\mathbb{R})$ onto $H_{\nu,-1}(\partial_0)$ for each $\nu \in \mathbb{R}\setminus\{0\}$ and we denote the extension again by ∂_0. In the same manner we obtain unitary operators

$$\begin{aligned} H_{\nu,k+1}(\partial_0) &\to H_{\nu,k}(\partial_0) \\ \varphi &\mapsto \partial_0\varphi \end{aligned}$$

for $k \in \mathbb{Z}$, as appropriate unitary extension/restriction of the originally discussed operator ∂_0 defined on $H_{\nu,0}(\mathbb{R})$.

2.1. On skew-selfadjoint operator

Let H_1 and H_2 be Hilbert spaces. For a densely defined, closed linear operator $C : D(C) \subseteq H_1 \to H_2$ and a block operator matrix B defined as follows

$$B = \begin{pmatrix} 0 & -C^* \\ C & 0 \end{pmatrix} \tag{5}$$

is skew-selfadjoint and so is the following diagonal operator matrix

$$A = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & B_n \end{pmatrix},$$

where each B_i, $i = \overline{1.n}$ is defined as in (5).

In a coupled system of the heat and Maxwell's equations with Dirichlet boundary condition and electric boundary condition A has the following form

$$\begin{pmatrix} 0 & \mathrm{div} & 0 & 0 \\ \overset{\circ}{\mathrm{grad}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{curl} \\ 0 & 0 & \overset{\circ}{\mathrm{curl}} & 0 \end{pmatrix}$$

where div, $\overset{\circ}{\mathrm{grad}}$, curl and $\overset{\circ}{\mathrm{curl}}$ are defined as follows. Let $\Omega \subseteq \mathbb{R}^3$ be an open set. Consider the following vector analytical differential operators

$$\mathrm{grad}_c : \overset{\circ}{C}_\infty(\Omega) \subseteq L^2(\Omega) \to \bigoplus_{k=1}^{n} L^2(\Omega)$$

$$\phi \quad \mapsto \quad (\partial_k \phi)_{k\in\{1,\dots,n\}}$$

and

$$\begin{aligned} \operatorname{div}_c \ : \ \bigoplus_{k=1}^{n} \overset{\circ}{C}_\infty(\Omega) \subseteq \bigoplus_{k=1}^{n} L^2(\Omega) \quad &\to \quad L^2(\Omega) \\ (\varphi_k)_{k\in\{1,\dots,n\}} \quad &\mapsto \quad \sum_{k=1}^{n} \partial_k \varphi_k. \end{aligned}$$

The operators grad_c and $-\operatorname{div}_c$ are formally adjoint to each other and closable. Denoting

$$\overset{\circ}{\operatorname{grad}} := \overline{\operatorname{grad}_c}, \ \overset{\circ}{\operatorname{div}} := \overline{\operatorname{div}_c}$$

and

$$\operatorname{grad} := \left(-\overline{\operatorname{div}_c}\right)^*, \ \operatorname{div} := \left(-\overline{\operatorname{grad}_c}\right)^*,$$

we can construct the following skew-selfadjoint operator

$$A_1^D := \begin{pmatrix} 0 & \operatorname{div} \\ \overset{\circ}{\operatorname{grad}} & 0 \end{pmatrix},$$

where $\overset{\circ}{\operatorname{grad}}$, div and grad, $\overset{\circ}{\operatorname{div}}$ are all together densely defined, closed linear operators. The operator A_1^D is not only skew-selfadjoint but also encode Dirichlet boundary condition, that is, φ being in $D\left(\overset{\circ}{\operatorname{grad}}\right)$ means that φ satisfies a generalized homogeneous Dirichlet boundary condition.

Due to the skew-selfadjointness of A, we have a long Sobolev chain with respect to $A+1$. Since $\pm 1 \in \rho(A)$, the domains of the operators A and $A+1$ coincide. There is the Sobolev chain

$$H_{k+1}(A+1) \hookrightarrow H_k(A+1) \text{ for } k \in \mathbb{Z}.$$

Here $H =: H_0(A+1)$, $H_k(A+1)$ is the domain of $(A+1)^k$ and it is a Hilbert space with the norm $|\cdot|_k := \left|(A+1)^k \cdot\right|_{0,0}$ for each $k \in \mathbb{N}$ and $H_{-k}(A+1)$ is the completion of H for each $k \in \mathbb{N}$ under the norm $|\cdot|_{-k} := \left|(A+1)^{-k} \cdot\right|_{0,0}$. For the sake of brevity, we also denote $H_{k,A} := H_k(A+1)$. Now we are in the position to construct the Sobolev lattices

$$\left(H_{\nu,k} \otimes H_{n,A}\right)_{k,n\in\mathbb{Z}}$$

for the chains $\left(H_{\nu,k}(\partial_0)\right)_{k\in\mathbb{Z}}$ and $\left(H_n(A+1)\right)_{n\in\mathbb{Z}}$ with respect to the operators $\partial_0 \otimes I_H$ and $I_{H_{\nu,0}} \otimes A$. Here $I_H \ : \ H \to H$ and $I_{H_{\nu,0}} \ : \ H_{\nu,0}(\mathbb{R}) \to H_{\nu,0}(\mathbb{R})$ are the identity operators. Note that $H_{\nu,k} \otimes H$ can be interpreted as the completion of the linear space generated by $H-$valued functions of the special form

$$t \mapsto \psi(t)\, w =: (\psi \otimes w)(t)$$

for each $k \in \mathbb{N}$, where $\psi \in \overset{\circ}{C}_\infty(\mathbb{R})$, $w \in H$. In fact, $H_{\nu,k} \otimes H$ is unitarily equivalent to $H_{\nu,k}(\mathbb{R}, H)$ for each $k \in \mathbb{N}$.

The operators $\partial_0 \otimes I_H$ and $I_{H_{\nu,0}} \otimes A$ are well-defined and have essentially the same properties as the operators ∂_0 and A, respectively. Therefore, we also write A and ∂_0 for their canonical extensions $A \otimes I_H$ and $I_{H_{\nu,0}} \otimes \partial_0$ in $H_{\nu,0}(\mathbb{R}, H)$.

2.2. The material law operator

The Fourier-Laplace transform

$$\mathcal{L}_\nu := \mathcal{F}\exp(-\nu m_0) : H_{\nu,0}(\mathbb{R},H) \to L^2(\mathbb{R},H)$$

given as a composition of the (temporal) Fourier transform $\mathcal{F}$ and the unitary weight operator $\exp(-\nu m_0)$, is a spectral representation associated with ∂_0. It is

$$\partial_0 = \mathcal{L}_\nu^*(im_0+\nu)\mathcal{L}_\nu.$$

This observation allows us to consistently define an operator function calculus associated with ∂_0 in a standard way and we can even extend this calculus to operator-valued functions by letting

$$M(\partial_0^{-1}) := \mathcal{L}_\nu^* M\left(\frac{1}{im_0+\nu}\right)\mathcal{L}_\nu.$$

Here the linear operator $M\left(\frac{1}{im_0+\nu}\right) : L^2(\mathbb{R},H) \to L^2(\mathbb{R},H)$ is determined uniquely via

$$\left(M\left(\frac{1}{im_0+\nu}\right)\varphi\right)(\lambda) := M\left(\frac{1}{im_0+\nu}\right)\varphi(\lambda) \text{ in } H$$

for every $\lambda \in \mathbb{R}$, $\varphi \in \overset{\circ}{C}_\infty(\mathbb{R},H)$ by an operator-valued function M. For a material law the operator-valued function M needs to be bounded and an analytic function $z \mapsto M(z)$ in an open ball $B_{\mathbb{C}}(r,r)$ with some positive radius r centered at r. Here we will concentrate on the following particular form of the material law

$$M(\partial_0^{-1}) = M_0 + \partial_0^{-1}M_1,$$

where M_0 is selfadjoint, bounded linear and $M_0 \geqslant c_0 > 0$ on the range $M_0[H]$, the null space $[\{0\}]M_0$ is non-trivial and $M_1 \in L(H)$ with $\Re M_1 \geqslant c_1 > 0$ on $[\{0\}]M_0$.

This is not an artificial assumption, rather a necessary constraint enforced by the requirement of causality and strictly positive definite condition

$$\Re\langle u \mid (\partial_0 M(\partial_0^{-1}))u\rangle_{H_{\nu,0}(\mathbb{R},H)} \geqslant c\langle u \mid u\rangle_{H_{\nu,0}(\mathbb{R},H)}$$

for $c \in \mathbb{R}_{>0}$ and all sufficiently large $\nu \in \mathbb{R}_{>0}$ and all $u \in D(\partial_0)$. The strict positive definite condition implies

$$\Re\langle u \mid (\partial_0 M_0 + M_1 + A)u\rangle_{H_{\nu,0}(\mathbb{R},H)} \geqslant c\langle u \mid u\rangle_{H_{\nu,0}(\mathbb{R},H)}$$

for $c \in \mathbb{R}_{>0}$ and all sufficiently large $\nu \in \mathbb{R}_{>0}$ and all $u \in D(\partial_0)$. Moreover,

$$\partial_0 M_0 + M_1 + A$$

has dense range in $H_{\nu,0}(\mathbb{R},H)$. For all sufficiently large $\nu \in \mathbb{R}_{>0}$, we have

$$\left\|\overline{(\partial_0 M_0 + M_1 + A)}^{-1}\right\|_{L(H_{\nu,0}(\mathbb{R},H))} \leq \frac{1}{c_\nu},\ 0 < c_\nu < c_1$$

and this also implies the solution theory of the following evolutionary problems

$$(\partial_0 M_0 + M_1 + A)u = S(u) + f,$$

where S is a suitable Lipschitz mapping.

2.3. Well-posedness of evolutionary problems with a non-linear perturbation term

After making some reformulations in the microwave heating problem, the problem gets the following shape

$$(\partial_0 M_0 + M_1 + A)\, u + \tilde{F}(u) = F,$$

where $\tilde{F}$ is a quadratic form and it is not Lipschitz continuous. For a Lipschitz continuous approximation of the quadratic form we recall the following Lemma and Theorem in [7].

Lemma 1. *Let $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be differentiable, and such that $\left(z \mapsto \left|\sqrt{z} f'(z)\right|\right)$ is bounded. Let $\mathcal{E} \in \mathbb{C}^{n\times n}$ be selfadjoint with $\mathcal{E} \geq 0$. Then there exists $C > 0$ such that*

$$\left|f\left(\langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^n}\right) - f\left(\langle v \mid \mathcal{E}v\rangle_{\mathbb{C}^n}\right)\right|_{\mathbb{R}} \leqslant C\,|u-v|_{\mathbb{C}^n}$$

for all $u, v \in \mathbb{C}^n$.

Theorem 4. *Let (Ω, μ) be a σ–finite measure space. Let $\mathcal{E} \in (L^\infty(\Omega))^{n\times n}$ and*

$$\begin{array}{rcl} \tilde{F} : D\left(\tilde{F}\right) & \to & H_{\nu,0}(\mathbb{R}) \otimes L^2(\Omega) \\ u & \mapsto & \left(\mathbb{R}\times\Omega \ni (t,\omega) \mapsto \langle u(t,\omega) \mid \mathcal{E}(\omega)\, u(t,\omega)\rangle_{\mathbb{C}^n}\right) \end{array}$$

with maximal domain. Here $D\left(\tilde{F}\right) \subseteq H_{\nu,0}(\mathbb{R}) \otimes \left(L^2(\Omega)\right)^n$. We assume that $\mathcal{E}(\omega) \in \mathbb{C}^{n\times n}$ is selfadjoint and positive for a.e. $\omega \in \Omega$. Let $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be differentiable with $f(0) = 0$ and such that

$$z \mapsto \left|\sqrt{z} f'(z)\right|$$

is bounded. Define

$$\begin{array}{rcl} F_f : D\left(F_f\right) \subseteq H_{\nu,0}(\mathbb{R}) \otimes \left(L^2(\Omega)\right)^n & \to & H_{\nu,0}(\mathbb{R}) \otimes L^2(\Omega) \\ u & \mapsto & \left(\mathbb{R}\times\Omega \ni (t,\omega) \mapsto f\left(\tilde{F}(u)(t,\omega)\right)\right) \end{array}$$

with maximal domain. Then $D\left(F_f\right) = H_{\nu,0}(\mathbb{R}) \otimes \left(L^2(\Omega)\right)^n$ and F_f is Lipschitz continuous.

Hence, it suffices to find a specific function f which satisfies all the assumptions in Lemma 1 and approximates the quadratic form.

Example 1. We consider the following function

$$\begin{array}{rcl} f_\xi : \mathbb{R}_{\geq 0} & \to & \mathbb{R}_{\geq 0} \\ x & \mapsto & \dfrac{2}{\xi}\left(\sqrt{1+\xi x} - 1\right) \end{array}$$

for $\xi \in \mathbb{R}_{>0}$. The function is infinitely differentiable for all $x > -\frac{1}{\xi}$ and as $\xi \to 0+$, it is approximated by the argument, that is,

$$f_\xi(x) \approx x.$$

The mapping $\left(z \mapsto \left|\sqrt{z} f_\xi'(z)\right|\right)$ is uniformly bounded. Indeed,

$$\left(z \mapsto f_\xi'(z)\sqrt{z}\right) \quad = \quad \left(z \mapsto \frac{\sqrt{z}}{\sqrt{1+\xi z}}\right)$$

$$\leqslant \quad \frac{1}{\sqrt{\xi}}.$$

Let $\mathcal{E}(\omega) \in \mathbb{C}^{n\times n}$ be selfadjoint and positive for a.e. $\omega \in \Omega$. By Lemma 1, the following holds

$$\left|f_\xi\left(\langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^n}\right) - f_\xi\left(\langle v \mid \mathcal{E}v\rangle_{\mathbb{C}^n}\right)\right| \leqslant C\,|u-v|_{\mathbb{C}^n}$$

for $C \in \mathbb{R}_{>0}$. Since $f_\xi(0) = 0$ for all $\xi \in \mathbb{R}_{>0}$, the mapping defined by

$$\begin{aligned} F_{f_\xi} \;:\; H_{\nu,0}(\mathbb{R}) \otimes \left(L^2(\Omega)\right)^n \quad &\to \quad H_{\nu,0}(\mathbb{R}) \otimes L^2(\Omega) \\ u \quad &\mapsto \quad \left((t,\omega) \mapsto f_\xi\left(\langle u(t,\omega) \mid \mathcal{E}(\omega)\,u(t,\omega)\rangle_{\mathbb{C}^n}\right)\right) \end{aligned}$$

is Lipschitz continuous for all $\xi \in \mathbb{R}_{>0}$ and $(t,\omega) \in \mathbb{R}\times\Omega$ (Theorem 4) and the Lipschitz constant of the mapping F_{f_ξ} is $\frac{2}{\sqrt{\xi}}\left\|\sqrt{\mathcal{E}}\right\|_\infty$. Furthermore, the following holds

$$F_{f_\xi}(u) = f_\xi\left(\langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^n}\right) \approx \langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^n}$$

for sufficiently small $\xi \in \mathbb{R}_{>0}$. We have obtained the Lipschitz continuous mapping $F_{f_\xi} = \left(u \mapsto f_\xi\left(\langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^n}\right)\right)$ which approximates $\tilde{F} = \left(u \mapsto \left(\langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^n}\right)\right)$ as $\xi \to 0+$.

Hence, the solution theory of the perturbed problem

$$(\partial_0 M_0 + M_1 + A)\,u + F_{f_\xi}(u) = F$$

provides an approximate solution of (4).

Theorem 5. *Let (Ω,μ) be a σ–finite measure space. Let $H = L^2(\Omega)^n$. Let $M_0 \in L(H)$ be selfadjoint, positive definite and $M_0|_{M_0[H]} \geqslant c_0 > 0$, and $M_1 \in L(H)$ with $\Re M_1|_{[\{0\}]M_0} \geqslant c_1 > 0$. Let $A : H_{1,A} \subseteq H \to H$ be a skew-selfadjoint operator. Assume that $\mathcal{E} \in (L^\infty(\Omega))^{n\times n}$ is selfadjoint and positive definite, and $0 \leq \frac{2}{c_\nu\sqrt{\xi}}\left\|\sqrt{\mathcal{E}}\right\|_\infty < 1$ for some $\xi \in \mathbb{R}_{>0}$. Let $u_0 \in D(A)$ and $F \in \chi_{\mathbb{R}_{\geq 0}}(m_0)\left[H_{\nu,0} \otimes H\right]$ be given data. Then there exists a unique solution $u \in H_{\nu,0} \otimes H$ of*

$$\overline{\left((\partial_0 M_0 + M_1 + A)|_{D(\partial_0)\cap D(A)}\right)}\,u = F_{f_\xi}(u) + F + \delta \otimes M_0 u_0$$

for all $\nu \geq \nu_0$ for some $\nu_0 \in \mathbb{R}_{>0}$. The solution depends continuously and causally on the data. Moreover, the initial condition

$$(M_0 u)(0+) = M_0 u_0$$

is attained in $H_{-1,A}$.

3. The microwave heating problem

The microwave heating problem has wide industrial applications and it has been studied theoretically and numerically in various situations (see e.g. [1–3] and the references therein). In the study of the microwave heating problem the electric conductivity and/or thermal conductivity are considered as an operator, which may depend on the temperature (see e.g. [3]). But we will study the microwave heating problem with temperature independent thermal conductivity, electric conductivity, magnetic permeability and electric permittivity. These material coefficients are defined as 3×3 matrix-valued functions depending only on the spatial variables. This may describe material properties more substantially. The equations are introduced in the introduction. The set of originally given equations turns into the following equations

$$\rho C_\rho \partial_0 \vartheta + \operatorname{div} q = \langle E \mid \sigma E\rangle_{\mathbb{C}^3}$$

$$\begin{aligned}
&\mathrm{grad}\vartheta + \lambda^{-1}q = 0\\
&-\mathrm{curl}H + \sigma E + \partial_0 \varepsilon E = J_1\\
&\mathrm{curl}E + \partial_0 \mu H = 0.
\end{aligned}$$

A formal reformulation of these equations yields

$$\left(\partial_0 M_0 + M_1 + \begin{pmatrix} 0 & \mathrm{div} & 0 & 0\\ \mathrm{grad} & 0 & 0 & 0\\ 0 & 0 & 0 & -\mathrm{curl}\\ 0 & 0 & \mathrm{curl} & 0 \end{pmatrix}\right)\begin{pmatrix} \vartheta\\ q\\ E\\ H \end{pmatrix} = F + \begin{pmatrix} \langle E \mid \sigma E\rangle_{\mathbb{C}^3}\\ 0\\ 0\\ 0 \end{pmatrix}, \tag{7}$$

where

$$M_0 = \begin{pmatrix} \rho C_\rho & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & \varepsilon & 0\\ 0 & 0 & 0 & \mu \end{pmatrix},\ M_1 = \begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & \lambda^{-1} & 0 & 0\\ 0 & 0 & \sigma & 0\\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and $\langle E \mid \sigma E\rangle_{\mathbb{C}^3} = \langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^{10}}$ with

$$\mathcal{E} := \begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & \sigma & 0\\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Now we reformulate (7) to the proper evolutionary problem. Assume that ϑ satisfies the (generalized) Dirichlet boundary condition and E satisfies the (generalized) homogeneous electric boundary condition. Then we have the following skew-selfadjoint operator

$$A := \begin{pmatrix} 0 & \mathrm{div} & 0 & 0\\ \overset{\circ}{\mathrm{grad}} & 0 & 0 & 0\\ 0 & 0 & 0 & -\mathrm{curl}\\ 0 & 0 & \overset{\circ}{\mathrm{curl}} & 0 \end{pmatrix}$$

in $H_{1,A} := H\left(\overset{\circ}{\mathrm{grad}}, \Omega\right) \oplus H(\mathrm{div}, \Omega) \oplus H\left(\overset{\circ}{\mathrm{curl}}, \Omega\right) \oplus H(\mathrm{curl}, \Omega)$, where $\Omega \subseteq \mathbb{R}^3$ is an open set. As in the preceding application, we assume that $\rho C_\rho : L^2(\Omega) \to L^2(\Omega)$, $\varepsilon : \left(L^2(\Omega)\right)^3 \to \left(L^2(\Omega)\right)^3$ and $\mu : \left(L^2(\Omega)\right)^3 \to \left(L^2(\Omega)\right)^3$ are selfadjoint, bounded linear and strictly positive definite operators. Hence, M_0 is selfadjoint, bounded linear and strictly positive definite in $M_0\left[\left(L^2(\Omega)\right)^{10}\right]$. Let λ^{-1} be in $L\left(\left(L^2(\Omega)\right)^3\right)$ and $\Re\lambda^{-1}$ strictly positive definite. Furthermore, assume that σ is selfadjoint, positive definite and it is in $(L^\infty(\Omega))^{3\times 3}$. Then, M_1 is in $L\left(\left(L^2(\Omega)\right)^{10}\right)$ and $\Re M_1$ is strictly positive definite on $[\{0\}]M_0 = \{0\} \oplus \left(L^2(\Omega)\right)^3 \oplus \{0\}^6$. Since $\sigma \in (L^\infty(\Omega))^{3\times 3}$ is selfadjoint and positive definite, so is $\mathcal{E} \in (L^\infty(\Omega))^{10\times 10}$. Moreover, there exists a selfadjoint, positive definite operator $\sqrt{\mathcal{E}}$. Hence, the quadratic

form $\langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^{10}}$ is approximated by the Lipschitz continuous mapping $F_{f_\xi} = \left(u \mapsto f_\xi\left(\langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^{10}}\right)\right)$ as $\xi \to 0+$, where f_ξ is defined in Example 1. Let $\vartheta_0 \in H\left(\overset{\circ}{\text{grad}}, \Omega\right)$, $E_0 \in H\left(\overset{\circ}{\text{curl}}, \Omega\right)$, $H_0 \in H(\text{curl}, \Omega)$ and $V_0 := \left(\rho C_\rho \vartheta_0, 0, \varepsilon E_0, \mu H_0\right)$. Then $u_0 := (\vartheta_0, 0, E_0, H_0) \in D(A)$, $V_0 = M_0 u_0$ and $V_0 \in M_0[D(A)]$. The initial boundary-value problem with respect to the microwave heating problem is presented in our framework as follows

$$\left(\partial_0 M_0 + M_1 + \begin{pmatrix} 0 & \overset{\circ}{\text{div}} & 0 & 0 \\ \text{grad} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\text{curl} \\ 0 & 0 & \overset{\circ}{\text{curl}} & 0 \end{pmatrix}\right)\begin{pmatrix} \vartheta \\ q \\ E \\ H \end{pmatrix} = F + \begin{pmatrix} \langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^{10}} \\ 0 \\ 0 \\ 0 \end{pmatrix} + \delta \otimes \begin{pmatrix} \rho C_\rho \vartheta_0 \\ 0 \\ \varepsilon E_0 \\ \mu H_0 \end{pmatrix}. \quad (8)$$

This problem yields the following initial value evolutionary problem with the Lipschitz continuous perturbation

$$\left(\partial_0 M_0 + M_1 + \begin{pmatrix} 0 & \text{div} & 0 & 0 \\ \overset{\circ}{\text{grad}} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\text{curl} \\ 0 & 0 & \overset{\circ}{\text{curl}} & 0 \end{pmatrix}\right)\begin{pmatrix} \vartheta \\ q \\ E \\ H \end{pmatrix} = F + \begin{pmatrix} F_{f_\xi}(u) \\ 0 \\ 0 \\ 0 \end{pmatrix} + \delta \otimes \begin{pmatrix} \rho C_\rho \vartheta_0 \\ 0 \\ \varepsilon E_0 \\ \mu H_0 \end{pmatrix} \quad (9)$$

for sufficiently small $\xi \in \mathbb{R}_{>0}$. In the next theorem we sum up the solution theory of (9), which concerns the approximation solution of (8).

Theorem 6. *Let $\lambda^{-1} \in L\left(\left(L^2(\Omega)\right)^3\right)$, and $\Re\lambda^{-1}$ be strictly positive definite. Let $\varepsilon, \mu \in L\left(\left(L^2(\Omega)\right)^3\right)$ be selfadjoint and strictly positive definite operators. Let $\rho C_\rho \in L\left(L^2(\Omega)\right)$ be selfadjoint and strictly positive definite. Let $\sigma \in \left(L^\infty(\Omega)\right)^{3\times 3}$ be selfadjoint, positive definite. Let $(\vartheta_0, 0, E_0, H_0) \in D(A)$ and $F \in \chi_{\mathbb{R}_{\geq 0}}(m_0)\left[H_{\nu,0}(\mathbb{R}) \otimes \left(L^2(\Omega)\right)^{10}\right]$ be given data. Let $\nu_0 \in \mathbb{R}_{>0}$. Furthermore, assume that $0 \leq \frac{2}{c_\nu\sqrt{\xi}}\left\|\sqrt{\mathcal{E}}\right\|_\infty < 1$ for all $\nu \geq \nu_0$ and for some parameter $\xi \in \mathbb{R}_{>0}$ and some $c_\nu > 0$. Then there exists a unique solution $u \in H_{\nu,0}(\mathbb{R}) \otimes \left(L^2(\Omega)\right)^{10}$ of (8) for all $\nu \geq \nu_0$. The solution depends continuously and causally on the given data.*

4. Conclusions

We have obtained a Lipschitz continuous function approximating the quadratic form

$$\left(u \mapsto \langle u \mid \mathcal{E}u\rangle_{\mathbb{C}^n}\right)$$

for a selfadjoint, positive definite operator $\mathcal{E}$ in $\left(L^\infty(\Omega)\right)^{n\times n}$, $n \in \mathbb{N}$. This gives us an opportunity to conclude the well-posedness and causality of the evolutionary problems with non-linear term consisting of the quadratic form with the help of the solution theory associated to the evolutionary problems with a Lipschitz continuous perturbation.

The quadratic form can be found in the heat equation coupled with Maxwell's equations. One of these coupled systems is the microwave heating problem. Here we assumed that the physical coefficients describing the properties of the underlying material, $\varepsilon, \mu \in L\left(\left(L^2(\Omega)\right)^3\right)$, $\lambda, \sigma, \alpha \in \left(L^\infty(\Omega)\right)^{3\times 3}$ are 3×3 matrix-valued functions depending only on the spatial variables.

Author Contributions: The author has read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing is not applicable.

Acknowledgments: I would like to express my heartfelt gratitude to my supervisor, Prof. Dr. Rainer Picard for his continuous support, guidance and his established solution theory of evolutionary problems. My especial thanks to Prof. Dr. Marcus Waurick who made numerous comments and corrections on my research work. I would also like to thank academician O.Chuluunbaatar for advising and his suggestions correcting some errors in the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hill, J. M. & Marchant, T. R. Modelling microwave heating. *Appl. Math. Model.* **20,** 3–15 (1996).
2. Yin, H. M. Regularity of weak solution to Maxwell's equations and applications to microwave heating. *J. Differ. Equ.* **200,** 137–161 (2004).
3. Yin, H. M. Existence and regularity of a weak solution to Maxwell's equations with a thermal effect. *Math. Methods Appl. Sci.* **29,** 1199–1213 (2006).
4. Picard, R. A structural observation for linear material laws in classical mathematical physics. *Math. Methods Appl. Sci.* **32,** 1768–1803 (2009).
5. Picard, R. & McGhee, D. *Partial Differential Equations: A unified Hilbert Space Approach* 469 pp. (Berlin/New-York, 2011).
6. Weidmann, J. *Linear Operators in Hilbert Spaces* 402 pp. (Springer-Verlag, New-York, 1980).
7. Tsangia, B. *Evolutionary problems: Applications to Thermoelectricity* PhD thesis (TU Dresden, 2014).

Information about the authors

Baljinnyam Tsangia—Dr.rer.nat, Lecturer of Department of Mathematics, School of Applied Sciences, Mongolian University of Science and Technology (e-mail: Baljinnyam.Tsangia@must.edu.mn, ORCID: 0000-0002-3331-2516)

UDC 519.872, 519.217
PACS 07.05.Tp, 02.60.Pn, 02.70.Bf
DOI: 10.22363/2658-4670-2024-32-2-222–233

Корректность задачи о микроволновом нагреве

Балжинням Цангиа

Монгольский университет науки и технологий, Улан-Батор, Монголия

Аннотация. Ряд начально-краевых задач классической математической физики формулируется в виде линейного операторного уравнения, а его корректность и причинность в гильбертовом пространстве были установлены ранее. Если задача имеет единственное решение и решение постоянно зависит от заданных параметров, то задача называется корректной. Независимость дальнейшего поведения решения до определенного момента указывает на причинность решения. В данной работе установлены корректность и причинность решения эволюционных задач с возмущением, определяемым квадратичной формой. В качестве примера рассмотрена связанная система уравнений теплопроводности и Максвелла (задача микроволнового нагрева).

Ключевые слова: Эволюционные задачи, нелинейное возмущение, Липшицева непрерывность, квадратичная форма, связанные задачи

Discrete & Continuous Models
& Applied Computational Science
ISSN 2658-7149 (Online), 2658-4670 (Print)

2024, 32 (2) 234–241
http://journals.rudn.ru/miph

Research article

UDC 519.6
PACS 07.05.Tp,
DOI: 10.22363/2658-4670-2024-32-2-234–241 EDN: CUCXTY

Developing a computer system for student learning based on vision-language models

Eugeny Yu. Shchetinin[1], Anastasia G. Glushkova[2], Anastasia V. Demidova[3]

[1] *Financial University under the Government of the Russian Federation, 49 Leningradsky Ave, Moscow, 125993, Russian Federation*
[2] *Endeavor, London W4 5HR, Chiswick Park, 566 Chiswick High Road, United Kingdom*
[3] *RUDN University, 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation*

(received: November 6, 2023; revised: January 20, 2024; accepted: February 7, 2024)

Abstract. In recent years, artificial intelligence methods have been developed in various fields, particularly in education. The development of computer systems for student learning is an important task and can significantly improve student learning. The development and implementation of deep learning methods in the educational process has gained immense popularity. The most successful among them are models that consider the multimodal nature of information, in particular the combination of text, sound, images, and video. The difficulty in processing such data is that combining multimodal input data by different channel concatenation methods that ignore the heterogeneity of different modalities is an inefficient approach. To solve this problem, an inter-channel attention module is proposed in this paper. The paper presents a computer vision-linguistic system of student learning process based on the concatenation of multimodal input data using the inter-channel attention module. It is shown that the creation of effective and flexible learning systems and technologies based on such models allows to adapt the educational process to the individual needs of students and increase its efficiency.

Key words and phrases: deep learning, vision-language learning model, neural networks-transformers, through-channel attention module

Citation: Shchetinin E. Y., Glushkova A. G., Demidova A. V., Developing a computer system for student learning based on vision-language models. *Discrete and Continuous Models and Applied Computational Science* **32** (2), 234–241. doi: 10.22363/2658-4670-2024-32-2-234–241. edn: CUCXTY (2024).

1. Introduction

The history of the development and creation of computer learning models follows from the creation of models of machine translation of text, Hidden Markov chain models, then Recurrent Neural Networks (RNN) were used for this purpose for a long time [1–3]. However, as the flow of information increases, the quality of its processing using then recurrent neural networks decreased significantly, because to obtain a complete and coherent final text it is not enough to translate individual sentences, it is necessary to consider its overall context. Also, recurrent networks required sequential computations, which limited the ability to effectively use modern GPUs for model training.

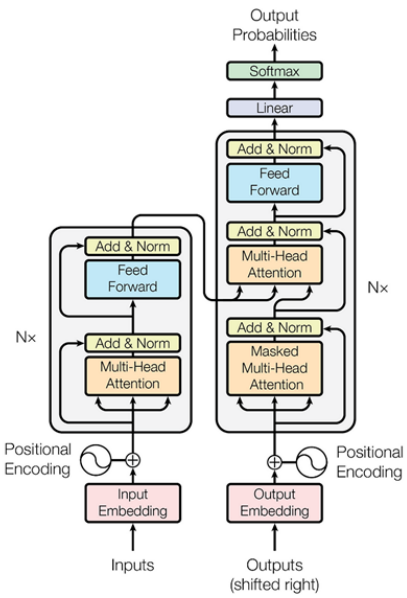


Figure 1. Transformer architecture

To fix the problem, an "attention mechanism" was developed to focus on important parts of the text [4, 5]. With its help, the neural network evaluates which position of the incoming sequence is the most important for a particular position of the sequence in the output. In the process of developing the architecture of recurrent networks, scientists from Google Research and Google Brain came up with a more technologically advanced family of deep learning architectures – transformer neural networks (transformers). It combines parallel data processing, the possibility of pre-training models and a wide application of the attention mechanism. A transformer neural network consists of two sets of layers – encoders and decoders, which contain multiple layers.

This paper presents a computer system-based vision-language model of student learning process based on combining the multimodal input data through a through-channel attention module. It is shown that the creation of effective and flexible learning systems and technologies based on such models allows to adapt the educational process to the individual needs of students and increase its efficiency.

2. Modeling the learning process using Vision-Language Models

To create a vision-language computer system of student learning, this paper used the Transformers architecture, that's most recent and successful approach in the field of natural language processing. Our model consists of several layers of transformers, each layer consisting of multiple attention mechanisms – input, internal and output. Input attention mechanisms allow the model to model and assign importance to different parts of the input data. Internal attention mechanisms allow the model to analyze the interactions between different data elements. Output attention mechanisms allow the model to generate responses based on the information received. The result is shown in Fig. 1.

The complexity and multimodality of learning data is such that instead of combining the multimodal input data through channel concatenation, which ignores the heterogeneity of different modalities, we propose a cross-channel attention module. This paper presents a computer system-based vision-language model of student learning process based on combining the multimodal input data through a through-channel attention module. Proposed model was trained on a large set of texts containing information from different subject areas. Each text is broken down into character sequences of several hundred (e.g., 40000) characters in length to feed the model. This approach allows the model to be trained on longer contexts and to consider the full amount of information when generating responses.

3. Model training

A teacher-guided learning method is used to train the model, where for each input sequence of symbols, there is a corresponding output sequence of symbols that represents the correct answer or the next step in student learning. While training the model, the input symbols are gradually fed into the model one by one, and the model generates predictions for the next symbol in the sequence. For each symbol generated, the model calculates the error and adjusts its parameters to reduce the error. This process is repeated over several epochs until the model reaches a given level of accuracy.

Advantages of a transformer-based model:

1. Flexibility: Transformers allow the modeling of long sequences of characters, which is especially important when processing texts and contexts in an educational setting. A Transformer-based model is able to capture the relationships between different data elements and generate responses that are tailored to the learning context.
2. Better context processing: Transformers allow the model to process many consecutive characters simultaneously using the attention mechanism. This allows the model to consider all previous symbols when generating a response, which is especially important in the case of student learning, where context can be critical to understanding and aiding learning.
3. Generating high quality responses: Transformers have the advantage of generating accurate and grammatically correct output sequences of symbols. This is especially important for educational tasks where accuracy and clarity of responses are critical.
4. Adaptability and Personalization: The Transformer-based model easily adjusts its parameters based on individual student needs to provide personalized learning. The model can adaptively respond to the student's proficiency level and progress, providing more appropriate and individualized responses and prompts.

Additional steps and research include determining the appropriate architecture and hyperparameters of the model, selecting the right data, optimizing training, and evaluating and validating the model using quality indicators of the original student learning. Despite the challenges, developing a Transformer-based vision-language model for student learning has great potential to improve the educational process [6]. This model can help students in getting more personalized learning, quick access to information, and effective comprehension of the material. Developing a language model for teaching students English based on Transformers can be very useful in an educational setting. Such a model can help students to master various aspects of English including grammar, vocabulary, reading, listening, and writing.

4. An example of a vision-language model for teaching students to learn English

Let's present an example of the language model of teaching students English based on transformers:

1. Problem Statement: The model was trained on a large corpus of English texts that can be collected from various sources - books, articles, blogs, etc. The goal of the model is to teach the student correct grammatical constructions, improve his/her vocabulary, develop reading and comprehension skills in English text, and help develop writing skills in English.
2. Input data: for model training, we use English text data, which we split into character sequences of several hundred (e.g., 40000) characters in length. Thus, each input sequence is a set of characters preceding a certain point in the text.
3. Model Architecture: The model consists of several layers of transformers, where each layer contains attention mechanisms. Input symbols are fed into the model sequentially, and at each step the model generates a prediction for the next symbol in the sequence.
4. Model training: The model is trained using a teacher approach, where for each input sequence of symbols, there is a corresponding correct output sequence of symbols. The model parameters are adjusted based on the error computed between the generated and correct output sequences. The goal is to minimize this error over multiple training epochs.
5. Answer generation and evaluation: after training the model, students can ask questions or provide texts in English and the model will generate answers or sentences using their knowledge of English. These responses can be evaluated based on correct grammatical constructions, lexical variety, comprehensibility and cohesion.

Such a model based on transformers can help students improve their English skills. Here are some possible examples of how to use the model:

- Generating answers to grammar questions: A student can ask a question about the correct use of a certain grammatical construction or tense, and the model can generate the correct answer with explanations and examples.
- Help with reading and comprehension of English texts: Students can provide a text to read, and the model is able to offer explanations of difficult words or phrases and provide translations or additional contextual materials for better understanding.
- Support for writing skills: Students can provide their English texts for revision and commentary. The model can offer corrections, suggestions for improving style and variation of phrases to help the student develop their writing skills.
- Exam Preparation: The model can help students prepare for English exams by providing practice in writing essays, composing narratives, answering grammar exercises, and reading and listening tests.

One practical example of using a Transformer-based model to teach students English could be to use it in online courses or educational platforms. The model can be integrated into a system where students can ask questions, write essays or do exercises and the model will provide feedback and hints to improve skills. For example, students can submit their essays for revision and the model will analyze the text, detect errors in grammar, and suggest correction options. It can point out difficult grammatical constructions and provide explanations about their usage. The model can also suggest different phrases or expressions to improve the writing style and expand the student's vocabulary.

In addition, the model can help students perceive and understand English text. By providing reading text, students can get explanations for difficult words, phrases, and expressions. The model can suggest synonyms, antonyms, or examples of how words are used in different contexts. Using the attention mechanism, the model can highlight key points in the text and help students understand its content. The model can offer supplementary materials related to topics from the text so that students can extend their knowledge and improve their comprehension.

Also, the model can offer reading and listening exercises, analyze students' responses, and provide feedback on their performance. It can help students develop listening skills, detect errors in grammar

or vocabulary by automatically checking answers. In addition to educational platforms, the model can be integrated into mobile apps that students can use to study and practice anywhere and anytime they want. This allows students to access English language learning instruction and support even outside the classroom.

One of the important functions of the computer vision-language model for English language learning is the ability to personalize learning [7, 8]. The model can consider the proficiency level and needs of each student by providing individualized recommendations and feedback. It can analyze student's mistakes and suggest appropriate material to correct them. This allows students to receive individualized instruction and focus on their weaknesses [9]. In addition, the model can have interactive dialog features, allowing students to ask questions and receive direct answers from the model. This can be useful for clarifying unclear concepts, overcoming difficulties, or getting additional explanations on topics that students find challenging.

Interactive dialog can also be used to practice speaking skills. Students can ask questions or ask the model to comment on their statements in English. This will help them practice grammar, correct pronunciation, and the ability to express themselves in English. Overall, the Transformer-based model for teaching students English can greatly enrich the learning process and improve the effectiveness of language learning. It can provide personalized support, give feedback, offer supplementary materials, and help to develop both writing, reading, listening, and speaking skills.

5. Discussion of the results

The development of a computerized language model for teaching English language learners based on Transformers is complex and requires a large amount of training data. In addition, continuous refinement and updating of the model is an important part of the process so that it can adapt to the changing needs and requirements of students. It is also important to account for the diversity of language skill levels and student needs so that the model can provide individualized support and adapt to each student. Despite the challenges in development, the Transformer-based model for teaching English language learners has significant potential. It can provide individualized and effective feedback, help students master grammar and vocabulary, understand, and analyze English texts, and develop writing skills. In the future, with further research and development, such models can become an integral part of the educational process. They can help students overcome language barriers, provide more flexible and personalized learning, and increase the effectiveness of English language learning.

Artificial intelligence extends the possibilities of the educational process from the creation of a digital university to the training of specific disciplines. Artificial intelligence becomes the basis for the creation of new means and tools for learning, in particular chatbots that can be used as simulators in the learning system. The most popular and convenient in teaching a foreign language in a higher education institution is the GPT chatbot based on a language model (GPT-3,5; GPT-4, etc. [9]). Implementation of GPT models is carried out not only by OpenAI company, but also by other developers both foreign (BERT by Google, Turing NLG by Microsoft) and domestic (Yandex company announced in February 2023 the release of ChatGPT-style – "YaLM 2.0", "Sber" company trained GPT-3 model on the Russian language corpus, creating ruGPT, widely used in the Russian-speaking field). GOOGLE has also launched its analog of the Bard chatbot, based on the principle of artificial intelligence.

This chatbot has all chances to become a valuable tool for teaching English to students. It can be used to practice speaking, grammar, vocabulary, and writing skills, learn and use cultural context, and foster group collaboration skills. With the help of technology such as GPT, English language

learning can become more accessible and engaging. However, GPT should be used in conjunction with traditional language learning methods, such as classroom instruction and language immersion, and with clear guidelines for students. When using ChatGPT in the classroom, it is important to set rules and expectations for student behavior. Students should be encouraged to use ChatGPT as a tool for learning, but not rely on it as the sole source of language learning. Teachers should also monitor students' interactions with ChatGPT to ensure that students are using it productively and that its use does not lead to the substitution of an understanding of AI-assisted language learning with the use of chatbot capabilities to enhance students' performance in their own language acquisition [10].

6. Conclusion

The development of a computer vision-language model for student learning is a powerful tool for improving the educational process. This model can analyze and understand the learning context, generating accurate and grammatically correct answers, and adapting its parameters according to individual student needs. However, it should be noted that developing and optimizing such a model is a challenging task that requires large computational power and voluminous training data. Also, this is only a general concept of developing a vision-language model based on transformer architecture for student learning. Such projects require deep knowledge in natural language processing, machine learning and neural networks [10].

Author Contributions: Conceptualization, Shchetinin E.; methodology, Shchetinin E.; software, Shchetinin E.; writing—review and editing Demidova A.; Demidova A.; supervision, Demidova A.; project administration, Glushkova A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Devlin, J., Chang, M., Lee, K. & K., T. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* 2018.
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. *Attention is All you Need* in *Advances in Neural Information Processing Systems* (eds Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. & Garnett, R.) **30** (Curran Associates, Inc., 2017), 5998–6008.
3. Liu Y. andOtt, M., Goyal N. andDu, J., Joshi, M., Chen, D., Levy, O., Lewis M. andZettlemoyer, L. & V., S. *RoBERTa: A Robustly Optimized BERT Pretraining Approach* 2019.
4. Clark, E. & Gardner, M. *Simple and Effective Multi-Paragraph Reading Comprehension* 2018.
5. Klein, G., Kim, Y., Deng, Y., Senellart, J. & Rush, A. *OpenNMT: Open-Source Toolkit for Neural Machine Translation* in *Proceedings of ACL 2017, System Demonstrations* (eds Bansal, M. & Ji, H.) **28** (Association for Computational Linguistics, Vancouver, Canada, July 2017), 67–72. doi:10.18653/V1/P17-4012.
6. Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. *Improving language understanding by generative pre-training* https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
7. Nogueira, R. & Cho, K. *Passage Re-ranking with BERT* 2019.
8. Schröder, S., Niekler, A. & Potthast, M. *Revisiting Uncertainty-based Query Strategies for Active Learning with Transformers* 2021.

9. Yang, F., Wang, X., Ma, H. & Li, J. Transformers-sklearn: a toolkit for medical language understanding with transformer-based models. *BMC Medical Informatics and Decision Making* **21,** 141–157. doi:10.1186/s12911-021-01459-0 (2021).
10. Rashid, M., Höhne, J., Schmitz, G. & Müller-Putz, G. A Review of Humanoid Robots Controlled by Brain-Computer Interfaces. *Frontiers in Neurorobotics,* 1–28 (2020).

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: 0000-0003-3651-7629, ResearcherID: O-8287-2017, Scopus Author ID: 16408533100)

Glushkova, Anastasia G.—researcher (e-mail: aglushkova@endeavorco.com, ORCID: 0000-0002-8285-0847, Scopus Author ID: 57485591900)

Demidova, Anastasia V.—Candidate of Physical and Mathematical Sciences, Assistant professor of Department of Probability Theory and Cyber Security (e-mail: demidova-av@rudn.ru, ORCID: 0000-0003-1000-9650)

UDC 519.6
PACS 07.05.Tp,
DOI: 10.22363/2658-4670-2024-32-2-234–241 EDN: CUCXTY

Разработка компьютерной системы обучения студентов на основе визуально-лингвистических моделей

Е. Ю. Щетинин[1], А. Г. Глушкова[2], А. В. Демидова[3]

[1] *Финансовый университет при Правительстве Российской Федерации, Ленинградский пр-т, д. 49, Москва, 125993, Российская Федерация*
[2] *Эндевор, ш. Чизвик, д. 566, Чизвик Парк, Лондон W4 5HR, Великобритания*
[3] *Российский университет дружбы народов, ул. Миклухо-Маклая, д. 6, Москва, 117198, Российская Федерация*

Аннотация. В последние годы методы искусственного интеллекта получили большое развитие в различных областях, в частности в образовании. Разработка компьютерных систем для обучения студентов является важной задачей и может значительно улучшить процесс обучения студентов. Разработка и внедрение методов глубокого обучения в образовательный процесс приобрели огромную популярность. Наиболее успешными среди них являются модели, учитывающие мультимодальный характер информации, в частности сочетание текста, звука, изображений и видео. Сложность обработки таких данных состоит в том, что объединение мультимодальных входных данных различными методами конкатенации каналов, игнорирующих неоднородность разных модальностей, является неэффективным подходом. Для решения этой проблемы в работе предложен междуканальный модуль внимания. В статье представлена компьютерная визуально-лингвистическая система процесса обучения студентов, основанная на объединении мультимодальных входных данных с использованием междуканального модуля внимания. Показано, что создание эффективных и гибких систем и технологий обучения на основе таких моделей позволяет адаптировать образовательный процесс к индивидуальным потребностям обучающихся и повысить его эффективность.

Ключевые слова: глубокое обучение, модель визуально-лингвистического обучения, нейронные сети-трансформеры, модуль сквозного внимания