ISSN 2500-316X (Online)

RUSSIAN TECHNOLOGICAL JOURNAL

РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ ЖУРНАЛ

TÜV International Certification
TÜV THÜRINGEN CERT
ISO 9001:2008
15 100 96652
TIC

Information systems.
Comruter sciences.
Issues of information security

Multiple robots (robotic centers) and systems.
Remote sensing and non-destructive testing

Modern radio engineering and telecommunication systems

Micro- and nanoelectronics.
Condensed matter physics

Analytical instrument engineering and technology

Mathematical modeling

Economics of knowledge-intensive and high-tech enterprises and industries.
Management in organizational systems

Product quality management. Standardization

Philosophical foundations of technology and society

12+
www.rtj-mirea.ru

МИРЭА
Российский технологический университет

9(6) 2021

ISSN 2500-316X (Online)

RUSSIAN TECHNOLOGICAL JOURNAL

РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ ЖУРНАЛ

- Information systems. Computer sciences. Issues of information security
- Multiple robots (robotic centers) and systems. Remote sensing and non-destructive testing
- Modern radio engineering and telecommunication systems
- Micro- and nanoelectronics. Condensed matter physics
- Analytical instrument engineering and technology
- Mathematical modeling
- Economics of knowledge-intensive and high-tech enterprises and industries. Management in organizational systems
- Product quality management. Standardization
- Philosophical foundations of technology and society

- Информационные системы. Информатика. Проблемы информационной безопасности
- Роботизированные комплексы и системы. Технологии дистанционного зондирования и неразрушающего контроля
- Современные радиотехнические и телекоммуникационные системы
- Микро- и наноэлектроника. Физика конденсированного состояния
- Аналитическое приборостроение и технологии
- Математическое моделирование
- Экономика наукоемких и высокотехнологичных предприятий и производств. Управление в организационных системах
- Управление качеством продукции. Стандартизация
- Мировоззренческие основы технологии и общества

Russian Technological Journal
2021, Vol. 9, No. 6

Russian Technological Journal
2021, том 9, № 6

https://www.rtj-mirea.ru

МИРЭА
Российский технологический университет

https://doi.org/10.32362/2500-316X

ISSN 2500-316X (Online)

Russian Technological Journal
2021, Vol. 9, No. 6

Publication date November 30, 2021.

The peer-reviewed scientific and technical journal highlights the issues of complex development of radio engineering, telecommunication and information systems, electronics and informatics, as well as the results of fundamental and applied interdisciplinary researches, technological and economical developments aimed at the development and improvement of the modern technological base.

Periodicity: bimonthly.
The journal was founded in December 2013. The titles were "Herald of MSTU MIREA" until 2016 (ISSN 2313-5026) and "Rossiiskii tekhnologicheskii zhurnal" from January 2016 until July 2021 (ISSN 2500-316X).

Founder and Publisher:
Federal State Budget
Educational Institution
of Higher Education
"MIREA – Russian Technological University"
78, Vernadskogo pr., Moscow, 119454 Russia.

The journal is included into the List of peer-reviewed science press of the State Commission for Academic Degrees and Titles of Russian Federation.
The Journal is included in RSL (Russian State Library), Russian citation index, elibrary, Socionet, Directory of Open Access Journals (DOAJ), Directory of Open Access Scholarly Resources (ROAD), Google Scholar, Open Archives Initiative, Ulrich's International Periodicals Directory.

Registration Certificate ПИ № ФС 77 - 81733, issued in August 19, 2021 by the Federal Service for Supervision of Communications, Information Technology, and Mass Media of Russia.

Russian Technological Journal
2021, том 9, № 6

Дата опубликования 30 ноября 2021 г.

Научно-технический рецензируемый журнал освещает вопросы комплексного развития радиотехнических, телекоммуникационных и информационных систем, электроники и информатики, а также результаты фундаментальных и прикладных междисциплинарных исследований, технологических и организационно-экономических разработок, направленных на развитие и совершенствование современной технологической базы.

Периодичность: один раз в два месяца.
Журнал основан в декабре 2013 года. До 2016 г. издавался под названием «Вестник МГТУ МИРЭА» (ISSN 2313-5026), а с января 2016 г. по июль 2021 г. под названием «Российский технологический журнал» (ISSN 2500-316X).

Учредитель и издатель:
федеральное государственное бюджетное
образовательное учреждение
высшего образования
«МИРЭА – Российский технологический университет»
119454, РФ, г. Москва, пр-т Вернадского, д. 78.

Журнал входит в Перечень ведущих рецензируемых научных журналов ВАК РФ, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук и доктора наук, индексируется в РГБ, РИНЦ, elibrary, Соционет, Directory of Open Access Journals (DOAJ), Directory of Open Access Scholarly Resources (ROAD), Google Scholar, Open Archives Initiative, Ulrich's International Periodicals Directory.

Свидетельство о регистрации СМИ ПИ № ФС 77 - 81733 от 19.08.2021 г. выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор).

https://www.rtj-mirea.ru

Editorial Board

Редакционная коллегия

Contents

Содержание

Russian Technological Journal. 2021;9(6):7–15

ISSN 2500-316X (Online)

Information systems. Computer sciences. Issues of information security

Информационные системы. Информатика. Проблемы информационной безопасности

UDC 004.051
https://doi.org/10.32362/2500-316X-2021-9-6-7-15

REVIEW ARTICLE

Comparative analysis of software optimization methods in context of branch predication on GPUs

Igor Yu. Sesin @,
Roman G. Bolbakov

MIREA – Russian Technological University, Moscow, 119454 Russia
@ Corresponding author, e-mail: isesin@protonmail.com

Abstract. General Purpose computing for Graphical Processing Units (GPGPU) technology is a powerful tool for offloading parallel data processing tasks to Graphical Processing Units (GPUs). This technology finds its use in variety of domains—from science and commerce to hobbyists. GPU-run general-purpose programs will inevitably run into performance issues stemming from code branch predication. Code predication is a GPU feature that makes both conditional branches execute, masking the results of incorrect branch. This leads to considerable performance losses for GPU programs that have large amounts of code hidden away behind conditional operators. This paper focuses on the analysis of existing approaches to improving software performance in the context of relieving the aforementioned performance loss. Description of said approaches is provided, along with their upsides, downsides and extents of their applicability and whether they address the outlined problem. Covered approaches include: optimizing compilers, JIT-compilation, branch predictor, speculative execution, adaptive optimization, run-time algorithm specialization, profile-guided optimization. It is shown that the aforementioned methods are mostly catered to CPU-specific issues and are generally not applicable, as far as branch-predication performance loss is concerned. Lastly, we outline the need for a separate performance improving approach, addressing specifics of branch predication and GPGPU workflow.

Keywords: general-purpose computing for graphical processing units, optimizing compilers, predication

• **Submitted:** 22.03.2021 • **Revised:** 26.05.2021 • **Accepted:** 25.07.2021

For citation: Sesin I.Yu., Bolbakov R.G. Comparative analysis of software optimization methods in context of branch predication on GPUs. *Russ. Technol. J.* 2021;9(6):7–15. https://doi.org/10.32362/2500-316X-2021-9-6-7-15

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

ОБЗОР

Сравнительный анализ методов оптимизации программного обеспечения для борьбы с предикацией ветвлений на графических процессорах

И.Ю. Сесин @,
Р.Г. Болбаков

МИРЭА – Российский технологический университет, Москва, 119454 Россия
@ Автор для переписки, e-mail: isesin@protonmail.com

Резюме. Технология GPGPU (General Purpose computing for Graphical Processing Units – расчеты общего назначения на графических процессорах) является мощным инструментом для переноса задач параллельной обработки информации на GPU (Graphical Processing Unit – графический процессор). Эта технология находит применение практически в любой области, требующей проведения массы параллельных расчетов, и применяется как в научной и коммерческой, так и в любительской среде. Разработчики программ общего назначения, запускаемых на GPU, неизбежно сталкиваются с падением производительности ввиду предикации ветвления кода. В условиях предикации ветвления исполняются обе ветви условного оператора вне зависимости от истинности условия, но посредством маскирования выполняемых инструкций программа учитывает только результат работы верной ветви. Из-за этого программы общего назначения, имеющие большие участки кода, скрытые за условными операторами, становятся существенно менее производительными на графических процессорах. В статье рассматриваются существующие в предметной области методы и подходы к увеличению производительности программного обеспечения в рамках их применимости к решению проблемы падения производительности при предикации. Приводится описание методов, их сильных и слабых сторон, а также рамок их применимости, на базе чего делается заключение о возможности их использования на GPU. В число рассмотренных методов и подходов вошли следующие: оптимизирующие компиляторы, JIT-компиляция, предсказатель переходов, спекулятивное исполнение, адаптивная оптимизация, специализация алгоритма во время исполнения, оптимизация на основе профилирования. Показано, что указанные аппаратные и программные подходы к увеличению производительности программного обеспечения преимущественно ориентированы на решение проблем специфичных для CPU (Central Processing Unit – центральный процессор) и в целом неприменимы для разрешения потерь производительности при предикации на GPU. Указывается на необходимость создания отдельного подхода, ориентированного именно на решение проблемы предикации ветвления на GPU.

Ключевые слова: расчеты общего назначения на графических процессорах, оптимизирующие компиляторы, предикация

• **Поступила:** 22.03.2021 • **Доработана:** 26.05.2021 • **Принята к опубликованию:** 25.07.2021

Для цитирования: Сесин И.Ю., Болбаков Р.Г. Сравнительный анализ методов оптимизации программного обеспечения для борьбы с предикацией ветвлений на графических процессорах. *Russ. Technol. J.* 2021;9(6):7–15. https://doi.org/10.32362/2500-316X-2021-9-6-7-15

Прозрачность финансовой деятельности: Никто из авторов не имеет финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Graphics processors, or Graphics Processing Unit (GPU), are specialized hardware that performs the processing of graphical information. Unlike a central processing unit (CPU), graphics processors are specialized for parallel processing of large amounts of data, which causes certain differences in their architectural design:

- CPUs have a small number of physical cores (from 1 to 32), GPUs have orders of magnitude more physical cores (can be hundreds or thousands[1, 2], depending on the compromise between the number of cores and their power chosen by the hardware manufacturer);
- cores in the CPU have additional hardware functionality, such as their own caches, instruction pipelines, and branch prediction modules. GPU kernels are very simple arithmetic kernels focused on fast processing of floating-point numbers. A number of modern GPUs also include specialized cores for ray tracing and tensor computing [1, 2];
- CPU is separated from random access memory; GPUs have direct access to video memory located on the same board;
- GPU cores cannot write data in the memory area allocated for the executable code;
- CPU falls into the Multiple Instruction, Multiple Data (MIMD) classification, GPU falls into the Single Instruction, Multiple Data (SIMD) classification according to Flynn [3].

Using General Purpose Computing on Graphics Processing Units (GPGPU) technology, it is possible to run programs on GPUs other than highly specialized shader programs. This technology finds application in many areas—from mining crypto currencies to calculating the protein folding[3].

Applying the GPGPU technology, one has to deal with certain peculiarities of the construction of programs and their behavior in the conditions of execution on a graphics processor. For example, in the course of writing a GPGPU program, the authors of this article noticed a downward trend in its performance as more functionality was added. At the same time, the performance losses were much more significant than could be assumed based on the complexity of the algorithm of the added functionality.

When looking into the above problem, it was found that this happens due to the specifics of the SIMD architecture of graphics processors: when the program executes a conditional operator, both branches will be executed, but the operations of the wrong branch will not be applied. This feature is called predication [4, p. 168] of branches of execution, and it is needed, first of all, to replace the dependence on the flow of execution with the dependence on data. The very need for such measures is justified by the fact that most of the hardware part of the GPU architecturally falls into the SIMD class of Flynn's taxonomy [3], and individual modules of the system cannot have their own threads of execution and, accordingly, cannot follow along various branches of the conditional operator.

The degree of influence of the above feature on the program differs from program to program and correlates with the number of possible settings. So, if the program is initially designed to solve one maximally specific task without the possibility of customization (for example, calculating a certain hash function), then the influence of predication will be minimized.

Programs that perform more general tasks, and, therefore, have a large list of plug-in or optional functional that will be used in the process of the program only with certain input data, are more susceptible to performance degradation.

Let us take a 3D scene renderer as an example of such a program. Rendering in the field of computer graphics is a term used to describe the process of obtaining an image from certain data characterizing the objects of the displayed scene. A program that implements such a process is called a renderer. Professional renderers can include thousands of different options that change the behavior of the program—from adjusting the angle of the camera's field of view to the detailed configuration of the bidirectional reflectance distribution function for each surface.

Taking a renderer program as an example, imagine a situation in which a certain scene is rendered with a set of primitives that can only scatter or emit light, but not reflect it. In this case, the code responsible for calculating the reflections should not be called, but its very presence will slow down the program due to predication.

To enable or disable the desired program behavior, in the vast majority of cases, one would use conditional statements that depend on some input data of the program, regardless of whether it just data or configuration settings.

As program's functionality grows, the number of conditional statements will increase, and accordingly,

[1] Advanced Micro Devices, Inc, Graphics Specifications, 2021. URL: https://www.amd.com/en/products/specifications/graphics. Accessed March 1, 2021.

[2] NVIDIA. Comparison of specifications of RTX 30 video cards. URL: https://www.nvidia.com/ru-ru/geforce/graphics-cards/30-series/compare/?section=compare-specs. Accessed March 1, 2021.

[3] Houston M. General Purpose Computation on Graphics Processors (GPGPU). ATI HD 2000 Series. Launch, Tunis, Tunisia; 2007. URL: https://graphics.stanford.edu/~mhouston/public_talks/R520-mhouston.pdf. Accessed March 1, 2021.

the costs spent by the GPU on processing conditional program branches will also increase, which, in turn, leads to a certain ceiling of program complexity—after a certain critical mass of branches, the program will slow down so much that will trigger the internal protection of the driver (if any) and abort the execution.

It should be noted that from a programmer's point of view, the mere presence of additional code slows down the program, despite the fact that the code should not be executed. This behavior is highly uncommon for programs running on the CPU; Moreover, the correct use of conditional jumps is often the key to writing more efficient programs, thanks to the branch predicator built into the cores of modern CPUs.

RESEARCH OF EXISTING SOFTWARE AND HARDWARE FOR INCREASING PERFORMANCE

Let us consider the existing technologies for optimizing programs and increasing performance, assessing their applicability for solving the described problem.

It should be noted that within this article there will be no attempts to numerically compare the methods under consideration. This is due to the fact that almost all methods of increasing software performance are based on a certain characteristic of the optimized program, hardware platform or programming language. Techniques that work for some programs may be useless or even harmful to others. The quality of implementation of a particular method also has a direct impact on the result obtained, and the same approach, implemented in different ways, can give strikingly different results. The specific metric of the success of the application of methods depends on many factors, both quantitative and qualitative, which cannot be excluded from the study without jeopardizing the reproducibility of the study itself. On the other hand, the inclusion of these factors will narrow the study down to comparing particular cases, namely, comparing specific programs on specific hardware in a specific configuration, which is not representative for describing the overall picture.

In view of the above, the authors analyze mainly the qualitative characteristics to determine applicability of methods in lieu of software performance improvement.

Optimizing compilers continue to play a leading role in improving software performance. In general, an optimizing compiler is any compiler that performs special operations on compiled code to improve its performance.

Initially, this meant replacing certain operations with equivalent, but more efficient ones, such as replacing multiplication or integer division by numbers that are powers of two, with bit shift operations, but with the development of the scientific field, software development methodologies, and hardware capabilities, compilers have acquired an extensive arsenal of optimization techniques.

They are usually divided into low-level and high-level optimizations.

Low-level optimizations involve changing the generated machine code to make the best use of hardware platform features. This includes the use of more efficient machine code constructs, including the use of special commands available in the target architecture, vectorization of operations, function inlining, etc.

High-level optimizations operate at the level of abstract algorithmic elements that make up a program—loops, branches, and basic blocks. They use data about the structure of the program to transform the intermediate representation of the program into a more efficient form.

In practice, the line between these groups is blurred, since many methods involve elements of both high-level code analysis and low-level control over the generation of machine instructions.

Optimizations widely used by compilers include:

- constant folding—if a certain expression consists only of constants, then it is calculated at the compilation stage and its result is substituted instead of the original expression;
- eliminating the "dead code" [5], i.e., code sections that cannot be reached by the program;
- eliminating the dead stores [6], i.e., removal of operations storing value in the variable that goes unread further in the code;
- optimizing the register allocation [7] reorganizes the code in such a way as to minimize the number of memory accesses during the program operation by keeping the most frequently used variables in certain general-purpose registers of the processor;
- operation parallelization—changing the order of operations so that they can be run in parallel at the level of threads, memory or instructions;
- strength reduction—replacing slow operations with equivalent, but faster ones on the target architecture;
- loop optimization—a wide group of methods focused on working with loops, including such approaches as moving invariants out of the loop, inversion of loops, loop unrolling, dividing and merging loop bodies, removing conditional statements from the loop, etc.;
- instruction selection [8] allows the compiler to pick the most efficient machine code combination for the target processor;

- Instruction Scheduling [9, 10]—reorganization of instructions so as not to cause downtime of the central processor pipeline as a result of long memory accesses, exhaustion of processor resources, or branching.

Most of the traditional methods of program optimization used in optimizing compilers can be excluded from consideration, since they are focused on CPU features that are not inherent in arithmetic cores used in graphics processors. The methods that can be applied are either already implemented by compilers of programs for GPUs, or do not solve the presented problem.

Let us consider more dynamic methods for increasing the performance.

JIT-compilation technology [11] (just-in-time) improves the performance of programs in languages that are compiled into the bytecode. The bytecode is called an intermediate representation of the program code [12–14], which is executed by a virtual machine, which is its key difference from machine code executed directly by the processor.

In the framework of the JIT-compilation technology, the bytecode of programs is compiled into machine code as needed, right at the time of the program execution[4]. This allows you to speed up the work of programs in several ways at once:

- the startup delay, which without this technology is caused by lengthy processing of the source code, is significantly reduced; compilation from source code to bytecode is significantly slower than compilation from the bytecode to machine code;
- using the features of specific hardware to improve the performance;
- optimizing the program using data obtained during program operation;
- the ability to dynamically link libraries without the overhead inherent in compiled languages.

However, the use of the JIT compilation within programs running on a GPU currently does not appear to be feasible due to architectural restrictions on dynamically changing the executable machine code on GPUs.

In addition, embedding compilation into the operation cycle of the GPU itself will require a lot of both hardware and software changes, and, in general, will not give the same performance gain as on the central processor; it is because there is no problem of connecting libraries on the GPU (due to the static compilation of programs), and the intermediate representation of the program will be fully compiled into machine code in the process of transferring the program to the GPU. In general, these problems make the use of JIT impractical on the GPU.

One of the key elements ensuring the performance of modern CPUs is the branch predictor[5]. A branch predictor allows a pipelined processor to begin loading instructions from one of the branches of a branching statement into the pipeline before a condition is determined to be true.

This plays an important role in improving performance due to the parallel execution of instructions operating on mutually independent data. In the general case, without a predictor, the processor pipeline will not know which instructions from which branch should be used, and will be able to start loading them only after evaluating the truth of the condition in the conditional statement, which would entail pipeline stalling for each of the conditions. This can be avoided by predicting the result of the condition and starting loading instructions from the corresponding branch into the pipeline in advance. If the predictor fails, the pipeline will idle until the correct instructions are loaded.

The branch prediction itself is carried out heuristically [15−17], commonly, based on the statistics of the execution of a given section of the code. The time cost of a predictor error is inevitable due to imperfect prediction mechanisms, but with a sufficiently accurate prediction mechanism, the time loss from incorrect predictions becomes insignificant compared to the gain from correct predictions.

The branch prediction approach was developed in the form of speculative execution[6]—in addition to loading instructions onto the pipeline; the predicted branch is also executed before the condition is established. In case of an error, time losses increase, because you need to flush the entire pipeline and load the correct branch instructions into it.

The branch prediction and speculative execution approaches are implemented in hardware and are completely transparent to the programmer. Nevertheless, with certain program constructions, it is possible to get performance degradation, which is most noticeable in the case when the program consistently forces the module to mispredict the branch. The existing practice of optimizing low-level programs for

[4] Croce L. Just in Time Compilation. Columbia University. URL: http://www.cs.columbia.edu/~aho/cs6998/Lectures/14-09-22_Croce_JIT.pdf. Accessed March 1, 2021.

[5] Fog A. The microarchitecture of Intel, AMD and VIA CPUs. An optimization guide for assembly programmers and compiler makers. Technical University of Denmark. URL: https://www.agner.org/optimize/microarchitecture.pdf. Accessed March 1, 2021.

[6] Gabbay F. Speculative execution based on value prediction. Research Proposal towards the Degree of Doctor of Sciences. Technion-Israel Institute of Technology (IIT), Department of Electrical Engineering. 1996. 65 p. URL: http://citeseerx.ist.psu.edu/viewdoc/download?-doi=10.1.1.29.5397&rep=rep1&type=pdf. Accessed March 1, 2021.

better interaction with the branch predictor is a direct consequence of the imperfection of the used prediction methods.

The very presence of a branch prediction and speculative execution mechanism has also opened the way for a whole class of hardware vulnerabilities that allow a program to gain unauthorized access to information. A specially designed program can determine the contents of memory cells using side channels (typically access latency) arising from erroneous branch predictions [18–20].

The authors of this article believe that predicting transitions to the GPU does not make sense due to the absence of a pipeline on its cores, as well as the absence of branches in the executable machine code in their typical understanding due to predication. Instead, the code is structured linearly and no conditional statements are executed when executing conditional statements' "jumps" of the instruction counter register.

Even if we assume that this obstacle will be overcome, implementing the prediction module into the GPU would be extremely difficult due to the complexity of such a module. A typical GPU includes a large number of weak arithmetic cores, orders of magnitude larger than the number of cores on a typical market CPU. Adding a predictor to each of them will increase their size and power consumption to unacceptable levels.

The adaptive optimization method [21] relies on a special toolkit, similar to JIT compilers, to recompile the program code while it is running. This method applies different optimization techniques to a program depending on the productivity of the code. For example, adaptive optimization will apply more aggressive optimization techniques in bottlenecks where the code is spending the most CPU time. Such aggressive optimization techniques are time-consuming and wasteful to apply to the entire program. Therefore, sections of the program that are rarely called will be optimized with more primitive, but faster methods, allowing you to save time in general, albeit at the expense of the fact that this code will run slower.

The application of the above approach to GPUs is impossible for a reason similar to JIT; for it to work, the program requires the ability to rewrite its code, which is not supported by modern GPUs.

Let us consider the run-time algorithm specialization, an approach that comes from the field of automatic theorem proving [22]. This approach implies the creation of specific implementations of complex functions for certain inputs, for which it is possible to represent the original function in a simpler form. This can sometimes translate into creating table values for the function and caching individual results for frequently repeated values. In more complex cases, several code variants can be generated corresponding to particular variants of the function, where, for example, part of the calculations is excluded due to specific values of the input data.

The aforementioned approach and methods similar to it [23] require not only writing to the memory area of the executable code, but also a large amount of additional memory for the code of the obtained specializations, which makes its application within the GPGPU technology impossible.

Let us consider now the profile-guided optimization technology. It is an approach to program optimization in which the optimization process is controlled by the performance results of the program [24, 25]. This approach often requires the use of a special compiler, which takes on the task of instrumentation and taking measurements of the program execution time. Such a special compiler translates, builds, and runs the program many times, analyzing during these test runs the frequency of use of various sections of the program and the speed of their invocation. Using this data, the compiler applies various optimization strategies to code sections that showed poor results during test runs. These strategies include selecting the optimal register allocation, function embedding, as well as a lot of techniques tied to the successful grouping of executable code in memory pages to speed up the work of caches at several levels at once—from the mechanism for managing virtual memory at the operating system level to the CPU instruction cache.

Due to the fact that the statistics collected by such a compiler is representative only of those actions that were performed by the program during the testing process, a typical set of tests should include the most common scenarios for working with the program. It could be much easier implemented for GPU programs, due to their lack of interactivity.

The authors of the article believe that the use of this method for GPU programs is impractical, since in the conditions of working with a GPU it is extremely difficult or even impossible to instrument the code and, accordingly, to obtain measurements of the performance of individual sections of the code necessary for the method to work.

From the analysis of existing technologies for increasing performance, it is clear that the problem of loss of performance with increasing functionality, which is typical for graphics processors, is atypical and non-trivial.

None of the approaches discussed solve the problem of branch predication on GPUs, but they are an important starting point in the process of finding a solution.

On the one hand, to solve the problem, the use of dynamic methods is required to maintain the flexibility of the program functionality for the GPU, but on the other hand, the considered existing solutions applied on the CPU cannot be used in the context of GPU in practice due to the particular features of the architecture focused on massively parallel calculations. Changes to the established GPU architecture are obviously impractical due to a number of factors, including a multiple increase in production costs and an inevitable decrease in performance.

By virtue of the stated provisions, it can be established that there is a need to find an approach to optimizing programs for GPU that is compatible with the limitations of graphic processors, but at the same time flexible enough to ensure the preservation of the program's functionality.

CONCLUSIONS

The article discussed the problem of decreasing the performance of programs for general-purpose calculations on GPU, which arises in the course of increasing their functionality. The connection of this problem with branching predication—an essential feature of the organization of the hardware platform of graphics processors—was established.

A number of existing approaches and technologies for increasing the performance of programs on the CPU were considered and the low degree of their applicability to programs using the GPGPU technology was shown.

It was noted that the existing methods, even though their direct application is impossible, are an important starting point for further research.

Further development of the problems outlined in the article implies the development of a specialized method for optimizing programs for GPU.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Markidis S., Chien S.W.D., Laure E., Peng I.B., Vetter J.S. NVIDIA Tensor Core Programmability, Performance & Precision. In: *2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. Vancouver, BC, Canada; 2018, p. 522–531. https://doi.org/10.1109/IPDPSW.2018.00091
2. Sanzharov V.V., Gorbonosov A.I., Frolov V.A., Voloboy A.G. Examination of the Nvidia RTX. *CEUR Workshop Proceedings*. 2019;2485:7–12. http://dx.doi.org/10.30987/graphicon-2019-2-7-12
3. Flynn M.J. Very high speed computing systems. *Proceedings of the IEEE*. 1966;54(12):1901–1909. https://doi.org/10.1109/PROC.1966.5273
4. Fisher J.A., Faraboschi P., Young C. *Embedded computing: A VLIW approach to architecture, compilers, and tools*. Elsevier; 2004. ISBN: 978-1-55860-766-8. URL: https://www.researchgate.net/publication/220690439_Embedded_computing_a_VLIW_approach_to_architecture_compilers_and_tools
5. Knoop J., Rüthing O., Steffen B. Partial dead code elimination. In: *Proceedings of the ACM SIGPLAN 1994 conference on Programming language design and implementation (PLDI '94)*. 1994, p. 147–158. https://doi.org/10.1145/178243.178256
6. Fink S., Knobe K., Sarkar V. Unified analysis of array and object references in strongly typed languages. In: Palsberg J. (Ed.). *Static Analysis. SAS 2000. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer. 2000. V. 1824. P. 155–174. https://doi.org/10.1007/978-3-540-45099-3_9
7. Runeson J., Nyström S.-O. Retargetable graph-coloring register allocation for irregular architectures. In: Krall A. (Ed.). *Software and Compilers for Embedded Systems (SCOPES 2003). Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer. 2003. V. 2826. P. 240–254. https://doi.org/10.1007/978-3-540-39920-9_17
8. Blindell G.H. *Instruction Selection: Principles, Methods, & Applications*. Springer; 2016. 171 p. ISBN 978-3-319-34017-3. http://dx.doi.org/10.1007/978-3-319-34019-7
9. Gibbons P.B., Muchnick S.S. Efficient instruction scheduling for a pipelined architecture. *ACM SIGPLAN Notices*. 1986;21(7):11–16. https://doi.org/10.1145/13310.13312
10. Su Ch.-L., Tsui Ch.-Y., Despain A.M. Low power architecture design and compilation techniques for high-performance processors. In: *Proceedings of COMPCON '94*. 1994, p. 489–498. https://doi.org/10.1109/CMPCON.1994.282878
11. Aycock J. A brief history of just-in-time. *ACM Comput. Surv*. 2003;35(2):97–113. https://doi.org/10.1145/857076.857077
12. Ogihara M. *Fundamentals of Java Programming*. Springer; 2018. 532 p.
13. Sage K. The Origins of Programming. In: *Concise Guide to Object-Oriented Programming. Undergraduate Topics in Computer Science*. Springer, Cham.; 2019, p. 1–9. https://doi.org/10.1007/978-3-030-13304-7_1
14. Saabith A.S., Fareez M.M.M., Vinothraj T. Python current trend applications-an overview. *IJAERD*. 2019;6(10):6–12. URL: http://ijaerd.com/papers/finished_papers/IJAERDV06I1085481.pdf
15. McFarling S. *Combining Branch Predictors*. Digital Western Research Lab (WRL). Technical Report, TN-36. 1993. 29 p. URL: https://www.hpl.hp.com/techreports/Compaq-DEC/WRL-TN-36.pdf

16. Skadron K., Martonosi M., Clark D.W. A Taxonomy of branch mispredictions, and alloyed prediction as a robust solution to wrong-history mispredictions. In: *Proceedings of the 2000 International Conference on Parallel Architectures and Compilation Techniques*. Philadelphia. 2000. https://doi.org/10.1109/PACT.2000.888344
17. Vintan L.N., Iridon M. Towards a high performance neural branch predictor. In: *IJCNN'99. International Joint Conference on Neural Networks Proceedings*. 1999. https://doi.org/10.1109/IJCNN.1999.831066
18. Kocher P., Horn J., Fogh A., Genkin D., et al. Spectre attacks: Exploiting speculative execution. In: *2019 IEEE Symposium on Security and Privacy (SP)*. 2019. 19 p. https://doi.org/10.1109/SP.2019.00002
19. Bhattacharyya A., Sandulescu A., Neugschwandtner M., Sorniotti A., et al. SMoTherSpectre: exploiting speculative execution through port contention. In: *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19)*. 2019, p. 785−800. https://doi.org/10.1145/3319535.3363194
20. Chen G., Chen S., Xiao Y., Zhang Y., Lin Z., Lai T.H. SgxPectre: Stealing intel secrets from SGX enclaves via speculative execution. In: *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE. 2019, p. 142−157. https://doi.org/10.1109/EuroSP.2019.00020
21. Arnold M., Fink S., Grove D., Hind M., Sweeney P.F. Adaptive optimization in the Jalapeno JVM. In: *Proceedings of the 15th ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications*. 2000, p. 47–65. https://doi.org/10.1145/353171.353175
22. Riazanov A. *Implementing an Efficient Theorem Prover*. PhD thesis. The University of Manchester; 2003. 210 p. URL: https://www.researchgate.net/publication/2906405_Implementing_an_Efficient_Theorem_Prover
23. Grant B., Mock M., Philipose M., Chambers C., Eggers S.J. Annotation-directed run-time specialization in C. *ACM SIGPLAN Not.* 1997;32(12):163−178. https://doi.org/10.1145/258994.259016
24. Pettis K., Hansen R.C. Profile guided code positioning. In: *Proceedings of the ACM SIGPLAN 1990 conference on Programming language design and implementation (PLDI '90)*. 1990, p. 16−27. https://doi.org/10.1145/93542.93550
25. Wicht B., Vitillo R.A, Chen D., Levinthal D. *Hardware Counted Profile-Guided Optimization*. 24 November 2014. URL: https://arxiv.org/pdf/1411.6361.pdf

About the authors

Igor Yu. Sesin, Postgraduate Student, Department of the Tool and Applied Software, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: isesin@protonmail.com. https://orcid.org/0000-0002-7323-9595

Roman G. Bolbakov, Cand. Sci. (Eng.), Associate Professor, Head of the Department of the Tool and Applied Software, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: bolbakov@mirea.ru. https://orcid.org/0000-0002-4922-7260

Об авторах

Сесин Игорь Юрьевич, аспирант, кафедра инструментального и прикладного программного обеспечения Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: isesin@protonmail.com. https://orcid.org/0000-0002-7323-9595

Болбаков Роман Геннадьевич, к.т.н., доцент, заведующий кафедрой инструментального и прикладного программного обеспечения Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: bolbakov@mirea.ru. https://orcid.org/0000-0002-4922-7260

Translated by E. Shklovskii

ISSN 2500-316X (Online)

UDC 004.9
https://doi.org/10.32362/2500-316X-2021-9-6-16-25

RESEARCH ARTICLE

The development of models of an analytical data processing system for monitoring information security of an informatization object using cloud infrastructure

Valerii A. Sizov,
Aleksey D. Kirov @

Plekhanov Russian University of Economics, Moscow, 117997 Russia
@ *Corresponding author, e-mail: kirov.ad@rea.ru*

Abstract. The article is devoted to the problem of developing an analytical data processing system (ADPS) for monitoring information security within the information security management system of modern companies conducting their main activities in cyberspace and using cloud infrastructure. Based on the analysis of modern information technologies related to ensuring information security of cloud infrastructure and the most popular products for ensuring information security of cloud infrastructures, as well as existing scientific approaches, a formalized approach to the synthesis of an ADPS for monitoring the information security of an informatization object using cloud infrastructure is proposed. This approach takes into account the usefulness of the used information technologies from the viewpoint of information security. A general model of the structure of information support of an analytical data processing system for monitoring information security, as well as a model of the dependence of the usefulness of information technology on time and the ratio of the skill level of an information security specialist and an attacker are presented. The quality of the information security monitoring system is used as a criterion in the first optimization model. The following limitations are suggested: limitation on the time of making a decision on an incident; limitation on the degree of quality of analysis of information security events by the analytical data processing system and limitation on the compatibility of data analysis functions with data types about information security events. The cited results of the study of the second model show a logically consistent dependence of the usefulness of information technology on time and the ratio of the skill level of an information security specialist to the skill level of an attacker. The particular models of the structure of the information support of ADPS are presented. They make it possible to determine the rational structure information support of ADPS according to particular criteria. The following particular criteria are used: the maximin criterion of the usefulness of the information support of ADPS for monitoring the information security of an informatization object in the cloud infrastructure; the criterion for the maximum relevance of information support distributed over the nodes of the cloud infrastructure for systems with a low degree of centralization of management.

Keywords: information security, analytical data processing system, cloud infrastructure, computer modeling, mathematical modeling, information security monitoring, object of informatization

• **Submitted:** 25.03.2021 • **Revised:** 11.05.2021 • **Accepted:** 25.07.2021

For citation: Sizov V.A., Kirov A.D. The development of models of an analytical data processing system for monitoring information security of an informatization object using cloud infrastructure. *Russ. Technol. J.* 2021;9(6):16–25. https://doi.org/10.32362/2500316X-2021-9-6-16-25

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Разработка моделей аналитической системы обработки данных для мониторинга ИБ объекта информатизации, использующего облачную инфраструктуру

В.А. Сизов,
А.Д. Киров @

Российский экономический университет имени Г.В. Плеханова, Москва, 117997 Россия
@ *Автор для переписки, e-mail: kirov.ad@rea.ru*

Резюме. Статья посвящена разработке аналитической системы обработки данных (АСОД) для мониторинга информационной безопасности (ИБ) в рамках системы менеджмента ИБ современных компаний, ведущих свою основную деятельность в киберпространстве и использующих облачную инфраструктуру. На основе анализа современных информационных технологий (ИТ) и наиболее востребованных продуктов обеспечения ИБ облачных инфраструктур, а также существующих научных подходов предложен формализованный подход к синтезу АСОД для мониторинга ИБ такого объекта информатизации. Этот подход учитывает полезность используемых ИТ с позиции ИБ. Представлена общая модель структуры информационного обеспечения АСОД для мониторинга ИБ, а также модель зависимости полезности ИТ от времени и соотношения уровня квалификации специалиста по ИБ и злоумышленника. В качестве критерия в первой оптимизационной модели используется качество системы мониторинга ИБ. В качестве ограничений предлагаются следующие: ограничение на время принятия решения на инцидент; ограничение на степень качества анализа событий ИБ аналитической системой обработки данных и ограничение на совместимость функций анализа данных с типами данных о событиях ИБ. Приведенные результаты исследования второй модели показывают логически непротиворечивую зависимость полезности ИТ от времени и соотношения уровня квалификации специалиста по ИБ к уровню квалификации злоумышленника. Представлены частные модели структуры информационного обеспечения АСОД, позволяющие определить рациональную структуру информационного обеспечения АСОД по частным критериям. В качестве частных критериев используются следующие: максиминный критерий полезности информационного обеспечения АСОД для мониторинга ИБ объекта информатизации в облачной инфраструктуре и критерий максимума актуальности информационного обеспечения, распределенного по узлам облачной инфраструктуры для систем с невысокой степенью централизации управления.

Ключевые слова: информационная безопасность, аналитическая система обработки данных, облачная инфраструктура, компьютерное моделирование, математическое моделирование, мониторинг ИБ, объект информатизации

• **Поступила:** 25.03.2021 • **Доработана:** 11.05.2021 • **Принята к опубликованию:** 25.07.2021

Для цитирования: Сизов В.А., Киров А.Д. Разработка моделей аналитической системы обработки данных для мониторинга ИБ объекта информатизации, использующего облачную инфраструктуру. *Russ. Technol. J.* 2021;9(6):16–25. https://doi.org/10.32362/2500-316X-2021-9-6-16-25

Прозрачность финансовой деятельности: Никто из авторов не имеет финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Analytical data processing systems (ADPS) for information security (IS) monitoring are a large emerging class of systems operating in the IS management system of modern companies.

Modern companies with main activities in cyberspace, using cloud infrastructure built on the technologies of cloud infrastructure service providers, use these systems to manage IS and to improve processes that implement IS methods, including monitoring processes in real time. The peculiarities of the technical implementation of such analytical systems determine special requirements for the convenience of their use by an IS engineer, efficiency, error-free operation, and quality of data processing in real time. In addition, the potential costs of a cloud service user for the transition from one service provider to another are characterized by the relationship between the consequences in the field of IS arising from cyber attacks, system failures, user errors, and other external factors and the difference in the levels of cybersecurity technologies of providers. Therefore, the level of IS provided by a cloud service provider is subject to the same tough conditions of bilateral market competition, which determines the prices of a cloud service provider and its strategy for providing customers with IS.

New information technologies (IT) actualize the problem of ensuring the security and protection of information for their developers, vendors and consumers. Currently, it is believed that the most effective way to ensure IS for the objects of informatization of the company and/or the cloud infrastructure used by it is the path of development and implementation of secure IT. Their use in cloud infrastructures can further reduce the cost of maintaining corporate information systems. A large number of modern IT used in cloud infrastructure is aimed at a comprehensive solution to the problem of increasing the efficiency of using available resources (Table 1).

The cloud infrastructure used to host corporate information systems is increasingly taking on a hybrid form. Part of the data most critical for the functioning of business processes is placed inside the company's own infrastructure using universal cloud technologies, forming a "private cloud of the company." Other data, in turn, are processed in "public clouds," where infrastructure is provided by cloud providers in an "Infrastructure as a Service," "Platform as a Service," or "Software as a Service" model. This necessitates the use of modern data security technologies in cloud infrastructures. Being integrated with modern data processing technologies, such technologies turn into full-fledged high-tech products offered by the leading companies in the market of IS products for cloud infrastructures. Examples of such products are presented in Table 2.

The analysis of the presented technologies showed their complexity. For their development and improvement, it is necessary to apply a scientific approach, including methods of mathematical and computer simulation.

In [5], an overview of the maturity models of the capabilities and indicators of the security of systems is presented. It is noted that these models are mostly reactive rather than proactive and, therefore, do not

Table 1. The content of modern IT related to the provision of IS in the cloud infrastructure

No.	IT name	Content
1	Data Center Network (DCN) [1]	Network equipment that allows one to automatically allocate the load on local area networks (LANs) of the cloud infrastructure (both on the cloud network equipment itself and on communication channels), provided as a service.
2	Proactive resource allocation [2]	A technology that allows one to automatically allocate cloud infrastructure resources (hardware resources in the form of cloud servers and workstations and software resources in the form of software, including tools for its development and debugging), provided as a service, based on predicting the expected load.
3	Self-diagnosis software [3]	A technology designed to automatically detect and correct errors that occur during the operation of software and the implementation of its functions.
4	Data encryption in cloud infrastructures [4]	A technology designed for distributed data encryption without using a single distribution and certification center.

Table 2. Purpose of modern products for IS of cloud infrastructures

No.	Product name for IS in cloud infrastructures	Manufacturer of IS product in cloud infrastructures	Purpose
1	MVision Cloud	McAfee	Creation of a "dynamic protection perimeter" capable of adapting to the dynamic conditions of the external environment. Create security policies and apply them to SaaS, PaaS, IaaS solutions, containers and virtual cloud components.
2	Cisco Cloud Security	Cisco	Protect users, data, and applications in the cloud from compromised account attacks, malware, and data leaks, no matter where they are or where they access the Internet. Neutralizing malware before it spreads to the network or endpoints and shortens recovery time from infection.
3	FortiGate-VM	FortiNet	High-speed VPN connections are used to protect data. Security policies are enforced in all environments. Cloud Security Center provides a centralized and consistent implementation of corporate network security and communications, and maintains secure connectivity of networks, locations, clouds, and data centers.
4	Kaspersky Security for virtual and loud environments	Kaspersky Laboratory	Simultaneous implementation of both the protection of the work environment and the concept of "security as code," which ensures continuous integration of workflows and bridging the cybersecurity gap in development environments.
5	CheckPoint CloudGuard	CheckPoint	Support of the widest range of cloud infrastructures including AWS, Microsoft Azure and Azure Stack, Google Cloud Platform, VMware Cloud on AWS, and more. Automatic provisioning and automatic scaling, along with automatic policy updates, ensures that defenses keep up with changes in your cloud. One unified console provides consistent visibility, policy management, logging, reporting, and control across all cloud environments and networks.

provide adequate measures to assess the overall security of the cloud system. Therefore, this paper proposes a Cloud Security Capabilities Maturity Model that augments existing cybersecurity models with a security metric. In [6], a methodology is presented to determine the best security measures for multi-cloud applications whose components are deployed in heterogeneous clouds. The methodology is based on application decomposition and threat modeling on components, followed by risk analysis along with capturing cloud business and security requirements. However, in these works, insufficient attention is paid to the processes and technologies of IS monitoring, which significantly affect the quality of its output information (completeness, relevance, timeliness, etc.).

The output information generated by ADPS is a collection of data obtained as a result of performing sets of data analytics functions included in its composition with a certain structure or several possible types of structures. Such collections of data can be used either in the current system, or redirected to other systems present in this cloud infrastructure. The quality of the ADPS output information is the degree of its suitability for use in detecting and investigating IS incidents by a security officer.

Computer modeling is one of the main methods proposed to improve the quality of data obtained as a result of the analysis of IS monitoring data using ADPS.

PROBLEM STATEMENT

Modern IS monitoring systems are created to ensure the possibility of an adequate and timely response to cyber attacks aimed at the information infrastructure of the electronic document management system and IS management of the company, which is the main component necessary for a prompt and strategic response to current threats, in accordance with the company's IS policy. Such infrastructure can be built on the basis of LAN technologies, constituting a "private cloud of the company" and/or in the form of cloud infrastructures provided by cloud service providers, constituting a "public cloud of the company."

To create an effective IS management system, a company needs to automate its management functions as much as possible, taking into account modern requirements for IS monitoring and the characteristics of hybrid cloud infrastructures.

Thus, the task of developing models of an analytical data processing system for monitoring the IS of an informatization object using cloud infrastructure is urgent. It is advisable to divide this task into a number of subtasks.

The first subtask is to develop a general model of the structure of the information support of the analytical data processing system for IS monitoring. The solution of this problem allows one to determine rational solutions close to the optimal, which are certain sets of information elements distributed over the nodes of the ADPS computer network.

The second subtask is to determine the nature and degree of dependence of the usefulness of information technology on time and the ratio of the qualification level of an IS specialist and the skill level of an attacker. The solution to this problem allows us to identify the conditions under which the utility of IT will be the best.

The third subtask is the development of a set of private models of the ADPS information support structure for monitoring the IS of an informatization object in a cloud infrastructure. The solution to this problem allows us to determine the rational structure of information support for ADPS according to particular criteria [7, 8].

DEVELOPMENT OF A GENERAL MODEL OF THE INFORMATION SUPPORT STRUCTURE OF THE ADPS FOR MONITORING IS

Let:

$$\theta = F(\delta, t), \tag{1}$$

where θ is the quality of the IS monitoring system; F is the quality function of the IS monitoring system; $\delta = G(R)$, where δ is the quality of the analysis of events of the ADPS IS; t is the time required for the analytical data processing system to make a decision on how likely it is that a particular IS event or their combination is an IS incident; G is a function of the quality of the analysis of events of the ADPS IS;

$$R = H(S),$$

where R is the degree of compatibility of data analysis procedures with data types about IS events; H is a function of the degree of compatibility of data analysis procedures with data types about IS events; S is the structure of information support for ADPS (sets of information elements distributed over the nodes of the ADPS computer network: data on IS events, additional data for the analysis and investigation of IS incidents).

Then the general model of the structure of the information support of the analytical data processing system for IS monitoring is as follows:

Find $\max_{\{S\}} \theta$ under constraints:

$$t1 < t2;\ \delta 1 < \delta < \delta 2;\ R1 < R < R2, \tag{2}$$

where $t1$–$t2$ is the time period for which it is necessary to make a decision as to how likely it is that a particular IS event or their combination is an IS incident for the most effective response to this incident and its investigation;

$\delta 1$–$\delta 2$ are the limits of the degree of quality of the analysis of IS events by the analytical data processing system, which make it possible to make a decision as to how likely it is that a particular IS event or their combination is an IS incident with the required degree of reliability;

$R1$–$R2$ are the limits of the degree of compatibility of data analysis functions with types of data on IS events, in which it is possible to make a decision as to how likely it is that a particular IS event or their combination is an IS incident.

DEVELOPMENT OF A MODEL OF DEPENDENCE OF THE USEFULNESS OF INFORMATION TECHNOLOGY ON TIME AND THE RATIO OF THE QUALIFICATION LEVEL OF AN IS SPECIALIST AND AN ATTACKER

To develop a model of the dependence of usefulness of information technology on time and the ratio of the skill level of an IS specialist and an attacker, it is proposed to use an analytical model based on the Rayleigh statistical distribution [9]. The ratio of the skill level of an IS specialist who eliminates vulnerabilities in software and hardware to the skill level of an attacker who attacks software and hardware that implements IT is used as a parameter of the Rayleigh distribution scale.

The assessment of the dependence of the utility of information technology on time and the ratio of skill levels of these categories is as follows:

$$P(t,\sigma) = 1 - \frac{t}{\sigma^2} e^{-\frac{t}{2\sigma^2}},\ t \ge 0,\ \sigma > 0, \tag{3}$$

where P is the level of IT usefulness at time t; σ is the ratio of the quantitative evaluation of the qualification level of an IS specialist to the skill level of the attacker.

The correspondence between quantitative and qualitative assessment of the qualification levels of an IS specialist and an attacker is given in Tables 3 and 4, respectively. In the case when the opposing sides are groups of people, it is necessary to apply group quantitative assessments of the level of their qualifications to them [10]. The final matrix of relationships between quantitative assessments of the qualification levels of an IS specialist and the skill level of an attacker is presented in Table 5.

The analysis of graphs in Fig. 1 shows that with an increase in the skill level of an attacker, the usefulness of IT for users decreases faster, and with an increase in the level of qualification of an IS specialist, the usefulness of IT for users quickly returns to the maximum value, which corresponds to reality.

Table 3. Correspondence of quantitative and qualitative assessments of qualification levels of an IS specialist

IS specialist class	Qualitative assessment of the qualification level of an IS specialist	Quantitative assessment of the qualification level of an IS specialist
Highly qualified specialist	High	1
Ordinary specialist	Intermediate	2
Low-skilled specialist	Low	3

Table 4. Correspondence of quantitative and qualitative assessments of the attacker's skill levels

Attacker's class	Qualitative assessment of the qualification level of an attacker	Quantitative assessment of the qualification level of an attacker
Highly qualified specialist	High	3
Ordinary specialist	Intermediate	2
Low-skilled specialist	Low	1

Table 5. The final matrix of relationships between quantitative assessments of the qualification levels of an IS specialist and the skill level of an attacker

IS specialist's class / Attacker's class	Highly qualified specialist	Ordinary specialist	Low-skilled specialist
Highly qualified specialist	0.33	0.5	1
Ordinary specialist	0.66	1	2
Low-skilled specialist	1	1.5	3

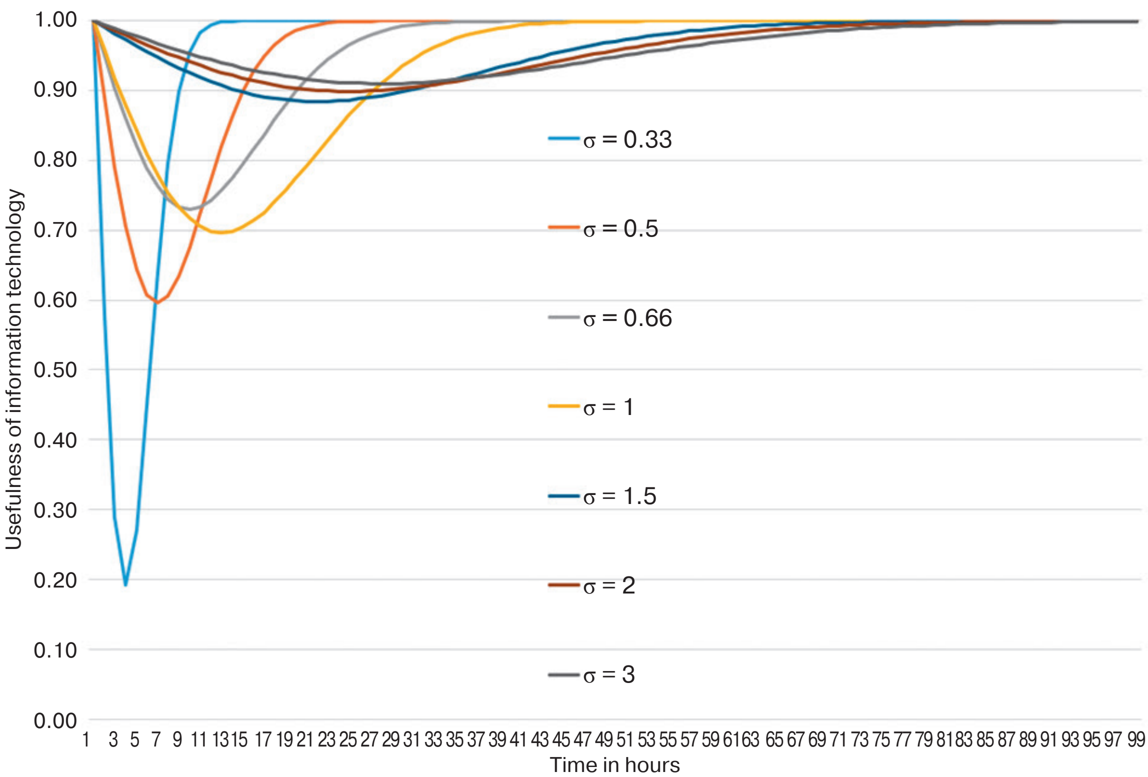


Fig. 1. Dependence of the usefulness of information technology on time and the ratio of the qualification levels of an IS specialist and an attacker

DEVELOPMENT OF PARTICULAR MODELS OF THE INFORMATION SUPPORT STRUCTURE OF THE ADPS FOR MONITORING SECURITY OF THE INFORMATIZATION OBJECT IN THE CLOUD INFRASTRUCTURE

Currently, in the field of IS management of modern informatization objects in cloud infrastructures, a group of IS incident management tasks is distinguished, which includes the following main tasks: monitoring IS events of informatization objects in cloud infrastructures and identifying IS incidents; registration of IS incidents; analysis of IS incidents; informing the administration of cloud infrastructure service providers about all cases of IS violations; collection of evidence for responding to incidents of IS in cloud infrastructures and others. From a practical point of view, one of the most effective ways to creating IS monitoring is the use of systems of the security information and event management (SIEM) class [11–14].

Solutions of the SIEM class are designed to provide the following functions:

- management of information and security events, including those in cloud infrastructures;
- collection and storage of registered security events;
- processing and analysis of IS events stored in internal databases using a system of rules created and managed by a security officer and unique for each cloud infrastructure.

The implementation of the above functions allows for the identification and analysis of incidents, as well as verification of the compliance of the IS management system with existing requirements and norms that are part of the standards, recommendations, orders and other regulatory and guidance documents in this area [15]. The implementation of the above functions assumes the presence of ADPS in systems of the SIEM class, the principles of construction of which were indicated earlier, or, at least, its key components.

However, the main disadvantage of such systems is the relatively long period of time required for them to analyze the data and make a decision about whether this IS event or their combination is an IS incident or not [16–18].

This drawback is based on the contradiction between the distributed nature of information sources about an IS event in cloud infrastructures (as a rule, these sources are IS tools integrated in cloud infrastructures) and a centralized way of deciding on actions with IS incidents. To resolve this contradiction, it is necessary, on the one hand, to provide the decision-making process with the most complete information, and on the other hand, this information must be relevant. Taking into account the large volumes of data on IS events in ADPS, it is necessary to optimize the structure of information support for ADPS, taking into account the structure and technical characteristics of the LAN.

The peculiarities of the functioning of ADPS for monitoring the IS of an informatization object based on cloud infrastructure allow for solving problems of improving the quality of output information using computer modeling, using the methods of utility theory, which make it possible to assess the useful effect of placing information elements in certain computing nodes of the cloud infrastructure [17]. Modern ADPS makes it possible to establish the degree of information reliability, taking into account not only the degree of completeness and accuracy of the data (the sufficiency of data for solving the problem and the correspondence of the structure and content of the data to the systems, the use of which is optimal for solving the assigned tasks), but also the degree of relevance of the data (the ability of information to reflect the real state of objects at the current time).

The method for determining the optimal computing nodes of the cloud infrastructure from the point of view of monitoring IS for placing certain types of information elements in them and distributing these elements among the nodes is based on computer modeling of the cloud infrastructure, in which IS is monitored and the application of the utility theory methods for assessing the degree of type correspondence information elements to one or another computing node of the cloud infrastructure. The specificity of IS incident management involves working with data arrays, which makes it possible to select the necessary data sets that can be distributed among the nodes of the cloud infrastructure based on the model [16–19]. Consequently, particular models should be used, among other things, to determine the structure and volumes of information exchange between nodes of the cloud infrastructure. The information support of ADPS for monitoring the IS of an informatization object in the cloud infrastructure can include both the data about the state of the IS of the informatization object in the cloud infrastructure, obtained directly from the IS tools deployed inside the cloud infrastructure, as well as their copies and/or prehistories received in the nodes cloud infrastructure in the places of their use ADPS for monitoring the IS of the object of informatization. It is advisable to use as particular criteria of these models: the maximin criterion of the usefulness of information support of ADPS for monitoring the IS of an informatization object in a cloud infrastructure; the criterion for the maximum relevance of information support distributed over the nodes of the cloud infrastructure for complex systems with a low degree of centralization of management [7, 8].

The result of solving the problem of developing the structure of the information support of the analytical data processing system for monitoring the IS of the informatization object is the optimal composition, according to given criteria, of the information support components of the ADPS and their placement on the nodes of the cloud infrastructure.

CONCLUSIONS

This article has analyzed the problem of determining a rational structure of information support for ADPS for monitoring the IS of an informatization object in a cloud infrastructure. At the same time, the ISMS, built on the basis of modern SIEM-systems, were considered, and a formalized approach based on the development and use of methods of mathematical and computer modeling was proposed.

The formalization and solution of this problem is based on the methods of utility theory and operations research, which allow using computer modeling to determine the optimal composition of the information support of ADPS and their distribution among the nodes of the cloud infrastructure in terms of the general usefulness of information, taking into account the analytical information technologies used in these nodes to identify IS incidents.

In general, this approach makes it possible to increase the efficiency of the procedure for identifying IS incidents by organizing a rational exchange of information between nodes of the cloud infrastructure (information protection means), taking into account the characteristics of analytical data processing procedures, and, in general, the quality of the IS monitoring system.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Liu Z., Zhao A., Liang M. A port-based forwarding load-balancing scheduling approach for cloud datacenter networks. *J. Cloud Comp*. 2021;10(1):13. https://doi.org/10.1186/s13677-021-00226-w
2. Chen J., Wang Y., Liu T. A proactive resource allocation method based on adaptive prediction of resource requests in cloud computing. *J. Wireless Com. Network*. 2021;24. https://doi.org/10.1186/s13638-021-01912-8
3. Wang J., Zhang G., Wang W., Zhang K., Sheng Y. Cloud-based intelligent self-diagnosis and department recommendation service using Chinese medical BERT. *J. Cloud Comp.: Advances, Systems and Applications*. 2021;10(1):4. https://doi.org/10.1186/s13677-020-00218-2
4. Chen Y., Liu H., Wang B., Sonompil B., Ping Y., Zhang Z. A threshold hybrid encryption method for integrity audit without trusted center. *J. Cloud Comp.: Advances, Systems and Applications*. 2021;10(1):3. https://doi.org/10.1186/s13677-020-00222-6
5. Ngoc T.L., Doan B.H. Capability maturity model and metrics framework for cyber cloud security. *Scalable Computing: Practice and Experience*. 2017;18(4):277–290. https://doi.org/10.12694/scpe.v18i4.1329
6. Afolaranmi S.O., Moctezuma L.E.G., Rak M., Casola V., Rios E., Lastra J.L.M. Methodology to Obtain the Security Controls in Multi-cloud Applications. In: *Proceedings of the 6th International Conference on Cloud Computing and Services Science* (*CLOSER 2016*). V.1. 2016, p. 327–332. http://doi.org/10.5220/0005912603270332
7. Sizov V.A. Development of models for improving the efficiency of data safety in a distributed computing environment based on dynamic data reservation. In: *Advances in Science and Technology*: *Collection of articles of the XXI International Scientific and Practical Conference*. 2019, p. 96–100. (in Russ.).
8. Sizov V.A. Models and methods of virtual-recovered redundancy of data of automatic information-control systems under extreme conditions. *Autom. Remote Control*. 1998;59(7):1047–1053.

СПИСОК ЛИТЕРАТУРЫ

1. Liu Z., Zhao A., Liang M. A port-based forwarding load-balancing scheduling approach for cloud datacenter networks. *J. Cloud Comp*. 2021;10(1):13. https://doi.org/10.1186/s13677-021-00226-w
2. Chen J., Wang Y., Liu T. A proactive resource allocation method based on adaptive prediction of resource requests in cloud computing. *J. Wireless Com. Network*. 2021;24. https://doi.org/10.1186/s13638-021-01912-8
3. Wang J., Zhang G., Wang W., Zhang K., Sheng Y. Cloud-based intelligent self-diagnosis and department recommendation service using Chinese medical BERT. *J. Cloud Comp.: Advances, Systems and Applications*. 2021;10(1):4. https://doi.org/10.1186/s13677-020-00218-2
4. Chen Y., Liu H., Wang B., Sonompil B., Ping Y., Zhang Z. A threshold hybrid encryption method for integrity audit without trusted center. *J. Cloud Comp.: Advances, Systems and Applications*. 2021;10(1):3. https://doi.org/10.1186/s13677-020-00222-6
5. Ngoc T.L., Doan B.H. Capability maturity model and metrics framework for cyber cloud security. *Scalable Computing: Practice and Experience*. 2017;18(4):277–290. https://doi.org/10.12694/scpe.v18i4.1329
6. Afolaranmi S.O., Moctezuma L.E.G., Rak M., Casola V., Rios E., Lastra J.L.M. Methodology to Obtain the Security Controls in Multi-cloud Applications. In: *Proceedings of the 6th International Conference on Cloud Computing and Services Science* (*CLOSER 2016*). 2016. V.1. p. 327–332. http://doi.org/10.5220/0005912603270332
7. Сизов В.А. Разработка моделей повышения эффективности сохранности данных в распределенной вычислительной среде на основе динамического резервирования данных. В сб.: *Advances in Science and Technology*: *сб. статей XXI международной научно-практической конференции*. М.: «Актуальность. РФ», 2019. С. 96–100.
8. Сизов В.А. Модели и методы виртуально-восстановительного резервирования данных автоматизированных информационно-управляющих систем в условиях

[Sizov V.A. Models and methods of virtual-recovered redundancy of data of automatic information-control systems under extreme conditions. *Automat. i Telemekh*. 1998;(7):176–184 (in Russ.).]

9. Arce D.G. Cybersecurity and platform competition in the cloud. *Computers & Security*. 2020;93:101774. https://doi.org/10.1016/j.cose.2020.101774
10. Dzhincharadze G.R. Methodological aspects of the organization of the personnel assessment procedure. *Inzhenernyi Vestnik Dona = Engineering journal of Don*. 2012;2(20):340–345 (in Russ.). Available from URL: https://cyberleninka.ru/article/n/metodicheskie-aspekty-organizatsii-protsedury-otsenki-personala
11. Sizov V.A., Kirov A.D. Problems of implementation SIEM-systems in the practice of managing information security of economic entities. *Otkrytoe obrazovanie = Open Education*. 2020;24(1):69–79 (in Russ.). https://doi.org/10.21686/1818-4243-2020-1-69-79
12. Lee J., Kim Y.S., Kim J.H., Kim I.K. Toward the SIEM architecture for cloud-based security services. In: *2017 IEEE Conference on Communications and Network Security (CNS)*. https://doi.org/10.1109/CNS.2017.8228696
13. Granadillo G.G., El-Barboni M., Debar H. New Types of Alert Correlation for Security Information and Event Management Systems. In: *2016 8^{th} IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. 2016. https://doi.org/10.1109/NTMS.2016.7792462
14. Kavanagh M., Rochford O. *Magic Quadrant for Security Information and Event Management.* Gartner technical report. 2015. 15 p.
15. Markov A.S., Tsirlov V.L. Structured content of information security requirements. *Monitoring pravoprimeneniya = Monitoring of Law Enforcement*. 2017;1(22):53–61 (in Russ.). https://doi.org/10.21681/2412-8163-2017-1-53-61
16. Nabil M., Soukainat S., Lakbabi A., Ghizlane O. SIEM selection criteria for an efficient contextual security. In: *2017 International Symposium on Networks, Computers and Communications (ISNCC)*. https://doi.org/10.1109/ISNCC.2017.8072035
17. Kirsanov K.K. The theory of utility in the period of change of conceptual provisions. *Naukovedenie* (*The Eurasian Journal*). 2015;7(2):38 (in Russ.). Available from URL: http://naukovedenie.ru/PDF/37EVN215.pdf
18. Kotenko I.V., Fedorchenko A.V., Saenko I.B., Kushnerevich A.G. Big data technologies for security event correlation based on event type accounting. *Voprosy kiberbezopasnosti = Cybersecurity issues*. 2017;5(24):2–16 (in Russ.). https://doi.org/10.21681/2311-3456-2017-5-2-16
19. Fedorchenko A.V., Levshun D.S., Chechulin A.A., Kotenko I.V. An Analysis of Security Event Correlation Techniques in SIEM-Systems. Part 2. *Trudy SPIIRAN = SPIIRAS Proceedings*. 2016;6(49):209–225 (in Russ.). https://doi.org/10.15622/sp.49.11

чрезвычайных ситуаций. *Автоматика и телемеханика*. 1998;7:176–184.

9. Arce D.G. Cybersecurity and platform competition in the cloud. *Computers & Security*. 2020;93:101774. https://doi.org/10.1016/j.cose.2020.101774
10. Джинчарадзе Г.Р. Методические аспекты организации процедуры оценки персонала. *Инженерный Вестник Дона*. 2012;2(20):340–345. URL: https://cyberleninka.ru/article/n/metodicheskie-aspekty-organizatsii-protsedury-otsenki-personala
11. Сизов В.А., Киров А.Д. Проблемы внедрения SIEM-систем в практику управления информационной безопасностью субъектов экономической деятельности. *Открытое образование*. 2020;24(1):69–79. https://doi.org/10.21686/1818-4243-2020-1-69-79
12. Lee J., Kim Y.S., Kim J.H., Kim I.K. Toward the SIEM architecture for cloud-based security services. In: *2017 IEEE Conference on Communications and Network Security* (*CNS*). https://doi.org/10.1109/CNS.2017.8228696
13. Granadillo G.G., El-Barboni M., Debar H. New types of alert correlation for security information and event management systems. In: *2016 8^{th} IFIP International Conference on New Technologies, Mobility and Security* (*NTMS*). 2016. https://doi.org/10.1109/NTMS.2016.7792462
14. Kavanagh M., Rochford O. *Magic Quadrant for Security Information and Event Management.* Gartner technical report. 2015. 15 p.
15. Марков А.С., Цирлов В.Л. Структурное содержание требований информационной безопасности. *Мониторинг правоприменения*. 2017;1(22):53–61.
16. Nabil M., Soukainat S., Lakbabi A., Ghizlane O. SIEM selection criteria for an efficient contextual security. In: *2017 International Symposium on Networks, Computers and Communications* (*ISNCC*). 2017. https://doi.org/10.1109/ISNCC.2017.8072035
17. Кирсанов К.К. Теория полезности в период смены концептуальных положений. *Науковедение* (*Вестник Евразийской науки)*. 2015;7(2):38. URL: http://naukovedenie.ru/PDF/37EVN215.pdf
18. Котенко И.В., Федорченко А.В., Саенко И.Б., Кушнеревич А.Г. Технологии больших данных для корреляции событий безопасности на основе учета типов связей. *Вопросы кибербезопасности*. 2017;5(24):2–16. https://doi.org/10.21681/2311-3456-2017-5-2-16
19. Федорченко А.В., Левшун Д.С., Чечулин А.А., Котенко И.В. Анализ методов корреляции событий безопасности в SIEM-системах. Часть 2. *Труды СПИИРАН*. 2016;6(49):208–225. https://doi.org/10.15622/sp.49.11

About the authors

Valerii A. Sizov, Dr. Sci. (Eng.), Professor, Department of Applied Informatics and Information Security, Institute of Mathematics, Information Systems and Digital Economy, Plekhanov Russian University of Economics (36, Stremyanny per., Moscow, 117997 Russia). E-mail: sizov.va@rea.ru. https://orcid.org/0000-0002-4844-4714

Aleksey D. Kirov, Specialist, Specialized Educational and Scientific Laboratory on Information Confrontation in Business, Department of Applied Informatics and Information Security, Institute of Mathematics, Information Systems and Digital Economy, Plekhanov Russian University of Economics (36, Stremyanny per., Moscow, 117997 Russia). E-mail: kirov.ad@rea.ru. https://orcid.org/0000-0002-8424-3071

Об авторах

Сизов Валерий Александрович, д.т.н., профессор, кафедра Прикладной информатики и информационной безопасности Института математики, информационных систем и цифровой экономики ФГБОУ ВО «Российский экономический университет имени Г.В. Плеханова» (117997, Россия, Москва, Стремянный пер., 36). E-mail: sizov.va@rea.ru. https://orcid.org/0000-0002-4844-4714

Киров Алексей Дмитриевич, специалист, специализированная учебно-научная лаборатория по информационному противоборству в бизнесе, кафедра Прикладной информатики и информационной безопасности Института математики, информационных систем и цифровой экономики ФГБОУ ВО «Российский экономический университет имени Г.В. Плеханова» (117997, Россия, Москва, Стремянный пер., 36). E-mail: kirov.ad@rea.ru. https://orcid.org/0000-0002-8424-3071

Translated by E. Shklovskii

Multiple robots (robotic centers) and systems. Remote sensing and non-destructive testing

Роботизированные комплексы и системы. Технологии дистанционного зондирования и неразрушающего контроля

UDC 681.158.3
https://doi.org/10.32362/2500-316X-2021-9-6-26-36

RESEARCH ARTICLE

Development of an information measuring and control system for a quadrocopter

Vladimir V. Sleptsov [1],
Vyacheslav L. Afonin [2],
Anna E. Ablaeva [1, @],
Dinh Ba Phuong [1]

[1] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[2] *Mechanical Engineering Research Institute, Russian Academy of Sciences, Moscow, 101990 Russia*
[@] *Corresponding author, e-mail: ablaeva@mirea.ru*

Abstract. The article deals with the issues of synthesis and analysis of information-measuring and control systems of quadcopters (QCs). The main sensors and modules used to determine the parameters of the coordinates of QCs are given. The speed-controlled electric drives used for control and the features of their choice are considered. The coordinate systems (fixed and mobile) and the kinematic scheme are given, according to which a system of differential equations is presented. The system describes the dynamics of the QC movement and takes into account the expected smooth movement of the QC with small roll and pitch angles. A functional scheme and a mathematical model of the information-measuring and control system of the QC in the form of a block diagram are developed taking into account the influence of delays in the receipt of information from the sensors of the QC parameters. A special feature of this work is to take into account the specific characteristics of the elements: adjustable electric drives (both direct and alternating current), parameter sensors (barometers, accelerometers, rangefinders, etc.). The paper studies an illustrative algorithm for the operation of the information-measuring and control system of the quadcopter. The type and parameters of the controllers of the QC control systems are determined. Special attention is paid to the settings for the control contours at the corresponding coordinates. The influence of the controllers of the coordinate control systems of the information-measuring and control systems of the QC on the effects of the interaction of coordinates is considered. The simulation results are presented. The optimal number of control loops for the coordinates of the information-measuring and control systems of the QC and the optimal type of settings for obtaining smooth transients (without overshoot) and for excluding the interaction of coordinates on quality indicators are determined.

Keywords: quadrocopter, adjustable electric drive, parameter sensors, optimum modulus

• **Submitted:** 17.05.2021 • **Revised:** 24.05.2021 • **Accepted:** 25.07.2021

For citation: Sleptsov V.V., Afonin V.L., Ablaeva A.E., Dinh Ba Phuong. Development of an information measuring and control system for a quadrocopter. *Russ. Technol. J.* 2021;9(6):26–36. https://doi.org/10.32362/2500-316X-2021-9-6-26-36

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Разработка информационно-измерительной и управляющей системы квадрокоптера

В.В. Слепцов [1],
В.Л. Афонин [2],
А.Е. Аблаева [1, @],
Динь Ба Фыонг [1]

[1] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[2] *Институт машиноведения им. А.А. Благонравова РАН, Москва, 101990 Россия*
[@] *Автор для переписки, e-mail: ablaeva@mirea.ru*

Резюме. В статье рассмотрены вопросы синтеза и анализа информационно-измерительных и управляющих систем квадрокоптеров. Приведены основные датчики и модули, применяемые для определения параметров координат квадрокоптеров. Рассмотрены регулируемые по скорости электроприводы, применяемые для управления, и особенности их выбора. Приведены системы координат (неподвижная и подвижная) и кинематическая схема, в соответствии с которыми представлена система дифференциальных уравнений, описывающая динамику движения квадрокоптера и учитывающая предполагаемое плавное движение квадрокоптера с малыми углами крена и тангажа. Разработаны функциональная схема и математическая модель информационно-измерительной и управляющей системы квадрокоптера (ИИУС КК) в виде структурной схемы, выполненные с учетом влияния запаздываний поступления информации с датчиков параметров квадрокоптера. Особенностью данной работы является учет конкретных характеристик элементов: регулируемых электроприводов (как постоянного, так и переменного тока), датчиков параметров (барометров, акселерометров, дальномеров и пр.). В работе исследован показательный алгоритм работы информационно-измерительной и управляющей систем квадрокоптера, определены тип и параметры регуляторов систем управления. Особое внимание уделено параметрам настройки для соответствующих контуров управления. Рассмотрено влияние указанных регуляторов информационно-измерительной и управляющей системы квадрокоптера на эффекты взаимовлияния координат. Представлены результаты моделирования. Определено оптимальное количество контуров управления координатами информационно-измерительной и управляющей системы квадрокоптера и оптимальный вид настроек для получения плавных переходных процессов (без перерегулирования) и исключения взаимовлияния координат на показатели качества.

Ключевые слова: квадрокоптер, регулируемый электропривод, датчики параметров, оптимум по модулю

• **Поступила:** 17.05.2021 • **Доработана:** 24.05.2021 • **Принята к опубликованию:** 25.07.2021

Для цитирования: Слепцов В.В., Афонин В.Л., Аблаева А.Е., Динь Ба Фыонг. Разработка информационно-измерительной и управляющей системы квадрокоптера. *Russ. Technol. J.* 2021;9(6):26–36. https://doi.org/10.32362/2500-316X-2021-9-6-26-36

Прозрачность финансовой деятельности: Никто из авторов не имеет финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

GLOSSARIUM

Yaw—angular movement of the aircraft around its vertical axis.

Pitch—angular motion in which its longitudinal axis changes its direction relative to the horizontal plane.

Roll—the deviation of the plane of symmetry from the vertical position.

INTRODUCTION

A modern quadcopter (QC) is an unmanned aerial vehicle with four controllable propellers providing movement along a given trajectory, designed for transporting usually light instruments and objects at a limited distance as well as for mineral exploration and other observations [1, 2]. General view of a typical quadcopter is shown in Fig. 1.

Fig. 1. General view of a typical QS

The use of QCs provides certain advantages [3–5]:

- cost-effectiveness;
- responsiveness;
- the capability of reaching high speed, gliding, and hovering over the control point;
- lower cost of manufacturing and operation (with comparable efficiency of fulfilling the assigned tasks);
- possibility of using without pilots. There is no danger of accident at that.

The disadvantages include the outstanding issues of integrating quadcopters into a common airspace as well as the certification, insurance, and registration issues, being already worked on.

MATERIALS AND METHODS

The QC motion is described in the fixed and moving coordinate systems shown in Fig. 2.

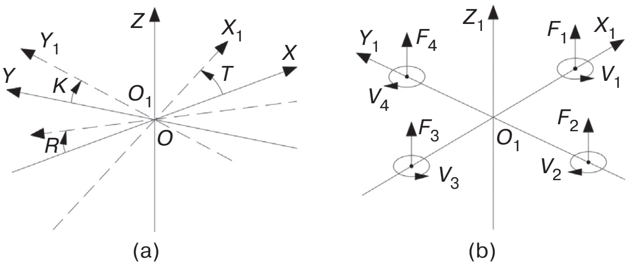


Fig. 2. (a) Fixed and moving coordinate systems, (b) kinematic scheme of a QC

Here, X, Y, and Z is the fixed coordinate system; X_1, Y_1, and Z_1 is the moving coordinate system, F_i are propeller thrust forces; V_i are rotation speeds of actuator motors of electric drives; R is the yaw angle; T is the pitch angle; K is the roll angle; g is the acceleration of gravity; ω_1 is the rotational speed of rotation of the ith propeller ($i = 1, ..., 4$).

The QC dynamics may be described by a system of differential equations in the following form [6]:

$$\ddot{X} = \frac{F_1 + F_2 + F_3 + F_4}{m} \times \left[\cos R \sin T \cos K + \sin R \sin K\right] - \frac{A_x}{m}\dot{X}, \quad (1)$$

$$\ddot{Y} = \frac{F_1 + F_2 + F_3 + F_4}{m} \times \left[\sin R \sin T \cos K + \cos R \sin K\right] - \frac{A_y}{m}\dot{Y}, \quad (2)$$

$$\ddot{Z} = \frac{F_1 + F_2 + F_3 + F_4}{m}\cos T \cos K - \frac{A_z}{m}\dot{Z} - g, \quad (3)$$

$$\ddot{T} = \frac{l}{J_{xx}}\left(F_4 - F_2\right), \quad (4)$$

$$\ddot{K} = \frac{l}{J_{yy}}\left(F_3 - F_1\right), \quad (5)$$

$$\ddot{R} = \frac{lb}{J_{zz}K_T}\left(F_1 - F_2 + F_3 - F_4\right), \quad (6)$$

$$F_i = K_T \omega_i^2. \quad (7)$$

Here, F_i stands for propeller thrust forces; J_{xx}, J_{yy}, J_{zz} are the QC inertia moments about corresponding axes; m is the QC weight; l is the distance from the QC center to the motor mounts; b is the process factor; A_x, A_y, and A_z are drag coefficients, and ω is the motor shaft speed.

The QC specific values are the following: m = 0.5 kg; l = 0.25 m; $K_T = 4 \cdot 10^{-5}$ (H·s^2)/rad^2; $b = 1.2 \cdot 10^{-7}$ (H·m·s^2)/rad^2; $A_x = A_y = A_z = 1$ kg/s; $J_{xx} = J_{yy} = 5 \cdot 10^{-3}$ kg·m^2; $J_{zz} = 9 \cdot 10^{-3}$ kg·m^2; $\omega_{i\max} = 300$ rad/s.

Since smooth motion of the QC with small roll and pitch angles is assumed, $\cos T \approx \cos K \approx \cos R \approx 1$, while $\sin T \approx T$, $\sin K \approx K$, and $\sin(R) \approx 0$. In addition, assuming smooth motion in the XOY plane, i.e., fulfillment of the condition $(F_1 + F_2 + F_3 + F_4) = mg$, equations (1)–(7) may be converted into the following form:

$$\ddot{X} = K_{\mathrm{T}x}\left(T\right) - \frac{A_x}{m}\dot{X}, \quad (8)$$

$$\ddot{Y} = -K_{\mathrm{K}y}\left(K\right) - \frac{A_y}{m}\dot{Y}, \quad (9)$$

$$\ddot{Z} = \frac{F_1 + F_2 + F_3 + F_4}{m} - \frac{A_z}{m}\dot{Z} - g, \quad (10)$$

$$\ddot{T} = \frac{l}{J_{xx}}\left(F_4 - F_2\right), \quad (11)$$

$$\ddot{K} = \frac{l}{J_{yy}}\left(F_3 - F_1\right), \tag{12}$$

$$\ddot{R} = \frac{lb}{J_{zz} \times K_{\mathrm{T}}}\left(F_1 - F_2 + F_3 - F_4\right). \tag{13}$$

Here, $K_{\mathrm{T}x}$ and $K_{\mathrm{K}y}$ are coefficients of mutual influence at corresponding coordinates (*X, Y*) of the QC.

To avoid possible uncertainties in understanding the following material, the concept of an information-measuring and control system (IMCS) of a QC should be introduced.

The QC IMCS implies a combination of hardware and software collecting, storing, and processing data on motion parameters, as well as the development of control actions for the control elements [7].

The following basic sensors are used for determining the QC coordinate parameters [8–11].

A barometer is a pressure measurement device which readings may give an indication of the flight altitude. The combination of a pressure sensor and GPS altitude sensor indicates the least error in determining altitude. The parameters of the commonly used MS5611 barometer (MEAS Switzerland SA) are the following:

- measured pressure ranges from 10 to 1200 hPa;
- accuracy is up to 0.1 m in the most accurate mode;
- measurement time is up to 10 ms.

An ultrasonic sensor is an instrument used for obtaining reliable information on the distance to large targets, even in environments with strong acoustic or electrical noise sources. The parameters of the commonly used I2XL-MaxSonar-EZ4 ultrasonic sensor are the following:

- reading interval is 67 ms (15 Hz);
- maximum distance is 765 cm;
- resolution within the range from 25 to 765 cm is 0.1 m.

A magnetometer is an electronic compass providing flight information relative to the Earth's magnetic field, located on the controller board.

A gyroscope is a device used for measuring the rate of angle change (usually measured along three axes).

An accelerometer is a device used for measuring the QC linear acceleration in the three-axis system. The parameters of the commonly used device combining gyroscope and MPU6050 accelerometer are the following:

- the range of measured angular velocities for gyroscope is ±250, ±500, ±1000, and ±2000 °/s;
- the range of measured accelerations for accelerometer is ±2, ±4, ±8, and ±16g.

A camcorder is a device used for recording the captured image. The parameters of the widely used Stack-X-1080P camcorder are the following:

- focal length is 2.8 mm;
- lens angle is *H*: 130°, *V*: 98°;
- camera sensor is 1/2.5-inch CMOS;
- DVR frame rate is 60 fps;
- NTSC or PAL video format.

GPS/GLONASS navigation module (e.g., H507A-05 Flight control PCBA) allows tracking and measuring distance, speed, and time. In fact, it is also a tracking system allowing identifying the exact location of the vehicle.

A flight controller module allows the QC to track its current location and speed. It also receives signals from the operator's transmitter. The flight controller interacts with the sensors onboard the QC to ensure a smooth flight, in particular with accelerometer, gyroscopes, and all that. In addition, the flight controller calculates the speed of each of the four engines and sends control signals to the electronic speed controllers (ESC).

Adjustable electric drives **(*AED*)** play a key role in QCs. They consist of an ESC and an electric motor. It is essential that both magnetoelectric and brushless (BL) direct current motors are used currently as electric motors in QCs. However, BL motors have become commonly used recently, which is due to their good performance characteristics [12], such as high specific torque, quick response, and control simplicity.

The parameters of the commonly used AED consisting of the ESC HW30A control unit and Walkera QR X350 motor are the following:

- supply voltage is 12 V;
- maximum current is 30 A;
- nominal rotor speed is 314 rad/s.

The development of the functional scheme and mathematical model for QC IMCS with special attention to its components will now be discussed [13–16]. Here, close attention should be paid to the values of delays in receiving information from corresponding sensors of QC parameters on the position of coordinates and their speeds. In the paper, these delays are considered as time constants for transfer functions of the corresponding sensors.

Taking into account the above equations, the functional and structural diagrams for the QC IMCS shown in Figs. 3 and 4 have been developed.

Here, the following designations are accepted: PC_z, PC_x, PC_y, PC_t, PC_k, and PC_r are position controllers for *Z*, *X*, *Y*, *T*, *K*, and *R* coordinates, respectively; AED1–AED4 are four adjustable electric drives; SC_t, SC_k, and SC_r are speed controllers for *T*, *K*, and *R* coordinates; PS_z, PS_x, PS_y, PS_t, PS_k, and PS_r are position sensors for *Z*, *X*, *Y*, *T*, *K*, and *R* coordinates, respectively; SS_t, SS_k, and SS_r are speed sensors for *T*, *K*, and *R* coordinates; CE1 and CE2 are correcting elements; $U_{\mathrm{SET}z}$, $U_{\mathrm{SET}x}$, $U_{\mathrm{SET}y}$, $U_{\mathrm{SET}t}$, $U_{\mathrm{SET}k}$, and $U_{\mathrm{SET}r}$ are position setting signals for *Z*, *X*, *Y*, *T*, *K*, and *R* coordinates, respectively; $U_{\mathrm{S}z}$, $U_{\mathrm{S}x}$, $U_{\mathrm{S}y}$, $U_{\mathrm{S}t}$, $U_{\mathrm{S}k}$, and $U_{\mathrm{S}r}$ are signals from position sensors for *Z*, *X*, *Y*, *T*, *K*, and *R* coordinates, respectively;

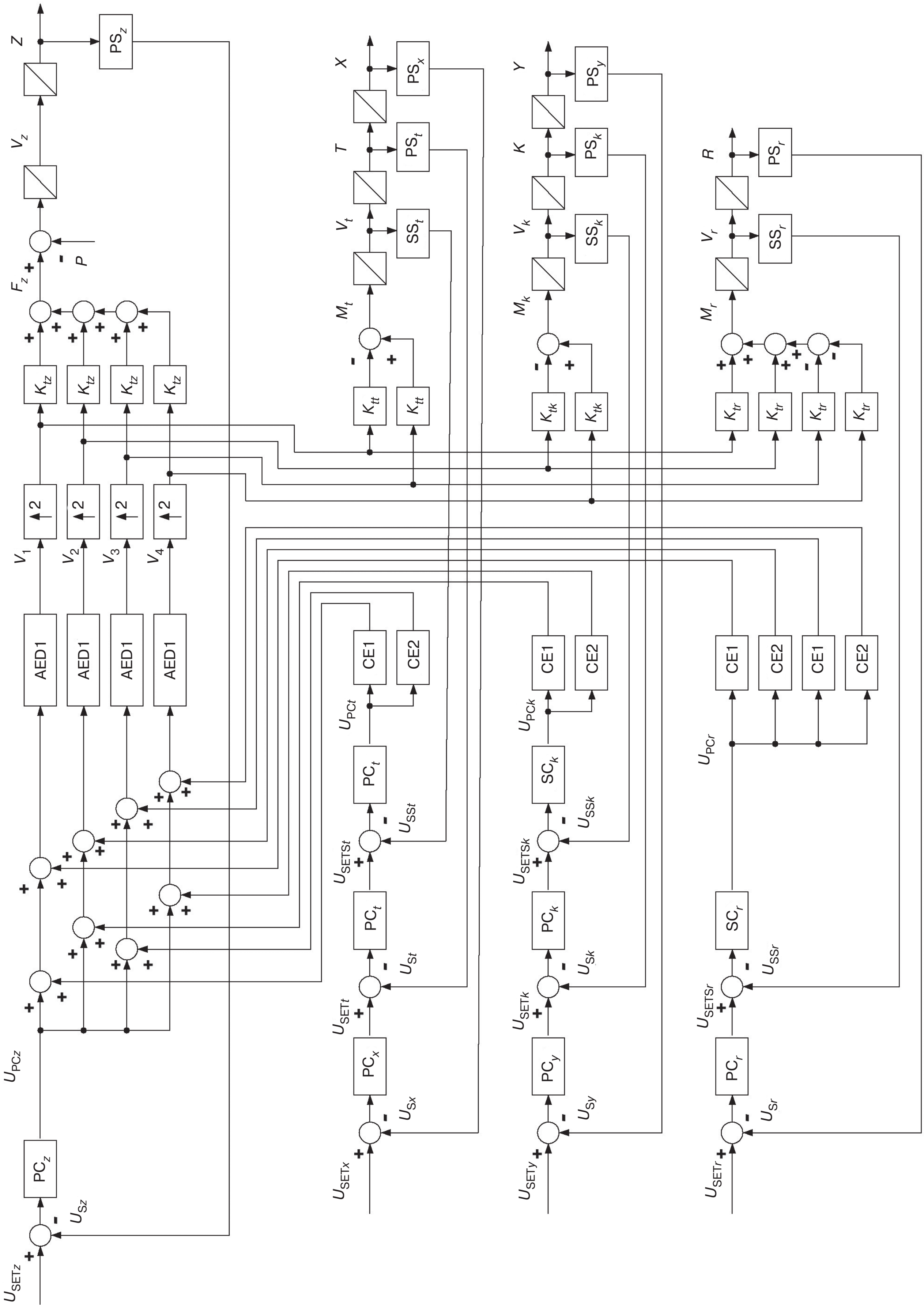


Fig. 3. A functional diagram for QC IMCS

U_{PCz}, U_{PCr} are output signals from Z and R coordinate position controllers, respectively; U_{SETt}, U_{SETk} are signals of setting T and K coordinates; U_{SETSt}, U_{SETSk}, and U_{SETSr} are signals of setting T, K, and R coordinate speeds, respectively; U_{SCt}, U_{SCk}, U_{SCr} are signals from controllers of T, K, and R coordinate speeds; F_z is the thrust force at Z coordinate; P is QC weight; V_z is the displacement velocity at Z coordinate; M_k, V_k are the torque and speed at T coordinate; M_k, V_k are the torque and speed at K coordinate; M_r, V_r are the torque and speed at R coordinate; K_{tz}, K_{tt}, K_{tk}, and K_{tr} are physical factors; $W_{PCz}(S)$, $W_{PCx}(S)$, $W_{PCy}(S)$, $W_{PCt}(S)$, $W_{PCk}(S)$, and $W_{PCr}(S)$ are transfer functions of position controllers for Z, X, Y, T, K, and R coordinates, respectively; $W_{SCt}(S)$ and $W_{SCk}(S)$ are transfer functions of speed controllers for T and K coordinates, respectively; S is the Laplace operator; K_{AED} and T_{AED} are the transfer factor and time constant for AED, respectively; K_{PSz} and T_{PSz} are the transfer factor and time constant for position sensor at Z coordinate, respectively; K_{PSx} and T_{PSx} are the transfer factor and time constant for position sensor at X coordinate, respectively; K_{PSy} and T_{PSy} are the transfer factor and time constant for position sensor at Y coordinate, respectively; K_{PSt} and T_{PSt} are the transfer factor and time constant for position sensor at T coordinate, respectively; K_{PSk} and T_{PSk} are the transfer factor and time constant for position sensor at K coordinate, respectively; K_{PSr} and T_{PSr} are the transfer factor and time constant for position sensor at R coordinate, respectively; K_{SSt} and T_{SSt} are the transfer factor and time constant for speed sensor at T coordinate, respectively; K_{SSk} and T_{SSk} are the transfer factor and time constant for speed sensor at K coordinate, respectively; K_{SSr} and T_{SSr} are the transfer factor and time constant for speed sensor at R coordinate, respectively; K_{vz} and T_{vz} are the transfer factor and time constant for velocity node at Z coordinate, respectively; K_{vx} and T_{vx} are the transfer factor and time constant for velocity node at X coordinate, respectively; K_{vy} and T_{vy} are the transfer factor and time constant for velocity node at Y coordinate, respectively; K_{vt} is the transfer factor for velocity node at T coordinate; K_{vk} is the transfer factor for velocity node at K coordinate; K_{vr} is the transfer factor for velocity node at R coordinate; F_{PCz}, F_{PCx}, F_{PCy}, F_{PCt}, F_{PCk}, F_{PCr}, and F_{PCz} are nonlinearities of position controllers at Z, X, Y, T, K, and R coordinates, respectively; $F1$, $F2$ are nonlinearities of correcting devices; F_{SCt} and F_{SCk} are nonlinearities of speed controllers at T and K coordinates.

The QC coordinate control loops are adjusted to technical and symmetrical optimums [17]. Specific values for the QC IMCS are given below.

***For the control loop at Z coordinate*:**

$K_{tz} = 4 \cdot 10^{-5}$ (H·s^2)/rad^2; $K_{vz} = 5$ (1/s); $T_{vz} = 2$ s; $K_{AED} = 2.5$ rad/(discrete·s); $T_{AED} = 0.001$ s; $K_{PSz} = 1$; $T_{PSz} = 0.01$ s. The loop is adjusted to the technical optimum; in this case,

$$W_{PCz}(S) = 5000(1 + 0.5S)/(1 + 0.05S).$$

***For the control loop at X coordinate*:**

$K_{tt} = 4 \cdot 10^{-5}$ (H·s^2)/rad^2; $K_{vt} = 50$ 1/(kg·m^2); $K_{vx} = 50$ (1/s); $T_{vx} = 2$ s; $K_{SSt} = 1$; $T_{SSt} = 0.01$ s; $K_{PSt} = 1$; $T_{PSt} = 0.05$ s; $K_{PSx} = 1$; $T_{PSx} = 0.05$ s. The speed loop at T coordinate is adjusted to the technical optimum; in this case,

$$W_{SCt}(S) = 500.$$

The position loop at T coordinate is adjusted to the technical optimum; in this case,

$$W_{PCt}(S) = 500.$$

The position loop at X coordinate is adjusted to the technical optimum; in this case,

$$W_{PCx}(S) = 0.1(1 + 0.5S)/(1 + 0.05S).$$

***For the control loop at Y coordinate*:**

$K_{tk} = 4 \cdot 10^{-5}$ (H·s^2)/rad^2; $K_{vk} = 50$ 1/(kg·m^2); $K_{vy} = 50$ (1/s); $T_{vy} = 2$ s; $K_{SSk} = 1$; $T_{SSk} = 0.01$ s; $K_{PSk} = 1$; $T_{PSk} = 0.05$ s; $K_{PSy} = 1$; $T_{PSy} = 0.05$ s.

The speed loop at K coordinate is adjusted to the technical optimum; in this case, $W_{SCk}(S) = 500$.

The position loop at K coordinate is adjusted to the technical optimum; in this case, $W_{PCk}(S) = 500$.

The position loop at Y coordinate is adjusted to the technical optimum; in this case,

$$W_{PCy}(S) = 0.1(1 + 0.5S)/(1 + 0.05S).$$

***For the control loop at R coordinate*:**

$K_{tr} = 4 \cdot 10^{-5}$ (H·s^2)/rad^2; $K_{vr} = 50$ 1/(kg·m^2); $K_{SSr} = 1$; $T_{SSk} = 0.01$ s; $K_{PSr} = 1$; $T_{PSr} = 0.01$ s.

The speed loop at R coordinate is adjusted to the technical optimum; in this case, $W_{SCr}(S) = 500$.

The position loop at R coordinate is adjusted to the technical optimum; in this case, $W_{PCr}(S) = 500$.

Since all coordinate control systems of QC perform their functions through 4 actuating motors, there is a strong mutual influence between them to be considered in using.

In the paper, rather the simple but illustrative operation algorithm for the QC IMCS has been investigated: the QC gains a height of 1 m; after 1 s, the QC moves to the right by 1 m; then the QC moves to the left by 1 m. In this case, the control systems for Z coordinate, X coordinate, and T coordinate (as intermediate one) are activated in the QC. When the control system is in operation, it is essential to observe the influence of a single coordinate (e.g., X) on the control system operating at another coordinate (e.g., Z), one after one. This influence should be minimal with the controllers selected properly.

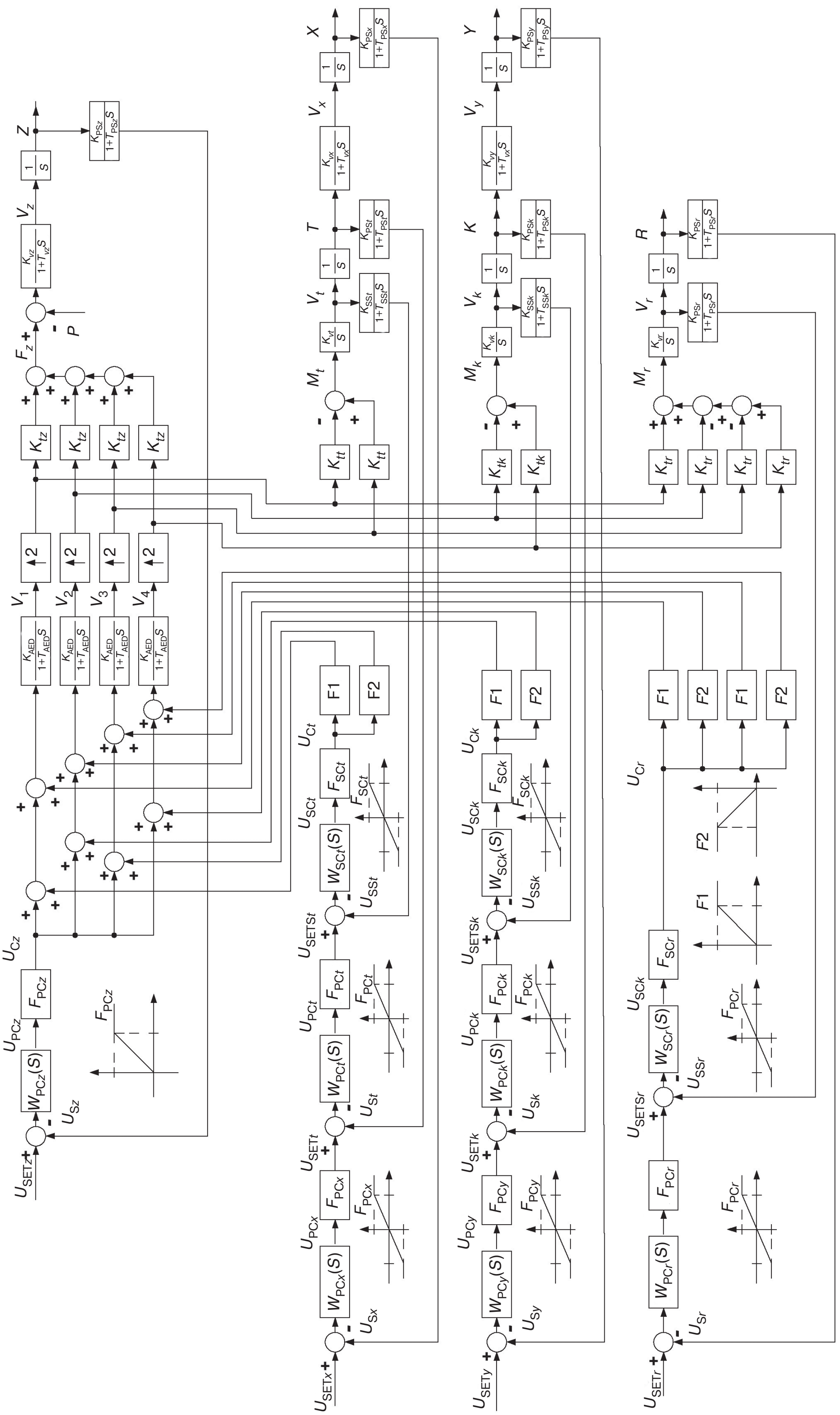


Fig. 4. A structural diagram for the QC IMCS

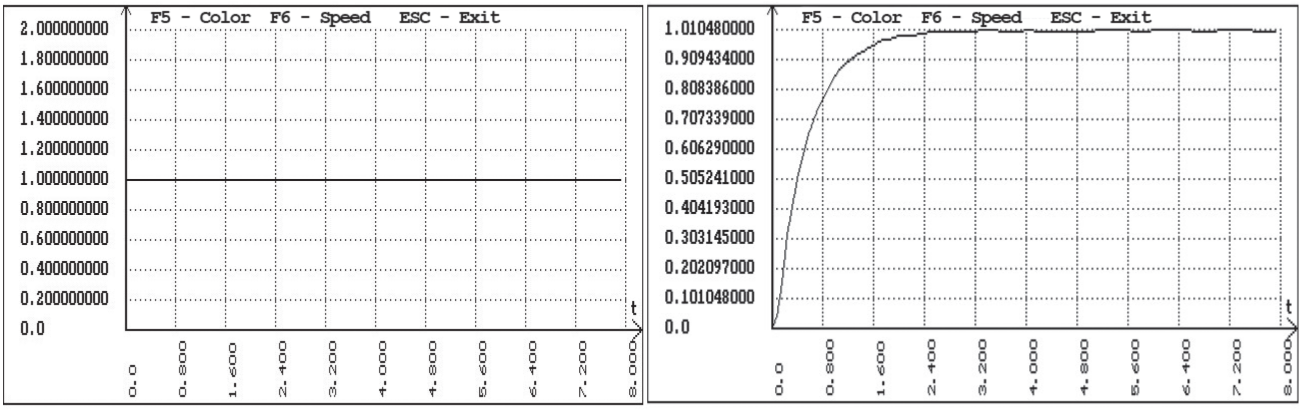
SIMULATION RESULTS

The behavior of the synthesized QC IMCS is simulated by applying the step signal having amplitude of 1 m to the input of the Z coordinate control system as well as the meander-type signal having amplitude of 1 m and frequency of 0.125 Hz to the input of the X coordinate control system after a time equal to 1 s. The simulation results are shown in Figs. 5–7.

The analysis of Figs. 5–7 shows that the synthesized IMCS as a part of the QC has the good control characteristics that are the absence of the overcontrol and the low static error (less than 10 mm). It should be noted that such good results have been obtained with the errors of the QC coordinate parameter sensors neglected. When they are taken into account, static errors increase significantly.

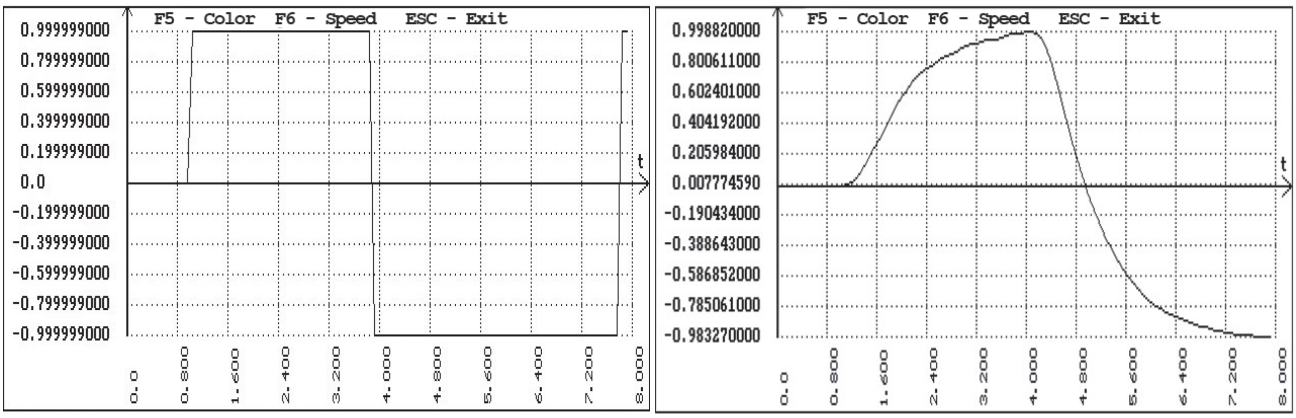
In addition, the regulators of the QC IMCS coordinate control system neutralize the mutual influence of coordinates.

With the parameters of coordinate controllers changing towards performance degradation of control



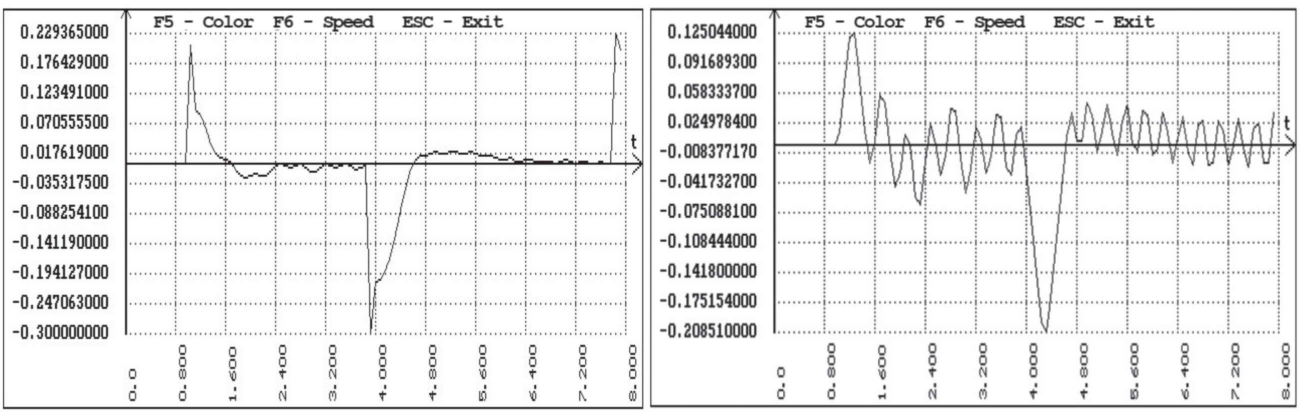
(a) (b)

Fig. 5. Position setting signals for Z coordinate (a) and actual position signals (b)



(a) (b)

Fig. 6. Position setting signals for X coordinate (a) and actual position signals (b)



(a) (b)

Fig. 7. Position setting signals for T (a) coordinate (a) and actual position signals (b)

loops, the effects of mutual influence of coordinates manifesting in shifts of some coordinates during operation of other coordinates, etc., are evidenced.

CONCLUSIONS

Based on the above, the following conclusions can be drawn:

- it would be advisable to implement the QC IMCS in the form of four control loops for the *X*, *Y*, *Z*, and *R* coordinates;
- control loops for the *T* and *K* coordinates should be subordinated to the position controllers of control loops for the *X* and *Y* coordinates;
- control loops for coordinates should be adjusted to the optimum modulus, thus obtaining smooth transients (without overcontrolling) and practically excluding the mutual influence of coordinates.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Bondarev A.N., Kirichek R.V. Overview of unmanned aerial apparatus for general use and regulation of air UAV movement in different countries. *Informatsionnye tekhnologii i telekommunikatsii = Telecom IT.* 2016;4(4):13–23 (in Russ.). Available from URL: https://www.sut.ru/doci/nauka/review/20164/13-23.pdf
2. Podrezov Yu.V. Features of the use of unmanned aerial vehicles to monitor the forest fire emergency situations. *Problemy bezopasnosti i chrezvychainykh situatsii = Safety and Emergencies Problems*. 2019;3:64–72 (in Russ.). Available from URL: http://lamb.viniti.ru/sid2/sid2free?sid2=J17785461
3. Juniper A. *The Complete Guide to Drones: Whatever your budget - Build + Choose + Fly + Photograph.* New York: Wellfleet Press; 2016. 144 p.
   [Juniper A. *Drones. Complete practical guide*. Moscow: KoLibri; 2019. 159 p. (in Russ.). ISBN 978-5-389-15611-1]
4. Gololobov V.N., Ul'yanov V.I. *Bespilotniki dlya lyuboznatel'nykh* (*Drones for the curious*). Saint-Petersburg: Nauka i Tekhnika (NiT); 2018. 249 p. (in Russ.). ISBN 978-5-94387-878-7
5. Yushenko A.S., Lebedev K.R., Zabihafar S.H. The adaptive neural network control of quadrotor helicopter. *Nauka i Obrazovanie: nauchnoe izdanie MGTU im. N.E. Baumana. = Science & Education. Scientific Education of Bauman MSTU*. 2017;7:262–267 (in Russ.).
6. Luukkonen T. *Modelling and Control of Quadcopter. School of Science*. Espoo; 2011. 26 p. Available from URL: https://sal.aalto.fi/publications/pdf-files/eluu11_public.pdf
7. Rokachevsky O.A., Sleptsov V.V., Ablaeva A.E., Dinh Ba Phuong. Synthesis of the height coordinate control loop of the information-measuring and control system of the quadrocopter. *Nauchno-tekhnicheskii vestnik Povolzh'ya. = Scientific and Technical Volga Region Bulletin*. 2021;1:38–43 (in Russ.).
8. Chulin N.A., Mironova I.V. Control system of unmanned aerial vehicle. *Inzhenernyi zhurnal: nauka i innovatsii = Engineering J: Science and Innovation*. 2018;9(81):1–11 (in Russ.). https://doi.org/10.18698/2308-6033-2018-9-1802
9. Kulikov G.V., Tambovsky S.S. Assessment of the quality of communication with the UAV in urban terrain. *Vestnik MGTU MIREA = Herald of MSTU MIREA*. 2015;1(6):205–217 (in Russ.).

СПИСОК ЛИТЕРАТУРЫ

1. Бондарев А.Н., Киричек Р.В. Обзор беспилотных летательных аппаратов общего пользования и регулирования воздушного движения БПЛА в разных странах. *Информационные технологии и телекоммуникации*. 2016;4(4):13–23. URL: https://www.sut.ru/doci/nauka/review/20164/13-23.pdf
2. Подрезов Ю.В. Особенности применения беспилотных летательных аппаратов для мониторинга чрезвычайных лесопожарных ситуаций. *Проблемы безопасности и чрезвычайных ситуаций*. 2019;3:64–72. URL: http://lamb.viniti.ru/sid2/sid2free?sid2=J17785461
3. Джунипер А. Дроны. *Полное практическое руководство*: пер. с англ. М.: КоЛибри; 2019. 159 с. ISBN 978-5-389-15611-1
4. Гололобов В.Н., Ульянов В.И. *Беспилотники для любознательных*. СПб: Наука и Техника (НиТ); 2018. 249 с. ISBN 978-5-94387-878-7
5. Ющенко А.С., Лебедев К.Р., Забихафар С.Х. Система управления квадрокоптером на основе адаптивной нейронной сети. *Наука и Образование: научное издание МГТУ им. Н.Э. Баумана*. 2017;7:262–267.
6. Luukkonen T. *Modelling and Control of Quadcopter. School of Science*. Espoo; 2011. 26 p. URL: https://sal.aalto.fi/publications/pdf-files/eluu11_public.pdf
7. Рокачевский О.А., Слепцов В.В., Аблаева А.Е., Динь Ба Фыонг. Синтез контура управления координатой высоты информационно-измерительной и управляющей системы квадрокоптера. *Научно-технический вестник Поволжья*. 2021;1:38–43.
8. Чулин Н.А., Миронова И.В. Система управления беспилотным летательным аппаратом. *Инженерный журнал: наука и инновации*. 2018;9(81):1–11. https://doi.org/10.18698/2308-6033-2018-9-1802
9. Куликов Г.В., Тамбовский С.С. Оценка качества связи с БПЛА в условиях городской застройки. *Вестник МГТУ МИРЭА*. 2015;1(6):205–217.
10. Соловьев В.В., Финаев В.И., Шадрина В.В., Пушкина И.В. Навигация для обеспечения коммуникаций в группе подвижных объектов. *Известия ЮФУ. Технические науки*. 2018;3(197):102–112. https://doi.org/10.23683/2311-3103-2018-3-102-112
11. Чугунов М.В., Полунина И.Н., Попков М.А. Проектирование квадрокоптера на базе интегрированной модельной среды. *Инженерные технологии и системы*. 2019;29(2):169–186. https://doi.org/10.15507/2658-4123.029.201902.169-186

10. Solov'ev V.V., Finaev V.I., Shadrina V.V., Pushkina I.V. Navigation for communication in a group of vehicles. *Izvestiya YUFU. Tekhnicheskie nauki = Izvestiya SFedU. Engineering Sciences*. 2018;3(197):102–111 (in Russ.). https://doi.org/10.23683/2311-3103-2018-3-102-112
11. Chugunov M.V., Polunina I.N., Popkov M.A. The quadrocopter design based on integrated model environment. *Inzhenernye tekhnologii i sistemy = Engineering technologies and systems.* 2019;29(2):169–186 (in Russ.). https://doi.org/10.15507/2658-4123.029.201902.169-186
12. Vysocky V.E., Voronin S.M., Gorshkov R.G. Simulation of electromagnetic and electromechanical processes of thyratron motor. *Vestnik IGEU = Vestnik* of *Ivanovo Power Engineering Institute*. 2011;1:54–57 (in Russ.). Available from URL: http://vestnik.ispu.ru/sites/vestnik.ispu.ru/files/publications/54-57_2.pdf
13. Savickij A.V., Pavlovsky V.E. Model of quadrocopter and algorithm of vehicle control based on neural network. *Preprinty IPM im. M.V. Keldysha = Keldysh Institute Preprints*. 2017;77. 20 p. (in Russ.). https://doi.org/10.20948/prepr-2017-77
14. Pavlovske V.E., Savickij A.V. The solving of the inverse problem for computation control action for quadrocopter. *Preprinty IPM im. M.V. Keldysha = Keldysh Institute Preprints*. 2017;17. 19 p. (in Russ.). http://doi.org/10.20948/prepr-2017-17
15. Krasovskij A.N., Suslova O.A. On the optimal control of the movement of a drone–quadrocopter by the criterion of the quality of energy consumption. *Uspekhi sovremennoi nauki i obrazovaniya = Achievements of modern science and education*. 2017;4(3):193–197 (in Russ.).
16. Savin S.I., Vorochaeva L.Yu., Savin A.I. Optimal trajectory generation for a quadrotor using convex programming. *Vestnik VGTU = The Bulletin of Voronezh State Technical University*. 2018;14(5):54–63 (in Russ.). Available from URL: https://cchgeu.ru/science/nauchnye-izdaniya/vestnik-voronezhskogo-gosudarstvennogo-tekhnicheskogo-universiteta-/bulletin-of-voronezh-state-technical-university/the-issues-of-the-journal-the-bulletin-of-voronezh-state-technical-university/14_5.pdf
17. Garipov V.K., Prokuratov I.I., Sleptsov V.V. *IIUS elektroprivodov dlya avtomatizirovannogo proizvodstva* (*Information-Measuring and Control Systems of Electric Drives for Automated Production*). Saarbrucken: Lap Lambert, Academic Publishing; 2014. 131 p. (in Russ.).

12. Высоцкий В.Е., Воронин С.М., Горшков Р.Г. Имитационное моделирование электромагнитных и электромеханических процессов вентильного двигателя. *Вестник ИГЭУ*. 2011;1:54–57. URL: http://vestnik.ispu.ru/sites/vestnik.ispu.ru/files/publications/54-57_2.pdf
13. Савицкий А.В., Павловский В.Е. Модель квадрокоптера и нейросетевой алгоритм управления. *Препринты ИПМ им. М.В. Келдыша*. 2017;77. 20 с. https://doi.org/10.20948/prepr-2017-77
14. Павловский В.Е., Савицкий А.В. Исследование обратной задачи для вычисления управляющих воздействий для квадрокоптера. *Препринты ИПМ им. М.В. Келдыша*. 2017;17. 19 с. http://doi.org/10.20948/prepr-2017-17
15. Красовский А.Н., Суслова О.А. Об оптимальном управлении движением дрона-квадрокоптера по критерию качества затрат энергии. *Успехи современной науки и образования*. 2017;4(3):193–197.
16. Савин С.И., Ворочаева Л.Ю., Савин А.И. Метод генерации оптимальных траекторий квадрокоптера посредством выпуклого программирования. *Вестник ВГТУ*. 2018;14(5):54–63. URL: https://cchgeu.ru/science/nauchnye-izdaniya/vestnik-voronezhskogo-gosudarstvennogo-tekhnicheskogo-universiteta-/bulletin-of-voronezh-state-technical-university/the-issues-of-the-journal-the-bulletin-of-voronezh-state-technical-university/14_5.pdf
17. Гарипов В.К., Прокуратов И.И., Слепцов В.В. *ИИУС электроприводов для автоматизированного производства*. Саарбрюккен: Lap Lambert, Academic Publishing; 2014. 131 с.

About the authors

Vladimir V. Sleptsov, Dr. Sci. (Eng.), Professor, Department of Instruments, Information and Measurement Systems, Institute of Integrated Safety and Special Instrument Engineering, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: sleptsov@mirea.ru. https://orcid.org/0000-0002-1003-8411

Vyacheslav L. Afonin, Dr. Sci. (Eng.), Chief Researcher, Mechanical Engineering Research Institute, Russian Academy of Sciences (4, Maly Kharitonyevsky pereulok, Moscow, 101990 Russia). E-mail: afoninwl@rambler.ru. https://orcid.org/0000-0003-3824-0707

Anna E. Ablaeva, Senior Lecturer, Department of Instruments, Information and Measurement Systems, Institute of Integrated Safety and Special Instrument Engineering, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: ablaeva@mirea.ru. https://orcid.org/0000-0002-1767-5413

Dinh Ba Phuong, Postgraduate Student, Department of Instruments, Information and Measurement Systems, Institute of Integrated Safety and Special Instrument Engineering, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: nhoht_ck49@yahoo.com. https://orcid.org/0000-0001-6153-4776

Об авторах

Слепцов Владимир Владимирович, д.т.н., профессор, кафедра «Приборы и информационно-измерительные системы» Института комплексной безопасности и специального приборостроения ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: sleptsov@mirea.ru. https://orcid.org/0000-0002-1003-8411

Афонин Вячеслав Леонидович, д.т.н., главный научный сотрудник, Институт машиноведения им. А.А. Благонравова Российской академии наук (101990, Россия, Москва, Малый Харитоньевский переулок, д. 4). E-mail: afoninwl@rambler.ru. https://orcid.org/0000-0003-3824-0707

Аблаева Анна Евгеньевна, старший преподаватель, кафедра «Приборы и информационно-измерительные системы» Института комплексной безопасности и специального приборостроения ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: ablaeva@mirea.ru. https://orcid.org/0000-0002-1767-5413

Динь Ба Фыонг, аспирант, кафедра «Приборы и информационно-измерительные системы» Института комплексной безопасности и специального приборостроения ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: nhoht_ck49@yahoo.com. https://orcid.org/0000-0001-6153-4776

Translated by K. Nazarov

Russian Technological Journal. 2021;9(6):37–45

ISSN 2500-316X (Online)

UDC 621.396.621
https://doi.org/10.32362/2500-316X-2021-9-6-37-45

RESEARCH ARTICLE

Very high frequency radio receiver preselector design

Egor V. Gurov [1],
Saygid U. Uvaysov [2, @],
Victoria V. Chernoverskaya [2],
Ruslan M. Uvaysov [2]

[1] *Siemens LLC, Moscow, 115184 Russia*
[2] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[@] *Corresponding author, e-mail: uvaysov@yandex.ru*

Abstract.
Objectives. The quality of a radio receiver preselector largely determines its main characteristics, including sensitivity. A preselector usually consists of linear elements: inductors, capacitors, low noise amplifiers, and switches. At high frequencies, the components cannot be considered as ideal ones, since active and reactive parasitic parameters significantly affect the frequency response of the components and, as a consequence, the network. Therefore, simulation of the networks requires more sophisticated component models, which take into account parasitic parameters. However, if refined components models are applied, it is still possible to obtain unsatisfactory results, since interconnections and footprints pads also affect the frequency response. This is true even if short traces with a length of about 5 mm are used at frequencies of about 100 MHz. These features must be taken into account for RF network design. The purpose of the work is to ensure the required characteristics of the preselector in the design process based on computer simulation.
Methods. Egor Gurov's methodology for analog VHF LC-filters was applied to radio receiver preselector design. The methodology contains the methods of discrete optimization, Monte-Carlo method, momentum analysis with Green's functions. Experimental results were obtained by prototype implementation and measurement with a vector network analyzer.
Results. The article presents the preselector design process. The preselector contains two analog switches, an analog band-pass filter, an analog high-pass filter, and a low-noise amplifier. Simulation and experimental results with their comparison are presented in the article.
Conclusions. Satisfactory results were obtained. It means that the Egor Gurov's method for analog filters design can be applied for more complex networks design such as radio receiver preselectors.

Keywords: radio receiver, preselector, SPICE model, scattering matrix, mathematical modeling, electronic component, microstrip line, frequency range, parasitic parameters

• **Submitted:** 28.09.2021 • **Revised:** 12.10.2021 • **Accepted:** 19.10.2021

For citation: Gurov E.V., Uvaysov S.U., Chernoverskaya V.V., Uvaysov R.M. Very high frequency radio receiver preselector design. *Russ. Technol. J.* 2021;9(6):37–45. https://doi.org/10.32362/2500-316X-2021-9-6-37-45

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Проектирование преселектора радиоприемника в диапазоне метровых волн

Е.В. Гуров [1]**,**
С.У. Увайсов [2, @]**,**
В.В. Черноверская [2]**,**
Р.М. Увайсов [2]

[1] *ООО «Сименс», Москва, 115184 Россия*
[2] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[@] *Автор для переписки, e-mail: uvaysov@yandex.ru*

Резюме.
Цели. Качество преселектора радиоприемника в значительной мере определяет его основные характеристики, в том числе чувствительность. Преселектор обычно состоит из линейных элементов – катушек индуктивности, конденсаторов, малошумящих усилителей и ключей. При работе на высоких частотах нельзя считать эти компоненты идеальными, так как активные и реактивные паразитные параметры вносят значительный вклад в частотную характеристику цепей. Поэтому для моделирования высокочастотных схем требуются более сложные модели компонентов, учитывающие паразитные составляющие. Однако при применении для всех компонентов уточненных моделей или S-параметров вероятность получения неудовлетворительных результатов сохраняется, поскольку соединительные линии и контактные площадки также вносят заметные искажения в частотную характеристику. Это наблюдается и для коротких линий длиной около 5 мм, которые оказывают влияние на частотах порядка 100 МГц. Поэтому подобные явления необходимо учитывать при разработке. Цель работы – обеспечение требуемых характеристик преселектора в процессе его автоматизированного схемотехнического проектирования на основе компьютерного моделирования.
Методы. Использована методика Гурова Е.В. для проектирования аналоговых фильтров диапазона метровых волн, но применительно к преселектору радиоприемника. Она включает в себя методы дискретной оптимизации, имитационного моделирования Монте-Карло, метод моментов для электромагнитного моделирования с использованием функции Грина. Экспериментальные результаты получены путем изготовления опытного образца и измерений с помощью векторного анализатора цепей.
Результаты. Представлен процесс проектирования преселектора радиоприемника диапазона метровых волн, состоящего из двух аналоговых ключей, полосового фильтра на сосредоточенных компонентах, малошумящего усилителя и фильтра высоких частот. Приведены результаты моделирования и экспериментальные результаты, их сравнение.
Выводы. Проведенные численные и натурные исследования позволяют говорить о том, что методика Гурова Е.В. для проектирования аналоговых LC-фильтров в диапазоне метровых волн применима и для более сложных цепей, таких как преселектор радиоприемника.

Ключевые слова: радиоприемник, преселектор, SPICE-модель, матрица рассеивания, математическое моделирование, электронный компонент, микрополосковая линия, диапазон частот, паразитные параметры

• **Поступила:** 28.09.2021 • **Доработана:** 12.10.2021 • **Принята к опубликованию:** 19.10.2021

Для цитирования: Гуров Е.В., Увайсов С.У., Черноверская В.В., Увайсов Р.М. Проектирование преселектора радиоприемника в диапазоне метровых волн. *Russ. Technol. J.* 2021;9(6):37−45. https://doi.org/10.32362/2500-316X-2021-9-6-37-45

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The maximum sensitivity of a radio receiver is determined by its noise level. Careful design of radio receiver units is required to achieve high sensitivity [1, 2]. The preselector is connected directly to the antenna and has the greatest influence on the intrinsic noise. This follows from the Friis formula for the noise figure in the case of a series connection of several stages:

$$F_{\text{total}} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \ldots + \frac{F_n - 1}{G_1 G_2 \ldots G_{n-1}}, \quad (1)$$

where F_{total} is the total output noise; n is the number of stages; F_1 is the noise figure of the first stage; G_1 is the power gain of the first stage, etc. The values of the noise coefficients and gains are dimensionless quantities.

It can be noted that the noise figure of the first stage is completely presented in the total output noise, and the noise figure of the second stage is divided by the gain of the first stage [1]. Hence, for maximum sensitivity, it is necessary to place a low-noise amplifier next to the antenna. However, as it was shown in [3], this approach does not always provide the minimum intrinsic noise level. In an environment with a large number of powerful emitting radio stations, such as large cities, a wideband low-noise amplifier may produce higher output noise than expected. The solution to this problem is a bandpass filter installation before the low-noise amplifier.

In the frequency range from several tens of megahertz to gigahertz, to obtain an arbitrary frequency response (arbitrary bandwidth and stopband), it is suitable to use passive inductor–capacitor (LC) filters. The design process for such filters is fully automated at the moment (for example, using *Nuhertz Technologies FilterSolutions®* software). Here, the well-known frequency response approximation method is used, which provides an electrical circuit of the filter with nominal values of the components, as well as a transfer function [4–7].

This method assumes that ideal passive components (inductors and capacitors) are applied and traces between the components do not affect the transfer function of the designed filter. However, at frequencies above 10 MHz, non-ideal components and topology on a printed circuit board have a noticeable influence on the frequency response [8]. *Nuhertz Technologies FilterSolutions®* software can take a quality factor (Q-factor) of the components into account, and even replace them with S-parameters. However, it is not possible to take into account the printed circuit board topology influence. *Keysight Genesys* software tried to automate the filter design process, taking into account both parasitic parameters of the components and printed circuit board topology. The method for designing LC filters in *Keysight Genesys* software is presented in [9]. It takes into account parasitic parameters of the components, their tolerances, and printed circuit board topology.

A preselector of a radio receiver, besides the filter, may contain a low-noise amplifier, which usually has an uneven (slope-down) frequency response, and its input and output impedances may differ from the standard 50 Ohm. It can also be used different bandpass filters selectable by analog switches.

This article presents an example of a radio receiver preselector design. The preselector consists of analog switches that switch the bandpass filters, one bandpass filter, a low-noise amplifier, and one high-pass filter. Based on the simulation results, a prototype was implemented. Experimental results were obtained with a vector network analyzer [10]. The obtained data can be considered as an application extension of the method [9] for the synthesis of more complex circuits than LC filters.

PROBLEM STATEMENT

The preselector for the superhetherodyne radio receiver with an operating bandwidth of 114…122 MHz is under consideration. The maximum sensitivity is required at the frequency of 121.5 MHz, and an intermediate frequency of 87 MHz on the radio frequency (RF) input must be additionally suppressed. It is supposed to use three sub-bands, but within the framework of this article, only one sub-band is considered taking into account the effect of the analog switches. The block diagram of the preselector is shown in Fig. 1.

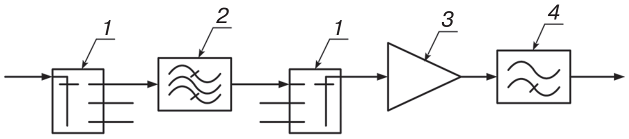


Fig. 1. Block diagram of the preselector: (*1*) analog switch, (*2*) bandpass filter, (*3*) low-noise amplifier, and (*4*) high-pass filter

Peregrine Semiconductor's (USA) PE42430 radio frequency switch was used for bandpass filter switching, and Qorvo's (USA) TQP3M9035 was used as a low-noise amplifier. A high-pass filter is required to suppress unwanted input signals at the intermediate frequency. It also increases the slope and improves the suppression of powerful broadcasting stations in the range of 88...108 MHz. The bandpass filter should also suppress the image frequency by at least 50 dB.

Thus, the requirements for the entire network can be described by the following expressions:

$$\begin{cases} A(f) > A_{\text{pass}}, & f_{\text{pass1}} \le f \le f_{\text{pass2}}, \\ A(f) < A_{\text{stop}}, & f < f_{\text{IF}},\ f > f_{\text{IM}}, \\ K(f) > 1, & \\ B_1(f) > 0, & \end{cases} \quad (2)$$

where $A(f)$ is the frequency response of the circuit (gain vs. frequency); f is the frequency; A_{pass} is the minimum allowable signal gain in the passband; A_{stop} is the maximum allowable signal gain in the stopband; f_{pass1} and f_{pass2} are the minimum and the maximum frequencies of the passband respectively, f_{IF} is the intermediate frequency; f_{IM} is the minimum frequency of the image signal; K and B_1 are the stability coefficients that are described on the Keysight Knowledge Center site[1]. Since the network has an active component, it is also necessary to make sure that self-oscillation will not occur; therefore, the extra stability requirements are imposed. Constraints on the stability coefficients must be met at all the frequencies. Here, considering the simulation frequencies only is enough, since there will be no self-oscillations of the networks if no signal gain. The stability coefficients application is limited by the networks with only one active amplifier stage.

SIMULATION OF THE PRESELECTOR

The preselector simulation was carried out in the *Keysight Genesys* 2018 software. The models of the RF switches and low-noise amplifier were downloaded from the manufacturer's official websites. The software allows network simulation with S-parameters, SPICE models, carrying out electromagnetic analysis of the printed circuit board topology using the method of moments. Thus, both the parasitic parameters of the components and the printed circuit board topology are taken into account.

The functional diagram of the preselector is shown in Fig. 2. To minimize the size of the preselector and meet the conditions (2), a third-order bandpass filter was applied. A signal with the intermediate frequency is suppressed using the LC-shunt connected between the capacitors at the output of the low-noise amplifier.

All the capacitors are of standard size of 0402, inductors are of the standard size of 0603 or 1206. Blocking capacitors C1 and C6 have a capacitance of 0.01 μF.

As the next step, the schematic and topology were set up in the *Keysight Genesys* environment. All the components were connected with microstrip lines. The topological model is shown in Fig. 3. In addition to metallization and contact pads, it also contains the contours of the components on the silkscreen layer.

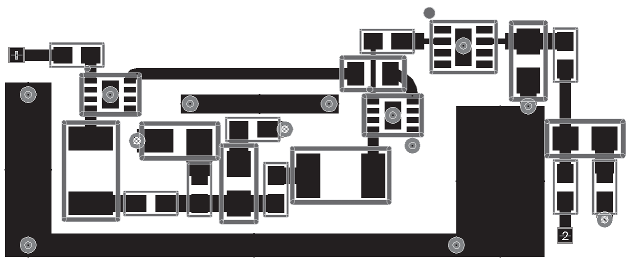


Fig. 3. The topology of the preselector

[1] URL: https://edadocs.software.keysight.com/genesys2018/simulation/getting-started-with-genesys-simulation/getting-started-with-measurements/stability-factor-and-measure. Accessed October 25, 2021.

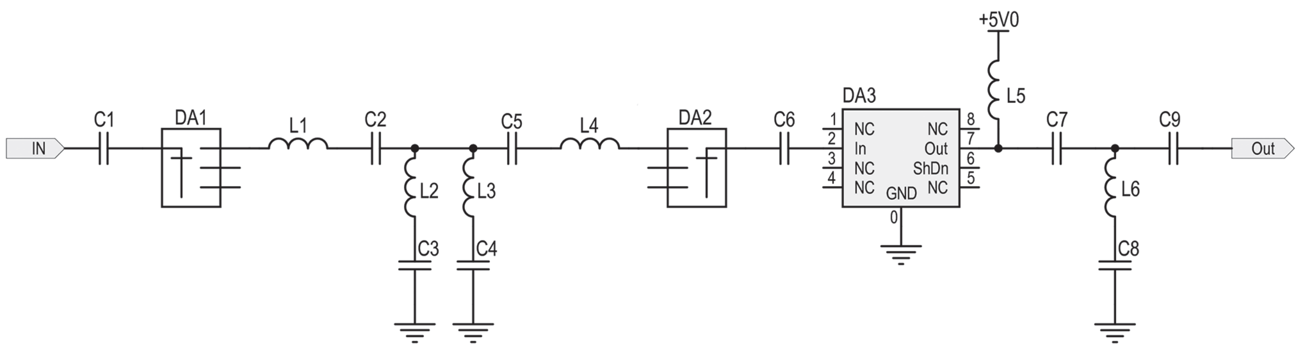


Fig. 2. Functional diagram of the preselector

Table. Preselector bill of components

Components	Nominal value	Part number	Manufacturer
C1, C6, C9	0.01 μF 5%	GRM1555C1E103JE01	Murata (Japan)
C2	3.9 ± 0.1 pF	GJM1555C1H3R9BB01	Murata
C3, C4, C8	43 pF 2%	GJM1555C1H430GB01	Murata
C5	3.3 ± 0.05 pF	GJM1555C1H3R3WB01	Murata
C7	33 pF 5%	GJM1555C1H330JB01	Murata
DA1, DA2	–	PE42430	Peregrine (USA)
DA3	–	TQP3M9035	Qorvo (USA)
L1	390 nH 2%	1206CS-391XGEB	Coilcraft (USA)
L2	75 nH 2%	LQW18AN75NG80	Murata
L3	8.2 ± 0.2 nH	LQW18AN8N2C80	Murata
L4	470 nH 2%	1206CS-470XGEB	Coilcraft
L5	47 nH 5%	LQW18AN47NJ80	Murata
L6	82 nH 2%	LQW18AN82NG80	Murata

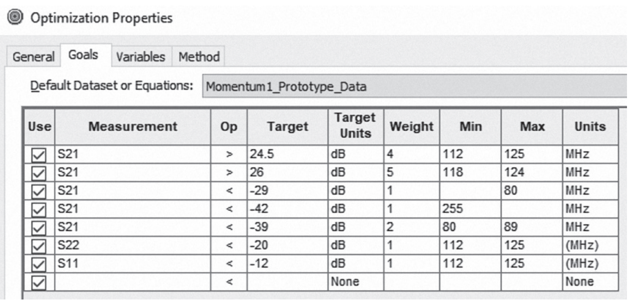
The models with Q-factor were initially chosen as the models of inductors and capacitors. The Q-factor model consists of an ideal capacitor or ideal inductor (for a capacitor and inductor, respectively) with a series connection of an ideal resistor (i.e., active resistance), it is also known as equivalent series resistance (ESR). This allows tuning the nominal values of the component, which may not be possible for the more complex and accurate models. Different components have different ESR values. At the current stage, it was assumed that the ESR is constant during the variation of the component's nominal values[2, 3] [11, 12]. The initial values of the components were obtained using the *Keysight Genesys* built-in automatic filter synthesis tools.

The nominal values of the components were calculated by optimization in accordance with the methodology [9]. This methodology first employs simple component models and selects component values. After the optimization, sensitivity analysis is conducted and the least sensitive component is replaced with the more complex model, which does not allow freely changing the value of the main parameter in a wide range. Then the optimization must be repeated. And so on until all the components are replaced with more complex models.

[2] Ceramic capacitors MLCC: application features. URL: https://www.compel.ru/lib/56541. Accessed October 25, 2021. (in Russ.).

[3] The physical meaning of the group delay of the filter. Digital filters with a linear phase-frequency characteristic. Theory and practice of digital signal processing. URL: http://www.dsplib.ru/content/filters/linphase/linphase.html. Accessed October 25, 2021. (in Russ.).

As a result, all models of the inductors and capacitors were replaced with the corresponding S-parameters provided by the manufacturers. The boundary conditions (2) were set up in the *Keysight Genesys* software, as shown in Fig. 4. Weights have been introduced to get the desired shape of the frequency response. Initially, they were all equal to 1 but then were manually changed during the optimization process. The criterion error was multiplied by the corresponding weight. The optimization goal was to minimize the total error.



Use	Measurement	Op	Target	Target Units	Weight	Min	Max	Units
☑	S21	>	24.5	dB	4	112	125	MHz
☑	S21	>	26	dB	5	118	124	MHz
☑	S21	<	-29	dB	1		80	MHz
☑	S21	<	-42	dB	1	255		MHz
☑	S21	<	-39	dB	2	80	89	MHz
☑	S22	<	-20	dB	1	112	125	(MHz)
☑	S11	<	-12	dB	1	112	125	(MHz)
☑		<		None				None

Fig. 4. Boundary conditions for the optimization

The simulation results are presented below. Figure 5 shows the frequency response (S21) and the return losses (S11, S22). The shaded areas display the boundary conditions indicated in Fig. 4. Figure 6 shows the stability coefficients (K, B1) and the group delay (GD) of the entire network[3]. In this case, there were no requirements imposed for the GD, and it is noted just for information.

The applied components for the simulation are listed in Table.

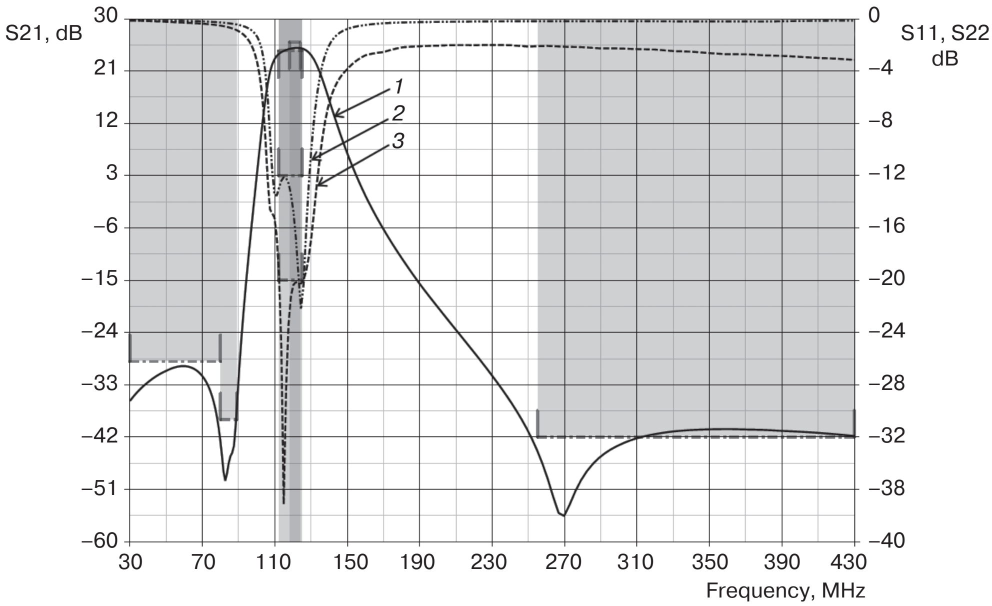


Fig. 5. Frequency response and return losses of the preselector: (*1*) frequency response, (*2*) return losses (S11), and (*3*) return losses (S22)

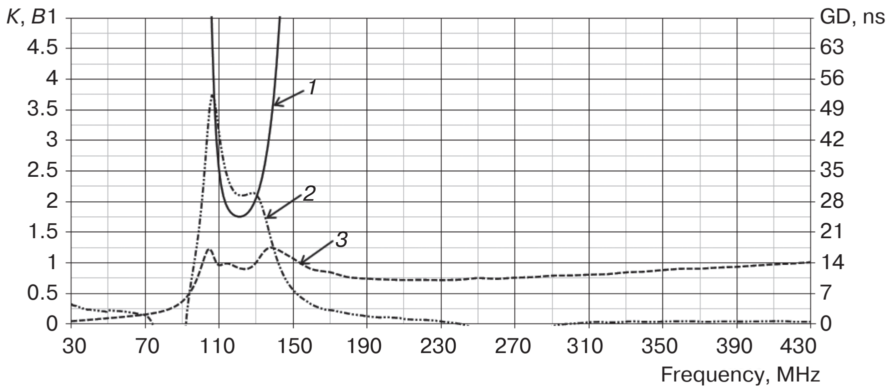


Fig. 6. Stability coefficients and GD of the preselector: (*1*) stability coefficient K (must be greater than 1), (*2*) GD, and (*3*) stability coefficient B1 (must be greater than 0)

EXPERIMENTAL RESULTS

The preselector was implemented on the 4-layer printed circuit board made of FR-4 fiberglass. The topology in Fig. 3 was performed on the top layer, the nearest inner layer was a solid polygon connected to the “ground” reference layer [8]. The input signal was delivered from the SMA connector mounted on the printed circuit board, the output signal was taken through the U.FL-R-SMT connector. The frequency response was obtained using Rohde and Schwarz ZVA8 vector network analyzer.

The capacitors of the GRM series were installed on the prototype instead of the GJM series capacitors because the right capacitors had not been purchased at that moment. GRM series capacitors have the status “not recommended for new designs.” These capacitors have a lower Q-factor; therefore, it could be expected more attenuation in the passband rather than in the simulation results, but without significant distortion of the frequency response curve.

The experimental results and their comparison to the simulated ones are presented in Fig. 7. The figure shows only the frequency response (S21). Return losses (S11 and S22), as well as stability factors, satisfied the required values.

The passband gain in the experimental curve was about 1 dB lower than in the simulation results, as was expected due to the lower Q-factor of the capacitors. In general, both results are in good agreement, especially at low frequencies. In the method [9], it is noted that a better frequency response can be achieved by tuning (i.e., replacing with another one with different nominal value) the most sensitive components on the prototype. It has not been done here, since the frequency response

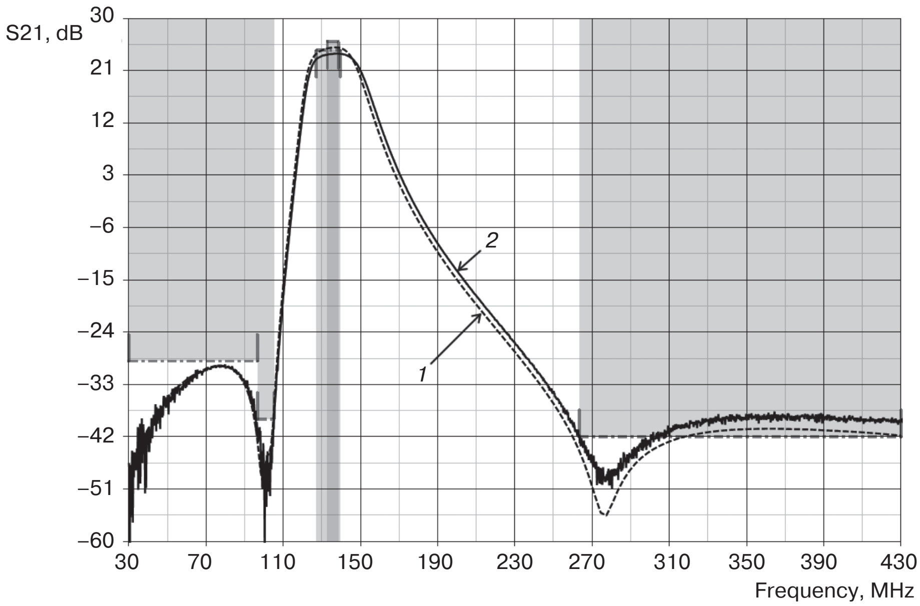


Fig. 7. Frequency response of the preselector: (*1*) simulation and (*2*) experimental data

of the preselector meets the specified requirements. Careful selection of components can be useful as part of a release device.

CONCLUSIONS

The radio receiver preselector is the most important part since it significantly affects its sensitivity. The article describes the design process of the radio receiver preselector in the very high frequency (VHF) range. The preselector consists of analog switches, one bandpass filter, a low-noise amplifier, and an additional high-pass filter. The design takes into account input and output impedances and slope-down frequency response of the analog switches and low-noise amplifier.

The designed preselector was implemented in the prototype, and the experimental data was obtained with the vector network analyzer. The simulated and experimental results are in good agreement. This is shown in Fig. 7. The discrepancy between the gain in the passband of about 1 dB is caused by the capacitors on the prototype with the lower Q-factor than in the model since the capacitors in the model were not available at the time of printed circuit board assembly. In addition, the frequency response can be improved by selecting the most sensitive components directly on the prototype. This operation has not been done, because it was only the prototype, and such selection is reasonable in the release device device.

In [9] it is shown that in the VHF range the frequency response of analog filters is significantly affected by parasitic parameters of the components and interconnection on the printed circuit board topology.

Following the LC-filter design method [9], the radio receiver preselector has been designed. It is shown that the method also allows taking into account the frequency responses of analog switches and low-noise amplifiers, as well as checking the stability of amplifying cascades.

In conclusion, the results presented in the article extend the applicability of the analog LC-filters design method of Egor Gurov for analog filters design [9]. It also allows the design of more complex networks, which can include active components as well.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Brannon B. Basics of designing a digital radio receiver (Radio 101). *Analog Devices*. 1998. URL: https://www.analog.com/en/technical-articles/basics-designing-digital-radio-receiver.html. Accessed October 25, 2021.
2. Gurov E.V., Rotkevich A.S. Experimental evaluation of the sensitivity of a radio receiver for varios structures of the input circuit. In: *Mezhvuzovskaya nauchno-tekhnicheskaya konferentsiya studentov, aspirantov i molodykh spetsialistov imeni E.V. Armenskogo: materialy konferentsii* (*Collection of materials of the Interuniversity scientific and Technical Conference of students, Postgraduates and Young Professionals named after E.V. Armensky*). Moscow: MIEM NIU VShE; 2018, p. 135–136. (in Russ.).
3. Gurov E.V. The influence of radio frequency filter on a low-noise amplifier signal to noise ratio degradation: experimental assessment. *Tekhnologii elektromagnitnoi sovmestimosti = Technologies of Electromagnetic Compatibility*. 2017;2(61):49–55 (in Russ.).
4. Paarmann L.D. *Design and Analysis of Analog Filters*: *A Signal Processing Perspective*. NY: Kluwer Academic Publishers; 2001. 440 p.
5. Gurov E.V., Uvaisov S.U., Uvaisova A.S., Uvaisova S.S. Discrete parametric optimization for very high frequency analog filters synthesis. *Kachestvo. Innovatsii. Obrazovanie. = Quality. Innovation. Education*. 2019;2(160):53–63 (in Russ.). https://doi.org/10.31145/1999-513x-2019-2-53-63
6. Smirnov A.V. Method of simultaneous optimization of radio devices performance in frequency and time domains. *Rossiiskii tekhnologicheskii zhurnal = Russian Technological Journal*. 2018;6(6):13–27 (in Russ.). https://doi.org/10.32362/2500-316X-2018-6-6-13-27
7. Hassan A.H. *Design and Size Reduction Analysis of Micro Strip HBP Filters*. Master's Thesis in Electronics. Master's Program in Electronics/Telecommunications. Sweden, Gävle: University of Gävle. January 2015. 68 p.
8. Kechiev L.N. *Proektirovanie pechatnykh plat dlya tsifrovoi bystrodeistvuyushchei apparatury* (*Design of printed circuit boards for digital high-speed equipment*). Moscow: Gruppa IDT; 2007. 616 p. ISBN 978-5-94833-024-2
9. Gurov E., Uvaysova A.S., Uvaysov S.U., Uvaysov R.M. Design Method for Non-Tunable LC-Filters. In: *2020 Moscow Workshop on Electronic and Networking Technologies (MWENT)*. 2020. https://doi.org/10.1109/MWENT47943.2020.9067472
10. Hiebel M. *Fundamentals of Vector Network Analysis*. 3rd ed. München: Rohde & Schwarz GmbH & Co.; 2008. 420 p.
[Hiebel M. *Osnovy vektornogo analiza tsepei*. Moscow: Izdatel'skii dom MEI; 2009. 501 p. ISBN 978-5-383-01305-2]
11. Gurov E.V., Uvaysov S.U., Uvaysova A.S., Uvaysova S.S. Nominal and actual values of inductor and capacitor parameters at high frequencies. *Rossiiskii tekhnologicheskii zhurnal = Russian Technological Journal*. 2019;7(4):44–53 (in Russ.). https://doi.org/10.32362/2500-316X-2019-7-4-44-53

СПИСОК ЛИТЕРАТУРЫ

1. Brannon B. Basics of designing a digital radio receiver (Radio 101). *Analog Devices*. 1998. URL: https://www.analog.com/en/technical-articles/basics-designing-digital-radio-receiver.html, дата обращения 25.10.2021
2. Гуров Е.В., Роткевич А.С. Экспериментальная оценка чувствительности радиоприемника при различных структурах входной цепи. В сб.: *Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов имени Е.В. Арменского*: *материалы конференции*. М.: МИЭМ НИУ ВШЭ; 2018. С. 135–136.
3. Гуров Е.В. Влияние радиочастотного фильтра на ухудшение соотношения сигнал/шум малошумящего усилителя: экспериментальная оценка. *Технологии электромагнитной совместимости*. 2017;2(61):49–55.
4. Paarmann L.D. *Design and Analysis of Analog Filters*: *A Signal Processing Perspective*. NY: Kluwer Academic Publishers; 2001. 440 p.
5. Гуров Е.В., Увайсов С.У., Увайсова А.С., Увайсова С.С. Алгоритм дискретной параметрической оптимизации для решения задач синтеза аналоговых фильтров диапазона очень высоких частот. *Качество. Инновации. Образование*. 2019;2(160):53–63. https://doi.org/10.31145/1999-513x-2019-2-53-63
6. Смирнов А.В. Метод одновременной оптимизации характеристик электрических фильтров в частотной и временной областях. *Российский технологический журнал*. 2018;6(6):13–27. https://doi.org/10.32362/2500-316X-2018-6-6-13-27
7. Hassan A.H. *Design and Size Reduction Analysis of Micro Strip HBP Filters*. Master's Thesis in Electronics. Master's Program in Electronics/Telecommunications. Sweden, Gävle: University of Gävle. January, 2015. 68 p.
8. Кечиев Л.Н. *Проектирование печатных плат для цифровой быстродействующей аппаратуры*. М.: Группа ИДТ; 2007. 616 с. ISBN 978-5-94833-024-2
9. Gurov E., Uvaysova A.S., Uvaysov S.U., Uvaysov R.M. Design method for non-tunable LC-filters. In: *2020 Moscow Workshop on Electronic and Networking Technologies (MWENT)*. 2020. https://doi.org/10.1109/MWENT47943.2020.9067472
10. Хибель М. *Основы векторного анализа цепей*: пер. с англ. М.: Издательский дом МЭИ; 2009. 501 с. ISBN 978-5-383-01305-2
11. Гуров Е.В., Увайсов С.У., Увайсова А.С., Увайсова С.С. Номинальные и эффективные значения параметров катушек индуктивности и конденсаторов на высоких частотах. *Российский технологический журнал*. 2019;7(4):44–53. https://doi.org/10.32362/2500-316X-2019-7-4-44-53
12. Cain J. Parasitic Inductance of Multylayer Ceramic Capacitors. KYOCERA AVX Components Corporation. February 2002. URL: https://www.avx.com/docs/techinfo/CeramicCapacitors/parasitc.pdf

12. Cain J. Parasitic Inductance of Multylayer Ceramic Capacitors. KYOCERA AVX Components Corporation. February 2002. Available from URL: https://www.kyocera-avx.com/docs/techinfo/CeramicCapacitors/parasitc.pdf

About the authors

Egor V. Gurov, Cand. Sci. (Eng.), Leading Engineer (Research and Design), Siemens LLC (9, ul. Bolshaya Tatarskaya, Moscow, 115184 Russia). E-mail: gurov.yegor@gmail.com.

Saygid U. Uvaysov, Dr. Sci. (Eng.), Professor, Head of the Department of Design and Production of Radio-electronic Means, Institute of Radio Engineering and Telecommunication Systems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: uvaysov@yandex.ru. Scopus Author ID 55931417100, ResearcherID H-6746-2015, https://orcid.org/0000-0003-1943-6819

Victoria V. Chernoverskaya, Cand. Sci. (Eng.), Associate Professor, Department of Design and Production of Radio-electronic Means, Institute of Radio Engineering and Telecommunication Systems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: chernoverskaya@mirea.ru.

Ruslan M. Uvaysov, 2nd year Undergraduate Student, Department of Design and Production of Radio-electronic Means, Institute of Radio Engineering and Telecommunication Systems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: uvajsov.r.m@edu.mirea.ru.

Об авторах

Гуров Егор Владимирович, к.т.н., ведущий инженер (Исследования и разработки) ООО «Сименс» (115184, Россия, Москва, ул. Большая Татарская, д. 9). E-mail: gurov.yegor@gmail.com.

Увайсов Сайгид Увайсович, д.т.н., профессор, заведующий кафедрой конструирования и производства радиоэлектронных средств Института радиотехнических и телекоммуникационных систем ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: uvaysov@yandex.ru. Scopus Author ID 55931417100, ResearcherID H-6746-2015, https://orcid.org/0000-0003-1943-6819

Черноверская Виктория Владимировна, к.т.н., доцент, кафедра конструирования и производства радиоэлектронных средств Института радиотехнических и телекоммуникационных систем ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: chernoverskaya@mirea.ru.

Увайсов Руслан Магомедович, магистрант 2-го года обучения, кафедра конструирования и производства радиоэлектронных средств Института радиотехнических и телекоммуникационных систем ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: uvajsov.r.m@edu.mirea.ru.

Translated by E. Shklovskii

Russian Technological Journal. 2021;9(6):46–56

ISSN 2500-316X (Online)

Modern radio engineering and telecommunication systems

Современные радиотехнические и телекоммуникационные системы

UDC 621.391.072
https://doi.org/10.32362/2500-316X-2021-9-6-46-56

RESEARCH ARTICLE

Optimal nonlinear filtering of MPSK signals against a background of harmonic interference with a random initial phase

Gennady V. Kulikov [@],
Do Trung Tien,
Elena V. Samokhina

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] Corresponding author, e-mail: kulikov@mirea.ru

Abstract.
Objectives. The widespread use of radio data transmission systems using signals with multiposition phase shift keying (MPSK) is due to their high noise immunity and the simplicity of constructing the transmitting and receiving parts of the equipment. The conducted studies have shown that the presence of non-fluctuation interference, in particular, harmonic interference, in the radio channel significantly reduces the noise immunity of receiving discrete information. The energy loss in this case, depending on the interference intensity, can range from fractions of dB to 10 db or more. Therefore, interference suppression is an important task for such radio systems. The aim of the work is to synthesize and analyze an algorithm for optimal nonlinear filtering of MPSK signals against a background of harmonic interference with a random initial phase.
Methods. The provisions of the theory of optimal nonlinear signal filtering and methods of statistical radio engineering are used.
Results. The synthesis and analysis of the algorithm of optimal nonlinear filtering of MPSK signals against the background of harmonic interference with a random initial phase are carried out. The synthesized receiver contains a discrete symbol evaluation unit, two phase-locked frequency circuits of reference generators that form evaluation copies of the signal and interference, and cross-links between them. Analytical expressions are obtained that allow calculating the dependences of the bit error probability on the signal-to-noise ratio and the interference intensity μ. It is established that uncompensated fluctuations of the initial phase of the useful signal have a greater effect on the receiver noise immunity than similar fluctuations of the phase of harmonic interference, especially with low positional signals.
Conclusions. Comparison of the obtained results with the results obtained in the case when there are no harmonic interference compensation circuits shows that the use of the obtained phase filtering algorithms allows for almost complete suppression of harmonic interference. Thus, if $\mu = 0.5$ and the probability of error is 10^{-2}, the energy gain at $M = 2$ is about 2.5 dB, at $M = 4$ is about 6 dB, at $M = 8$ and $M = 16$ is at least 10 dB.

Keywords: multi-position phase manipulation, harmonic interference, optimal nonlinear filtering, noise immunity, bit error probability

• **Submitted:** 08.02.2021 • **Revised:** 17.02.2021 • **Accepted:** 25.07.2021

For citation: Kulikov G.V., Do Trung Tien, Samokhina E.V. Optimal nonlinear filtering of MPSK signals against a background of harmonic interference with a random initial phase. *Russ. Technol. J.* 2021;9(6):46–56. https://doi.org/10.32362/2500-316X-2021-9-6-46-56

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Оптимальная нелинейная фильтрация сигналов М-ФМ на фоне гармонической помехи со случайной начальной фазой

Г.В. Куликов @,
До Чунг Тиен,
Е.В. Самохина

МИРЭА – Российский технологический университет, Москва, 119454 Россия
@ Автор для переписки, e-mail: kulikov@mirea.ru

Резюме.
Цели. Широкое распространение радиосистем передачи данных с использованием сигналов с многопозиционной фазовой манипуляции (М-ФМ) обусловлено их высокой помехоустойчивостью и простотой построения передающей и приемной частей аппаратуры. Проведенные исследования показали, что наличие в радиоканале, кроме шумовой, нефлуктуационных, в частности, гармонических помех существенно снижает помехоустойчивость приема дискретной информации, энергетический проигрыш в этом случае в зависимости от интенсивности помехи может составлять от долей дБ до 10 дБ и более, поэтому борьба с ними является важной задачей для таких радиосистем. Цель работы – синтез и анализ алгоритма оптимальной нелинейной фильтрации сигналов М-ФМ на фоне гармонической помехи со случайной начальной фазой.
Методы. Использованы положения теории оптимальной нелинейной фильтрации сигналов и методы статистической радиотехники.
Результаты. Проведены синтез и анализ алгоритма оптимальной нелинейной фильтрации сигналов М-ФМ на фоне гармонической помехи со случайной начальной фазой. Синтезированный приемник содержит блок оценки дискретного символа, две схемы фазовой автоподстройки частоты опорных генераторов, формирующих оценочные копии сигнала и помехи, и перекрестные связи между ними, получены аналитические выражения, позволяющие рассчитать зависимости вероятности битовой ошибки от отношения сигнал/шум и интенсивности помехи μ. Установлено, что нескомпенсированные флуктуации начальной фазы полезного сигнала оказывают большее влияние на помехоустойчивость приемника, чем аналогичные флуктуации фазы гармонической помехи, особенно при малой позиционности сигналов.
Выводы. Сравнение полученных результатов с результатами, полученными в случае отсутствия цепи компенсации гармонической помехи, показывает, что использование полученных алгоритмов фильтрации фаз позволяет обеспечить практически полное подавление гармонической помехи. Так, при $\mu = 0.5$ для вероятности ошибки $P_{eb} = 10^{-2}$ энергетический выигрыш при $M = 2$ составляет около 2.5 дБ, при $M = 4$ – около 6 дБ, при $M = 8$ и $M = 16$ – не менее 10 дБ.

Ключевые слова: многопозиционная фазовая манипуляция, гармоническая помеха, оптимальная нелинейная фильтрация, помехоустойчивость, вероятность битовой ошибки

• **Поступила:** 08.02.2021 • **Доработана:** 17.02.2021 • **Принята к опубликованию:** 25.07.2021

Для цитирования: Куликов Г.В., До Чунг Тиен, Самохина Е.В. Оптимальная нелинейная фильтрация сигналов М-ФМ на фоне гармонической помехи со случайной начальной фазой. *Russ. Technol. J.* 2021;9(6):46–56. https://doi.org/10.32362/2500-316X-2021-9-6-46-56

Прозрачность финансовой деятельности: Никто из авторов не имеет финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The widespread use of multiple phase-shift keyed (MPSK) signals in modern digital information transmission systems is due to their high energy and spectral characteristics. However, the quality of communication was reported [1–7] to decrease significantly, if, in the radio channel, there is not only a noise interference, but also nonfluctuating interferences of various types. This is particularly so where such an interference is narrowband (harmonic) and has the same frequency as that of the desired signal [1, 2]. Depending on the interference intensity, the energy loss in this case can be fractions of decibel to 10 dB and more.

Such a decrease in the quality of reception is explained by the fact that demodulator algorithms are optimized to receive signals against the background of only white Gaussian noise and do not take into account the characteristics of nonfluctuating interferences. Introduction of units of compensation of such interferences to the demodulator circuit can considerably increase the noise immunity of signal reception. The purpose of this work was to synthesize and analyze an algorithm of optimal nonlinear filtrating of MPSK signals against the background of a harmonic interference with a random initial phase.

SYNTHESIS OF AN ALGORITHM OF OPTIMAL NONLINEAR FILTERING

Let us consider the following receiver-input process over the time range $t \in (0, kT]$:

$$x(t) = s_\Sigma(\mathbf{C}_k, t, \varphi_{sig}) + s_{int}(t, \varphi_{int}) + n(t) = \\ = s_{sig,int}(\mathbf{C}_k, t, \varphi_{sig}, \varphi_{int}) + n(t). \quad (1)$$

It is an additive mixture of MPSK signal $s_\Sigma(\mathbf{C}_k, t, \varphi_{sig})$, which in a certain ($k$th) digit time slot has the form

$$s_k(C_k = i, t, \varphi_{sig}) = A_0 \cos(\omega_0 t + \varphi_i + \varphi_{sig}), \\ \varphi_i = i2\pi/M,\, t \in ((k-1)T, kT],\, i = 0, 1, \ldots, M-1; \quad (2)$$

a harmonic interference with a random initial phase:

$$s_{int}(t, \varphi_{int}) = \mu A_0 \cos(\omega_{int} t + \varphi_{int});$$

and noise interference $n(t)$ with the parameters:

$$\langle n(t) \rangle = 0; \langle n(t_1) n(t_2) \rangle = \frac{N_0}{2} \delta(t_2 - t_1).$$

Here, $\mathbf{C}_k$ is the vector of discrete information symbols.

Let us consider the MPSK signal as a discrete–continuous Markov process in which the states of a discrete parameter can change at certain times, multiple of T, and the random initial phases φ_{sig} and φ_{int} of the signal and interference, respectively, are Wiener processes [8]:

$$\dot{\varphi}_{sig}(t) = n_{\varphi_{sig}}(t), \quad \dot{\varphi}_{int}(t) = n_{\varphi_{int}}(t).$$

Here, $n_{\varphi_{sig}}(t)$ and $n_{\varphi_{int}}(t)$ are white Gaussian noises with zero means and single-sided spectral densities $N_{\varphi_{sig}}$ и $N_{\varphi_{int}}$, respectively. Such a representation of these random processes describes well the fluctuating phase shifts in self-excited oscillators [9]. In practice, especially in narrowband channels, the rates of the continuous processes $\varphi_{sig}(t)$ and $\varphi_{int}(t)$ typically vary slowly in comparison with the rates of change of information and channel symbols; i.e.,

$$\tau_{\varphi_{sig}} >> T, \; \tau_{\varphi_{int}} >> T,$$

where $\tau_{\varphi_{sig}}$ and $\tau_{\varphi_{int}}$ are the correlation times of these processes.

A priori information on each of the diffusion Markov processes, $\varphi_{sig}(t)$ and $\varphi_{int}(t)$, is given by the Fokker–Planck–Kolmogorov equation [8].

Thus, in view of all the above, the circuit diagram of a receiver of such discrete–continuous processes can be synthesized using a previously proposed filtrating algorithm [10]. In this case, the mixed *a posteriori* probability density of $s_{sig,int}(\mathbf{C}_k, t, \varphi_{sig}, \varphi_{int})$ containing the vector $\mathbf{C}_k$ is

$$p_{ps}(t, \mathbf{C}_k, \varphi_{sig}, \varphi_{int}) = \\ = w_{ps}(t, \varphi_{sig}, \varphi_{int}) p_{ps}(t, \mathbf{C}_k | \varphi_{sig}, \varphi_{int}),$$

where $w_{ps}(t, \varphi_{sig}, \varphi_{int})$ is the *a posteriori* probability density of the independent parameters φ_{sig} and φ_{int}, which is unconditional on the vector of discrete symbols; and $p_{ps}(t, \mathbf{C}_k|\varphi_{sig}, \varphi_{int})$ is the conditional *a posteriori* probability of the state of the vector of discrete symbols at fixed values of φ_{sig} and φ_{int}.

The conditional *a posteriori* probability $p_{ps}(t, \mathbf{C}_k|\varphi_{sig}, \varphi_{int})$ obeys the equation [10]

$$\dot{p}_{ps}(t,\mathbf{C}_k|\varphi_{sig},\varphi_{int}) = p_{ps}(t,\mathbf{C}_k|\varphi_{sig},\varphi_{int}) \times \left\{F(t,\mathbf{C}_k,\varphi_{sig},\varphi_{int}) - \left\langle F(t,\varphi_{sig},\varphi_{int})\right\rangle\right\}, \quad (3)$$

where

$$F(t,\mathbf{C}_k,\varphi_{sig},\varphi_{int}) = \sum_{j=1}^{k} F_j(t,\mathbf{C}_j,\varphi_{sig},\varphi_{int}),$$

$$F_j(t,\mathbf{C}_j,\varphi_{sig},\varphi_{int}) = -\frac{1}{N_0}[x(t) - s_{sig,int,j}(\mathbf{C}_j,t,\varphi_{sig},\varphi_{int})]^2, \quad (4)$$

$$\left\langle F(t,\varphi_{sig},\varphi_{int})\right\rangle = \sum_{C_1=0}^{M-1}\sum_{C_2=0}^{M-1}\dots\sum_{C_k=0}^{M-1} F(t,\mathbf{C}_k,\varphi_{sig},\varphi_{int}) \times p_{sig,int}(t,\mathbf{C}_k|\varphi_{sig},\varphi_{int}).$$

Under the assumption that the *a priori* probabilities of the values of information symbols (for MPSK) in radio communication are identical and equal to $1/M$, and also identical are the probabilities of the transition of a discrete symbol from one state to another, the solution of Eq. (3) at time $t = kT$ is written as

$$p_{ps}(kT,\mathbf{C}_k|\varphi_{sig},\varphi_{int}) = \frac{\exp\left[\sum_{j=1}^{k}\int_{(j-1)T}^{jT} F_j(\tau,C_j,\mathbf{C}_{j-1},\varphi_{sig},\varphi_{int})d\tau\right]}{\sum_{\mathbf{C}_k}\exp\left[\sum_{j=1}^{k}\int_{(j-1)T}^{jT} F_j(\tau,C_j,\mathbf{C}_{j-1},\varphi_{sig},\varphi_{int})d\tau\right]}.$$

Here, for the writing to be more informative, the discrete symbol C_j is separately indicated in the list of the arguments of the function $F_j(\times)$.

To find the *a posteriori* probability of one or another value of the symbol C_k, the latter expression should be averaged over M possible values $C_1, C_2, \ldots, C_{k-1}$. This gives

$$p_{ps}(kT,C_k|\varphi_{sig},\varphi_{int}) = \sum_{\mathbf{C}_{k-1}}\left\{\frac{\exp\left[\sum_{j=1}^{k}\int_{(j-1)T}^{jT} F_j(\tau,C_j,\mathbf{C}_{j-1},\varphi_{sig},\varphi_{int})d\tau\right]}{\sum_{\mathbf{C}_k}\exp\left[\sum_{j=1}^{k}\int_{(j-1)T}^{jT} F_j(\tau,C_j,\mathbf{C}_{j-1},\varphi_{sig},\varphi_{int})d\tau\right]}\right\}.$$

The expression can be significantly simplified under the additional assumption that the quality of reception is good. In this case, the symbols that have already been decided by the time $t = kT$ can be replaced by their estimated values; i.e., the vector $\mathbf{C}_{k-1}$ can be replaced by $\mathbf{C}^*_{k-1} = \{C^*_1, C^*_2, \ldots, C^*_{k-1}\}$. Combining the terms independent of C_k into coefficient K and taking into account the independence of symbols in the information sequence of the corresponding MPSK signal transmissions, one can obtain the formula

$$p_{ps}(T,C_k|\varphi_{sig},\varphi_{int}) = K\frac{\exp\left[\int_0^T F_k(\tau,C_k,\varphi_{sig},\varphi_{int})d\tau\right]}{\sum_{C_k=0}^{M-1}\exp\left[\int_0^T F_k(\tau,C_k,\varphi_{sig},\varphi_{int})d\tau\right]}.$$

An algorithm to estimate a discrete information symbol follows from the condition of the maximum of this *a posteriori* probability at time $t = T$:

$$(C_k = i) \Rightarrow \max\{p_{ps}(t, \mathbf{C}_k|\varphi_{sig}, \varphi_{int})\}. \quad (5)$$

Let us introduce the notation

$$J_0 = \int_0^T F_k(\tau,C_k = 0,\varphi_{sig},\varphi_{int})d\tau,$$
$$\dots \quad (6)$$
$$J_{M-1} = \int_0^T F_k(\tau,C_k = M-1,\varphi_{sig},\varphi_{int})d\tau.$$

The integrands of these formulas are found from expressions (4). Then, algorithm (5) can be rewritten in the form

$$(C_k = i) \Rightarrow \max\{\exp(J_i)\}, \quad (7)$$

or

$$(C_k = i) \Rightarrow \max\{J_i\}. \quad (8)$$

Let us transform integrals (6), taking into account expressions (1) and (2):

$$J_0 = \frac{2}{N_0}\int_0^T [x(t) - s_{\text{int}}(t,\varphi_{\text{int}})] s_k(t, C_k = 0, \varphi_{\text{int}}) dt,$$
$$\ldots \tag{9}$$
$$J_{M-1} = \frac{2}{N_0}\int_0^T [x(t) - s_{\text{int}}(t,\varphi_{\text{int}})] s_k(t, C_k = M-1, \varphi_{\text{sig}}) dt.$$

Algorithm (8) is structurally similar to the MPSK signal processing algorithm for the case where signals are received against the background of only white Gaussian noise [11]. The difference consists in the presence of a procedure of subtraction from the received mixture $x(t)$ a copy of the interference that is formed on the receiver side. The circuit contains a set of correlators (9), which determine the degree of similarity of the compensated mixture with reference signals corresponding to all the possible values of the information symbol C_k.

The reference signals and the copy of the interference contain information on the initial phases φ_{sig} and φ_{int}. The true values of these quantities can be replaced by their estimated values φ^*_{sig} and φ^*_{int}. Let us develop algorithms to make these estimates.

Let us solve this problem by the Gaussian approximation [8, 12] of the *a posteriori* probability density of these random parameters. This method is valid at high signal-to-noise ratios and long observation times. In this case, such an assumption is valid, which makes it possible to transition from difficult-to-solve differential equations for probability densities, which follow from the Fokker–Planck–Kolmogorov equation [8], to approximate relations for the expected values φ^*_α and the *a posteriori* variances $K_{\alpha\beta}$ of the approximating multidimensional (of dimension z) Gaussian distribution:

$$\dot{\varphi}^*_\alpha = a_\alpha(\varphi^*) + \sum_{\beta=1}^{z} K_{\alpha\beta} \frac{\partial < F(t,\varphi^*) >}{\partial \varphi^*_\beta},$$
$$\dot{K}_{\alpha\beta} = \sum_{\nu=1}^{z}\left[\frac{\partial a_\alpha(\varphi^*)}{\partial \varphi^*_\nu} K_{\nu\beta} + \frac{\partial a_\beta(\varphi^*)}{\partial \varphi^*_\nu} K_{\alpha\nu}\right] +$$
$$+ b_{\alpha\beta}(\varphi^*) + \sum_{\nu=1}^{z}\sum_{\gamma=1}^{z} K_{\alpha\nu} \frac{\partial^2 < F(t,\varphi^*) >}{\partial \varphi^*_\nu \partial \varphi^*_\gamma} K_{\gamma\beta}.$$

The drift and diffusion coefficients for the independent Wiener processes $\varphi_{\text{sig}}(t)$ and $\varphi_{\text{int}}(t)$ are the following [8, 13]:

$$a(\varphi_{\text{sig}}) = a(\varphi_{\text{int}}) = 0,$$
$$b(\varphi_{\text{sig}}) = \frac{N_{\varphi_{\text{sig}}}}{2};\; b(\varphi_{\text{int}}) = \frac{N_{\varphi_{\text{int}}}}{2},$$

and $K_{\varphi_{\text{sig}}\varphi_{\text{int}}} = K_{\varphi_{\text{int}}\varphi_{\text{sig}}} = 0.$

Assuming that the phase lock time is less than T, we obtain the equations of optimal nonlinear filtering of the parameters φ^*_{sig} and φ^*_{int}:

$$\dot{\varphi}^*_{\text{sig}} = K_{\varphi_{\text{sig}}\varphi_{\text{sig}}} \frac{\partial \left\langle F_k(t,\varphi^*_{\text{sig}},\varphi^*_{\text{int}})\right\rangle}{\partial \varphi^*_{\text{sig}}},$$
$$\dot{\varphi}^*_{\text{int}} = K_{\varphi_{\text{int}}\varphi_{\text{int}}} \frac{\partial \left\langle F_k(t,\varphi^*_{\text{sig}},\varphi^*_{\text{int}})\right\rangle}{\partial \varphi^*_{\text{int}}}, \tag{10}$$

where the values of the *a posteriori* variances $K_{\varphi_{\text{sig}}\varphi_{\text{sig}}}$ and $K_{\varphi_{\text{int}}\varphi_{\text{int}}}$ are determined from the system of equations

$$\dot{K}_{\varphi_{\text{sig}}\varphi_{\text{sig}}} = \frac{N_{\varphi_{\text{sig}}}}{2} + K^2_{\varphi_{\text{sig}}\varphi_{\text{sig}}} \frac{\partial^2 \left\langle F_k(t,\varphi^*_{\text{sig}},\varphi^*_{\text{int}})\right\rangle}{\partial \varphi^{*\,2}_{\text{sig}}},$$
$$\dot{K}_{\varphi_{\text{int}}\varphi_{\text{int}}} = \frac{N_{\varphi_{\text{int}}}}{2} + K^2_{\varphi_{\text{int}}\varphi_{\text{int}}} \frac{\partial^2 \left\langle F_k(t,\varphi^*_{\text{sig}},\varphi^*_{\text{int}})\right\rangle}{\partial \varphi^{*\,2}_{\text{int}}}, \tag{11}$$

and

$$< F_k(t,\varphi^*_{\text{sig}},\varphi^*_{\text{int}}) > =$$
$$= F_k(t, C_k = 0, \varphi^*_{\text{sig}},\varphi^*_{\text{int}}) \frac{\exp J_0}{\sum_{i=0}^{M-1}\exp J_i} +$$
$$+ F_k(t, C_k = 1, \varphi^*_{\text{sig}},\varphi^*_{\text{int}}) \frac{\exp J_1}{\sum_{i=0}^{M-1}\exp J_i} +$$
$$+ F_k(t, C_k = 2, \varphi^*_{\text{sig}},\varphi^*_{\text{int}}) \frac{\exp J_2}{\sum_{i=0}^{M-1}\exp J_i} + \ldots +$$
$$+ F_k(t, C_k = M-1, \varphi^*_{\text{sig}},\varphi^*_{\text{int}}) \frac{\exp J_{M-1}}{\sum_{i=0}^{M-1}\exp J_i}.$$

In view of expressions (4),

$$\frac{\partial < F_k(t,\varphi^*_{\text{sig}},\varphi^*_{\text{int}}) >}{\partial \varphi^*_{\text{sig}}} = \frac{2A_0}{N_0}[x(t) - s_{\text{int}}(t,\varphi^*_{\text{int}})] \times$$
$$\times \sum_{i=0}^{M-1} s^h_k(t, C_k = i, \varphi^*_{\text{sig}}) \frac{\exp J_i}{\sum_{l=0}^{M-1}\exp J_l},$$

$$\frac{\partial < F_k(t,\varphi_{sig}^*,\varphi_{int}^*) >}{\partial \varphi_{int}^*} = -\frac{2\mu A_0}{N_0}\sin(\omega_{int}t + \varphi_{int}^*)\times$$
$$\times\left[x(t) - \sum_{i=0}^{M-1} s_k(t, C_k = i, \varphi_{sig}^*)\frac{\exp J_i}{\sum_{l=0}^{M-1}\exp J_l}\right].$$

Here,

$$s_k^h(t, C_k = i, \varphi_{sig}^*) = \frac{ds_k(t, C_k = i, \varphi_{sig}^*)}{d\varphi_{sig}^*} =$$
$$= -A_0 \sin(\omega_0 t + \varphi_i + \varphi_{sig}^*).$$

At a high signal-to-noise ratio, the latter expressions are simplified. If the parity and symmetry of the constellation diagram of the MPSK signal are also taken into account, then,

$$\frac{\partial < F_k(t,\varphi_{sig}^*,\varphi_{int}^*) >}{\partial \varphi_{sig}^*} \approx \frac{2A_0}{N_0}[x(t) - s_{int}(t,\varphi_{int}^*)]\times$$
$$\times \sum_{i=0}^{M/2-1} s_k^h(t, C_k = i, \varphi_{sig}^*)\,\mathrm{th}\,J_i,$$

$$\frac{\partial < F_k(t,\varphi_{sig}^*,\varphi_{int}^*) >}{\partial \varphi_{int}^*} \approx -\frac{2\mu A_0}{N_0}\sin(\omega_{int}t + \varphi_{int}^*)\times$$
$$\times [x(t) - \sum_{i=0}^{M/2-1} s_k t, C_k = i, \varphi_{sig}^*)\,\mathrm{th}\,J_i].$$

Making statistical averaging and assuming that, in the steady-state mode, at small filtering error,

$$\cos(\varphi_{int} - \varphi_{int}^*) \approx 1,\ \cos(\varphi_{sig} - \varphi_{sig}^*) \approx 1,$$

we obtain

$$\overline{\frac{\partial^2 < F_k(t,\varphi_{sig}^*,\varphi_{int}^*) >}{\partial \varphi_{sig}^{*\,2}}} \approx -\frac{A_0^2}{2N_0}\left(1 + \mathrm{th}\frac{A_0^2 T}{2N_0}\right),$$

$$\overline{\frac{\partial^2 < F_k(t,\varphi_{sig}^*,\varphi_{int}^*) >}{\partial \varphi_{int}^{*\,2}}} \approx -\frac{\mu^2 A_0^2}{N_0}.$$

In the steady-state mode of filtering of the continuous parameters φ_{sig} and φ_{int}, the solution of Eqs. (11) tends to the steady-state values $\overline{K_{\varphi_{sig}\varphi_{sig}}}$ and $\overline{K_{\varphi_{int}\varphi_{int}}}$, the derivatives of which are zero. Hence,

$$\overline{K_{\varphi_{sig}\varphi_{sig}}} = \sqrt{\frac{N_{\varphi_{sig}} N_0}{A_0^2\left(1 + \mathrm{th}\frac{A_0^2 T}{2N_0}\right)}},$$

$$\overline{K_{\varphi_{int}\varphi_{int}}} = \sqrt{\frac{N_{\varphi_{int}} N_0}{2\mu^2 A_0^2}}.$$

Finally, algorithms (10) of filtering of the random initial phases of the signal and the interference take the form

$$\dot{\varphi}_{sig}^* = S_1 K_1 A_0\left[x(t) - s_{int}(t,\varphi_{int}^*)\right]\times$$
$$\times \sum_{i=0}^{M/2-1} s_k^h(t, C_k = i, \varphi_{sig}^*)\,\mathrm{th}\,J_i, \quad (12)$$

$$\dot{\varphi}_{int}^* = -S_2 K_2 \mu A_0 \sin(\omega_{int}t + \varphi_{int}^*)\times$$
$$\times [x(t) - \sum_{i=0}^{M/2-1} s_k(t, C_k = i, \varphi_{sig}^*)\,\mathrm{th}\,J_i], \quad (13)$$

where S_1 and S_2 are the transconductances of control elements (CEs) in each phase-lock channel, $K_1 = \frac{2\overline{K_{\varphi_{sig}\varphi_{sig}}}}{N_0 S_1}$, and $K_2 = \frac{2\overline{K_{\varphi_{int}\varphi_{int}}}}{N_0 S_2}$.

Figure 1 presents the circuit diagram of a quasi-coherent receiver in which interrelated algorithms (7)–(9), (12), and (13) are implemented. The receiver contains an estimator of discrete symbol C_k^*, two phase-lock circuits of reference oscillators O1 and O2, and cross couplings between them.

Oscillators O1 and O2 generate estimation copies of the signal and the interference, which are phase-locked with the corresponding oscillations contained in the received mixture $x(t)$. At the input of the receiver, two subtractors are installed. Subtractor S1 compensates the harmonic interference in the received mixture; after S1, the signal is transmitted to a demodulator to make a decision on the received symbol. Subtractor S2 compensates the signal in the received mixture, and the obtained oscillation in the ideal case contains only the harmonic interference and noise. This oscillation is used by the phase-lock circuit of the interference channel.

ANALYSIS OF THE NOISE IMMUNITY OF THE SYNTHESIZED RECEIVER

Let us analyze the noise immunity of the synthesized quasi-coherent receiver of MPSK signals using a

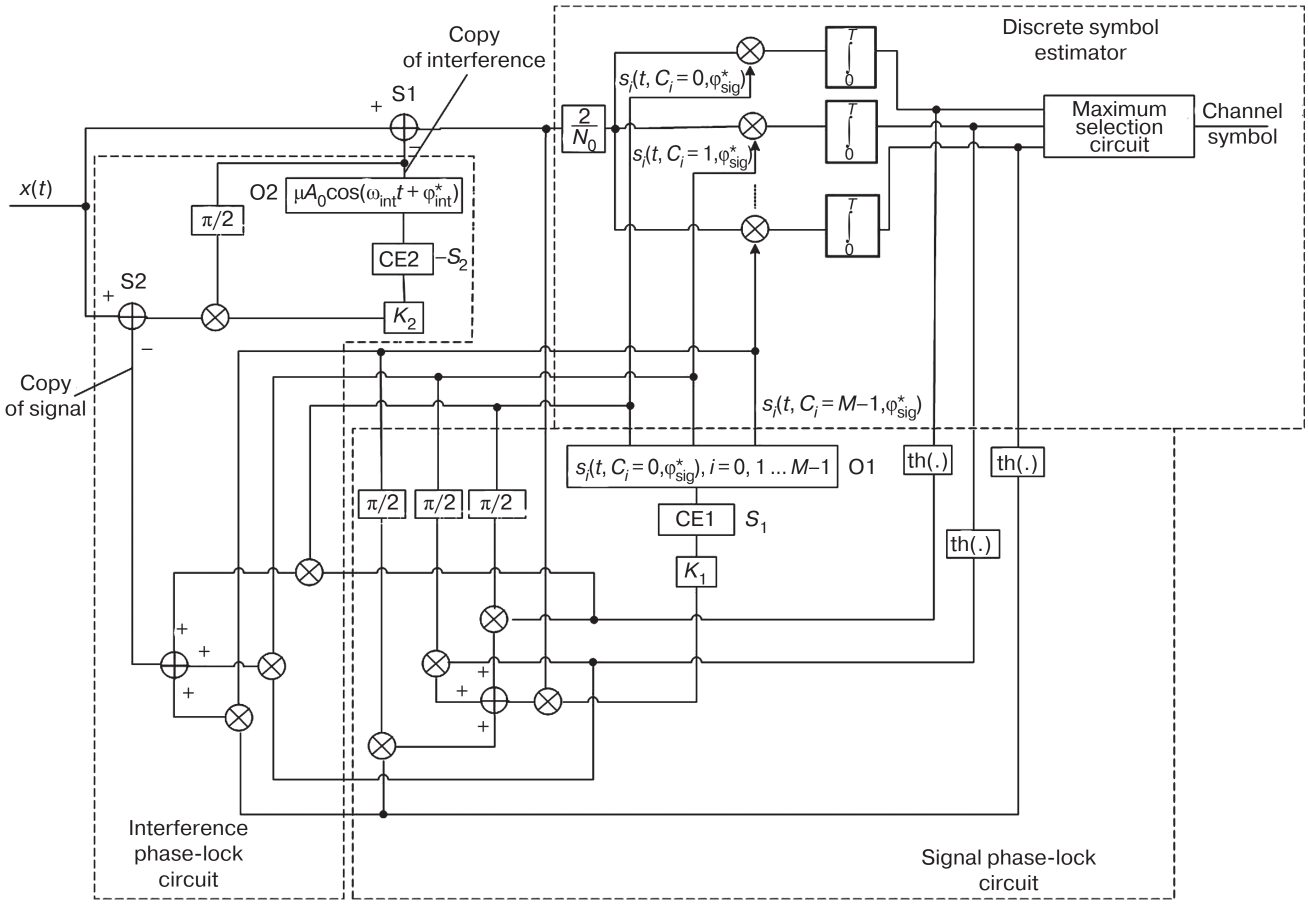


Fig. 1. Circuit diagram of a quasi-coherent receiver of an MPSK signal against a background of a harmonic interference with a random initial phase

published procedure [2]. The symbol and bit error probabilities can be defined as follows:

$$P_{se} = 1 - \prod_{i=1}^{M-1} p(u_i = J_0 - J_i > 0)\big|_0,$$

$$P_{be} = P_{se}/\log_2 M,$$

where $p(u_i = J_0 - J_i > 0)\big|_0 = 1 - \Phi\left(\frac{m_{ui}}{\sqrt{D_{ui}}}\right)$, and $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-t^2/2}dt$.

Using (1), (2), and (9), we obtain

$$m_{ui} = \langle J_0 - J_i \rangle = \frac{2E_s}{N_0}[\cos(\varphi_{sig} - \varphi_{sig}^*) \times$$
$$\times (1 - \cos(i2\pi/M)) - \sin(\varphi_{sig} - \varphi_{sig}^*)\sin(i2\pi/M)] +$$
$$+ \mu\frac{2E_s}{N_0}\frac{\sin x}{x}[(\cos\eta_{int} - \cos\eta_{int}^*)(1 - \cos(i2\pi/M)) -$$
$$- (\sin\eta_{int} - \sin\eta_{int}^*)\sin(i2\pi/M)],$$

$$D_{ui} = \frac{4E_s}{N_0}(1 - \cos(i2\pi/M)),$$

where $x = \Delta\omega_{int}T/2$, $\Delta\omega_{int} = \omega_{int} - \omega_0$, $\eta_{int} = \Delta\omega_{int}T/2 + \varphi_{int} - \varphi_{sig}^*$ and $\eta_{int}^* = \Delta\omega_{int}T/2 + \varphi_{int}^* - \varphi_{sig}^*$.

The parameters m_{ui} and D_{ui} are conditional on the values of φ_{sig}, φ_{sig}^*, φ_{int}, and φ_{int}^*. To obtain the unconditional error probabilities, one should make the corresponding averaging under the assumption that the *a posteriori* probability densities of the random phases φ_{sig} and φ_{int} are Gaussian:

$$w(\varphi_{sig}) = \frac{1}{\sqrt{2\pi \overline{K_{\varphi_{sig}\varphi_{sig}}}}}\exp\left[-\frac{(\varphi_{sig} - \varphi_{sig}^*)^2}{2\overline{K_{\varphi_{sig}\varphi_{sig}}}}\right],$$

$$w(\varphi_{int}) = \frac{1}{\sqrt{2\pi \overline{K_{\varphi_{int}\varphi_{int}}}}}\exp\left[-\frac{(\varphi_{int} - \varphi_{int}^*)^2}{2\overline{K_{\varphi_{int}\varphi_{int}}}}\right],$$

and that the quantities φ^*_{sig} and φ^*_{int} are uniformly distributed in the range $(0, 2\pi]$. Moreover, in this case, one can use the approximate formula [8]

$$<< p(u_i > 0)\Big|_0 >>_{\varphi_{\text{sig}}\varphi_{\text{int}}} = 1 - \Phi\left(<< \frac{m_{ui}}{\sqrt{D_{ui}}} >>_{\varphi_{\text{sig}}\varphi_{\text{int}}} \right) =$$

$$= 1 - \Phi\left(\frac{<< m_{ui} >>_{\varphi_{\text{sig}}\varphi_{\text{int}}}}{\sqrt{D_{ui}}} \right).$$

Averaging over the parameters φ_{sig} and φ_{int} gives [14, 15]

$$<< m_{ui} >>_{\varphi_{\text{sig}}\varphi_{\text{int}}} =$$

$$= \frac{2E_s}{N_0}[(1 - \cos(i2\pi/M))\exp(-\overline{K_{\varphi_{\text{sig}}\varphi_{\text{sig}}}}/2) +$$

$$+ \mu \frac{\sin x}{x}(\exp(-\overline{K_{\varphi_{\text{int}}\varphi_{\text{int}}}}/2) - 1)((1 - \cos(i2\pi/M)) \times$$

$$\times \cos\eta^*_{\text{int}} - \sin(i2\pi/M)\sin\eta^*_{\text{int}}))].$$

Analytical averaging over the parameter η^*_{int} failed; therefore, this averaging was performed numerically.

Figure 2 illustrates the dependences of the bit error probability P_{be} on the signal-to-noise ratio E_b/N_0 for the quasi-coherent receiver of MPSK signals at various M against the background of a harmonic interference with a random initial phase, $\Delta\omega_{\text{int}}$, and a relative intensity of $\mu = 0.5$. Here, the parameters are the *a priori* variances of the progressions of the phases φ_{sig} and φ_{int} in one digit time slot T:

$$\sigma^2_{\varphi_{\text{sig}}} = \frac{N_{\varphi_{\text{sig}}}T}{2}, \quad \sigma^2_{\varphi_{\text{int}}} = \frac{N_{\varphi_{\text{int}}}T}{2}.$$

For comparison, Fig. 2 presents curves *1* constructed at $\mu = 0$. The other curves are constructed at the following parameters: (*2*) $\sigma^2_{\varphi_{\text{int}}} = 0.25$ and $\sigma^2_{\varphi_{\text{sig}}} = 0$, (*3*) $\sigma^2_{\varphi_{\text{int}}} = 0$ and $\sigma^2_{\varphi_{\text{sig}}} = 0.25$, and (*4*) $\sigma^2_{\varphi_{\text{int}}} = 0.25$ and $\sigma^2_{\varphi_{\text{sig}}} = 0.25$.

Analysis of the graphs shows the validity of the obtained algorithms. It is seen that the uncompensated fluctuations of the initial phase of the desired signal affect more strongly the interference immunity of the receiver than similar fluctuations of the phase of the harmonic interference. This is particularly clear at small M. This is due to the fact that it is the phase structure of the MPSK signal that contains information on the discrete symbol.

Figure 3 describes the effect of the interference μ on the bit error probability.

In Fig. 3, one of the curves is constructed without using interference compensation circuits, and the two others are built by simultaneously using algorithms (12) and (13). One can see that, in the former case, with increasing μ, the error probability increases significantly (by one to several

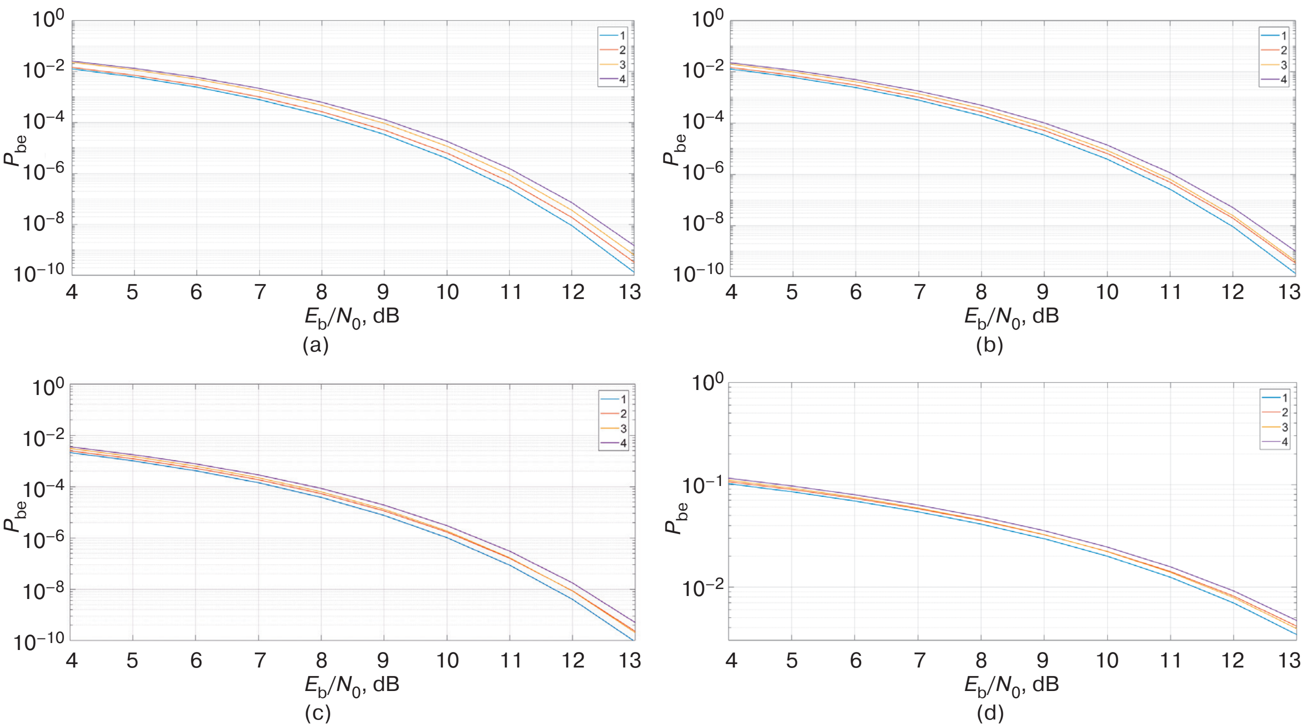


Fig. 2. Dependences of the bit error probability on the signal-to-noise ratio in the reception of MPSK signals against a background of a harmonic interference with a random initial phase at M = (a) 2, (b) 4, (c) 8, and (d) 16

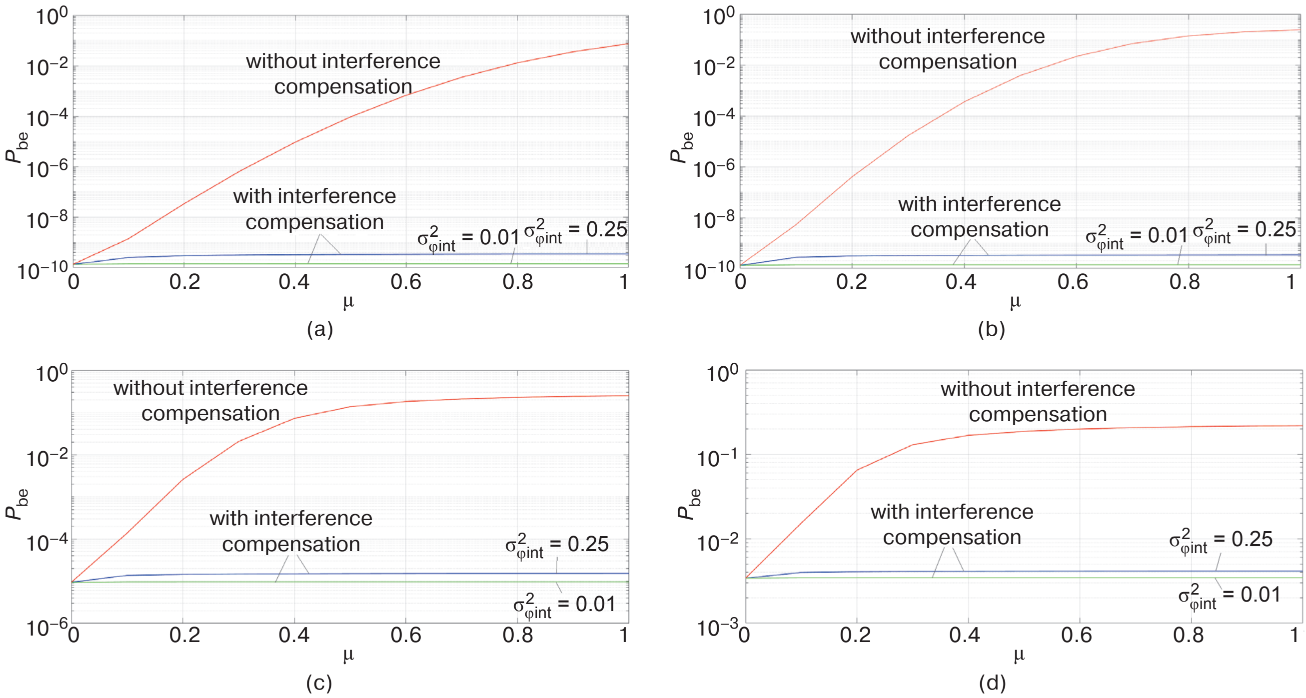


Fig. 3. Dependence of the bit error probability on the relative interference intensity at E_b/N_0 = 13 dB and M = (a) 2, (b) 4, (c) 8, and (d) 16

orders of magnitude), and in the latter cases, the harmonic interference is virtually completely suppressed at any μ.

Comparison of the results of this work with the previously obtained data [2] for the case of the absence of harmonic interference compensation circuits showed that the use of the above-derived phase-filtering algorithms considerably improves the noise immunity of the receiver. For example, at $\mu = 0.5$ and an error probability of $P_{be} = 10^{-2}$, the energy gain at $M = 2$ is about 2.5 dB; at $M = 4$, about 6 dB; and at $M = 8$ and $M = 16$, no less than 10 dB.

CONCLUSIONS

The synthesized algorithm of the optimal nonlinear filtering of MPSK signals against the background of harmonic interference with a random initial phase considerably improves the noise immunity of discrete information reception under complex interference conditions and ensures a significant energy gain of the radio system. At the same time, of note is the significant structural complexity of such a receiver. Furthermore, on the receiver side, the interference frequency and level should be known *a priori*, which is difficult to ensure in practice. Therefore, the determined characteristics should be regarded as potentially achievable, the ones at which the development of interference compensation devices should be targeted.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Fomin A.I., Yalin A.K. Noise immunity of receiving digital phase-shift keyed signals against the background of sinusoidal interference and Gaussian noise. *Telekommunikatsii = Telecommunications*. 2013;10:19–26 (in Russ.).
2. Kulikov G.V., Nguyen V.D., Nesterov A.V., Lelyuh A.A. Noise immunity of reception of signals with multiple phase shift keying in the presence of harmonic interference. *Naukoemkie tekhnologii = Science Intensive Technologies*. 2018;11:32–38 (in Russ.).
3. Kulikov G.V., Nguyen V.D., Do T.T. Effect of phase-shift interference on the noise immunity of correlation demodulator of signals with multiple phase shift keying. *Rossiiskii tekhnologicheskii zhurnal = Russian Technological Journal*. 2019;7(2):18–28 (in Russ.). https://doi.org/10.32362/2500-316X-2019-7-2-18-28

СПИСОК ЛИТЕРАТУРЫ

1. Фомин А.И., Ялин А.К. Помехоустойчивость приема цифровых фазоманипулированных сигналов на фоне синусоидальной помехи и гауссовского шума. *Телекоммуникации*. 2013;10:19–26.
2. Куликов Г.В., Нгуен В.З., Нестеров А.В., Лелюх А.А. Помехоустойчивость приема сигналов с многофазной фазовой манипуляцией при наличии гармонических помех. *Наукоемкие технологии*. 2018;11:32–38.
3. Куликов Г.В., Нгуен В.З., До Ч.Т. Влияние фазоманипулированной помехи на помехоустойчивость корреляционного демодулятора сигналов с многопозиционной фазовой манипуляцией. *Российский технологический журнал*. 2019;7(2):18–28. https://doi.org/10.32362/2500-316X-2019-7-2-18-28

4. Nguyen V.D. Noise immunity of a coherent reception of signals with multiple phase shift keying in the presence of a retranslated interference. *Zhurnal radioelektroniki = J. Radio Electronics.* 2019;3 (in Russ.). https://doi.org/10.30898/1684-1719.2019.3.4
5. Kulikov G.V., Nguyen V.D., Kulagin V.P. Simulation of a coherent demodulator of M-ary phase shift keying signals in the presence of non-fluctuation interferences. In: *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. 2019, p. 149–152. https://doi.org/10.1109/RCAR47638.2019.9044009
6. Petrov A.V. Anti-jam robustness of binary phase shift keyed signals receiving under the impact of chaotic impulsive jamming with random duration and phase of radio-impulse. *Radiotekhnika = J. Radioengineering*. 2018;8:28–33 (in Russ.).
7. Zvonarev V.V., Karabelnikov I.A., Popov A.S. Probability of BPSK signal reception error in presence of similar hindrance. *Voprosy radioelektroniki = Issues of Radio electronics*. 2019;3:55–59 (in Russ.). https://doi.org/10.21778/2218-5453-2019-3-55-59
8. Tikhonov V.I., Kul'man N.K. *Nelineinaya fil'tratsiya i kvazikogerentnyi priem signalov* (*Nonlinear filtering and quasi-coherent signal reception*). Moscow: Sovetskoe Radio; 1975. 704 p. (in Russ.).
9. Tikhonov V.I. *Statisticheskaya radiotekhnika* (*Statistical radio engineering*): 2nd ed. Moscow: Radio i svyaz'; 1982. 624 p. (in Russ.).
10. Tikhonov V.I., Kharisov V.I., Smirnov V.A. Optimal filtration of discrete-continuous processes. *Radiotekhnika i elektronika = Radio Engineering and Electronics*. 1978;23(7):1441–1452 (in Russ.).
11. Proakis J.G. *Digital communications*. 3th ed. McGraw-Hill Book Co.; 1995. 956 p.
[Prokis J. *Tsifrovaya svyaz'*. (*Digital communication*). Moscow: Radio i svyaz'; 2000. 800 p. (in Russ.). ISBN 5-256-01434-X]
12. Yarlykov M.S., Mironov M.A. On the applicability of the Gaussian approximation in the Markov theory of optimal nonlinear filtering. *Radiotekhnika i elektronika = Radio Engineering and Electronics*. 1972;17(11):2285–2294 (in Russ.).
13. Yarlykov M.S. *Primenenie markovskoi teorii nelineinoi fil'tratsii v radiotekhnike* (*Application of the Markov theory of nonlinear filtering in radio engineering*). Moscow: Sovetskoe Radio; 1980. 258 p. (in Russ.).
14. Gradshtein I.S., Ryzhik I.M. *Tablitsy integralov, summ, ryadov i proizvedenii* (*Tables of integrals, sums, series and products*). Moscow: Nauka; 1971. 1108 p. (in Russ.).
15. Bateman H., Erdeyi A. *Tables of integral transforms*. V. 1. New York: McGraw-Hill Book Co.; 1954. 344 p.
[Beitmen G., Erdeii A. *Tablitsy integral'nykh preobrazovanii* (*Tables of integral transforms*): in 2 v. Moscow: Nauka; 1968. V. 1. 344 p. (in Russ.).]

4. Нгуен В.З. Помехоустойчивость когерентного приема сигналов с многофазной фазовой манипуляцией при наличии ретранслированных помех. *Журнал радиоэлектроники*. 2019;3. https://doi.org/10.30898/1684-1719.2019.3.4
5. Kulikov G.V., Nguyen V.D., Kulagin V.P. Simulation of a coherent demodulator of M-ary phase shift keying signals in the presence of non-fluctuation interferences. In: *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. 2019, p. 149–152. https://doi.org/10.1109/RCAR47638.2019.9044009
6. Петров А.В. Помехоустойчивость приема сигналов с двоичной фазовой манипуляцией при воздействии хаотической импульсной помехи со случайной длительностью и фазой радиоимпульса. *Радиотехника*. 2018;8:28–33.
7. Звонарев В.В., Карабельников И.А., Попов А.С. Вероятность ошибки приема BPSK-сигнала в присутствии подобной помехи. *Вопросы радиоэлектроники*. 2019;3:55–59. https://doi.org/10.21778/2218-5453-2019-3-55-59
8. Тихонов В.И., Кульман Н.К. *Нелинейная фильтрация и квазикогерентный прием сигналов*. М.: Сов. Радио; 1975. 704 с.
9. Тихонов В.И. *Статистическая радиотехника*. 2-е изд., перераб. и доп. М.: Радио и связь; 1982. 624 с.
10. Тихонов В.И., Харисов В.И., Смирнов В.А. Оптимальная фильтрация дискретно-непрерывных процессов. *Радиотехника и электроника*. 1978;23(7):1441–1452.
11. Прокис Дж. *Цифровая связь*: пер. с англ.; под ред. Д.Д. Кловского. М.: Радио и связь; 2000. 800 с. ISBN 5-256-01434-X
12. Ярлыков М.С., Миронов М.А. О применимости гауссовой аппроксимации в марковской теории оптимальной нелинейной фильтрации. *Радиотехника и электроника*. 1972;17(11):2285–2294.
13. Ярлыков М.С. *Применение марковской теории нелинейной фильтрации в радиотехнике*. М.: Сов. Радио; 1980. 258 с.
14. Градштейн И.С., Рыжик И.М. *Таблицы интегралов, сумм, рядов и произведений*. М.: Наука; 1971. 1108 с.
15. Бейтмен Г., Эрдейи А. *Таблицы интегральных преобразований*: в 2 т.; пер с англ. М.: Наука; 1968. Т. 1. 344 с.

About the authors

Gennady V. Kulikov, Dr. Sci. (Eng.), Professor, Department of Radioelectronic Systems and Complexes, Institute of Radio Engineering and Telecommunication Systems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: kulikov@mirea.ru. Scopus Author ID 36930533000, http://orcid.org/0000-0001-7964-6653

Do Trung Tien, Postgraduate Student, Department of Radioelectronic Systems and Complexes, Institute of Radio Engineering and Telecommunication Systems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: dotrungtien1993@gmail.com.

Elena V. Samokhina, Cand. Sci. (Eng.), Associate Professor, Department of Telecommunications, Institute of Radio Engineering and Telecommunications Systems, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: samohina@mirea.ru.

Об авторах

Куликов Геннадий Валентинович, д.т.н., профессор, кафедра радиоэлектронных систем и комплексов Института радиотехнических и телекоммуникационных систем ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: kulikov@mirea.ru. Scopus Author ID 36930533000, http://orcid.org/0000-0001-7964-6653

До Чунг Тиен, аспирант, кафедра радиоэлектронных систем и комплексов Института радиотехнических и телекоммуникационных систем ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: dotrungtien1993@gmail.com.

Самохина Елена Викторовна, к.т.н., доцент, кафедра телекоммуникаций Института радиотехнических и телекоммуникационных систем ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: samohina@mirea.ru.

Translated by V. Glyanchenko

Russian Technological Journal. 2021;9(6):57–63 ISSN 2500-316X (Online)

Analytical instrument engineering and technology

Аналитическое приборостроение и технологии

UDC 004.932
https://doi.org/10.32362/2500-316x-2021-9-6-57-63

RESEARCH ARTICLE

The method of increasing the information content of microfocus X-ray images

Nikolay E. Staroverov [1, @],
Artem Y. Gryaznov [1],
Irina G. Kamyshanskaya [2, 3],
Nikolay N. Potrakhov [1],
Ekaterina D. Kholopova [1]

[1] *Saint Petersburg Electrotechnical University "LETI", Saint Petersburg, 197376 Russia*
[2] *St. Petersburg State University, Saint Petersburg, 199034 Russia*
[3] *City Mariinsky Hospital, Saint Petersburg, 191014 Russia*
[@] *Corresponding author, e-mail: nik0205st@mail.ru*

Abstract. A method for processing microfocus X-ray images is described. It is based on high-frequency filtration and morphological image processing, which increases the contrast of the X-ray details. One of the most informative X-ray techniques is microfocus X-ray. In some cases, microfocus X-ray images cannot be reliably analyzed due to the peculiarities of the shooting method. So, the main disadvantages of microfocus X-ray images are most often an uneven background, distorted brightness characteristics and the presence of noise. The proposed method for enhancing the contrast of fine image details is based on the idea of combining high-frequency filtering and morphological image processing. The method consists of the following steps: noise suppression in the image, high-frequency filtering, morphological image processing, obtaining the resulting image. As a result of applying the method, the brightness of the contours in the image is enhanced. In the resulting image, all objects will have double outlines. The method was tested in the processing of 50 chest radiographs of patients with various pathologies. Radiographs were performed at the Mariinsky Hospital of St. Petersburg using digital stationary and mobile X-ray machines. In most of the radiographs, it was possible to improve the images contrast, to highlight the objects boundaries. Besides, the method was applied in microfocus X-ray tomography to improve the information content of projection data and improve the reconstruction of the 3D image of the research object. In both the first and second cases, the method showed satisfactory results. The developed method makes it possible to significantly increase the information content of microfocus X-ray images. The obtained practical results make it possible to count on broad prospects for the method application, especially in microfocus X-ray.

Keywords: microfocus X-ray, digital image processing, image filtering

• **Submitted:** 18.03.2021 • **Revised:** 19.08.2021 • **Accepted:** 20.08.2021

For citation: Staroverov N.E., Gryaznov A.Y., Kamyshanskaya I.G., Potrakhov N.N., Kholopova E.D. The method of increasing the information content of microfocus X-ray images. *Russ. Technol. J.* 2021;9(6):57–63. https://doi.org/10.32362/2500-316X-2021-9-6-57-63

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Метод повышения информативности рентгеновских снимков

Н.Е. Староверов [1, @],
А.Ю. Грязнов [1],
И.Г. Камышанская [2, 3],
Н.Н. Потрахов [1],
Е.Д. Холопова [1]

[1] *Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина), Санкт-Петербург, 197376 Россия*
[2] *Санкт-Петербургский государственный университет, Санкт-Петербург, 199034 Россия*
[3] *СПб ГБУЗ «Городская Мариинская больница», Санкт-Петербург, 191014 Россия*
[@] *Автор для переписки, e-mail: nik0205st@mail.ru*

Резюме. Описан основанный на высокочастотной фильтрации и морфологической обработке изображения метод обработки микрофокусных рентгеновских снимков, повышающий контраст деталей рентгенограммы. Одной из наиболее информативных методик рентгенографии является микрофокусная рентгенография. В ряде случаев микрофокусные рентгеновские изображения не могут быть достоверно проанализированы из-за особенностей способа съемки. Так, основными недостатками микрофокусных рентгеновских изображений чаще всего являются неравномерный фон, искаженные яркостные характеристики и наличие шумов. Предлагаемый метод повышения контраста мелких деталей изображения основан на идее сочетания высокочастотной фильтрации и морфологической обработки изображений. Метод состоит из следующих шагов: подавление шумов на изображении, высокочастотная фильтрация, морфологическая обработка изображения, получение результирующего изображения. В результате применения метода усиливается яркость контуров на изображении. На полученном изображении все объекты будут иметь двойные контуры. Метод был апробирован при обработке 50 рентгенограмм органов грудной клетки пациентов с разнообразной патологией. Рентгенограммы были выполнены в городской Мариинской больнице Санкт-Петербурга на цифровых стационарных и передвижных рентгеновских аппаратах. На большей части рентгенограмм удалось улучшить контраст снимков, выделить границы объектов. Также метод был применен в микрофокусной рентгеновской томографии для улучшения информативности проекционных данных и улучшения восстановления 3D-образа объекта исследования. Как для первого, так и для второго случаев метод показал удовлетворительные результаты. Разработанный метод позволяет существенно повысить информативность микрофокусных рентгеновских снимков. Полученные практические результаты позволяют рассчитывать на широкие перспективы применения метода, особенно в микрофокусной рентгенографии.

Ключевые слова: микрофокусная рентгенография, цифровая обработка изображений, фильтрация изображений

• **Поступила:** 18.03.2021 • **Доработана:** 19.08.2021 • **Принята к опубликованию:** 20.08.2021

Для цитирования: Староверов Н.Е., Грязнов А.Ю., Камышанская И.Г., Потрахов Н.Н., Холопова Е.Д. Метод повышения информативности рентгеновских снимков. *Russ. Technol. J.* 2021;9(6):57−63. https://doi.org/10.32362/2500-316X-2021-9-6-57-63

Прозрачность финансовой деятельности: Никто из авторов не имеет финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

It is well known that X-ray radiography is a method for studying a research object using X-rays. One of the most informative techniques of its application is microfocus X-ray using radiation sources with focal spots less than 100 µm in size.

Using of these sources allows obtaining images with direct geometric (projection) magnification of the image by 5–20 times. Today, microfocus X-ray is widely used in medicine, in quality assessment of agricultural crops, and in nondestructive industrial control [1–6].

In some cases, however, microfocus X-ray images cannot be analyzed accurately due to the poor image quality. Thus, the main disadvantages of the microfocus X-ray images are most often the uneven background, distorted brightness characteristics, and the presence of noise. In some cases, the widely used contrast enhancement algorithms [4−10] may not allow increasing the contrast of X-ray images, while keeping the noise level at the acceptable level.

The developed method solves this problem by enhancing the contrast of the object structure details without a significant increase in the image noise.

METHOD FOR ENHANCING THE CONTRAST OF THE IMAGE FINE DETAILS

The proposed method for enhancing the contrast of the image fine details is based on the concept of combining high-pass filtering with morphological processing of images. The method consists of the following steps: suppression of the image noise, high-pass filtering, morphological processing of the image, and obtaining the resulting image.

In the paper, an adaptive median filter is used for noise reduction [4, 11].

High-pass filtering is performed by changing the Fourier image by multiplying it with the Gaussian high-pass filter (HPF) image.

The Gaussian filter is set in the frequency domain according to the following expression:

$$H(u,v) = 1 - \mathrm{e}^{-D^2(u,v)\,/\,2D_0^2},$$

where, D_0 stands for a given constant taking values greater than zero, while $D(u, v)$ is the distance in the frequency domain between point (u, v) and the origin of coordinates.

Thus, after this filter, low-frequency components are significantly attenuated in the image, while high-frequency components remain unchanged. In other words, the brightness of areas with slow changes in this parameter is significantly reduced, while the areas with sudden changes in brightness (object boundaries) remain unchanged.

The next step is morphological image processing [12]. A square-shaped structural element with the size of 9 pixels is used for morphologic expansion (dilatation). After this operation, the thickness of all objects in the image is increased by 2 pixels.

The expansion of image A by structural element B is denoted by $A \oplus B$ and is given by the following expression:

$$A \oplus B = \bigcup_{b \in B} A_b.$$

Structural element B is applied to all image pixels. Every time the origin of the structure element coordinates coincides with a single pixel, a carry and subsequent logical addition with corresponding image pixels is applied to the entire structural element.

In the next step, the image after HPF is multiplied by a constant (the value of the constant varies from 1.2 to 1.8, depending on the image brightness and contrast), and then the image obtained by morphological expansion and also multiplied by a constant is subtracted from it [13]. The resulting image is described by the following expression:

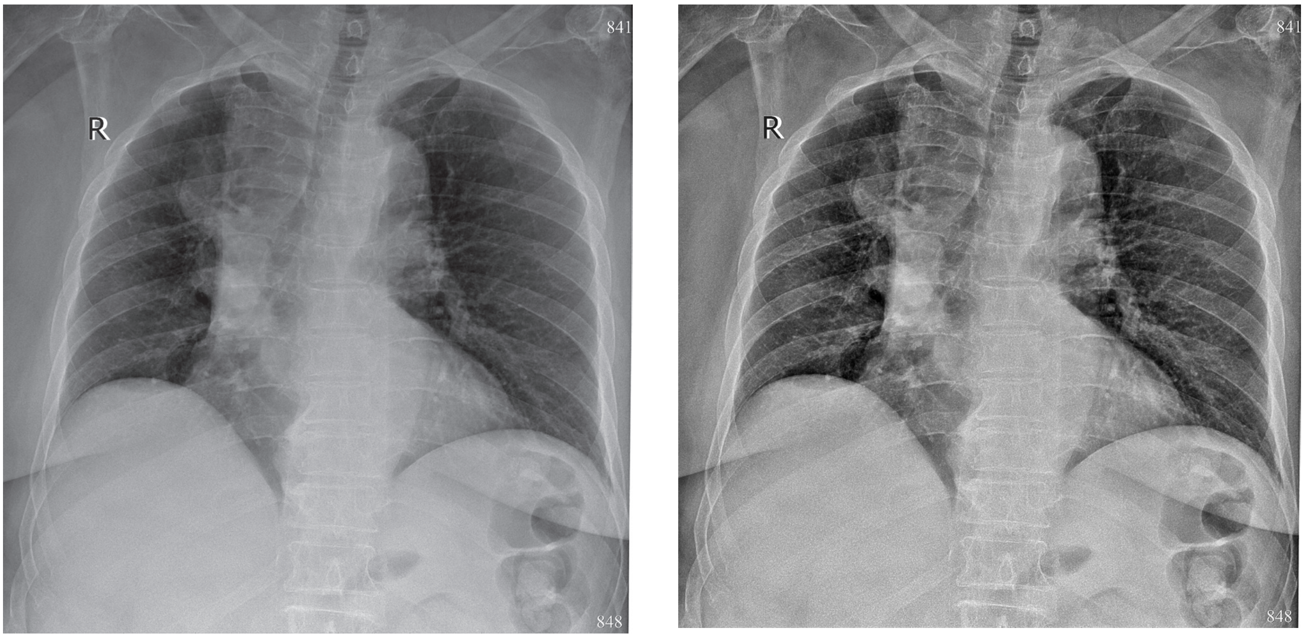
$$I_{\mathrm{res}} = I + I_{\mathrm{s}} \times C_1 - I_{\mathrm{m}} \times C_2,$$

where I is the original image; I_{s} is the image after high-pass filtering; C_1, C_2 are positive constants; I_{m} is the image obtained by morphological expansion.

This results in the brightness enhancement of contours in the image. In the resulting image, all objects have double contours; in case of the negative image (wherein denser structures have higher brightness), an inner contour which pixels have higher brightness than the object as well as an outer contour which pixels have lower brightness than the object appear [14, 15]. The algorithm for the positive image may be implemented in a similar way.

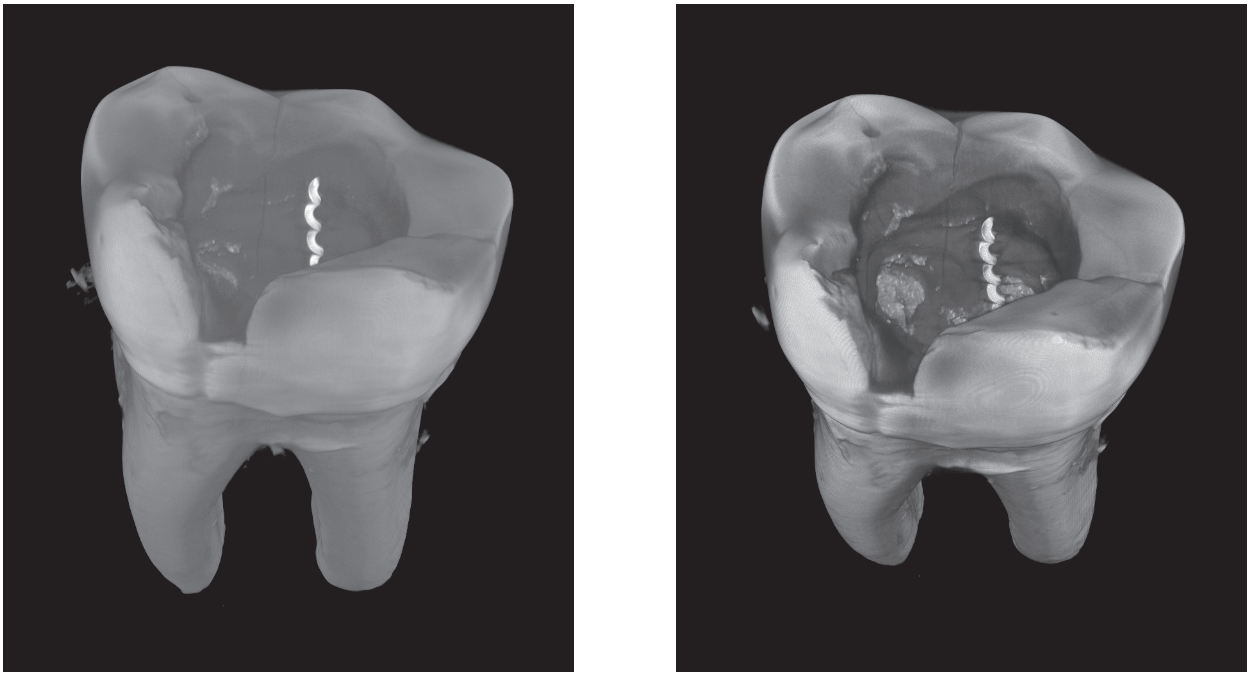
DISCUSSION

An example of the algorithm is illustrated in Fig. 1 showing chest X-rays before and after applying the method. The sharpness as well as the contrast of the image fine details enhance significantly. At the same time, there is no increase in the image noise, while the visual analysis becomes easier. Application of the proposed method allows detecting lung lesions more accurately.



(a) (b)

Fig. 1. Chest X-rays:
(a) before processing; (b) after processing



(a) (b)

Fig. 2. Recovered results of tomographic analysis of tooth specimens:
(a) before processing; (b) after processing

The method has been tested in processing 50 chest X-rays (CXRs) of patients with multiple pathologies. The CXRs have been done at the Mariinsky Municipal Hospital (Saint Petersburg, Russia) using the digital stationary and mobile X-ray machines. Digital processing of a series of X-ray images showing the extensive inhomogeneous opacification of both lungs by this method allows focalizing and identifying the following: circumscribed areas of the increased transparency, inhomogeneous thickening of the distorted lung pattern due to interstitial elements, inhomogeneous infiltration in lung tissue, and pleural effusion boundaries. In a series of X-ray images showing focal lung disease, the method allows visualizing clearly the boundaries of low-intensity foci identifying barely visible destruction sites in them. In CXRs of patients with COVID pneumonia, the postprocessing enhances the contrast and boundaries of polysegmental different-sized foci and infiltration sites. In CRSs of patients with traumatic injury to the chest organs, the method allows identifying barely visible signs of rib fracture, pneumomediastinum, and mantle pneumothorax.

The method has also been used in microfocus X-ray tomography for increasing the information content of projective data as well as for improving the recovery of the 3D image of the research object. The method gives satisfactory results for both the first and the second cases. The slices obtained with the use of the MRCT-04 microfocus X-ray tomograph in studying human tooth specimens have been used for analyzing tomographic data. The results of recovery without using the developed method are shown in Fig. 2a, while those of the recovery using the method are presented in Fig. 2b.

CONCLUSIONS

In the paper, the algorithmic support for the method of increasing the information content of X-ray images, first of all, the contrast, is considered. The method allows increasing the detection of low-contrast and fine details of the research object structure in its image significantly in a number of application areas for X-ray radiography.

Practical results allow expecting ample opportunities for applying the method, especially to microfocus X-ray radiography.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1 Arkhipov M.V., Priyatkin N.S., Gusakova L.P., et al. X-ray computer methods for studying the structural integrity of seeds and their importance in modern seed science. *Tech. Phys*. 2019;64(4):582–592. https://doi.org/10.1134/S1063784219040030
2. Potrakhov N.N., Gryaznov A.Yu., Zhamova K.K., Bessonov V.B., Obodovskii A.V., Staroverov N.E., Kholopova E.D. Microfocus X-ray diagnostics in medicine: physical and technical features and modern device. *Biotekhnosfera*. 2015;5(41):55–63 (in Russ.).
3. Staroverov N.E., Gryaznov A.Y., Kholopova E.D. Digital x-ray image processing with using adaptive histogram equalization and adaptive background correction. *Biomedicine Radioengineering = Biomeditsinskaya radioelectronika*. 2018;5:56–58.
4. Staroverov N.E., Kholopova E.D., Gryaznov A.Yu., Zhamova K.K. The methods of processing of microfocus X-ray images. In: *2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering* (*EIConRus*). 2017, p. 745–747. https://doi.org/10.1109/EIConRus.2017.7910665
5. Mazurov A.I., Potrakhov N.N. Possibilities and limitations of microfocus radiography in medicine. *Biotekhnosfera*. 2010;4(10):20–23 (in Russ).
6. Gryaznov A.Yu., Potrakhov E.N., Potrakhov N.N. Portable unit for X-ray express quality food products. *Biotekhnosfera*. 2009;6(6):26–28 (in Russ.).
7. Stark J.A. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process*. 2000;9(5):889–896. https://doi.org/10.1109/83.841534

СПИСОК ЛИТЕРАТУРЫ

1. Arkhipov M.V., Priyatkin N.S., Gusakova L.P., et al. X-ray computer methods for studying the structural integrity of seeds and their importance in modern seed science. *Tech. Phys*. 2019;64(4):582–592. https://doi.org/10.1134/S1063784219040030
2. Потрахов Н.Н., Грязнов А.Ю., Жамова К.К., Бессонов В.Б., Ободовский А.В., Староверов Н.Е., Холопова Е.Д. Микрофокусная рентгенография в медицине: физико-технические особенности и современные средства рентгенодиагностики. *Биотехносфера*. 2015;5(41):55–63.
3. Staroverov N.E., Gryaznov A.Y., Kholopova E.D. Digital x-ray image processing with using adaptive histogram equalization and adaptive background correction. *Biomedicine Radioengineering = Biomeditsinskaya radioelectronika*. 2018;5:56–58.
4. Staroverov N.E., Kholopova E.D., Gryaznov A.Yu., Zhamova K.K. The methods of processing of microfocus X-ray images. In: *2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*. 2017, p. 745–747. https://doi.org/10.1109/EIConRus.2017.7910665
5. Мазуров А.И., Потрахов Н.Н. Возможности и ограничения микрофокусной рентгенографии в медицине. *Биотехносфера*. 2010;4(10):20–23.
6. Грязнов А.Ю., Потрахов Е.Н., Потрахов Н.Н. Портативная установка для рентгеновского экспресс-контроля качества пищевой продукции. *Биотехносфера*. 2009;6(6):26–28.
7. Stark J.A. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process*. 2000;9(5):889–896. https://doi.org/10.1109/83.841534

8. Mukhopadhyay S., Chanda B. A multiscale morphological approach to local contrast enhancement. *Signal Process*. 2000;80(4):685–696. https://doi.org/10.1016/S0165-1684(99)00161-9
9. Parihar A.S., Verma O.P., Khanna C. Fuzzy-contextual contrast enhancement. *IEEE Trans. Image Process*. 2017;26(4):1810–1819. https://doi.org/10.1109/TIP.2017.2665975
10. Pogany A., Gao D., Wilkins S.W. Contrast and resolution in imaging with a microfocus x-ray source. *Rev. Sci. Instrum*. 1997;68(7):2774–2782. https://doi.org/10.1063/1.1148194
11. Pizer S.M., Amburn E.P., Austin J.D., et al. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process*. 1987;39(3):355–368. https://doi.org/10.1016/S0734-189X(87)80186-X
12. Vincent L. Morphological grayscale reconstructions in image analysis: Applications and efficient algorithms. *IEEE Trans. Image Process*. 1993;2(2):176–201. https://doi.org/10.1109/83.217222
13. Gryaznov A.Yu., Guk K.K., Staroverov N.E., Kholopova E.D. Method for sharpening and contrasting details of X-ray images. *Fizicheskie osnovy priborostroeniya = Physical Bases of Instrumentation*. 2019;8(4(34)):34–37 (in Russ.). https://doi.org/10.25210/jfop-1904-034037
14. Staroverov N.E., Gryaznov A.Yu., Potrakhov N.N., Kholopova E.D., Guk K.K. New methods of digital processing of microfocus X-ray images. *Biomed. Eng.* 2019;52:435–438. https://doi.org/10.1007/s10527-019-09864-6
[Staroverov N.E., Gryaznov A.Yu., Potrakhov N.N., Kholopova E.D., Guk K.K. New methods of digital processing of microfocus X-ray images. *Meditsinskaya tekhnika*. 2018;52(6):53–55 (in Russ.).]
15. Staroverov N.E., Gryaznov A.Y., Kholopova E.D., Guk K.K. Microelectronic devices digital X-ray image processing method development. *J. Phys.: Conf. Ser.* 2018;967(1):012016. https://doi.org/10.1088/1742-6596/967/1/012016

8. Mukhopadhyay S., Chanda B. A multiscale morphological approach to local contrast enhancement. *Signal Process*. 2000;80(4):685–696. https://doi.org/10.1016/S0165-1684(99)00161-9
9. Parihar A.S., Verma O.P., Khanna C. Fuzzy-contextual contrast enhancement. *IEEE Trans. Image Process*. 2017;26(4):1810–1819. https://doi.org/10.1109/TIP.2017.2665975
10. Pogany A., Gao D., Wilkins S.W. Contrast and resolution in imaging with a microfocus x-ray source. *Rev. Sci. Instrum*. 1997;68(7):2774–2782. https://doi.org/10.1063/1.1148194
11. Pizer S.M., Amburn E.P., Austin J.D., et al. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process*. 1987;39(3):355–368. https://doi.org/10.1016/S0734-189X(87)80186-X
12. Vincent L. Morphological grayscale reconstructions in image analysis: Applications and efficient algorithms. *IEEE Trans. Image Process*. 1993;2(2):176–201. https://doi.org/10.1109/83.217222
13. Грязнов А.Ю., Гук К.К., Староверов Н.Е., Холопова Е.Д. Метод повышения резкости и контрастности деталей рентгеновских изображений. *Физические основы приборостроения*. 2019;8(4(34)):34–37. https://doi.org/10.25210/jfop-1904-034037
14. Староверов Н.Е., Грязнов А.Ю., Потрахов Н.Н., Холопова Е.Д., Гук К.К. Новые методы цифровой обработки микрофокусных рентгеновских изображений. *Медицинская техника*. 2018;52(6):53–55.
15. Staroverov N.E., Gryaznov A.Y., Kholopova E.D., Guk K.K. Microelectronic devices digital X-ray image processing method development. *J. Phys.: Conf. Ser.* 2018;967(1):012016. https://doi.org/10.1088/1742-6596/967/1/012016

About the authors

Nikolay E. Staroverov, Postgraduate Student, Saint Petersburg Electrotechnical University (5, ul. Professora Popova, St. Petersburg, 197376 Russia). E-mail: nik0205st@mail.ru. Scopus Author ID 57193738290, https://orcid.org/0000-0002-4404-5222

Artem Y. Gryaznov, Dr. Sci. (Eng.), Professor, Saint Petersburg Electrotechnical University (5, ul. Professora Popova, St. Petersburg, 197376 Russia). E-mail ay.gryaznov@yandex.ru. Scopus Author ID 12142307400. https://orcid.org/0000-0003-2062-2213

Irina G. Kamyshanskaya, Cand. Sci. (Med.), Associate Professor, St. Petersburg State University (7/9, Universitetskaya nab., St. Petersburg, 199034 Russia); Radiologist of the X-ray Department, St. Petersburg State Budgetary Educational Institution "Mariinsky City Hospital" (56, Liteiny pr., St. Petersburg, 191014 Russia). E-mail: o.kafedra@mail.ru. Scopus Author ID 15834578000.

Nikolay N. Potrakhov, Dr. Sci. (Eng.), Professor, Head of Department, Saint Petersburg Electrotechnical University (5, ul. Professora Popova, St. Petersburg, 197376 Russia). E-mail: kzhamova@gmail.com. Scopus Author ID 8689381700. https://orcid.org/0000-0001-8806-0603

Ekaterina D. Kholopova, Postgraduate Student, Saint Petersburg Electrotechnical University (5, ul. Professora Popova, St. Petersburg, 197376 Russia). E-mail: wkholopova@gmail.com. Scopus Author ID 57193737033. https://orcid.org/0000-0002-9103-0066

Об авторах

Староверов Николай Евгеньевич, аспирант, ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» (197376, Россия, Санкт-Петербург, ул. Профессора Попова, д. 5). E-mail: nik0205st@mail.ru. Scopus Author ID 57193738290, https://orcid.org/0000-0002-4404-5222

Грязнов Артем Юрьевич, д.т.н., профессор, ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» (197376, Россия, Санкт-Петербург, ул. Профессора Попова, д. 5). E-mail ay.gryaznov@yandex.ru. Scopus Author ID 12142307400. https://orcid.org/0000-0003-2062-2213

Камышанская Ирина Григорьевна, к.м.н., доцент, ФГБОУ ВО «Санкт-Петербургский государственный университет» (199034, Санкт-Петербург, Университетская наб., д. 7–9); врач-рентгенолог рентгеновского отделения, СПб ГБУЗ «Городская Мариинская больница» (191014, Россия, Санкт-Петербург, Литейный пр., д. 56). E-mail: o.kafedra@mail.ru. Scopus Author ID 15834578000.

Потрахов Николай Николаевич, д.т.н., профессор, заведующий кафедрой, ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» (197376, Россия, Санкт-Петербург, ул. Профессора Попова, д. 5). E-mail: kzhamova@gmail.com. Scopus Author ID 8689381700. https://orcid.org/0000-0001-8806-0603

Холопова Екатерина Дмитриевна, аспирант, ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» (197376, Россия, Санкт-Петербург, ул. Профессора Попова, д. 5). E-mail: wkholopova@gmail.com. Scopus Author ID 57193737033. https://orcid.org/0000-0002-9103-0066

Translated by K. Nazarov

UDC 334.71: 656: 338.245
https://doi.org/10.32362/2500-316X-2021-9-6-64-72

RESEARCH ARTICLE

Method for assessing testing difficulty in educational sphere

Alexander S. Sigov,
Viktor Ya. Tsvetkov @,
Igor E. Rogov

MIREA – Russian Technological University, Moscow, 119454 Russia
@ Corresponding author, e-mail: cvj2@mail.ru

Abstract. The problem of testing in education is relevant for many countries. Testing solves three problems. The first task is to assess the quality of current training. The second task is to conduct a comparative analysis of learning outcomes. The third task is the management of the educational process in a particular educational institution and in the education sector. This determines the relevance of testing and the relevance of developing new methods for assessing test results. The article proposes a new method for assessing test results for different situations: "teacher–student," computer test, virtual testing model, test on a mixed reality model and others. To solve the problem, a special quasi-sigmoidal function is introduced. It is analogous to the logistic function, but takes into account the peculiarities of real testing of students. The logistic function ranges from minus infinity to plus infinity. There are no negative assessments in education. The introduced function lies only in the positive range of the argument. It describes actual positive scores when testing students. The authors called this function the complexity function. With its help, the complexity of the subject is assessed according to the test results. To substantiate the method, the function of the logarithms of the odds, logistic regression and the resulting Rasch method are investigated. The article notes two shortcomings of the Rasch method. The testing principle has been defined for the new function, which is used to estimate complexity. The article introduces two new concepts: the test difficulty function and the integral test score. Integral assessment of testing is a smooth function and makes it possible to go from a stepwise dependence to a continuous one. The cumulative test score translates the point test results into a continuous function and creates a correlation between the scores. The results of an experiment with the participation of RTU MIREA students are presented. The experimental results are analyzed. The possibility of using the method in educational processes is shown. The method is an alternative to the Rasch method.

Keywords: algorithm, education, testing, complexity, software components, logits, logistic equation, Rasch model

• **Submitted:** 27.08.2021 • **Revised:** 04.10.2021 • **Accepted:** 11.10.2021

For citation: Sigov A.S., Tsvetkov V.Ya., Rogov I.E. Method for assessing testing difficulty in educational sphere. *Russ. Technol. J.* 2021;9(6):64–72. https://doi.org/10.32362/2500-316X-2021-9-6-64-72

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Методы оценки сложности тестирования в сфере образования

А.С. Сигов,
В.Я. Цветков @,
И.Е. Рогов

МИРЭА – Российский технологический университет, Москва, 119454 Россия
@ Автор для переписки, e-mail: cvj2@mail.ru

Резюме. Проблема тестирования в сфере образования является актуальной для многих стран. Тестирование решает три задачи: проводит оценку качества текущего обучения, дает инструмент для сравнительного анализа результатов обучения, дает инструмент для управления образовательным процессом в отдельном учебном заведении и по отрасли. Это определяет актуальность тестирования и разработки новых методов оценки результатов тестирования. Статья предлагает новый метод оценки результатов тестирования для разных ситуаций: «преподаватель – учащийся», компьютерный тест, виртуальная тестирующая модель, тест на модели смешанной реальности и другие. Для решения этой задачи вводится квазисигмоидальная функция. Она является аналогом логистической функции, но учитывает особенности реального тестирования. Логистическая функция лежит в интервале от минус до плюс бесконечности. В образовании отрицательных оценок не бывает. Введенная функция лежит только в положительной области аргумента, то есть оценок тестирования. Эту функцию авторы называют функцией сложности. Предварительно исследуется функция логарифмов шансов, логистическая регрессия и вытекающий из этого метод Раша. Отмечены два недостатка метода Раша. Определяется принцип тестирования, который используется в функции сложности. Статья вводит два новых понятия: функция сложности тестирования и интегральная оценка тестирования. Интегральная оценка тестирования является интегральной функцией от точечных оценок. Она переводит точечные результаты тестирования в непрерывную функцию и создает корреляцию между оценками. Приводятся результаты эксперимента с участием студентов РТУ МИРЭА. Результаты эксперимента анализируются. Показана возможность применения метода в образовательных процессах. Метод является альтернативой методу Раша.

Ключевые слова: алгоритм, образование, тестирование, сложность, программные компоненты, логиты, логистическое уравнение, модель Раша

• **Поступила:** 27.08.2021 • **Доработана:** 04.10.2021 • **Принята к опубликованию:** 11.10.2021

Для цитирования: Сигов А.С., Цветков В.Я., Рогов И.Е. Методы оценки сложности тестирования в сфере образования. *Russ. Technol. J.* 2021;9(6):64–72. https://doi.org/10.32362/2500-316X-2021-9-6-64-72

Прозрачность финансовой деятельности: Никто из авторов не имеет финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The problem of assessing the complexity of testing and testing programs [1–3] is one of the most important in the field of education. At the same time, the complexity of the software components [4] as a computational object and the complexity of the test as a technological object, taking into account cognitive factors, are distinguished. Testing algorithms are qualitatively different from computational ones. They are designed to include a person in the information processing chain in the "person–test" system. Testing algorithms include two groups: algorithms for informational interaction "student–test" and algorithms for processing test results. There is a difference between the two. Algorithms of information interaction "student–test" are interactive or the algorithms of the second kind. Algorithms for processing test results are direct or the algorithms of the first kind. Testing itself can be viewed as human–machine interaction. Accordingly, the test structure or system can be considered as a human–machine system (HMS). Currently, the HMS comprises various systems, including an informational HMS. An example of the informational HMS is a geographic information system, in which information is interactively processed, and the solution to the problem is carried out by the methods of human and computer iterations. The testing system can be considered as an information man–machine system. Human resources and cognitive components are especially important for testing systems designed to work with students of a certain level of knowledge and certain forms of cognitive perception. Therefore, it is important in the process of using testing systems to evaluate the complexity of tests, to adapt them to certain levels of students' knowledge. Such adaptation is possible based on the analysis of testing results and the development of methods for comparative assessment of the complexity of testing programs based on the results of direct testing. This article is devoted to this issue.

1. RESEARCH METHODOLOGY

The research is based on logical, statistical, qualitative, and comparative analysis. The materials used were publications in the development of testing algorithms and the application of testing methods in education.

2. RESULTS OF STUDIES

2.1. Theoretical justification

In 1934, Chester Itner Bliss used the cumulative normal distribution function to map it and called it "probit" [7] as a kind of probability unit. In 1944, Joseph Berkson used the log of odds and called the function "logit," an abbreviation for "log istic un it," after the analogy of "probit." This unit has a logarithmic connotation and can be viewed as a logarithmic unit. In statistics, the logit function or log-odds is the logarithm of the odds or logarithmic odds, and the ratio [8] $p/(1-p)$ is called the chance, where p is the probability of an event occurring, $(1-p)$ is the probability non-occurrence of the event. Chance is a probabilistic characteristic, therefore logit connects the logarithmic and probabilistic components.

Logit-function (log-odds) plays an important role in logistic regression. By finding the $p/(1-p)$ relation, each probability can easily be transformed to logit. Notwithstanding a simple transformation, the physical meaning of log-odds can not always be comprehended. Jaccard, [9, p.10] called them "…illogical and complicated for interpretation, particularly in cases when large amount of statistical data is not available. However, corresponding formulas for their calculation are relatively simple, even though the results of calculation are difficult for deciphering."

Chance is the ratio of the likelihood of success to the likelihood of failure. As an equation, this is $p(\mathrm{A})/p(-\mathrm{A})$, where $p(\mathrm{A})$ is the probability of event A, and $p(-\mathrm{A})$ is the probability of "not A" (that is, the complement to A). The log odds (logit) give us the log odds of A, which can be written as

$$\log(\mathrm{A}) = \log(p(\mathrm{A})/p(-\mathrm{A})) \quad (1)$$

Generically, we can describe logarithmic chances (logit) in the form:

$$\text{log-odds} = \log[p/(1-p)]. \quad (2)$$

Mathematically, logit is the inversion of a standard logistic function $\sigma(x) = 1/(1 + e^{-x})$, therefore, logit is defined as follows:

$$\mathrm{logit}(p) = \sigma^{-1}(p) = \log[p/(1-p)] \quad (3)$$

for $p \in (0, 1)$.

Thus, logit-function $f(x)$ is a kind of function (Fig. 1), which maps probability values from (0, 1) on real numbers in set $(-\mu, +\mu)$ [10].

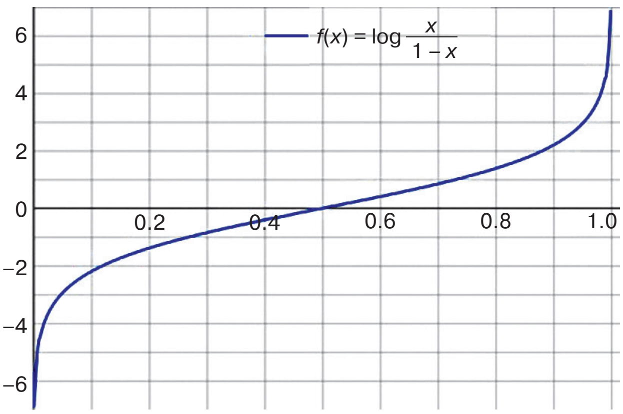


Fig. 1. Plot of the logit function

In Fig. 1, probabilities are placed on the horizontal axis. Real numbers on a conventional scale are shown on the vertical axis. The scale is chosen according to the conditions of a specific task.

Associated with the concept of logits is the concept of logistic regression. Simple logistic regression is ideologically close to linear regression. There are two differences. Linear regression uses a linear relationship between two measurements (x, y). Logistic regression uses a logistic relationship or logistic function between one measurement x and its probability p.

The logistic function can be one-parameter, two-parameter, three-parameter, and four-parameter [11]. A one-parameter function is used more often, which is in good agreement with the theory of logits.

The logistic function belongs to the class of sigmoidal functions. Sigmoid or sigmoidal function is usually called a smooth monotonic function [12], shaped like the letter "S," which has two asymptotes and describes the reaction and saturation (Fig. 1). The canonical example of a sigmoid is the one-parameter logistic function

$$f(x) = \frac{1}{1+e^{-x}}. \quad (4)$$

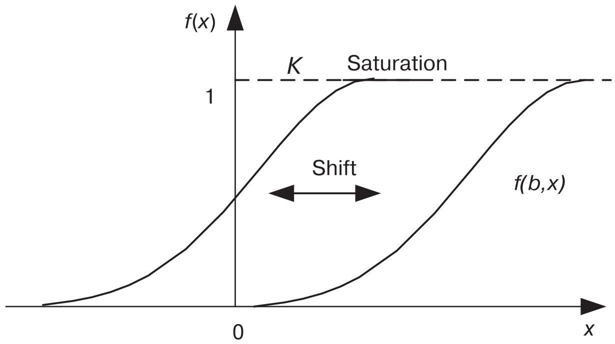


Fig. 2. Ascending sigmoid

The sigmoid is bounded by two horizontal asymptotes (Fig. 2), to which it approaches as the argument approaches to ±∞. Usually, these asymptotes (for an one-parameter function) are "0" (at $x = -\infty$) and some constant (at $x = +\infty$). In many cases, the constant at $x = +\infty$ is 1. This simplifies the relationship between probability and logistic functions. In Rasch's assessment theory, the model in Fig. 2 is used to evaluate the performance of groups. This type of sigmoid can be called ascending. Such a sigmoid has positive function values for negative argument values. To shift it, you need to enter a special parameter b (Fig. 2). There is another type of sigmoid, which is called descending (Fig. 3). For this sigmoid, the reaction shows a transition from the upper asymptote to the lower one.

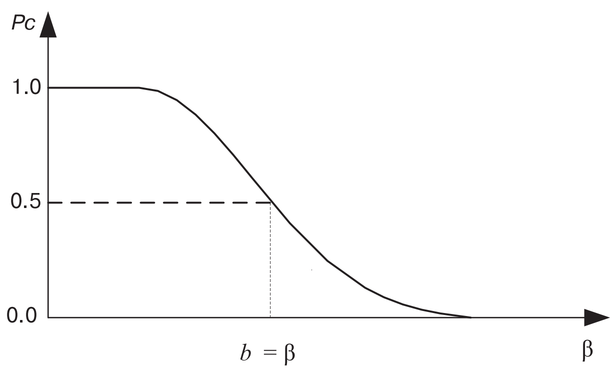


Fig. 3. Descending sigmoid

Figure 3 shows a descending sigmoid for which the probability values are plotted along the vertical axis and the test complexity values—along the horizontal axis.

The derivative of any sigmoid is a "Gaussian-like" curve with a maximum at zero, asymptotically approaching to zero at $x = \pm\infty$. The sigmoid family includes arctangent, hyperbolic tangent, and other functions.

If we compare the logit function (Fig. 1) and the descending sigmoid (Fig. 3), then the rotation of Fig. 1 by 90° clockwise will lead to complete similarity of the graphs in Fig. 1 and Fig. 3. This is what we mean when talking about the inversion of the logit function. The decreasing sigmoid (Fig. 3) is usually used to assess the difficulty of testing.

Thus, to assess the difficulty of testing, functions are used that have two asymptotes (upper and lower) and a smooth section connecting them.

2.2. Experimental research

Experimental researches were conducted in the study groups of the Institute of Information Technologies of the RTU MIREA. The shortcoming of the Rasch model when assessing the complexity of testing is that its argument exists from −∞ to +∞, while in real testing there are no negative values of the argument. This is mathematically explicable if we analyze the logit function (Fig. 1). This leads to the fact that most of the asymptote (Fig. 3) lies in the negative region of the argument, and the shift of the curve itself in Fig. 3 is used as a comparative characteristic of the test difficulty.

The basic concept of the Rasch model is that functions are needed that have two asymptotes and a transition region between them. To implement this idea, the authors introduce the "complexity function" $f_c(x)$ or "test complexity function," which exists in the positive domain of the argument and has an asymptotic complexity constraint only from below. One can speak of a quasi-sigmoidal function, since it has a restricted domain of the argument, only in the positive region. Its initial value at $x = 0$ is equal to the constant "c." This function allows us to assess the difficulty of different subjects based on the results of testing one group. There are many such functions that can be built. The authors have investigated a number of functions and chosen a simple form of the complexity function

$$f_c(x) = c\frac{1}{1+kx}. \quad (5)$$

In Eq. (5) x is a real argument, $f_c(x)$ is an assessment function. For $x > 0$, $k > 1$, $c \geq 1$. For the normalized function, f_c is equal to 1 at $x = 0$ and asymptotically approaches to 0 as $x \to +\infty$. This function is a quasi-sigmoid in the positive domain of the argument.

The ordered number of the student in the study group was chosen as an argument. This is a strictly point value. The function was determined by the assessment of the given student. The ordered student number means that the students in the group were pre-ranked by their grade level in ascending order from lowest to highest. Conditionally, these can be the numbers that correspond to real names. If we compare the ordered numbers of students to their grades, then we get an approximate picture, shown in Fig. 4.

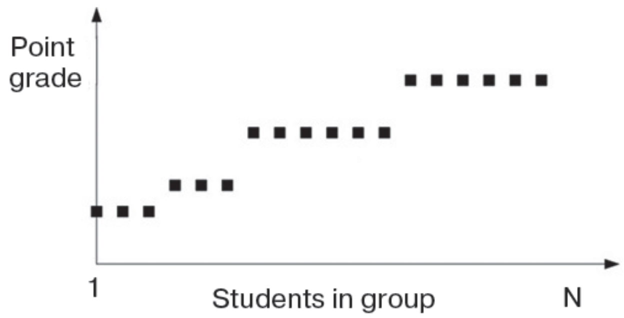


Fig. 4. Grouping grades within a group

Figure 4 shows the discrete values of the grades for the discrete value of the argument (student number in the group). There is no need to talk about any continuity in such a situation. Students can have the same grade, which corresponds to the rows of fixed grades on the graph. In the course of the research, it turned out that for sustainable assessment it is advisable to choose a function in the form of a cumulative or integral value. The formation of x is carried out as follows.

1. The grades of students in group z, obtained during testing, are recorded. Then, the statistics in the form of a "comb" is collected.
2. Grades z are ranked, ranked values z^* and stepwise increasing statistics are obtained (Fig. 4).
   $z \to z^*$.
3. After that, the integral grades x are formed according to the rule

$$x_i = \sum_{j=1}^{i} z_j^*.$$

Figure 5 shows integral and point grades.

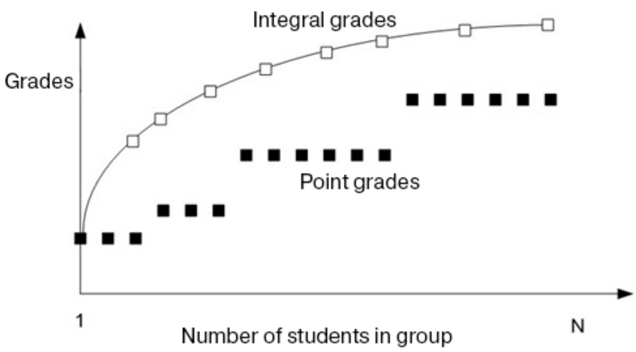


Fig. 5. Integral and point grades

In Fig. 5, integral grades (selectively) are shown by open squares. The formation of integral grades, as a method, to some degree resembles the smoothening of discrete values by the moving average method. Even the same point grades have different meanings in the integral grade. This creates correlation and continuity in the formation of grades.

The complexity function is constructed according to expression (5). The name of the function is due to the fact that it characterizes the complexity of the test or subject for a given study group. Table 1 shows the initial data of point grades for four subjects of students "Subject 1–Subject 4." In groups, 21 students were selected.

Table 1. Initial grades z^* in ascending order for four subjects in one group

Number	Subject 1	Subject 2	Subject 3	Subject 4
1	4	4.5	5	5.5
2	5	5.5	6	6.5
3	5.5	6	6.5	7
4	6	6.5	7	7.5
5	6	6.5	7	7.5
6	6	6.5	7	7.5
7	6.5	7	7.5	8
8	7	7.5	8	8.5
9	7	7.5	8	8.5
10	7.5	8	8.5	9
11	7.5	8	8.5	9
12	8	8.5	9	9.5
13	8	8.5	9	9.5
14	8	8.5	9	9.5
15	8.5	9	9.5	10
16	8.5	9	9.5	10
17	8.5	9	9.5	10
18	9	9.5	10	10
19	9	9.5	10	10
20	9	9.5	10	10
21	9	9.5	10	10

Students were assessed on a 10-point system. A 20-point assessment and even a 100-point one, as recommended by Khlebnikov [1], is acceptable. Our research has shown that it is better to normalize the grades, that is, to set the grade interval for processing from 0 to 1. Table 2 shows integral normalized grades for the same groups.

Table 2. Normalized integral grades for four subjects

Number	Subject 1	Subject 2	Subject 3	Subject 4
1	0.4	0.45	0.5	0.55
2	0.9	1	1.1	1.2
3	1.45	1.6	1.75	1.9
4	2.05	2.25	2.45	2.65
5	2.65	2.9	3.15	3.4
6	3.25	3.55	3.85	4.15
7	3.9	4.25	4.6	4.95
8	4.6	5	5.4	5.8
9	5.3	5.75	6.2	6.65
10	6.05	6.55	7.05	7.55
11	6.8	7.35	7.9	8.45
12	7.6	8.2	8.8	9.4
13	8.4	9.05	9.7	10.35
14	9.2	9.9	10.6	11.3
15	10.05	10.8	11.55	12.3
16	10.9	11.7	12.5	13.3
17	11.75	12.6	13.45	14.3
18	12.65	13.55	14.45	15.3
19	13.55	14.5	15.45	16.3
20	14.45	15.45	16.45	17.3
21	15.35	16.4	17.45	18.3

Table 3. Values of the complexity function for four subjects

Number	Subject 1	Subject 2	Subject 3	Subject 4
1	0.714286	0.689655	0.666667	0.645161
2	0.526316	0.5	0.47619	0.454545
3	0.408163	0.384615	0.363636	0.344828
4	0.327869	0.307692	0.289855	0.273973
5	0.273973	0.25641	0.240964	0.227273
6	0.235294	0.21978	0.206186	0.194175
7	0.204082	0.190476	0.178571	0.168067
8	0.178571	0.166667	0.15625	0.147059
9	0.15873	0.148148	0.138889	0.130719
10	0.141844	0.13245	0.124224	0.116959
11	0.128205	0.11976	0.11236	0.10582
12	0.116279	0.108696	0.102041	0.096154
13	0.106383	0.099502	0.093458	0.088106
14	0.098039	0.091743	0.086207	0.081301
15	0.090498	0.084746	0.079681	0.075188
16	0.084034	0.07874	0.074074	0.06993
17	0.078431	0.073529	0.069204	0.065359
18	0.07326	0.068729	0.064725	0.06135
19	0.068729	0.064516	0.06079	0.057803
20	0.064725	0.06079	0.057307	0.054645
21	0.061162	0.057471	0.054201	0.051813

Table 3 shows the calculation of the complexity function for four subjects taken by the same group.

The values given in Table 3 are used to build the function shown below (Fig. 6).

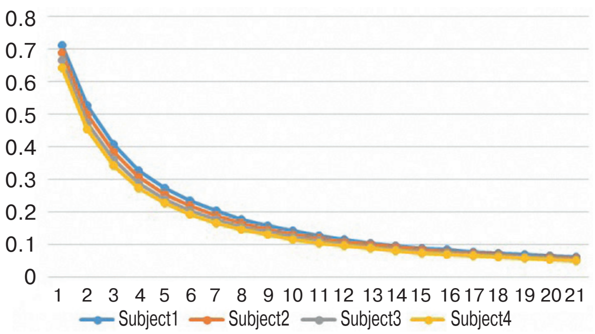


Fig. 6. Plots of the complexity function for 4 subjects

A feature of real testing (Fig. 6) is that the graph displays only a part of the sigmoid (Fig. 3), since RTU MIREA students do not receive zero grades and low grades of type 1 and 2. A value of $f_c = 1$ corresponds to a zero result in the test assessment. In reality, students always gain some points, and therefore the value of the function $f_c < 1$ for learning conditions at RTU MIREA (Fig. 6).

In the research, it was shown that the graphical proximity of the complexity functions to each other is not clear. Therefore, the true comparison is based on tabular data. Tables such as Table 3 allow us to find numerical characteristics featuring the comparative complexity of objects for a given group. There are two such characteristics: integral and point. The integral grade is defined as the integral under the curves. The point grade is determined as the average of the columns in Table 3.

The difference between the dynamics of the complexity function and the function in the Rasch model is that the Rasch model is shifted horizontally, but the complexity function is shifted vertically; the more difficult the subject, the lower the curve. Unlike the logit-based Rasch model, the complexity function is calculated with a minimum of computational resources. Using the data of Table 3, the grades for a group performance were obtained (Table 4).

Table 4. Final comparative characteristics of the complexity of 4 subjects for one test group

Indicators	Subject 1	Subject 2	Subject 3	Subject 4
Integral indicator of complexity	4.139	3.904	3.695	3.510
Relative increment of integral indicator	n/a	0.235	0.443	0.629
Point complexity indicator	0.197	0.186	0.176	0.167
Relative increment of point indicator	n/a	0.011	0.021	0.030

It follows from Table 4 that the most difficult subject for this experiment turned out to be Subject 1. The easiest subject for testing for RTU MIREA students is Subject 4 (Mathematical logic). These indicators are comparative and can be used for different subjects, but for a given group within one university. Each group is a bearer of a collective intelligence. It is a self-organizing system in which group members "pull" each other up to a common level. For another group, a qualitative correlation between the complexity of subjects is possible, but it may have quantitative difference. This method is an alternative to the Rasch model and serves as the basis for constructing fairly simple software components and algorithms for processing test results. This technique allows conducting comparative analysis with robust results.

3. DISCUSSION

A functional approach [13] and a statistical approach can be distinguished in testing. An example of a statistical approach is the application of the Rasch model [14, 15]. A significant shortcoming of this model in assessing the complexity of testing is that in it the concept of probability is incorrectly used, that is incomprehensible for specialists who do not know statistics. And it is unknown whether Rasch himself knew it. Many humanitarians use the formula without understanding its meaning and limitations. In statistics, only the frequency of observations is usually equated with probability. Usually, a statistical value is equated to a statistical value. But in the Rasch model, the probability is determined not by the statistics of observations (interval grade), but by one (point) student's grade. In this approach, the probability is considered to be 1 if a student (one-time grade) gets the highest grade, that is, a certain standard is chosen that equates to a probability equal to 1. The ratio of one student's grade to the normative grade is declared a probability and the Rasch model is applied. In fact, the grade characterizes the intellectual level of the student, taking into account his mental state (anxiety, physical fatigue, mental fatigue, etc.). In a different state, a student can receive different grades at the same intellectual level. Accordingly, the application of the Rasch method to the same group will give different grades, meaning that it is a statistically unstable characteristic. The difference in this assessment is explained by the presence of an objective causal relationship, and not by a probabilistic characteristic, as in the Rasch model. Therefore, the Rasch's model is simple and good, but completely unreliable.

The obtained results have shown that the complexity function is applicable for assessing the complexity of subjects and is easy to use. Application of the function gives relative grades [16] and group grades [17].

CONCLUSIONS

The proposed method is applicable for the majority of tests of the following kinds: "teacher–student," "software program–student," "virtual test simulator–student," "mixed reality simulator–student," and others. Testing assessment quality increases with the increase of the grading scale up to a 10-score system and higher. This method differs from the majority of the assessment methods by introduction of an integral grade that relates to the grades in a group thereby reducing random errors. The method and the algorithm based on this method are fairly clear if compared to the Rasch model which is conditional. In the Rasch model, a single-parameter logistic function is used that has one asymptote at the zero value of the function and another asymptote at the value of the function equal to 1. Such a function can suitably be related with probabilities. However, the meaning of the sigmoid slope value stays behind the frames of this work. The study conducted by Arnol'd [18] has shown that in some cases the slope can change depending on the value of a parameter of the two-parametric logistic model, and characterizes the rate of the resource consumption. The proposed method includes simple algorithms of ranging, computation of the integral variable, computation of the complexity function, and statistical processing of the results. The method is simple and available for the majority of teachers/professors in universities. At the same time, the method is accessible for researching other complexity functions.

REFERENCES

1. Neiman Yu.M., Khlebnikov V.A. *Vvedenie v teoriyu modelirovaniya i parametrizatsii pedagogicheskikh testov* (*Introduction to the theory of modeling and parametrization of pedagogical tests*). Moscow: Prometei; 2000. 168 p. (in Russ.). ISBN 5-7042-1068-6
2. Safiulin R.Z. Technology development of testing in education. *Upravlenie obrazovaniem: teoriya i praktika* = *Education management review*. 2015;1(17):139−149 (in Russ). Available from URL: https://emreview.ru/index.php/emr/issue/view/17
3. Kolster R. Structural ambidexterity in higher education: excellence education as a testing ground for educational innovations. *Eur. J. High. Educ*. 2021;11(1):64−81. https://doi.org/10.1080/21568235.2020.1850312
4. Kudzh S.A., Tsvetkov V.Ya., Rogov I.E. Life cycle support software components. *Rossiiskii tekhnologicheskii zhurnal* = *Russian Technological Journal*. 2020;8(5):19−33 (in Russ.). https://doi.org/10.32362/2500-316X-2020-8-5-19-33
5. Mgbemena E. Man-machine systems: A review of current trends and applications. *FUPRE J. Sci. Ind. Res*. (*FJSIR*). 2020;4(2):91−117.
6. Bogoutdinov B.B., Tsvetkov V.Ya. Application of the model of complementary resources in investing activities. *Vestnik Mordovskogo universiteta* = *Engineering Technologies and Systems*. 2014;24(4):103−116 (in Russ.). https://doi.org/10.15507/VMU.024.201403.103
7. Bliss C.I. The method of probits. *Science*. 1934;79(2037):38−39. https://doi.org/10.1126/science.79.2037.38
8. Viera A.J. Odds ratios and risk ratios: what's the difference and why does it matter? *South. Med. J.* 2008;101(7):730–734. https://doi.org/10.1097/smj.0b013e31817a7ee4
9. Jaccard J. Interaction Effects in Logistic Regression. USA: Sage Publishing; 2001. V. 135. 80 p. https://dx.doi.org/10.4135/9781412984515
10. Cramer J.S. The origins and development of the logit model. In book: *Logit models from economics and other fields*. Cambridge University Press; 2003. Ch. 9. https://doi.org/10.1017/CBO9780511615412.010
11. Jokar M., Subbey S., Gjøsæter H. A logistic function to track time-dependent fish population dynamics. *Fisheries Research*. 2021;236:105840. https://doi.org/10.1016/j.fishres.2020.105840
12. Muñoz García D.A., Cardona Gómez D.C., Hoyos Mendez Y.C. Vólvulo del sigmoide: revisión narrative. *Revista Facultad Ciencias de la Salud*: *Universidad del Cauca*. 2020;22(1):36−44. https://doi.org/10.47373/rfcs.2020.v22.1575
13. Tsvetkov V.Ya., Rogov I.E. Functional approach to the assessment of academic performance of study groups. *Obrazovatel'nye resursy i tekhnologii* = *Educational Resources and Technologies*. 2021;1(34):61−68 (in Russ.).
14. Seif N.A.S.M.A. *Building and standardizing a school readiness battery for kindergarten children in light of classical theory and item response theory* (*A comparative study*). MA Thesis. Fayoum University; 2018. Available from URL: http://www.fayoum.edu.eg/english/Education/Psychology/pdf/MsnadaMscE.pdf

СПИСОК ЛИТЕРАТУРЫ

1. Нейман Ю.М., Хлебников В.А. *Введение в теорию моделирования и параметризации педагогических тестов*. М.: Прометей; 2000. 168 с. ISBN 5-7042-1068-6
2. Сафиулин Р.З. Развитие технологий тестирования в образовании. *Управление образованием: теория и практика*. 2015;1(17):139−149. URL: https://emreview.ru/index.php/emr/issue/view/17
3. Kolster R. Structural ambidexterity in higher education: excellence education as a testing ground for educational innovations. *Eur. J. High. Educ*. 2021;11(1):64−81. https://doi.org/10.1080/21568235.2020.1850312
4. Кудж С.А., Цветков В.Я., Рогов И.Е. Поддержка жизненного цикла программных компонент. *Российский технологический журнал*. 2020;8(5):19−33. https://doi.org/10.32362/2500-316X-2020-8-5-19-33
5. Mgbemena E. Man-machine systems: A review of current trends and applications. *FUPRE J. Sci. Ind. Res*. (*FJSIR*). 2020;4(2):91−117.
6. Богоутдинов Б.Б., Цветков В.Я. Применение модели комплементарных ресурсов в инвестиционной деятельности. *Вестник Мордовского университета*. 2014;24(4):103−116. https://doi.org/10.15507/VMU.024.201403.103
7. Bliss C.I. The method of probits. *Science*. 1934;79(2037):38−39. https://doi.org/10.1126/science.79.2037.38
8. Viera A.J. Odds ratios and risk ratios: what's the difference and why does it matter? *South. Med. J.* 2008;101(7):730−734. https://doi.org/10.1097/smj.0b013e31817a7ee4
9. Jaccard J. Interaction effects in logistic regression. USA: Sage Publishing; 2001. V. 135. 80 p. https://dx.doi.org/10.4135/9781412984515
10. Cramer J.S. The origins and development of the logit model. In book: *Logit models from economics and other fields*. Cambridge University Press; 2003. Ch. 9. https://doi.org/10.1017/CBO9780511615412.010
11. Jokar M., Subbey S., Gjøsæter H. A logistic function to track time-dependent fish population dynamics. *Fisheries Research*. 2021;236:105840. https://doi.org/10.1016/j.fishres.2020.105840
12. Muñoz García D.A., Cardona Gómez D.C., Hoyos Mendez Y.C. Vólvulo del sigmoide: revisión narrative. *Revista Facultad Ciencias de la Salud*: *Universidad del Cauca*. 2020;22(1):36−44. https://doi.org/10.47373/rfcs.2020.v22.1575
13. Цветков В.Я., Рогов И.Е. Функциональный подход к оценке успеваемости учебных групп. *Образовательные ресурсы и технологии*. 2021;1(34):61−68.
14. Seif N.A.S.M.A. *Building and standardizing a school readiness battery for kindergarten children in light of classical theory and item response theory* (*A comparative study*). MA Thesis. Fayoum University; 2018. URL: http://www.fayoum.edu.eg/english/Education/Psychology/pdf/MsnadaMscE.pdf
15. Цветков В.Я., Войнова Е.В. Модификация модели Раша для оценки свободного тестирования. *Вестник Рязанского государственного радиотехнического университета*. 2018;1(63):90−94. https://doi.org/10.21667/1995-4565-2018-63-1-90-94
16. Кудж С.А. Методы сравнительного анализа. *Славянский форум*. 2019;3(25):140−150.

15. Tsvetkov V.Ya., Voinova E.V. Modification of Rasch model for free testing assessment. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta = Vestnik of RSREU*. 2018;1(63):90–94 (in Russ.). https://doi.org/10.21667/1995-4565-2018-63-1-90-94
16. Kudzh S.A. Methods for comparative analysis. *Slavyanskii forum = Slavic Forum*. 2019;3(25):140–150 (in Russ.).
17. Kudzh S.A. Assessment of group cognitive complexity. *Slavyanskii forum = Slavic Forum*. 2018;2(20):36–43 (in Russ.).
18. Arnol'd V.I. Catastrophe theory. *Itogi nauki i tekhniki*. "*Seriya Sovremennye problemy matematiki. Fundamental'nye napravleniya*". 1986;5:219–277 (in Russ.).

17. Кудж С.А. Оценка групповой когнитивной сложности. *Славянский форум*. 2018;2(20):36–43.
18. Арнольд В.И. Теория катастроф. *Итоги науки и техники. Серия «Современные проблемы математики. Фундаментальные направления»*. 1986;5:219–277.

About the authors

Alexander S. Sigov, Academician of RAS, Dr. Sci. (Phys.–Math.), Professor, President, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: sigov@mirea.ru. ResearcherID L-4103-2017, Scopus Author ID 35557510600.

Viktor Ya. Tsvetkov, Dr. Sci. (Eng.), Dr. Sci. (Econ.), Professor, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: cvj2@mail.ru. Scopus Author ID 56069916700.

Igor E. Rogov, Director of the Institute of Pre-University Training, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: rogov@mirea.ru.

Об авторах

Сигов Александр Сергеевич, академик РАН, д.ф.-м.н., профессор, президент ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: sigov@mirea.ru. ResearcherID L-4103-2017, Scopus Author ID 35557510600.

Цветков Виктор Яковлевич, д.т.н., д.э.н., профессор, советник ректората ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: cvj2@mail.ru. Scopus Author ID 56069916700.

Рогов Игорь Евгеньевич, директор Института довузовской подготовки ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: rogov@mirea.ru.

Translated by E. Shklovskii

Russian Technological Journal. 2021;9(6):73–87 ISSN 2500-316X (Online)

UDC 001.5+621.3
https://doi.org/10.32362/2500-316x-2021-9-6-73-87

REVIEW ARTICLE

A discussion of S.M. Krylov's book "Neocybernetics" (2008)

Evgeny S. Gornev [1], Igor V. Matyushkin [1, 2, @]

[1] *Molecular Electronics Research Institute, Moscow, Zelenograd, 124460 Russia*
[2] *National Research University of Electronic Technology, Moscow, Zelenograd, 124498 Russia*
[@] *Corresponding author, e-mail: imatyushkin@niime.ru*

Abstract. A comparative analysis of the "general formal technology (GFT)" by S. M. Krylov is carried out in the context of the published book of the authors "General Theory of Technologies and Microelectronics" (2020) and on the basis of his work of 2008. Despite the abstractness of the algebraic-algorithmic approach, Krylov offers a number of specific constructions that are in demand during the fourth industrial revolution and for the future development of industrial technology in nanoelectronics and biotechnology. Industrial technology is considered as a complex object of management, i.e., it is the object of study of the new discipline "neocybernetics". Although the foundations of this approach were laid in 1930s–1960s within the framework of logical and mathematical research, its expansion is inevitable when using self-organization processes to obtain functional supramolecular structures in technological processes of nanoelectronics (for example, DNA origami engineering). The issues of complexity quantification for a product itself (structure) and its manufacturing technology, or, according to Krylov, the complexity of technological automata, have become even more relevant than before. The theoretical issues of self-organization, the development of artificial life, and the creation of self-replicating technical systems also seem promising for solution. In our opinion, Krylov's formal technology is an important "block" in the advancement of general theory of technologies (GTT) useful for describing the technology at the levels: operation, route, and process. We would like to encourage a wide range of readers to study the book and form a steady interest in general technological issues. The value of GTT and GFT extends beyond the sphere of technology and, in a narrow sense, factory production, but also into the area of "fine" regulation of physiology in biological objects and pharmacy, as well as into the problem field of cognitive sciences, psychology, and education. when the focus is on the personality structure and heterogeneous constructs "floating in the sea of the unconscious". Both S.M. Krylov and we demonstrate that the issues of industrial technology cannot be considered without abstract formalization and without reference to philosophy.

Keywords: general technology theory, formal technology, neocybernetics, finite-state machines, synergetics, complexity

• **Submitted:** 26.04.2021 • **Revised:** 20.09.2021 • **Accepted:** 27.09.2021

For citation: Gornev E.S., Matyushkin I.V. A discussion of S.M. Krylov's book "Neocybernetics" (2008). *Russ. Technol. J.* 2021;9(6):73–87. https://doi.org/10.32362/2500-316X-2021-9-6-73-87

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

РЕЦЕНЗИЯ

Дискуссия по книге С.М. Крылова «Неокибернетика» (2008)

Е.С. Горнев [1], **И.В. Матюшкин** [1, 2, @]

[1] АО «НИИ молекулярной электроники», Москва, Зеленоград, 124460 Россия
[2] НИУ «Московский институт электронной техники», Москва, Зеленоград, 124498 Россия
[@] Автор для переписки, e-mail: imatyushkin@niime.ru

Резюме. В контексте вышедшей книги авторов «Общая теория технологий и микроэлектроника» (2020) проводится сравнительный анализ «общей формальной технологии (ОФТ))» С.М. Крылова на основе его работы 2008 года. Несмотря на абстрактность алгебраико-алгоритмического подхода, Крылову удалось предложить ряд конкретных конструкций, востребованных в ходе четвертой промышленной революции и для будущего развития промышленной технологии наноэлектроники и биотехнологии. Промышленная технология рассматривается как сложный объект управления, т.е. выступает объектом изучения новой дисциплины «неокибернетика». Хотя основы данного подхода закладывались в 1930–1960 годах в рамках логико-математических исследований, но его расширение неизбежно при использовании процессов самоорганизации для получения функциональных надмолекулекулярных структур в технологических процессах наноэлектроники (например, инженерия ДНК-оригами). Вопросы исчисления сложности самого изделия (конструкции) и технологии его изготовления, или, по Крылову, сложности технологических автоматов, стали еще более актуальными, чем раньше. Теоретические вопросы самоорганизации, разработки искусственной жизни, создания самовоспроизводящихся технических систем также представляются перспективными для решения. Формальная технология Крылова выступает, по нашему мнению, важным «блоком» при построении общей теории технологий (ОТТ), полезным для описания технологии на уровне технологических: операции, маршрута и процесса. Хочется побудить широкий круг читателей к прочтению книги и сформировать устойчивый интерес к общетехнологической проблематике. Значение ОТТ и ОФТ простирается за пределы сферы техники и, в узком смысле, заводских производств, в область «тонкой» регуляции физиологии в биологических объектах и фармацевтика, а также в проблемное поле когнитивных наук, психологии, образования, когда в центре внимания оказываются структура личности и гетерогенные конструкты, «плавающие в море бессознательного». Мы, как и С.М. Крылов, демонстрируем, что вопросы промышленной технологии нельзя рассматривать без абстрактной формализации и без обращения к философии.

Ключевые слова: общая теория технологий, формальная технология, неокибернетика, конечные автоматы, самоорганизация, сложность

• **Поступила:** 26.04.2021 • **Доработана:** 20.09.2021 • **Принята к опубликованию:** 27.09.2021

Для цитирования: Горнев Е.С., Матюшкин И.В. Дискуссия по книге С.М. Крылова «Неокибернетика» (2008). *Russ. Technol. J.* 2021;9(6):73–87. https://doi.org/10.32362/2500-316X-2021-9-6-73-87

Прозрачность финансовой деятельности: Никто из авторов не имеет финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

In the paper, we would like to pay tribute to the memory of Sergey Mikhailovich Krylov (1948–2020), Doctor of Technical Sciences, Professor of the Computer Department at Samara State Technical University, who has contributed to the creation of the general theory of technologies (GTT) [1]. Krylov's system of ideas on the formal theory of technologies (GFT, sometimes simply FT), as he called it, is presented in several papers [2], his dissertation [3], and three books [4–6]. Although the paper aims to compare the GTT and GFT, it would be appropriate to refer directly to only one book entitled "Neocybernetics" [5] (Fig. 1) fallen into the category of rare books. To understand the genesis and relationship between Krylov's ideas beyond the scope of the GFT and GTT, the synopsis review form is chosen by the authors. The book itself is written in a popular science style rather than in a scientific one. It should be noted that the term "neocybernetics" is increasingly used by different authors [7] while the distinction with Wiener cybernetics is along the lines of the control object complexity and, accordingly, incomplete information on its state. Any technological system can be considered just as a control object of higher complexity.

Fig. 1. S.M. Krylov and his book cover [5][1]

To clarify the subject of the GTT and GFT study, it would be appropriate to quote one passage used by Krylov and other authors as well. In the article "Technology" published in 1901 in the Brockhaus and Efron Encyclopedic Dictionary, D.I. Mendeleev characterizes technology as the science dealing with producing artificial objects, i.e., objects that do not exist in nature, from natural raw materials:

"The technology is the doctrine of advantageous (i.e., absorbing the least human labor and energy of nature) methods of processing natural products into the ones needed (necessary or useful, or convenient) for use in human life. Although technology is profoundly different from socioeconomic doctrines by its subject, it has rather many direct and indirect relations with them, since the economy (saving) of labor and material (raw material), as well as time and forces through them, is the first priority task of any production. And the essence of the factory production doctrine loses its ground completely if the production advantageousness (economy) is overlooked... The task of technology shows that it does not have those highest and absolute requirements which the abstract sciences related to the visible or internal nature have, that it contains in itself the application of other more abstract knowledge to life, and that its content must change according to the circumstances and conditions of place and time. However, these, so to speak, negative aspects of technology are redeemed, firstly, by the direct and vital importance, which factories and plants already have nowadays and which must strengthen more and more in the future. Secondly, by the fact that the doctrine of the methods used by plants and factories illuminates via scientific principles what is produced by practice. Through this not only the production is improved but also the scope of scientific understanding of things and phenomena is expanded" [8].

BOOK COMPOSITION

The Preface and Introduction provide the following three sources inspiring Krylov: "Tektology" by Bogdanov [9], "Notes by the Translator" by Augusta Ada Lovelace (1843), and works on synergetics. It should be reminded that Bogdanov has defined tektology as a general organizational science with emphasis placed on forms of organization (thus, "regulating forms" anticipate cybernetics). Why is the world organized? – this question arises for Krylov as well, so he writes: "The main goal of GFT is to try to explain logically (and technologically) the structure of the world around us at the level, first of all, of our perception, that is at macroscopic level, which 'scale center' is man himself. And the way this macrolevel can be realized in terms of different possible physical microlevels is kind of a secondary issue that is not considered here" [5, p. 11]. At the same time, "formal technology, on the contrary, starts from studying the most general properties of the most general operations on both real and imaginary (abstract) objects, moving down in the hierarchy towards more and more specific properties of more and more specific objects and operations" [5, pp. 6–7]. For better understanding, reference should be made to Bogdanov himself [9, p. 79]: "With the infinitely rich material of the Universe and the infinite variety of its forms, where can these persistent and systematic, repeating, and increasing, with the growth of knowledge, analogies come from? To regard them as 'incidental coincidences' is to introduce a great arbitrariness into the world-view and even to come into obvious contradiction with the theory of probabilities. There can be only one scientifically justified conclusion: the actual unity of organizational methods is found everywhere – in mental and physical complexes, in living and dead nature, in the work of spontaneous forces and the conscious activity of people." According to Bogdanov, Krylov represented formative forms as the development of systems: "The development of any system—material or informational—is a technological process to some extent. Therefore, a very important application area for formal technology is the most various development processes which we shall further call evolutionary: both natural existing in the surrounding environment and artificial realized by man or with his participation" [5, p. 9]. It should be

[1] URL: http://vt.samgtu.ru/index.php. Accessed: October 25, 2021 (in Russ.).

noted that Krylov pursues the ideas of evolution more consistently, compared to the GTT, although evolution is considered here [1] not only in terms of technology but also in terms of technological operation (TO). Evolutionarity, manifesting itself in the technosphere through the creative activity of the spirit [1, p. 380], is a universal natural feature:

"In nature, through the chaos of innumerable variations, the same dialectical laws of motion that dominate the seemingly random events in history as well, being the same laws that run like a golden thread through the evolution of human thinking to gradually sink into the mind of thinking people, are paving their ways" [10].

Borrowed from Ada Lovelace, Byron's daughter, the view of TO captures the systematicity of the object being modified, in addition to abstractness: "By the word operation we mean any procedure changing mutual relationship between two or more things, whatever kind of relationship they may be. This is the most general definition including all objects in the Universe. But the science of such operations being specifically derived from mathematics is an independent science having its own theoretical validity and significance, just as logic has its own validity, regardless of the objects to which its explanations and methods are applied." In the GTT, TO is introduced as a change in the state of the substratum, but Krylov specifies any change as a change in the mutual relationship between parts. Then, Krylov identifies informational and "ordinary" material technologies according to the object types. The former is studied by the classical theory of algorithms [5, p. 11]: "Being the flesh and blood of mathematics (theory of algorithms, to be more exactly), GFT may adapt any of its methods developed so far to its needs without too much trouble. Moreover, the formal technology defines also the place of mathematics itself (as a computing technology) among all the many technologies very clearly." This is in compliance with our definition of technology as "the objectification of widely used algorithms" [1, p. 30].

We paid attention to the complexity when considering the level of technological route (TR). Krylov goes a little further, linking quite rightly the complexity with self-organization and thus, with synergetics (represented by Prigozhin primarily). He writes about synergetics and GFT: "Mutual recognition and 'exchange of ideas' may benefit both scientific areas, whereas a possible confrontation is likely to reflect only the typical social situation of 'power seizure.'" It should be noted that synergetics connects holistic perception with the search for order parameters in the research object; i.e., there must be order in any object, and the task of synergetics is the search for it. Krylov also writes about the constructive nature of GFT up to the construction of new "physical worlds", and, in our opinion, this only further focuses on the design technologies as twin ones in relation to the "manufacturing" technologies, as A.A. Ivlev has written [11]. In any case, studying self-organization problems cannot be separated from those of the complexity calculus.

The focus on constructiveness represents, perhaps, the most original feature of Krylov's approach consisting in "ultimate abstracting the notions of operation objects and mechanisms of linking objects to each other into various constructs and structures, whether it would be the words consisting of symbols, units and aggregates consisting of parts, or 'molecules' built of 'atoms'... It should be noted that it took about ten years to develop this seemingly very crude and very simplified approach" [5, pp. 15–16].

Chapter 1 entitled "Formal Technology and Metaphysics" is historical and bibliographic in nature, while Chapter 2 entitled "What Does Functionality Give?" focuses on defining GFT. In our book, we have done exactly the same, while the notions of "product," "target product," and thus functionality, are put at the forefront of the GTT as well. Chapter 3 entitled "Up the Ladder of Evolutionary Technologies" introduces the dynamic and evolutionary nature into GFT, but at the same time, the concept of hierarchy is naturally introduced [1, p. 69]. Not only changes in structures implementing each operation but also certain technologies and their evolutionary transformations, operations and working transitions of these operations, the role and nature of operations, and the change of this role during evolution may be analyzed. Thus, the evolution of microelectronics [1, Chapter IV] illustrates a general case of the technology evolution. Here, the evolution of processes as well as the evolution of the means implementing these processes are considered.

The next two chapters entitled "Why and How Life Emerges" and "Evolution of Systems" deal with the problems being fashionable in the 1960s, the discussions on which will be, certainly, continuing. Krylov refers to the concepts of entropy, negentropy, and adaptation mechanisms; and it is easy to see parallels with compositions of the fourth and fifth chapters [1] in this. Adaptability and homeostasis are important concepts when discussing the industrial technology. Chapters 6 and 7 on program-controlled technological systems of the future and Church's thesis, respectively, illustrate technical aspects of the production organization. Here, there is also some parallel with the chapters of our book on levels of production process and technology. Chapter 8, "How and Where Do We Develop To? Social and Economic Systems," where the focus is made on human aspects, e.g., beauty, is analogous to the last, seventh chapter [1], at least as far as the considered problem is concerned.

CONTENT AND DISCUSSION OF CHAPTER 1

Chapter 1 begins with a lengthy discussion of man's knowledge of the world gained through religion and then philosophy as well as through the unity of metaphysics and mathematics in terms of the number concept: "This was how the Pythagoreans emerged. Enthralled by the beauty and orderliness of arithmetical and geometrical constructions, they tried to deprive the number of its main advantage: the abstractness." Today, mathematics has become a theory of algorithms, and a return to metaphysics is required on a new level to explain the world. So, the question is asked: "But there is no unified mathematical theory on the origin of life. And why not? What is so special about mathematics as a generalized universal method for describing and solving all problems, that radically distinguishes it from those methods used by nature?" And the answer is given: "In all cases without exception, mathematics operates exclusively with artificial objects: numbers or some of their "substitutes" (symbols and/or codes). Nature "works" only with physical objects... In short, real physical objects being in physical contact with each other always manifest their inherent functionality related to their inherent physics, that is, to the physics of the Universe in which the interaction takes place" [5, pp. 24–25].

The problem raised by Krylov is quite delicate and belongs to sophisticated metamathematics. In practical terms, Krylov is right referring to the hidden axiom of mathematics: "For any mathematical transformations, the primitive objects of mathematical operations (numbers and codes) have only one abstract property that is to represent some quantities." However, it is commonly known that in Cantorian set theory being the qualitative theory of differential equations, numbers are secondary. This difference in approaches to mathematics occurs even in mathematical logic itself, if the works of the formalist Hilbert and his contemporary Gottlob Frege (predicate calculus) are compared. Syntactic thinking is characteristic of the Kolmogorov tradition as well as of the work "Algorithmic Theories of Everything" (2000) by Jürgen Schmidhuber cited by Krylov [12]. The physicosemantic approach is closer to us [1, p. 204]; and the same position is also shared by Krylov: "It is always possible to find objects whose description in 'Schmidhuber universes' turns out to be not only incomplete but completely inadequate" [5, p. 28]. On the other hand, any study, including that of technology, inevitably encounters problems of language and symbolism.

Here, the notion of functionality is used, but Krylov understands it unconventionally. In this case, attention is drawn to the object-oriented programming (OOP) paradigm and the "Object–Properties–Functionality" triad. A formula is written out [5, p. 30] stating the following: 1) an object can have several properties; 2) each property depends on the properties of other objects (and other properties of the object itself), while functionality is the functional relationship itself. Krylov warns: "It makes no sense to consider only functionality without related physical properties: it would be just a part of a long-standing branch of mathematics called 'the general theory of computable functions.'" Two levels of GFT are identified: simple level based on the "Object–Properties" dyad; and the more developed one considering the entire triad. Two types of formal TOs are also immediately identified: synthesis similar to addition, and decomposition, i.e., separation. It is noteworthy that Krylov also introduces the TO analysis coinciding in meaning with the Measurement and Attestation (M&A) TO introduced by us [1, p. 116]; this is expressed mathematically as $F_i(x) \rightarrow < x, \beta >$, where β stands for the desired characteristic (parameter) of object x. Also, the seemingly incorrect case when $\beta \in \{0, 1\}$ stands for the sign of absence or presence of an object as well as the general case of measurement as a comparison of the particular object form with a standard is considered.

The TO definition is given in [5, p. 35]: "Objects in FT may in fact be understood as everything to which its 'technological operations' can be applied, that is, operations affecting in some way the objects themselves or/and their properties, or/and their mutual arrangement." By analogy with the alphabet in mathematical logic, Krylov introduces "naturally," in his own words, basic elements or the base for technology. And after all, considering the history of microelectronics, this theoretical idea seemingly taken from a mathematical mind has been indeed embodied in a list of materials, structural elements (of electronic component base), and typical technological processes. At the same time, modern technologies are open to innovations; they evolve changing their own "axiomatics." An algebraic, in spirit, definition of formal technology as a set of base and operations on its elements and structures, i.e., obtained from the elements of the base of complex objects, is given in [5, p. 36]. Here, the analogy with the carrier and signature of an algebraic system may be easily seen.

Then, the notion of creative technology is introduced, which is excessive in our opinion, since any practical technology always creates something new, setting aside the copying technology being an exotic case for production (e.g., polymerase chain reaction or the analogy with image transfer in microelectronics). Creativity, it seems, should be correlated with "usefulness." In this case, Stanislaw Lem's thought about classifying technology as science may be recalled, and it may be said that there is no useless science in the most pragmatic meaning of the word "useful," since it is never known in advance what information on nature would be useful and, moreover, would be extremely important and necessary. On the other hand, the distinction between creative and non-creative technologies gives Krylov grounds to speak on self-replication of biological cells and on amplification of life forms during evolution (up to the emergence of intelligence) as well as to introduce the notion of "evolutionary technology" [5, p. 41]. When analyzing the sum of technologies and using the notion of conjugate points [1, p. 355], Krylov, like us, comes to the idea of evolution through the development and complication of the technology base, but in an alternative way using the notions of creativity and base.

Krylov draws attention to the fact that "the novelty of structures should be determined exclusively by means of technology itself." This idea seems profound and important for analyzing the inventive activity and innovations, i.e., project activity becomes subordinate to the production and technological one. In practice, the opposite is usually true; at first, the structure, and then only the production route is determined. In a broader sense, it would be better to move to the level of the production process rather than to that of the route [1, p. 31]. Eventually, not "what to do" but rather "how to do it" becomes the main thing. Only then, the novelty would get the right to exist. At the same time, the inseparable unity of technology and structure has been pointed out, and Krylov strengthens and singles out the second side of this unity based on algebraic considerations. As for self-replication, Krylov's thought refers to Codd's cellular automata and sufficiency of a low level of functionality determined by the number of possible local transition functions and/or the number of cell states.

By analogy with a function of several variables, Krylov introduces a formal-technological function (FT-scheme of synthesis) that results in a new object/structure. Here, the set of acceptable TO parameters is explicitly introduced; in our terminology, this is the *tekhne* of the TO level [1, p. 115]. In addition to the schemes of synthesis, the "fixed disintegration" operation similar in meaning to "dosing" in food technology is introduced. Mathematical definitions for a number of private types of TO are given. And it should be said here that the TO described by Krylov is rather equivalent to microoperations [1, p. 108].

The following theorem on "the efficiency of accumulated knowledge," i.e., on the possibility to number all TRs, is formulated. Krylov explains: "Any technology being a finite algorithmic system by definition (i.e., a system with a finite number of types of initial operation objects and a finite number of operations on objects themselves) can generate only a finite number of objects (syntheses or structures) for a finite number of stages." The theorem is true abstractly and mathematically only. Some operations change the properties of objects while the combination of such properties, even with a finite number of operations, may be infinite in practical terms. This implies a somewhat fantastic, in our opinion, idea of the constructor–analyzer representing a Turing machine connected to a technological unit. According to Krylov, the technology is complete when it is infinitely creative and the technology of its synthesis can be recovered from the product design. The recoverability is seemingly understood as existence with an accuracy of unambiguity, since, for example, a joint can be made by different technological welding methods (laser, electric, ionic, etc.) exclusive of various additives. Krylov considers the coincidence of the term with completeness in logic accidental, but it seems that the analogy is quite adequate, since a formula is provable when there is a chain of formulas linking it with axioms. Chapter 1 is concluded by several mathematized definitions and statements as well as by a summary figure showing subtypes of technologies [5, p. 68], (Fig. 2).

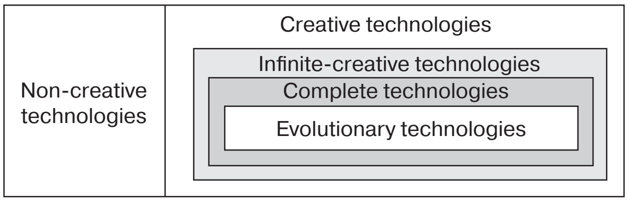


Fig. 2. Types of technologies selected by S.M. Krylov

CONTENT AND DISCUSSION OF CHAPTER 2

Chapter 2 deals with abstract notions of property and functionality, being rather difficult to understand; it invades the realm of philosophy, and is based on works by M. Bunge and E.M. Karpov. Krylov is aware of the complexity of the issue: "Interpreting the meaning of the term 'property' in each specific case in relation to each version of representation of systems, their components, and relations between components is one of the most complexities in studying systems... Referring to properties of individual objects and systems of objects, the fundamental difference in properties related to the dynamics of their behavior and in properties statically connected with objects themselves, that is those that allow distinguishing one object from another regardless of their dynamics, are not usually emphasized" [5, p. 74]. Krylov embarks upon the path of G. Klir and E.M. Karpov and gives the following definition: "A property of an object (element or structure) is its some numerical, geometrical, physical, or any other characteristic, including the imaginary one, remained steady under certain stable conditions, by which they (elements, structures) may differ from each other or from their state preceding in time and/or space."

Krylov emphasizes the technical meaning of this definition and rightly points out that "the presence of object properties may not necessarily imply the presence, among analytical operations, of the corresponding technology for operations directly determining the 'presence' or 'absence' of certain properties, and even 'measuring' their value" [5, p. 78]. Krylov also extends the predicative meaning of the property and writes openly that it is quite possible to display the property using an integer or finite real number, not only a Boolean number. It should be noted that this approach exists in the probability theory when a random event is compared with a random variable. According to Krylov, "in this notation, the possibility to limit the real (i.e., quite complete) list of object properties by its reduced version representing a full picture of their behavior in a particular context becomes quite obvious" [5, p. 80]. This problem of incompleteness is rather essential for practicing M&A TO of industrial technology and interpreting their results (in this regard, see our comments on the physical–statistical approach in the reliability theory [1, p. 233]). With regard to the importance of OOP, Krylov writes: "For us, the significant thing in this concept is its most important, central link, namely, that any object should be represented by a single set of parameters describing its features (properties!) as well as methods (techniques and functions) describing the use (interaction) of parameters inside and outside the object, i.e., with other objects. Indeed, any real object of the real world is not just a set of its static characteristics. It is also a process of its existence in the world around, its dynamics, and its evolution" [5, p. 81].

The object state is defined as a set of specific values of its properties and is interpreted as a vector in a multidimensional "space" (in our understanding, this space is analogous to the phase space in statistical physics). Noteworthy is the "method of assigned functionalities" consisting in decomposition of a property having multiple values into several properties with few ones, e.g., binary values [5, p. 83]. And, as Krylov rightly notes, this method is often used in designing cellular automata.

It should be repeated that the notion of "functionality" is introduced nonconventionally. In the GTT, functionality is the property of the final product to perform a useful function, but it may be undoubtedly possible to postulate the local functionality of the element of product or semi-finished product responding to external impact under some other given conditions in a "correct" manner, in the developer's view. Such a need does arise in microelectronics when characterizing basic structural elements, while the degree of "correctness" is fixed in the "Platform Development Kit." Krylov, though also mentioning external impact (through change), takes a different approach [5, p. 84]: "The property of an object (element or structure) is called functional when it allows changing the state of another object which this property of this object interacts with when these objects physically converge or connect." In our opinion, it would be more correct to speak of linking by introducing the abstraction "linkage," since it is hard to imagine physical convergence for information objects. Then, Krylov discusses the example of a *2AND–NOT* gate where its two inputs are called input properties while its output, i.e., the output property, becomes functional only when it is connected to another gate. However, functionality does exist without a connection, being specified by the truth table and its "correctness." Then, the operation of the four-component flip-flop is discussed in the language of functionality [5, p. 90]. As a more complex example, Krylov introduces the so-called "semistructured heterogeneous self-timed machine with individual object names" in the context of calculations using chemical reactions (in our opinion, a rather good example is given here by the Kirdin kinetic machine). Krylov refers to the Kauffman Boolean automaton being another network implementation.

Several statements on emergent properties concludes Chapter 2. Whereas the system-wide statement on emergence is considered in the GTT only in the sense that each level of technology description generates new specific concepts, Krylov links emergence with functionality in a more classical way. "An emergent property is the property of object x (with a non-empty value) been absent at previous steps of technology in those objects which object x is obtained from" [5, p. 100]. Here is an example of statements: "Statement 2.1. The property 'to have a given form' is emergent (in technologies, with dimensionality at least 2)." In our understanding, the dimensionality of the technology is identified with the number of properties of its final structure. For the von Neumann 29-state cellular automaton, two statements are formulated in the language of functionality. "Functionality" turns out to be a notion analogous to "method" in OOP. "Kinematic emergence" is the last notion introduced in Chapter 2.

CONTENT AND DISCUSSION OF CHAPTERS 3–5

Chapters 3–5 deal, on the one hand, with the development of the notions of "functionality" and "complexity" and, on the other hand, with physicochemical processes of prebiological and biological evolution. The property of self-replication is emergent when compared to organic molecules; this transition from the second to the third chapter is natural and subordinate to GFT logic. However, Chapter 3 begins with strange postulates for evolutionary technology: "There is 'space' and 'time'" [5, p. 111], with a reference to the everyday understanding (as far as we can see, this understanding corresponds to the Newtonian one). Statement 3.3 given on p. 102 is clearly physicalist, which reduces its generality but specifies and fixes Krylov's atomistics. Krylov's quasiatoms are analogous to atoms in chemistry having mass and inertia while their properties are quantized (Statement 3.4) to enable replication of base elements; samples of elements can be replenished from an external source. Perhaps, analogies with the works of Democritus are relevant here. The notions of "fields" and correspondingly of "forces" are of first importance: "More complex, nonlinear or nonmonotonic, or even alternating-sign versions of the functional dependence of the attractive force on the distance between elements are extremely difficult to analyze due to the infinitely large number of elements and the infinity of space itself" [5, p. 115].

Physicalism and even mechanicalism are also perceptible in the appearance of the inertia law in Statement 3.5. However, the following more general formulation would be interesting: "In order for the technology satisfying Statements 3.3–3.4 to be evolutionary, it is necessary that elements of the base have the property of preserving <u>the properties</u> ~~the direction of motion~~ in the absence of external forces (~~attractive and/or repulsive~~)." (*Krylov's text is crossed out while our one is underlined*). Constructing a type of forces, Krylov necessarily embarks on the path of molecular dynamics methods; to give concrete expression, we shall give an effective type, in Krylov's opinion, of the interaction force between two elements of the base:

$$F = \frac{K\left(d_{ij} - \left(r_i + r_j\right)\right)}{d_{ij}{}^{n}},$$

where i, j are element indexes, $r_{i,j}$ is their radii, d_{ij} is the distance between their centers, n is degree of interaction, and K is constant. For justification, Krylov refers to his own computer experiments demonstrating the presence of stable structures at $n = 3$ (which corresponds to the physics of our world) but refers only to one literature source published in 1997. It is noted that the introduction of the medium viscosity and heteropolarity of elements, i.e., the "charge" property, allows obtaining results that are more interesting. In our opinion, the book would have benefited has Krylov included a more detailed description beyond illustrations of lumping balls (Fig. 3). The summary is given on p. 121: "For example, ball-cells can be provided with a set of states and functionalities similar to those of von Neumann cellular self-replicating automata. This results in a structure wherein self-replication of structures may occur. Thus, for the first time we have approached the first formal–technological model of the 'universe' in which evolutionary (or rather, similar to them) technologies and processes are theoretically feasible."

Krylov is skeptical of the probability of self-replicating configurations as in the von Neumann automaton recording the following in the Statement 3.6: "The slightest inhomogeneity in the environment surrounding the automaton may result in a failure in the self-replicating process, and the later stops" [5, p. 122]. However, by equipping quasiatoms with the new "number of electrons" property resulting in an analogue of covalent bonding, Krylov has been able to obtain stable configurations preserving the structure under various collisions.

Chapter 4 begins with discussing the concept of entropy. We tend to agree with the following remark (p. 125, emphasis added): "... (*The conventional approach*) considers any system as already given, existing in nature, whether it is actually so or whether *it is created in a constructive manner* from some initial components. Hence the following paradox arises: as long as probabilities of appearing particular states of a system (i.e., its different 'configurations' in the broadest sense of the term) are unknown, nothing can be said either about the value of its entropy or about the nature of its possible changes under particular transformations (conversions) of the system." The redundancy of the Shannon formula for the amount of information in a message as well as the great simplicity chosen by nature in the genetic code (64 codons per 20 amino acids) and used by the developers of information systems are noted here.

Functional complexity as an all possible number of states of the system is introduced. Structural complexity, or structural functionality, is introduced first for linear structures understood through left-right relations. It is stated on p. 131 that eventually the minimum number of functional states providing "three-dimensional" structural completeness of linear structures (analogue is the curve in three-dimensional space) is equal to five. Then, one of the basic objects is selected to be called an automaton (in the GTT, it is the TO processor): "According to this scheme, the required minimum number of states of the synthesis automaton itself can be easily identified. It turns out to be small and equal to 12. That is, the (molecular) automaton ... capable of synthesizing linear structures may have only 12 functional states." As for any object-structure, the automaton functionality is introduced.

Krylov draws attention to the role of the energy factor underestimated, in his opinion, in Langton's works on artificial life. The scheme shown in Fig. 3 resembles the scheme of protein synthesis in ribosomes taken from some biology textbook. It should be noted that nature has really chosen linear structures, i.e., RNA molecules or protein primary sequences, but articulations of parts is more diverse in human engineering, not to mention the ways of association/connection in mental constructs.

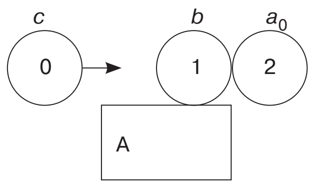


Fig. 3. Fragment of the schematic diagram for the random synthesis automaton A similar to the synthesis of a linear biopolymer. Circles show monomers, while the numbers in the circles are the values for the monomer states

The following four types of functionalities are listed on p. 135: dynamic (as the ability to move in space), structural (as the ability to articulate), logical (as the ability to control the assembly through the automaton), and energetic (introduced by Krylov vaguely). The first and the fourth types are closely related. Complementing the base with "automata–decomposition catalysts" makes the technology more productive. Krylov provides the results of computer simulation again but without a reference to the primary source. After appealing to the Kolmogorov complexity, the really fundamental *theorem on a fixed point of self-replication* is established (p. 145):

"Among linear technologies with the number of functional states per one base element greater than one, there are technologies in which the functional complexity of structures increases constantly with increasing length, whereas the functional complexity of the process (automaton) of their synthesis remains a constant value. Therefore, there are bound to be structures (or their combinations) whose functional complexity is equal to or even greater than the functional complexity of their synthesis procedure (automata) and, consequently, there are bound to be structures capable of replicating themselves."

In our GTT version, we postulate and even provide non-strict justification for the statement that the complexity of the product does not exceed the complexity of its production route (generalizing, of technological process). We confirm indirectly by this that the number of technological alternatives and variations [1, p. 115] may be very large as opposed to the finite number of structural versions. Thus, both statements come into intuitive contradiction, although expressed in different terms. There is no attempt to solve this contradiction here, but simply record it; discussing all nuances of correlation of "*tekhne*," "TR complexity," and hidden information in TO prescriptions would require more than one paper. In our opinion, the emergence of these contradictions in different GTT versions is fruitful and bears the impress of dialectics.

Chapter 4 ends with the description of more complex structures of automata enhanced by the possibility of complementarity (in order to "capture" DNA synthesis). The possibility of their combined action is close to Manfred Eigen concept of an hypercycle, and its features may be seen in our GTT as well; first, in the duality thesis [1, p. 107], and second, in the origin of Moore's law and scientific and technological progress [1, pp. 362, 373]. In our opinion, the idea of searching for ways of self-replication, despite its attractiveness, has had a somewhat detrimental effect on the development of science, only diverting intellectual forces from more achievable practical goals. Falling under the spell of this idea, Krylov continues trying to "attack self-replication" in Chapter 5 by introducing the concept of a partially self-replicating bio-like system (PSBS) along with two principles of Darwinian systems (mutation and selection). Three works by Eigen [13] translated into Russian in the 1970s have been apparently unnoticed by Krylov; otherwise, at least one of them would have been listed in references. While Eigen has used the apparatus of ordinary differential equations and nonlinear dynamics, Krylov is a vivid representative of the structural-algorithmic approach. The structural and functional features of the PSBS having a spatially distributed character, i.e., being intrinsically connectionist, are discussed on p. 173; the presence of control signals for activation and inhibition reinforces this analogy as well [14] (see Rumelhart's principles).

COMMENTS ON THE CONTENT AND DISCUSSION OF CHAPTERS 6–8

Chapter 6 begins with the following semi-definition: "Future technologies are, first of all, technologies integrated into multipurpose (ideally, extremely multifunctional, i.e., universal) automatically operating systems capable of solving a huge number of tasks without human intervention. Flexible automated manufacturing systems (FAMS) and, strange though it may seem, computers may be considered prototypes of such systems" [5, p. 178]. With regard to microelectronics, for example, it is seen how industrial technology is transformed into "high technology" [1, p. 262], which is associated with the transition to the so-called "Industry 4.0." Then, Krylov declares: "Based on the formal technological analysis, we shall try to develop some new basic model of the extremely multifunctional (universal) program-controlled device to perform any technological operations from a given set of operations F of some technology on any objects from the combined set A ∪ B of base elements and structures of the same technology." Then the problem of specifying a "universal technological system" arises. For solving it, Krylov informally uses the abstraction of the Turing machine, and the following definition appears: "It is natural to understand the universality of the system in any technology T = < A ∪ B, F > as the possibility to implement in it any sequence of any operations from F on any elements and structures from A ∪ B allowed for these operations including all possible results of analysis operations from F that can change further sequences of operations." In terms of the TO level, the analysis operations are control-measuring ones, and the GFT and GTT positions coincide here; however, further on the difference arises: for Krylov it is, first of all, controllability (the if... else conditional-branching test) while in the GTT, it is an element of *tekhne* at the production process level. A certain practical compromise here is the elimination of the defective semi-finished product from further technological process.

Then, the notion of "recursive scheme" similar to recursive function in the theory of algorithms is introduced in [5, p. 181]. "Storage cells" are introduced in a somewhat unsophisticated way, i.e., Krylov tries to turn to and transcend some computing model [5, p. 183]: "This assumes that all such (structures – *auth.*) are fairly uniform (i.e., homogeneous) objects that can be stored in single-type storage cells and moved by single-type transport routes." For conditions of the real production, the assumption of homogeneity seems rather shaky. It should be noted that the "computing model" is a relevant notion [15] from the computability theory requiring a separate analysis. Here, Krylov inserts some material and technical aspects. The very notion of a "storage cell" is used as a generalization of the memory cell in computer science; the difference between access by address and access by pointer in FAMS scheme is discussed (Fig. 6.2 in [5]). Krylov's logic is very much affected by the notion of assembly production or assembly "chemistry," but for us microelectronics serves as an example of a different, non-assembly technology. Krylov notes the following on p. 189: "To achieve 'algorithmic completeness', i.e., to create a sufficiently powerful information basis in various technologies equivalent to mathematics in its potential capabilities (and, therefore, providing a solution, among other things, to any problems related to controlling the universal technological system), the formal mechanism should not be that complicated. There should be at least 'something' that may be distinguished from 'nothing' along with the ability to locate this 'something' in right places of space (in right storage cells) in the right sequence so that this sequence could then be processed again."

In the final part of Chapter 6, Krylov turns to specific designs of schemes; here, the abstractness of presentation disappears giving way to Krylov's scientific specialization in the design of "systems-on-a-chip" as well as programmable logic integrated circuits and matrices. Although the general terminology is preserved, it is completely forgotten that an integrated circuit, first of all, converts a signal [1, p. 55] rather than something atomic and molecular. Thus, FAMS feasibility remains questionable, although for the GTT production process level, Krylov's idea seems rather interesting and, perhaps, relevant in terms of process flow. Separation of storage cells and process cells in the context of smart chemical synthesis [5, p. 216] corresponds, in our view, to the logic of the pharmaceutical and genetic engineering development while the idea of controlling demonstrates the great potential of chip synthesis with biotechnology. In this respect, the book seems extremely useful for developers of "labs-on-a-chip" and microsystem technics (Fig. 4) being an excellent example of descending from abstract algebraic abstractions to the real production, even though biopharmaceutical for now. On the other hand, we are skeptical of disseminating these ideas to other technologies.

Krylov also rightly notes that technology may not always create linear objects (Chapter 4) and turns to planar objects estimating the complexity of their synthesis on p. 225. Technological cells and Krylov's synthesis machine resemble in some way the cells of the cellular automaton but have a more complex internal structure inherited from the stored objects; there is also an analogy in the synthesis results (e.g., by S.M. Achasova [16]).

Chapter 7 deals with adaptation of Church's thesis for technological systems. The question is formulated as follows (p. 235): "What are the effective applicability limits of the computing technology operating with natural and rational numbers (i.e., the technology of partial recursive functions) in comparison with more complex technologies operating with objects of the real physical world?" The subordinate character of mathematics is noted once again (p. 240): "Mathematics is a tektology of neutral complexes (in Bogdanov's understanding, complex is something very close to the notion of structure in FT)." The notions of "partially unknown technology" and "complexity threshold" are introduced. The proofs of 10 statements presented in Chapter 7 seem not to be rigorous enough.

In chapter 8, Krylov attempts expanding the GFT to cover social systems, and the intertwined nature of

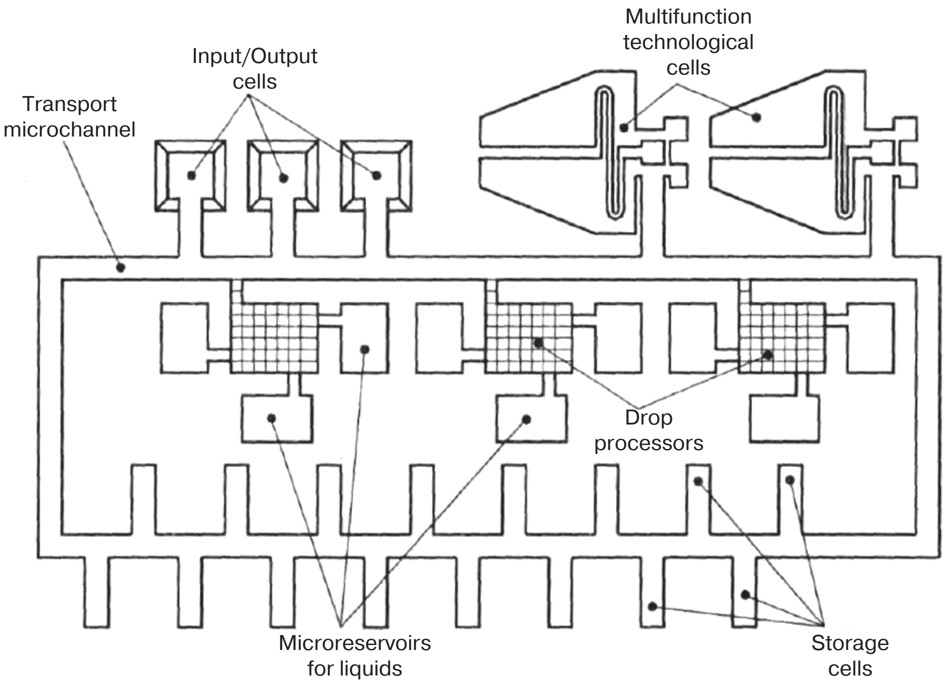


Fig. 4. Topography of the hypothetical nanofactory-on-chip having recursive structure and containing three microobject Input/Output devices (Fig. 6.13 in [5])

personality traits that cannot always be separated along parameter axes is pointed out on p. 252. Recognizing the complexity of the problems, Krylov in his inexact reasoning relies on the model for implementing the shortest path of evolutionary development discussed in Section 5.1 as well as on the Kolmogorov complexity criterion of the multicomponent object (in FT interpretation) discussed in Chapter 4. The above views cannot avoid comments.

As rightly pointed out in Neocybernetics, "the main ('basic') component of various social, industrial and other groups of people is the human being (the individual)." While the natural and social sciences study nature and society, technologies create the human environment. According to P.K. Engel'meier [17], "...technology is a part of the social history of mankind inextricably linked to nature. Through technology man transforms the environment adapting it to his needs." The manufacturing or social technology cannot develop without human activity. Technology (so far, at least) is not manless; it is a unitary whole only when "augmented by humanity," and this is where the difference, maybe the most significant one, lies, since bioevolution is undoubtedly a non-moral process, which is not the case of technological evolution. Technology is not a separate part of the process. Someone or something implements it. This is an indispensable condition for its existence. Without this, it is dead.

Reasoning about the impact of climate on the development of social structures and, especially, on the development of technology including currently are highly subjective. On the other hand, according to the classics of dialectics, "if man, by dint of his knowledge and inventive genius, has subdued the forces of nature, the latter avenge themselves upon him by subjecting him, in so far as he employs them, to a veritable despotism independent of all social organization" [18]. Krylov writes on p. 258: "The most important 'internal' mutagenic factor determining the possible rate of evolutionary development of a given society is 'the degree of permitted free action of certain social subsystems.' Simply stated, this is the degree of 'inner freedom' of a given society. If the level of this freedom is low, i.e., the system is rather rigidly organized (e.g., 'totalitarian'), then, obviously, the rate of evolution of such system would be minimum, and it would inevitably lag behind in its development compared with its more 'internally free' neighbors. This, in turn, would result in the loss of efficiency as well as in losing the competitive struggle." These conclusions are unsupported. In contrast, technological breakthroughs of the thirties in Germany and in the USSR should be noted. These countries are considered totalitarian regimes. In addition, what about the pace of technological advance in China today? The concept of the Megamachine proposed by L. Mumford may be also referred to.

Krylov reveals the difference between a technocratic organization and a democratic one: "In the first case, restraints are developed by the ruling elite while in the second one they are the product of the social group activities and citizens interested in them." The United States is a technocratic society despite its advanced democracy. Democratic France pursues the same path [19]. If only...democracy in form ceases to be democracy in substance.

It is stated on p. 262: "The tragic mistake of the founders of Marxism was to accept the most unfortunate result of the development of social systems as the norm, as the 'driving force of history,' even though the social systems existed at that time deserved their fate," and then on p. 263: "...therefore, it would hardly make sense to deny nature its 'reasonableness.' It is simply a reason of a different plane, different organization, and different destination than our own reason being the product of the former's activity furthermore." Then, Krylov pursues Vernadsky's ideas of the noosphere: "All types of reasons are capable of solving (creatively! – i.e., with creation of new 'inventions') various problems, including those they face unexpectedly. In case, of course, these tasks are fundamentally solvable within the resources and technologies that are available to these reasons at the current stage of development."

In the context of the fourth industrial revolution, the Marxist view of the driving force of history that is primarily the technology development acquires even greater relevance, but at the same time, Engels's warning [20] should not be overlooked: "Let us not, however, flatter ourselves overmuch on account of our human victories over nature. For each such victory nature takes its revenge on us. Each victory, it is true, in the first place brings about the results we expected, but in the second and third places it has quite different, unforeseen effects which only too often cancel the first." We would like to pursue Krylov's thoughts on beauty and social reason [5, pp. 264–265] and tell the ideologists of revolutions (Klaus Schwab) that the world is more complicated than they think, and its self-organization should be trusted. Apparently, this is why Krylov has made critical remarks addressing Marxism, although the debate over the essence of true Marxism is still continuing [21].

CONCLUSIONS

We hope that in the summary review form, we have managed to encourage the reader becoming directly acquainted with the books by S.M. Krylov. It should be noted that the monograph [6] is publicly available. Comparing both GTT and GFT theories (however, they are rather still sketches being far away from the status of a real "theory"), the first one is more general in terms of objects of consideration and may include GFT at three levels of six: technological operation, technological route, and technological process. In matters relating to the complexity measure and entropy, GFT has advanced further demonstrating an interesting approach inspired by algorithmic and atomic assembly processes of molecular "living" structures, but anyway, many issues are still far from being solved. In our book [1], Krylov's works have not escaped our attention, but we have focused less on them than they deserve.

Nevertheless, the book may encourage a fundamental view of science—from arithmetic to quantum physics—as opposed to the informational "pills" often used in teaching to "improve" education. These "improvements" ensuring seemingly immediate success and favoring technological "production" of personality structures suitable for society (and who is so brave as to define the criteria of suitability?) may result eventually in the destruction of that education awakening the desire to compete and overcome obstacles.

The immediate success of such "improvements" overshadows further harms, which they are paid for with. The richer society is, the more obviously it embarks on this path. Who knows whether "synthetic obstacles" would ever begin to be implemented so that they return value to pursuing goals that are excessively facilitated?

In our opinion, the excessive regulation of educational goals and content by the state is an inevitable process, especially during the technology transition to the industrial stage [1] of its development. At the same time, this is a negative process reducing the diversity of personalities whose structures are formed through upbringing and education. Exactly in Krylov's book, general formally algebraic principles of formation including mental-psychological constructs are given. Eliminating the possibility of self-organization, manifestation of human freedom, and the chaos inherent, albeit to a small extent, in nature, the social development path variability is eventually reduced. Contemporary philosophers note the chaos aspect: "From a traditionalist (deterministic) society we are moving to a stochastic (technological) one" [22].

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Krasnikov G.Ya., Gornev E.S., Matyushkin I.V. *Obshchaya teoriya tekhnologii i mikroelektronika* (*General theory of technology and microelectronics*). Moscow: TEKhNOSFERA; 2020. 434 p. (in Russ.). ISBN 978-5-94836-611-1. Available from URL: https://www.technosphera.ru/lib/book/656
2. Krylov S.M. Formally-technological models in the general systems theory. *Izvestiya Samarskogo nauchnogo tsentra Rossiiskoi akademii nauk = Izvestia of Samara Scientific Center of the Russian Academy of Sciences*. 2003;5(1):83−90 (in Russ.).
3. Krylov S.M. *Theory and methods of development of multifunctional programmable technological systems based on formal technological analysis*. Thesis. Samara; 2005. 451 p. (in Russ.).
4. Krylov S.M. *Formal'naya tekhnologiya i evolyutsiya* (*Formal technology and evolution*). Samara: SamGTU; 2012. 325 p. (in Russ.). ISBN 978-5-7964-1547-4
5. Krylov S.M. *Neokibernetika: algoritmy, matematika evolyutsii i tekhnologii budushchego* (*Neocybernetics: algorithms, mathematics of evolution, and future technologies*). Moscow: URSS; 2008. 288 p. (in Russ.). ISBN 978-5-382-00651-2
6. Krylov S.M. *Metamatematicheskie osnovy nauki budushchego: monografiya* (*Metamathematical foundations of the science of the future: monograph*). Samara: SamGTU; 2014. 247 p. (in Russ.). ISBN 978-5-7964-1686-0
7. Verzilin D.N., Sokolov B.V., Yusupov R.M. Neocybernetics: state of research and development prospects. In: *"Sistemnyi analiz v proektirovanii i upravlenii": sb. nauchnykh trudov XXIII Mezhdunar. nauch.-praktich. konf.* (*System analysis in design and management. Collection of scientific papers XXIII Intern. scientific and practical conf.*) V. 1. St. Petersburg: Politekh-press; 2019, p. 81−98 (in Russ.).
8. Mendeleev D.I. Tekhnologiya. In: *Sochineniya. T. XXI. Ekonomicheskie raboty* (Technology. In: *Writings. V. XXI. Economic works)*. Leningrad – Moscow: Izdatel'stvo Akademii nauk SSSR; 1952. P. 333.
9. Bogdanov A.A. *Tektologiya: vseobshchaya organizatsionnaya nauka*: *v 2-kh kn*. (*Tectology: general organizational science*: *in 2 v.*) Moscow: Ekonomika; 1989. V. 1. 304 p. (in Russ.).
10. Engel's F. *Anti-Dyuring*. In: Marks K., Engel's F. *Polnoe sobranie sochinenii* (Engels F. *Anti-Dühring*. In: Marx K., Engels F. *Full composition of writings*). Moscow: Gospolitizdat; 1961. V. 20. P. 11. (in Russ.).
11. Ivlev A.A., Artemenko V.B. Ontology of military technologies: bases, structure, visualisation and application. *Vooruzhenie i Ekonomika = Armament and Economics*. 2011;4(16):35−52 (in Russ.). http://www.viek.ru/16/35-52.pdf
12. Schmidhuber J. Algorithmic Theories of Everything. https://arxiv.org/abs/quant-ph/0011122
13. Eigen M., Shuster P. *Gipertsikl. Printsipy organizatsii makromolekul* (*Hypercycle. Principles of the organization of macromolecules*). Moscow: Mir; 1982. 270 p. (in Russ.).

СПИСОК ЛИТЕРАТУРЫ

1. Красников Г.Я., Горнев Е.С., Матюшкин И.В. *Общая теория технологий и микроэлектроника*. М.: ТЕХНОСФЕРА; 2020. 434 с. ISBN 978-5-94836-611-1. URL: https://www.technosphera.ru/lib/book/656
2. Крылов С.М. Формально-технологические модели в общей теории систем. *Известия Самарского научного центра Российской академии наук*. 2003;5(1):83−90.
3. Крылов С.М. *Теория и методы разработки многофункциональных программируемых технологических систем на основе формально-технологического анализа*: дис. … докт. техн. наук. Самара; 2005. 451 с.
4. Крылов С.М. *Формальная технология и эволюция*. Самара: Самарский гос. технический ун-т; 2012. 325 с. ISBN 978-5-7964-1547-4
5. Крылов С.М. *Неокибернетика: алгоритмы, математика эволюции и технологии будущего*. М.: URSS; 2008. 288 с. ISBN 978-5-382-00651-2
6. Крылов С.М. *Метаматематические основы науки будущего: монография*. Самара: Самарский гос. технический ун-т; 2014. 247 с. ISBN 978-5-7964-1686-0
7. Верзилин Д.Н., Соколов Б.В., Юсупов Р.М. Неокибернетика: состояние исследований и перспективы развития. В сб.: «*Системный анализ в проектировании и управлении»: сб. научных трудов XXIII Междунар. науч.-практич. конф.* Ч. 1. СПб.: Политех-Пресс; 2019. С. 81−98.
8. Менделеев Д.И. Технология. В кн.: *Сочинения.* Т. XXI. Экономические работы. Л.-М.: Изд. АН СССР; 1952. С. 333.
9. Богданов А.А. *Тектология: всеобщая организационная наука*: *в 2-х кн.* Абалкин Л.И. (отв. ред.). М.: Экономика; 1989. Кн. 1. 304 с.
10. Энгельс Ф. Анти-Дюринг. В кн.: Маркс К., Энгельс Ф. *Полное собрание сочинений*. Т. 20. М.: Госполитиздат; 1961. С. 11.
11. Ивлев А.А., Артеменко В.Б. Онтология военных технологий: основы, структура, визуализация и применение. Ч. 1. *Вооружение и Экономика*. 2011;4(16):35−52. http://www.viek.ru/16/35-52.pdf
12. Schmidhuber J. *Algorithmic Theories of Everything*. https://arxiv.org/abs/quant-ph/0011122
13. Эйген М., Шустер П. *Гиперцикл. Принципы организации макромолекул*: пер. с англ.; под ред. Волькенштейна М.В., Чернавского Д.С. М.: Мир; 1982. 270 с.
14. Rumelhart D., McClelland J. *Parallel Distributed Processing.* V. 1. Explorations in the Microstructure of Cognition: Foundations. PDP Research Group (Eds.). MIT Press, Cambridge, MA; 1986. ISBN 0-262-63113-X
15. Матюшкин И.В. Коннекционистское расширение минимальной модели вычислений (Часть 1). *Философские проблемы информационных технологий и киберпространства*. 2016;1(11):103−120. https://doi.org/10.17726/philIT.2016.11.1.004.27
16. Ачасова С.М. Клеточно-автоматная самовоспроизводящаяся матрица из искусственных биологических клеток. *Проблемы информатики*. 2016;3(32):13−25.
17. Энгельмейер П.К. Творчество в технике. В кн.: *Философия техники*. Вып. 1. СПб.: Военмех; 2006. С. 150−185.

[Eigen M., Schuster P. *The Hypercycle. A Principles of Natural Self-Organization.* Berlin, Heidelberg, NY: Springer; 1979]

14. Rumelhart D., McClelland J. *Parallel Distributed Processing.* V. 1. Explorations in the Microstructure of Cognition: Foundations. PDP Research Group (Eds.). MIT Press, Cambridge, MA; 1986. ISBN 0-262-63113-X
15. Matyushkin I.V. Connectionist extension of irreducible computation model (Part 1). *Filosofskie problemy informatsionnykh tekhnologii i kiberprostranstva = Philosophical Problems of Information Technology and Cyberspace*. 2016;1(11):103–120 (in Russ.). https://doi.org/10.17726/philIT.2016.11.1.004.27
16. Achasova S.M. Cellular automata self-replicating matrix of artificial biological cells. *Problemy informatiki = Problems of Informatics*. 2016;3(32):13–25 (in Russ.).
17. Engel'meier P.K. Creativity in technology. In: *Filosofiya tekhniki* (*Philosophy of technology*). St. Petersburg: Voenmekh; 2006. V. 1. P. 150–185. (in Russ.).
18. Engel's F. About authority. In: Marx K., Engels F. *Polnoe sobranie sochinenii* (*Full composition of writings*). Moscow: Gospolitizdat; 1961. V. 18. P. 11. (in Russ.).
19. Gelbreit D.K. *Novoe industrial'noe obshchestvo* (*New industrial society*). Moscow: AST; 2004. 602 p. (in Russ.). ISBN 5-17-024777-X
    [Galbraith J.K. *The New Industrial State*. Princeton University Press; 1985. 520 p.]
20. Engel's F. Dialectics of nature. In: Marx K., Engels F. *Polnoe sobranie sochinenii* (*Full composition of writings*). Moscow: Gospolitizdat; 1961. V. 20. P. 495–499. (in Russ.).
21. Penzin A., Podoroga V., Chukhrov K. Marx against Marxism, Marxism against Marx. *Stasis*. 2017;5(2):530–552 (in Russ.). https://doi.org/10.33280/2310-3817-2017-5-2-266-288
22. Budanov V.G., Arshinov V.I., Filatova O.E., Popov Yu.M. Third paradigm and laws of development of societies. *Slozhnost'. Razum. Postneklassika = Complexity. Mind. Postnonclassic*. 2020;1:38–45 (in Russ.). https://doi.org/10.12737/2306-174X-2020-37-44

18. Энгельс Ф. Об авторитете. В кн.: Маркс К., Энгельс Ф. *Полное собрание сочинений*. М.: Госполитиздат; 1961. Т. 18. С. 303.
19. Гэлбрейт Д.К. *Новое индустриальное общество*: пер. с англ. М.: АСТ; 2004. 602 с. ISBN 5-17-024777-X
20. Энгельс Ф. Диалектика природы. В кн.: Маркс К., Энгельс Ф. *Полное собрание сочинений*. М.: Госполитиздат; 1961. Т. 20. С. 495–499.
21. Пензин А., Подорога В., Чухров К. Маркс против марксизма, марксизм против Маркса. *Стасис*. 2017;5(2):530–552. https://doi.org/10.33280/2310-3817-2017-5-2-266-288
22. Буданов В.Г., Аршинов В.И., Филатова О.Е., Попов Ю.М. Третья парадигма и законы развития социума. *Сложность. Разум. Постнеклассика*. 2020;1:38–45. https://doi.org/10.12737/2306-174X-2020-37-44

About the authors

Evgeny S. Gornev, Dr. Sci. (Eng.), Professor, Corresponding Member of the RAS, Deputy Head of the Priority Technological Direction for Electronic Technologies, JSC Molecular Electronics Research Institute (6/1, ul. Akademika Valieva, Zelenograd, Moscow, 124460 Russia). E-mail: egornev@niime.ru. https://orcid.org/0000-0003-1706-4142

Igor V. Matyushkin, Cand. Sci. (Phys.-Math.), Associate Professor, Senior Researcher, Laboratory of Neuromorphic Systems, JSC Molecular Electronics Research Institute (6/1, ul. Akademika Valieva, Zelenograd, Moscow, 124460 Russia); Associate Professor, Department of Design and Development of Integrated Circuits, National Research University of Electronic Technology (1, pl. Shokina, Zelenograd, Moscow, 124498 Russia). E-mail: imatyushkin@niime.ru. https://orcid.org/0000-0002-2988-8803

Об авторах

Горнев Евгений Сергеевич, д.т.н., профессор, член-корр. РАН, заместитель руководителя приоритетного технологического направления по электронным технологиям АО «НИИ молекулярной электроники» (124460, Россия, Москва, Зеленоград, ул. Академика Валиева, д. 6/1). E-mail: egornev@niime.ru. https://orcid.org/0000-0003-1706-4142

Матюшкин Игорь Валерьевич, к.ф.-м.н., доцент, с.н.с., лаборатория нейроморфных систем АО «НИИ молекулярной электроники» (124460, Россия, Москва, Зеленоград, ул. Академика Валиева, д. 6/1); доцент, кафедра проектирования и конструирования интегральных схем НИУ «Московский институт электронной техники» (124498, Россия, Москва, Зеленоград, пл. Шокина, д. 1). E-mail: imatyushkin@niime.ru. https://orcid.org/0000-0002-2988-8803

Translated by K. Nazarov

MIREA – Russian Technological University.
78, Vernadskogo pr., Moscow, 119454, Russian Federation.
Publication date November 30, 2021.

МИРЭА – Российский технологический университет.
119454, РФ, г. Москва, пр-т Вернадского, д. 78.
Дата опубликования 30.11.2021 г.

https://www.rtj-mirea.ru

МИРЭА
Российский
технологический
университет