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

12+
www.rtj-mirea.ru

МИРЭА
Российский технологический университет

10(6) 2022

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
2022, Vol. 10, No. 6

Russian Technological Journal
2022, том 10, № 6

https://www.rtj-mirea.ru

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

ISSN 2500-316X (Online)

Russian Technological Journal
2022, Vol. 10, No. 6

Publication date November 30, 2022.

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 Russian State Library (RSL), Russian Science Citation Index, eLibrary, Socionet, Directory of Open Access Journals (DOAJ), Directory of Open Access Scholarly Resources (ROAD), Google Scholar, Ulrich's International Periodicals Directory.

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

The subscription index of *Pressa Rossii*: **79641**.

Russian Technological Journal
2022, том 10, № 6

Дата опубликования 30 ноября 2022 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, Ulrich's International Periodicals Directory.

Регистрационный номер и дата принятия решения
о регистрации СМИ ПИ № ФС 77 - 81733 от 19.08.2021 г.
СМИ зарегистрировано Федеральной службой
по надзору в сфере связи, информационных технологий
и массовых коммуникаций (Роскомнадзор).

Индекс по объединенному каталогу «Пресса России» **79641**.

https://www.rtj-mirea.ru

Editorial Board

Редакционная коллегия

Contents

Содержание

Information systems. Computer sciences. Issues of information security
Информационные системы. Информатика. Проблемы информационной безопасности

UDC 004.4'2 + 004.51
https://doi.org/10.32362/2500-316X-2022-10-6-7-19

RESEARCH ARTICLE

Computational complexity when constructing rational plans for program execution in a given field of parallel computers

Valery M. Bakanov @

MIREA – Russian Technological University, Moscow, 119454 Russia
@ *Corresponding author, e-mail: e881e@mail.ru*

Abstract

Objectives. The construction of rational plans (schedules) for parallel program execution (PPE) represents a challenging problem due to its ambiguity. The aim of this work is to create methods for developing such plans and specialized software for implementing these methods, which are based on the internal properties of algorithms, primarily on the property of internal (hidden) parallelism.

Methods. The main method for developing PPE plans was the construction, analysis, and purposeful transformation of the stacked-parallel form (SPF) of information graphs of algorithms (IGA). The SPF was transformed by transferring operators from tier to tier of the SPF (this event was taken as an elementary step in determining the computational complexity of scenario execution). As a transformation tool, a method for developing transformation scenarios in the scripting programming language Lua was used. Scenarios were created by a heuristic approach using a set of Application Programming Interface (API) functions of the developed software system. These functions formed the basis for a comprehensive study of the parameters of the IGA and its SPF representation for the subsequent construction of a PPE plan applying to a given field of parallel computers.

Results. Features of the internal properties of the algorithms that affect the efficiency of SPF transformations were identified during the course of computational experiments. Comparative indices of the computational complexity of obtaining PPE plans and other parameters (including code density, etc.) were obtained for various SPF transformation scenarios. An iterative approach to improving heuristic methods favors developing optimal schemes for solving the objective problem.

Conclusions. The developed software system confirmed its efficiency for studying the parameters of hidden parallelism in arbitrary algorithms and rational use in data processing. The approach of using a scripting language to develop heuristic methods (scenarios) for the purposeful transformation of IGA forms showed great flexibility and transparency for the researcher. The target consumers of the developed methods for generating schedules for parallel execution of programs are, first of all, developers of translators and virtual machines, and researchers of the properties of algorithms (for identifying and exploiting the potential of their hidden parallelism). The developed software and methods have been successfully used for a number of years for increasing student competence in data processing parallelization at Russian universities.

Keywords: algorithm graph, fine information structure of program, stacked-parallel form of graph, rational execution parameters of parallel program, execution plan of parallel program

• **Submitted:** 30.01.2022 • **Revised:** 29.03.2022 • **Accepted:** 05.09.2022

For citation: Bakanov V.M. Computational complexity when constructing rational plans for program execution in a given field of parallel computers. *Russ. Technol. J*. 2022;10(6):7–19. https://doi.org/10.32362/2500-316X-2022-10-6-7-19

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

The author declares no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Вычислительная сложность построения рациональных планов выполнения программ на заданном поле параллельных вычислителей

В.М. Баканов [@]

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: e881e@mail.ru*

Резюме

Цели. Построение рациональных планов (расписаний) выполнения параллельных программ (ВПП), вследствие неоднозначности, является сложной задачей. Цель работы – создание методик разработки таких планов и специализированного программного обеспечения для реализации этих методик, полагающихся на внутренние свойства алгоритмов, в первую очередь на свойство внутреннего (скрытого) параллелизма.

Методы. Основными методами при разработке планов ВПП являются построение, анализ и целенаправленное преобразование ярусно-параллельной формы (ЯПФ) информационных графов алгоритмов (ИГА). Преобразование ЯПФ осуществляется путем переноса операторов с яруса на ярус ЯПФ (именно это событие и принято за элементарный шаг при определении вычислительной сложности выполнения сценария). В качестве инструмента преобразования применен метод разработки сценариев преобразования на скриптовом языке программирования Lua. Сценарии создаются на основе эвристического подхода и используют набор API-функций (API – Application Programming Interface) разработанной программной системы, позволяющих всесторонне изучить параметры ИГА и его ЯПФ-представления для последующего построения плана ВПП на заданном поле параллельных вычислителей.

Результаты. Результаты вычислительных экспериментов выявили особенности внутренних свойств алгоритмов, влияющих на эффективность преобразований ЯПФ. Получены сравнительные показатели вычислительной сложности получения планов ВПП и иных параметров (включая плотность кода и др.) при применении различных сценариев преобразования ЯПФ. Итерационный подход к улучшению эвристических методов позволит приблизиться к оптимальным схемам решения целевой задачи.

Выводы. В целом разработанный программный комплекс подтвердил эффективность в исследовании параметров скрытого параллелизма в произвольных алгоритмах и рационального его использования при обработке данных. Подход применения скриптового языка для разработки эвристических методов (сценариев) целенаправленного преобразования форм ИГА показал большую гибкость и прозрачность для исследователя. Целевыми потребителями разработанных методов генерации расписаний параллельного выполнения программ в первую очередь являются разработчики трансляторов и виртуальных машин, исследователи свойств алгоритмов (в направлении нахождения и использования потенциала скрытого их параллелизма). Разработанное программное обеспечение и методики несколько лет применяются при обучении студентов в университетах России, что позволило повысить компетенции учащихся в области параллелизации обработки данных.

Ключевые слова: граф алгоритма, тонкая информационная структура программы, ярусно-параллельная форма графа, рациональные параметры выполнения параллельной программы, план выполнения параллельной программы

• **Поступила:** 30.01.2022 • **Доработана:** 29.03.2022 • **Принята к опубликованию:** 05.09.2022

Для цитирования: Баканов В.М. Вычислительная сложность построения рациональных планов выполнения программ на заданном поле параллельных вычислителей. *Russ. Technol. J.* 2022;10(6):7–19. https://doi.org/10.32362/2500-316X-2022-10-6-7-19

Прозрачность финансовой деятельности: Автор не имеет финансовой заинтересованности в представленных материалах или методах.

Автор заявляет об отсутствии конфликта интересов.

INTRODUCTION

A currently popular method for reducing computation time is parallelization, involving the simultaneous processing of data in parts on a set of multiple computers with subsequent combination of the obtained results. While the possibility of using extra hardware to overcome existing fundamental and purely technological limitations of increasing performance is intuitively apparent, the possibility of identifying analogous parallelism potentials in algorithms is less so [1].

A separate and nontrivial problem when organizing parallel computing involves the construction of a plan (schedule) for the execution of parallel programs (PPE). Here, almost every sequential program (algorithms forming the basic components of any program conventionally represented in sequential form) can be represented in a parallel form using methods that preserve the set of operations and causal relationships between them to support different execution efficiencies on parallel computers of a given architecture. According to this formulation of the problem, it is the algorithms themselves, comprising the real building blocks of programs, whose internal properties become an increasingly important area of study.[1]

Each of the considered PPE plans is associated with certain program execution quality parameters (time, required computational resources, memory load, etc.). The formulation of the objective function and solution to the multi-parameter optimization problem is dependent on the stated problem.

Despite the differences in the architecture and systems of machine instructions of various parallel computing systems and parallel programming technologies, there are scientifically substantiated general approaches to the construction of enlarged plans (schedules) for program execution. Abstracting from the specifics of parallel programming technologies, it is logical to refer to such plans in terms of a *framework* for the execution of a parallel program.

[1] Voevodin V., Dongarra J. (Eds.). *AlgoWiki: Open Encyclopedia of Parallel Algorithmic Features.* http://algowiki-project.org. Accessed October 20, 2022 (in Russ.).

This paper proposed methods and their implementation (in the form of problem-solving program scenarios) for developing rational PPE plans on a given (possibly heterogeneous) field of parallel computers. The developed scenarios are intended for embedding as parallelizing blocks in newly developed systems for creating executable program code. In this case, the emphasis is on achieving maximum computational speed, since program debugging requires multiple translations with complex optimization while ensuring required quality.

METHODS

The problem of finding methods for constructing rational PPE plans was solved by creating a specialized software system of the instrumental level, whose generalized flow diagram and information flow diagram are shown in Fig. 1.

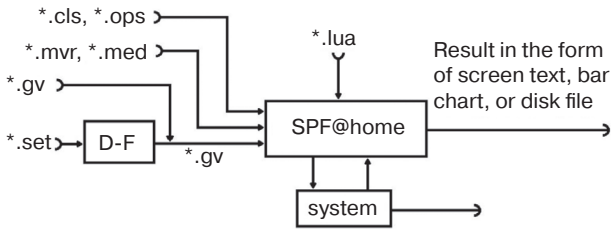


Fig. 1. Instrumental software complex for constructing PPE plans [2]

The input of the software system (Fig. 1) receives a description of an algorithm to be analyzed in the form of a conventional sequential program in an imperative style or its formal description in the form of a directed acyclic information graph of the algorithm (IGA), i.e., an ‘operators → operands’ dependence. In this case, the graph vertices are associated with operators (groups of operators) of the program; and the graph edges, with data transmission lines. In Fig. 1, the files of types *.set and *.gv are the program and information graph files of the analyzed program, respectively; *.mvr and *.med, the files of the metrics of vertices and edges of the graph of the algorithm, respectively; *.cls and *.ops, the files of parameters of computers and program operators, respectively; and *.lua, the Lua text file containing scripts

for PPE plans. When the results of the development are used as an algorithmic basis for the functioning of the components of parallelizing translators, the IGA is built on the basis of the results of parsing the source code by a translator (and the library functions are processed separately). When the application is implemented as a separate *Code Morphing Software* component, the IGA is built on the basis of the generated sequential executable code. A useful feature of using the representation of programs as IGA is that the graph can be processed in separate blocks (on the level of subroutines and files) with subsequent assembly of the graph of the complete application.

The internal logical parallelism in algorithms is identified and analyzed by simulating an actor model (D-F (Data Flow) module) and constructing special sections of the IGA in the stacked-parallel form (SPF) [3] (SPF@home module) [4, 5]. Both modules, which are developed in the C/C++ language in a GUI style for Win'32 model (a command line mode was additionally implemented for massive computations), are completely open-source and can be downloaded for free use (installation file format).[2] The D-F module builds an PPE plan for the operator execution asynchronism model; schedules in the SPF@home module style are designed for synchronous calling of groups of operators.

The D-F module is a universal computer of the *Symmetric MultiProcessing* (SMP) architecture (shared-memory systems) [6], whose input receives a sequential program in an imperative assembler-like language (3-character command mnemonics and a 3-address system with the AT & T operand order). The program is executed in a simulator of a computer of the Data-Flow static architecture; here, the order of execution of processor instructions is determined not by the order in which they appear in the machine code, but by the readiness of the operands [7, 8].

The inverse problem of optimizing the parameters of a computer from the characteristics of the computation process can be solved by varying the number of parallel computers and the rules for fetching instructions from the buffer [4]. Export of IGA to third-party programs in DOT format is provided along with a detailed simulation protocol.

Conditional execution is implemented by the predicate method [9, 10], with cycles being implemented using a system of macros, which unfold cyclic structures. The convenience of visualizing the solution is contributed by the output of the program execution data as a function of the computation intensity (the number of simultaneously executed operators as a function of time) and Gantt charts.

The SPF@home module [5] is designed to model and select the best (in a given sense) scenarios for transforming the SPF as a plan for the parallel execution of operators on a computational system of a given architecture. A significant advantage of using the SPF is the satisfactory temporal computational complexity of its derivation, which is quadratic polynomial with respect to the number of graph vertices. The SPF can initially be constructed in "upper" or "lower" form (where all the operators are located on the tiers of the SPF as close as possible to the beginning or end of the program execution, respectively). An important user advantage of the SPF is its improved visualization of the representation of the identified parallelism in the algorithm: even the initially obtained SPF of the algorithm is already a certain initial (usually far from optimal) PPE plan.

Groups (bundles) of operators located on the tiers of the SPF are executed in parallel within each tier, while bundles of operators on each tier are executed sequentially, beginning with the initial tier. The minimum possible height of the SPF (the number of tiers), which is determined by the length of the critical path in the IGA [11, 12], determines the shortest execution time of the algorithm. In the SPF@home module, the obtained SPF is visualized in text and graphical forms (the bar diagram of the SPF widths is an easy representation of the height distribution function of the SPF widths). The mechanism for setting operator execution metrics is used to set parameters for operators (for example, their execution time or required resources for a heterogeneous field of computers), computers (types of operators that can be executed on a given computer), as well as data transmission lines (transmission time, data sizes). The practice of using the metrics mechanism is described in detail in the subsection that presents PPE scheduling on a given heterogeneous field of computers.

Additionally, the SPF@home module provides the ability to obtain information on the data lifetime required to execute a given algorithm. These data, which exist as a consequence of the execution of individual operators, serve as input operands for other operators of the algorithm. This information (in fact, an estimate of the local *capacitive complexity* of the program execution) is important for determining the required parameters of the internal registers of the processor and/or solving questions about the optimal placement of data between the processor registers and RAM.

The main method of transforming the SPF is the purposeful movement of operators between the tiers of the SPF while maintaining information links in the IGA. In general, this is an *NP*-complete problem, which belongs to the class of constrained scheduling

[2] http://vbakanov.ru/dataflow/content/install_df.exe. Accessed October 24, 2022 (in Russ.); http://vbakanov.ru/spf@home/content/install_spf.exe. Accessed October 24, 2022 (in Russ.).

problems [13] and can be used in the construction of rational (iteratively tending to optimal) PPE plans. In the present work, a heuristic approach was used to obtain a solution whose scenarios (in this case, SPF transformations in the required direction) are implemented using the Lua scripting language [14]. Lua was chosen because it is completely open-source, close in syntax to common programming languages (C/C++ style), and compact when embedded in a parent application.

Each Lua function call is actually a wrapper over the corresponding Application Programming Interface (API) call of the parent program. The set of APIs of the SPF@home system, which covers almost all foreseeable actions on the IGA, can be used to analyze a graph of any complexity (limited by the resources of the computing device). In this sense, the use of the SPF graph is only one of the possible solution methods. A total of three types of calls can be distinguished:

- Information calls—serve to obtain information on the IGA and its SPF, on the basis of which data the specific IGA processing method is selected for solving the problem. Examples include determining the total number of tiers of the SPF, the number of operators in a given tier, the range of possible locations of a given operator in the tiers of the SPF, etc.
- Action calls—serve to implement specific methods for solving the problem of constructing an PPE schedule. Examples are to build the upper or lower form of the SPF, add an empty tier under the data, move an operator from tier to tier, etc.
- Auxiliary calls—serve to output the computed data in text and graphical forms for data exchange with other applications, work with the file system, etc.

The information graph of the algorithm in the SPF form can formally be represented by a 2D list of identifier elements (e.g., unique numbers) of operators $[a_{ij}]$, where $i = 1 \ldots W$ is the number of a row; the quantity W or SPF height is determined by the length of the critical path in the IGA; and $j = 1 \ldots j_i^{\max}$ is the number of operators in row i. The quantity $H = \max_i (1 \ldots j_i^{\max})$ is referred to as the SPF width. In a real IGA, the position of each operator on the tiers is limited by the presence of information links in the algorithm and by the range $i^{\min} \le i \le i^{\max}$, where $i^{\min}$ and $i^{\max}$ are the admissible numbers of tiers of placement of a given operator in the SPF; the range $i^{\min} \le i \le i^{\max}$ can be logically described in terms of the *variability* of positions on the tiers of the ith operator. The SPF is actually the (initial, naive) PPE plan (schedule). For a given description, the direction of the unit vector of time coincides with the increase in the number of a tier of the SPF.

In general, the proposed approach is fully consistent with the Explicitly Parallel Instruction Computing (EPIC) style [15] intended for software implementation of parallelizing translator blocks. At the same time, for the Very Long Instruction Word (VLIW) computing architecture [10], the term *operator* should be understood as a machine instruction in order to fully adhere to the Instruction-Level Parallelism (ILP) concept [4]. For multithreaded systems on multicore processors, it is logical to correlate the "operator" with a *parallelism granule* of a much larger size, e.g., on the level of an operator/operators or procedures of a high-level programming language [16]. The latter fits well with the concept of interpreters. In both cases, the presented general methodology for constructing a rational PPE plan remains unchanged.

The internal implementation of the data, which does not have to be provided for the explicit construction of the SPF in the form of a 2D array, can be in any format convenient for computer implementation. For example, in the naive case, it can be that establishing a one-to-one correspondence between the IGA in the form of a set of directed edges $\{k, l\}$ (adjacency matrix) identified by pairs of vertex numbers i_k, j_k and i_l, j_l, where i and j are the numbers of a row and a column in the SPF.

The examples for the study comprised popular data processing algorithms (linear algebra, statistics, array operations, etc.). Additional, artificial (not corresponding to any of the applied algorithms, but generated in accordance with the specified parameters) IGAs were prepared. A disadvantage of the experimental material was the relatively small dimensions of the processed data due to the considerable difficulty of manual programming. However, the performed experiments developed an increase in the identified trends including an increase in the dimension of the processed data throughout the studied dimension range.

The computational complexity of executing scenarios of SPF transformations was determined using an analog (applied in array sorting operations) of the classical method of estimating the target parameter, namely, determination of the number of *elementary steps* (permutations of two elements of an array being sorted) that is required to complete the operation. In our case, it is logical to define the elementary step as a permutation of an operator from tier to tier of the SPF. This approach has all the advantages and disadvantages of the classical method, including the failure to take into account the complexity of analyzing the situation and making decisions about taking a specific elementary step.

In this study, the estimated *code density* characterizes the resource use of a parallel computing system (number of computers) when executing this algorithm (formally,

the deviation of the widths of the tiers of the transformed JPF from a given value). When computing resources are not fully used, the translator has to insert NOP instructions into empty places in bundles of parallel-executed instructions, leading to decreased code efficiency.

The following series of experiments was carried out using the SPF@home module as offering the greatest flexibility in transforming the SPF of algorithm graphs; IGAs were generated by the D-F module based on the program code. During the computational experiments, the SPF@home module saves the most detailed simulation protocol for subsequent analysis. For this work, the following parameters obtained in the course of the target transformation of the SPF are of particular interest:

- The height H and width W of the obtained SPF (width constraints were set as a parameter that ensures the execution of the algorithm on a given number of parallel computers).
- The uniformity of the tier width distribution (*code density*) in this SPF was estimated by the coefficient of variation $CV = \frac{1}{\bar{W}}\sqrt{\frac{\sum(W-\bar{W})^2}{H-1}}$, where $\bar{W}$ is the arithmetic mean of the tier widths over the SPF.
- The computational complexity of the performed SPF transformation (in units of the number of permutations of operators from tier to tier of the SPF).

The efficiencies of SPF transformations were compared by using two heuristic methods (scenarios), which were represented in the form of Lua scripts and based on different general approaches to SPF transformations. The first (*Strategy_01*) used a dichotomy-based approach of massive transfer of operators from tier to tier of the SPF, whereas the second (*Strategy_02*) uses a gradual transfer of operators from "more loaded" tiers to "less loaded" ones. In both cases, if necessary, additional (initially empty) tiers are created in necessary places to ensure the execution of the algorithm on a given number of parallel computers.

RESULTS

Scheduling of program execution on a fixed number of parallel computers with the possibility of increasing the program execution time

This subsection considers the most general case corresponding to the condition $\bar{W} >> P$, where P is the number of parallel computers. It is in this case that the SPF height (program execution time) has to be increased.

The effectiveness of the above heuristic methods was tested by successively applying them to the SPF of the studied algorithms in the range of the numbers P of parallel computers from W_0 (the width of the initial SPF) to 1 (fully sequential execution of the algorithm).

Figures 2–5 show the change in the target quantities, specifically, (a) computational complexity, (b) SPF height, and (c) coefficient of variation of the widths of the SPF tiers (ordinate axis) as functions of a given number P of parallel computers (abscissa axis). The SPF transformations of the corresponding algorithms were carried out according to the scenarios *Strategy_01* and *Strategy_02* (curves 1 and 2, respectively).

As can be seen from the data in Figs. 2–5, the application of both methods to the studied algorithms gives similar results: the *Strategy_02* scenario is faster than its rival (panels (a)), whereas the algorithm execution times differ little (panels (b)). In code density (panels (c)), both scenarios show similar tendencies toward minimizing the objective function using a small (much smaller than the SPF width) number of parallel computers. The bizarre shapes of the curves are a consequence of the complexity of the scripts used and the processing of integer values, while the curves using the *Strategy_02* script are visually more monotonic.

Quantitatively, the *Strategy_02* scenario has a lower (by a factor of approximately 2–4 in the studied range of processed data sizes) computational complexity than *Strategy_01*, although the converse was expected at first glance. However, the *Strategy_02* script has more complex internal logic compared to *Strategy_01* (in the latter case, it is primitive), which cannot be taken into account by the accepted computational complexity estimation system. In view of the above, it may be more logical to use the *Strategy_01* script in the components of parallelizing systems for fast, but rather rough construction of PPE plans, while the *Strategy_02* method should be used for constructing these plans in the optimization mode.

Scheduling of program execution on the minimum number of parallel computers with the possibility of increasing the program execution time

In the case $\bar{W} \approx P$ (the arithmetic mean of the widths of the initial SPF is comparable to the number of parallel computers), the problem arises of PPE scheduling with the maximum code density without increasing the SPF height ("balancing" of operators over the tiers of the SPF). The developed empirical methods for balancing the SPF gave contradictory results: in some cases, it was possible to achieve almost 100% code density, whereas some algorithms were virtually unmodifiable (due to restrictions on moving operators between the tiers because of the need to preserve the information dependences in the algorithm).

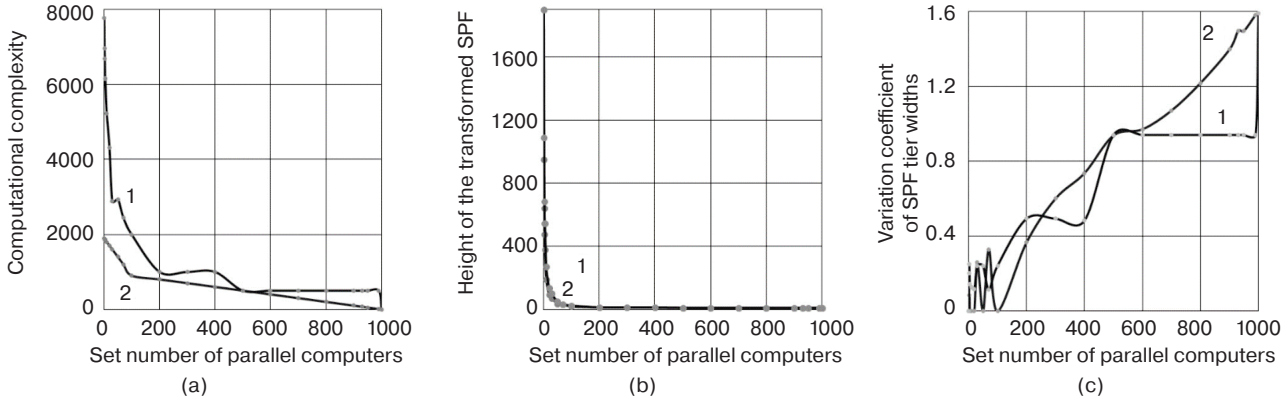


Fig. 2. Algorithm for multiplying 10th-order square matrices by the classical method

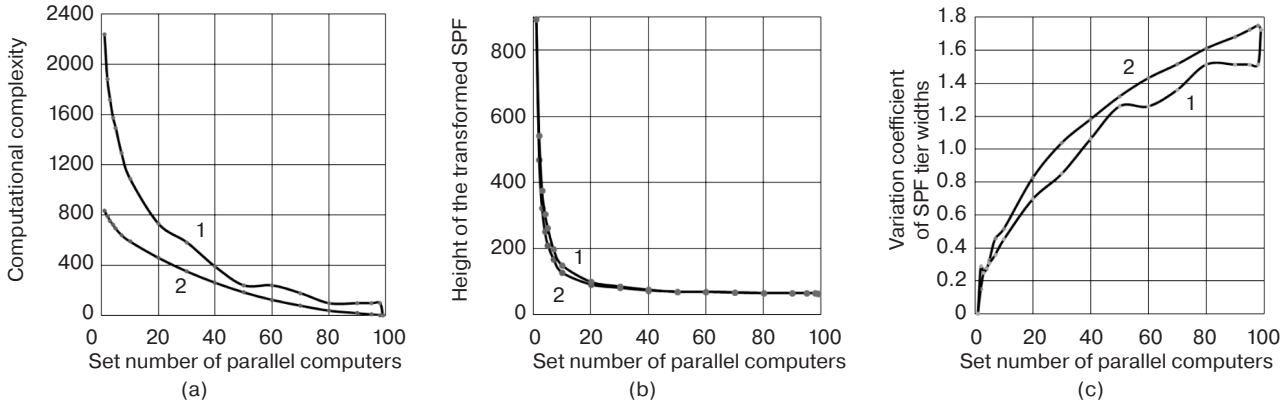


Fig. 3. Algorithm for solving systems of linear algebraic equations (SLAE)
of the 10th order by the direct (non-iterative) Gauss method

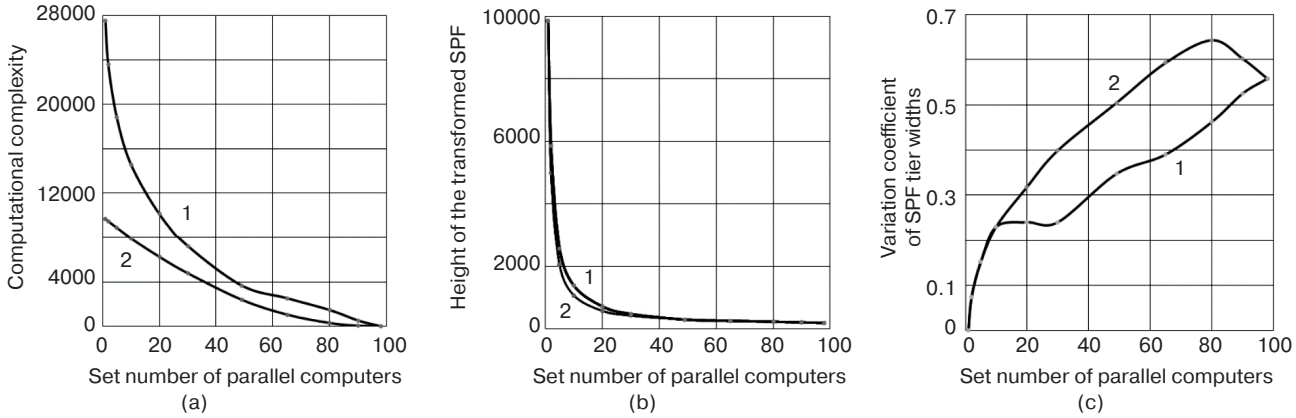


Fig. 4. Artificially generated algorithm e19039_o9853_t199

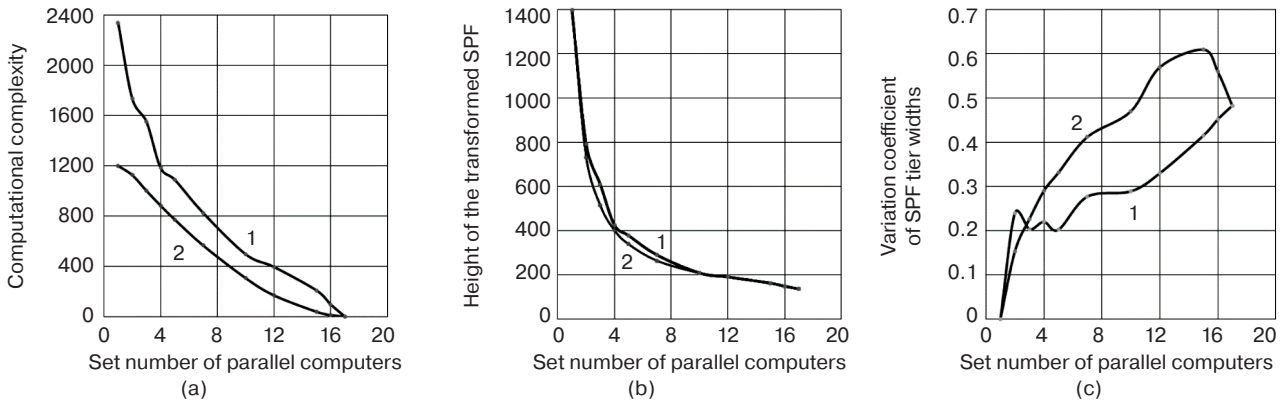


Fig. 5. Artificially generated algorithm e2367_o1397_t137

The SPF@home software module can also be used for solving an inverse problem, e.g., the determination of the parameters of the parallel computing system from the performance–cost balance of the system itself.

Promising PPE scheduling methods

In all the above experiments, the SPF was initially computed in the upper form (all the operators were located as high as possible in the tiers of the SPF). In this case, the operators are moved between the tiers mainly from top to bottom: from the initial tiers to the final ones (this does not, of course, prohibit multiple movements with repeated changes in the direction of movement of operators between the tiers).

The described sequence of actions is logically justified by the characteristic shape of the computation intensity curves under the conditions of unlimited parallelism, in which a sharp increase in the initial part is followed by a peak and a smooth decrease by the end of the program execution. Such a curve shape, which occurs at significant amounts of input data, once again confirms that the execution of operators on a field of parallel computers belongs to one of the varieties of multichannel queuing systems. In this case, since the variability of the operators located in the region of the peak of the computation intensity function is high, the efficiency of obtaining a satisfactory PPE plan by moving down the SPF operators is also significant. We will refer to the algorithms whose initial (unmodified) form of width distribution in the upper SPF corresponds to the one described above as belonging to the Π class.[3]

As a matter of discussion, it was of interest to consider the lower form of the initial SPF (in this case, all the operators are moved as far as possible toward the end of the program execution). Such an SPF can be obtained from the upper form by moving the operators through the tiers as low as possible or, more easily, by constructing the SPF in the direction from the end of the program to its beginning (in the latter case, the computational complexity of obtaining the SPF remains the same as when obtaining the upper SPF). The distributions of the widths of the SPF in the upper and lower forms are compared below.

Figure 6 shows the bar diagram of the distribution of the SPF widths at given sizes of processed data (indicated in the captions) that was obtained by the SPF@home module. Each of the four columns of a row presents two diagrams: the one on the left is for the upper form of the SPF of this algorithm and the one on the right is for its lower SPF. Here, H, W, and $\bar{W}$ are the height, width, and arithmetic mean width of the SPF (the last is shown in Fig. 6 by the dotted line; the forward slash symbol separates the parameters for the upper and lower SPF).

The data in Fig. 6 are interesting because of the possibility of significant balancing of the SPF without using complex heuristic algorithms for its reorganization. In fact, all the possible solutions for the reorganization of the SPF are within the range between the upper and lower forms; however, when choosing the lower form as the initial form, the priority movement of operators is upward.

The last statement raises the question of whether or not there are states that are balanced still better than the two boundary (upper and lower) forms. To answer this question, an experiment was carried out to study the stepwise transformation of the SPF from the upper to the lower form (in Fig. 7, the numbers on the abscissa axis are the numbers of movements of operators from tier to tier), in which along with the CV index (solid lines, left-hand ordinate axis), the irregularity of the widths of the SPF tiers was estimated by the ratio of the width of the widest tier to that of the narrowest one (dashed line, right-hand ordinate axis).

Despite the entire set of SPF states not being completely covered in these experiments (possible movements of operators along the SPF tiers were downward to the maximum variability), it can be argued with a high degree of probability that the target values do not increase in the range of existence and that it is in the region of the lower SPF that they are minimal.

When determining rational PPE plans in the case of obtaining easily determined conditions where the algorithm belongs to the Π class, significant savings in computations are possible: instead of implementing the above rather complex scenarios, it is enough to construct the lower SPF. The previously used quantitative characteristics of the irregularity of the tier widths do not provide information on the shape of the curve with this irregularity. As an additional estimate of the irregularity of the distribution of operators over the SPF tiers, the well-known graphic-analytical method can be used to determine the income inequality, which consists in calculating the numerical parameters of the stratification (Lorenz curve and Gini coefficient)[4], despite the mirror-opposite shape of the analyzed curves.

The example shows the importance of studying the properties (including classification) of algorithms from the side of their essential facet that consists in their internal parallelism in order to make the best practical use of these properties.

[3] We will refer to the Π class such algorithms that are characterized by the distribution of the tier widths in the SPF of the information graph in the upper form (when all operators are as close as possible to the beginning of execution) with the presence of a pronounced maximum at the beginning and of a gentle decrease at the end.

[4] Gini coefficient: Are everyone equal? *Open Journal*, an investment and finance medium. https://journal.open-broker.ru/economy/koefficient-dzhini/. Accessed March 31, 2022 (in Russ.).

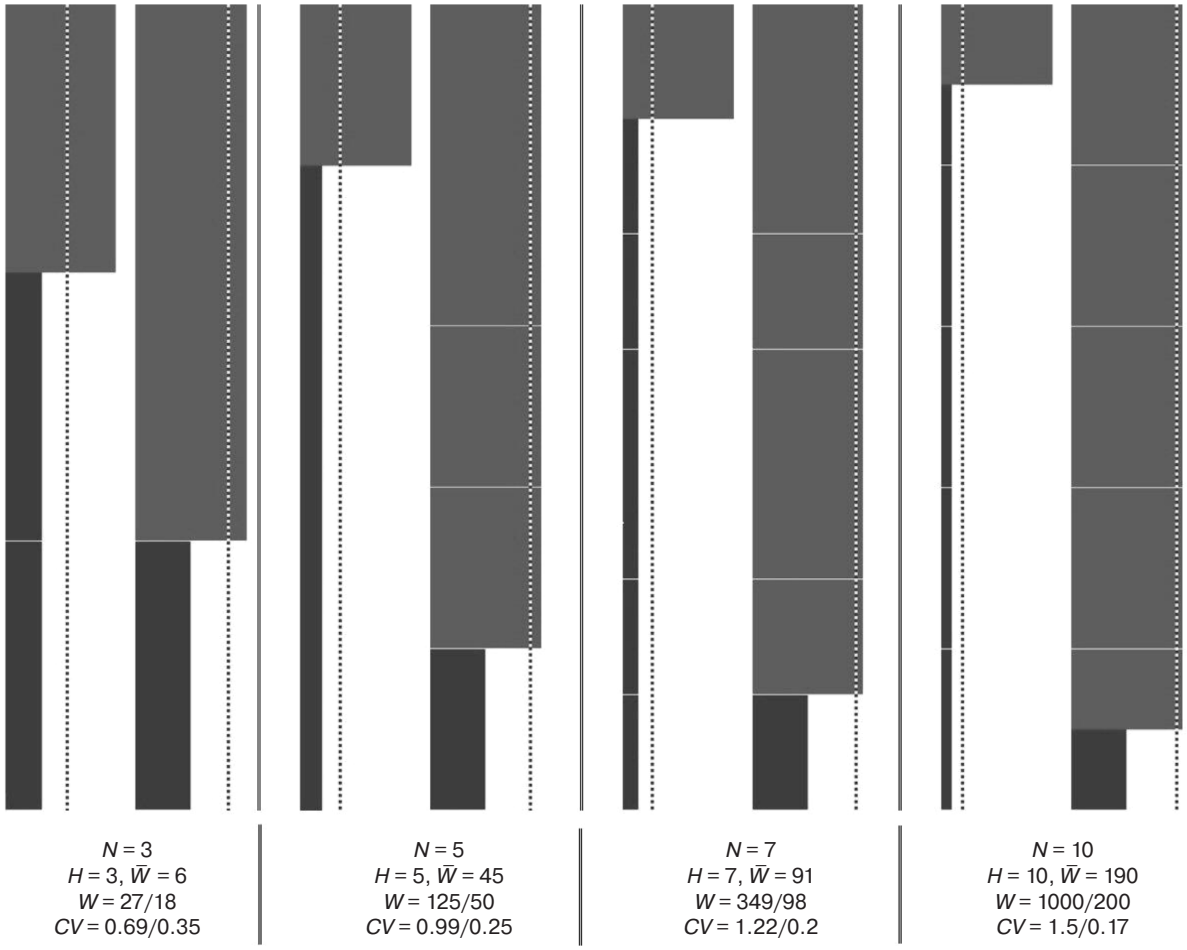


Fig. 6. Algorithm for multiplying square matrices by the classical method (Abscissa axis is the tier width for $N = 3, 5, 7, 10$ orders of the matrices). The light- and dark-gray areas contain tiers of maximum and minimum widths, respectively

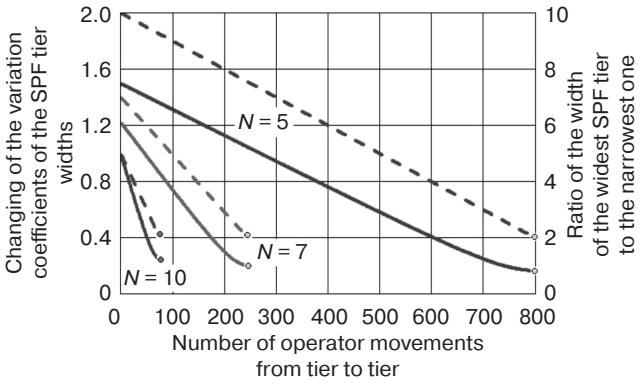


Fig. 7. Irregularity of the distribution of the widths of the SPF tiers for the algorithm of matrix multiplication by the classical method (*N* is the order of the multiplied matrices)

PPE scheduling on a given number of heterogeneous computers

Modern multicore processors are more and more often developed with computing cores of various capabilities. Therefore, it is practically useful to be able to schedule the PPE for such systems (with a heterogeneous field of parallel computers).

The SPF@home module supports this feature by comparing information from two metrics files for operators and computers (*.ops and *.cls, respectively, Fig. 1). It is possible to set a match on a set of freely assignable attributes for any range of operators/computers. The condition for the feasibility of a given operator on a given computer is the relation $minVal\_i \le Val\_i \le maxVal\_i$ at the same i, where Val_i, $minVal\_i$, and $maxVal\_i$ are the numerical values of this parameter for the operator and the computer, respectively.

Since PPE scheduling on a heterogeneous field of parallel computers is a more complex procedure than those described above, the emphasis here is on Lua programming. Since one SFP tier may contain operators the execution of which require different computers, it can be useful to apply a metaphor for splitting the SFP tiers into families of subtiers, each of which corresponds to a block of computers with certain capabilities. All the operators on a given tier have the same execution capabilities, the sequence of processing them within a tier/subtier is, to a first approximation, arbitrary. Figure 8a illustrates the splitting of operators on one of the tiers of the SPF in the case of 11 parallel computers of

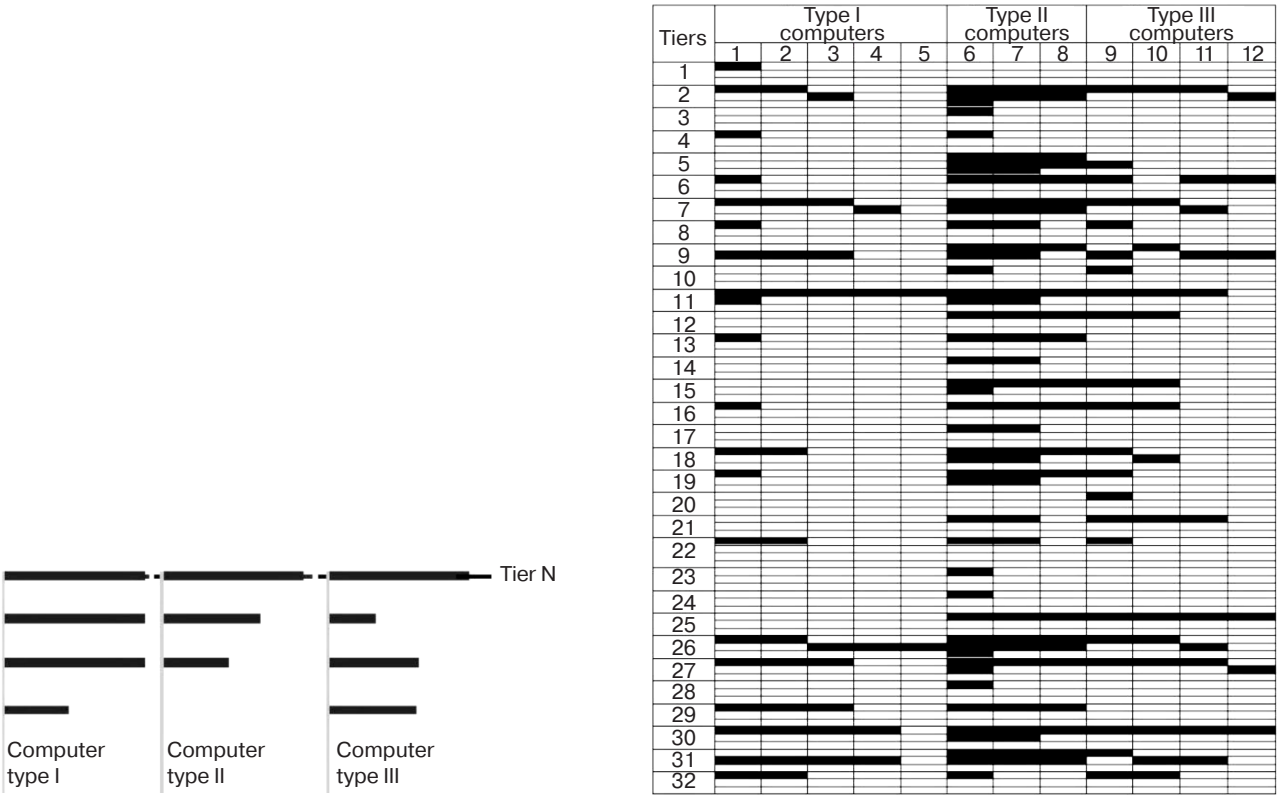


Fig. 8. (a) Splitting of the SPF tiers into families of subtiers in solving the problem of scheduling for a heterogeneous field of parallel computers and (b) the result of computing the schedule for the execution of a real parallel program

three types, and Fig. 8b presents the result of computing the real PPE plan on a heterogeneous field of parallel computers (three types of computers, 5, 3 and 4 units; the numbers of executable operators are hidden).

In this case, the total time T of solving the problem is determined by the sum over all the tiers of the maximum values of the execution times of operators on the subtiers of this tier:

$$T = \sum_j \left(\max_{k_j} \sum_i t_{ik} \right),$$

where j is the number of tiers, i is the number of subtiers on a given tier, k_j are the types of computers on the jth tier, and t_{ik} is the execution time of an operator of type i on a computer of type k.

If the goal is to achieve maximum performance, the optima number of computers of a particular type can be determined by minimizing the parameter T (e.g., by solving the inverse optimization problem to determine the ratio between the numbers of computing g devices of different types). The problem of minimizing the total solution time T becomes more complicated if each operator can be executed on several computers because of the ambiguity of the parameter t_{ik} in the above expression; here, additional balancing over substages is needed.

DISCUSSION

This study confirmed the possibility of gradual iterative improvement (in a given direction) of heuristic scenarios for transforming the initial SPF of various algorithms. By and large, it is possible to develop faster scenarios with a slightly inferior quality for an intended purpose and relatively slower scenarios with a higher quality (in fact, on the level of optimization).

Although this work is focused on the computational complexity of scenarios for obtaining PPE plans (schedules) for the PPE, the presented software system should also show its efficiency in solving multidimensional optimization problems using the known algorithms (this software system in this case is a *mathematical model of the subject of optimization*). To implement this case of the use, the discussed computer system permits working in command line mode.

Despite the low computational complexity of obtaining the SPF from the IGA, the method of using the SPF as the basis for constructing PPE plans has the disadvantage that it is impossible to easily take into account the execution time of operations, as a result of which the execution time of a bundle of operators on one tier of the SPF has to be considered equal to the execution time of the slowest of them. By taking an approach consisting in the purposeful movement of operators through the tiers of the SPF, operators can be

sorted on each tier according to the closest possible time for their execution.

The experiments detected a significant dispersion of the properties of the algorithms represented by information graphs in the possibility of forming PPE plans with the maximum code density. Different algorithms require different methods for their efficient transformation. It seems important to a priori (even before the reorganization of the SPF, at the time of its receipt) define scenarios for its effective purposeful modification with a given goal. The tool here should be the creation of a system for classifying algorithms according to some parameters that determine effective methods for their transformation. A promising approach to solving this problem involves formal methods of artificial intelligence.

All the performed experiments showed that the identified trends increase with increasing dimension of the processed data. This provides confidence that the trends determined by modeling are preserved when scaling by the volume of data being processed.

In accordance with the iterative principle (inherent in the heuristic approach) of gradually approaching the best solution of the problem under consideration, we may have confidence in the possibility of quantitative improvement (with respect to the above parameters) in methods for scheduling of program execution on a field of parallel computers, which is given or determined by the solution of the optimization problem.

An important productive property of the developed software system is the possibility of solving inverse problems of determining the parameters of a computing system in accordance with the specified requirements for the very process of program execution, e.g., for execution time.

CONCLUSIONS

In general, the developed software system confirmed its efficiency in studying the parameters of hidden parallelism in arbitrary algorithms and its rational use in data processing. The approach using a scripting language for the development of heuristic methods (scenarios) for purposeful transformation of the forms of the information graph of algorithms showed greater flexibility and transparency for the researcher. The necessary flexibility is achieved by using an interpreted scripting language, while the processing speed is achieved according to the capabilities of the executable code of the compiled language of the parent application.

The target consumers of the developed methods for generating PPE schedules are, in the first place, developers of translators and virtual machines, as well as researchers into the properties of algorithms in order to identify and exploit the potential of their hidden parallelism. A practical application of the proposed methodology has to take into account a number of applied implementation issues, including known problems of using pointers, conflicts of memory operations in Load/Store instruction bundles, etc., which do not fundamentally change the proposed methodology.

The developed software, comprising methods for detecting hidden parallelism and its parameters in arbitrary algorithms, as well as constructing rational plans (schedules) for the PPE on a given field of computers, has been used for several years teaching students at Russian universities in order to improve their competences in data processing parallelization processes.

REFERENCES

1. Voevodin V.V., Voevodin Vl.V. *Parallel'nye vychisleniya* (*Parallel computing*). St. Petersburg: BHV-Petersburg; 2004. 608 p. (in Russ.).
2. Bakanov V. Research and selection of rational methods for obtaining framework of schedules for the parallel programs execution. In: Silhavy R., Silhavy P., Prokopova Z. (Eds.). *Data Science and Intelligent Systems. CoMeSySo 2021. Lecture Notes in Networks and Systems*. V. 231. Springer, Cham. https://doi.org/10.1007/978-3-030-90321-3_22
3. Fedotov I.E. *Parallel'noe programmirovanie. Modeli i priemy* (*Parallel programming. Models and techniques*). Moscow: SOLON-Press; 2018. 390 p. (in Russ.). ISBN 978-5-91359-222-4
4. Bakanov V.M. Dynamics control computing in the processors data flow architecture for different types of algorithms. *Programmnaya inzheneriya = Software Engineering*. 2015;9:20–24 (in Russ.).
5. Bakanov V.M. Software complex for modeling and optimization of program implementation on parallel calculation systems. *Open Comput. Sci*. 2018;8(1): 228–234. https://doi.org/10.1515/comp-2018-0019
6. Padua D. (Ed.). *Encyclopedia of parallel computing*. NY: Springer; 2012. 2195 p. https://doi.org/10.1007/978-0-387-09766-4
7. Dennis J.B., Misunas D.P. A preliminary architecture for a basic data-flow processor. In: *Proc. Second Annual Symp. Computer Architecture* (*ISCA 75*). 1975. P. 126–132. https://doi.org/10.1145/642089.642111
8. Kukunas J. *Power and performance: Software analysis and optimization*. Morgan Kaufman, Elsevier Inc.; 2015. 300 p.
9. McNairy C., Soltis D. Itanium 2 processor microarchitecture. *IEEE Micro*. 2003;23(2):44–55. https://doi.org/10.1109/MM.2003.1196114
10. Tanenbaum E., Ostin T. *Arkhitektura komp'yutera* (*Computer architecture*). Transl. from Eng. St. Petersburg: Piter; 2019. 816 p. (in Russ.). ISBN 978-5-4461-1103-9
11. Cormen T.H., Leiserson C.E., Rivest R.L., Stein C. *Introduction to algorithms*. 3rd ed. London: MIT Press; 2009. 1292 p.
12. McConnell J.J. *Analysis of algorithms: An active learning approach*. Jones & Bartlett Publishers; 2008. 451 p.
13. Garey M., Johnson D. *Vychislitel'nye mashiny i trudnoreshaemye zadachi* (*Computing machines and difficult tasks*). Transl. from Eng. Moscow: Kniga po trebovaniyu; 2012. 420 p. (in Russ.). ISBN 978-5-458-26100-5
    [Garey M., Johnson D. *Computers and intractability*. San Francisco; 1979. 338 p.]
14. Ierusalimschy R. *Programming in Lua*. 3rd ed. PUC-Rio, Brasil, Rio de Janeiro; 2013. 348 p.
15. Fisher J.A. Very long instruction word architectures and the ELI-512. In: *Proceedings of the 10th Annual International Symposium on Computer Architecture* (*ISCA '83*). 1983. P. 140–150. https://doi.org/10.1145/800046.801649
16. Babb R.I. Parallel processing with large-grain data flow techniques. *Computer*. 1984;17(7):55–61. https://doi.org/10.1109/MC.1984.1659186

СПИСОК ЛИТЕРАТУРЫ

1. Воеводин В.В., Воеводин Вл.В. *Параллельные вычисления*. СПб.: БХВ-Петербург; 2004. 608 с.
2. Bakanov V. Research and selection of rational methods for obtaining framework of schedules for the parallel programs execution. In: Silhavy R., Silhavy P., Prokopova Z. (Eds.). *Data Science and Intelligent Systems. CoMeSySo 2021. Lecture Notes in Networks and Systems*. V. 231. Springer, Cham. https://doi.org/10.1007/978-3-030-90321-3_22
3. Федотов И.Е. *Параллельное программирование. Модели и приемы*. М.: СОЛОН-Пресс; 2018. 390 с. ISBN 978-5-91359-222-4
4. Баканов В.М. Управление динамикой вычислений в процессорах потоковой архитектуры для различных типов алгоритмов. *Программная инженерия*. 2015;9:20–24.
5. Bakanov V.M. Software complex for modeling and optimization of program implementation on parallel calculation systems. *Open Comput. Sci*. 2018;8(1): 228–234. https://doi.org/10.1515/comp-2018-0019
6. Padua D. (Ed.). *Encyclopedia of parallel computing*. NY: Springer; 2012. 2195 p. https://doi.org/10.1007/978-0-387-09766-4
7. Dennis J.B., Misunas D.P. A preliminary architecture for a basic data-flow processor. In: *Proc. Second Annual Symp. Computer Architecture* (*ISCA 75*). 1975. P. 126–132. https://doi.org/10.1145/642089.642111
8. Kukunas J. *Power and performance: Software analysis and optimization*. Morgan Kaufman, Elsevier Inc.; 2015. 300 p.
9. McNairy C., Soltis D. Itanium 2 processor microarchitecture. *IEEE Micro*. 2003;23(2):44–55. https://doi.org/10.1109/MM.2003.1196114
10. Таненбаум Э., Остин Т. *Архитектура компьютера*: пер с англ. СПб.: Питер; 2019. 816 с. ISBN 978-5-4461-1103-9
11. Cormen T.H., Leiserson C.E., Rivest R.L., Stein C. *Introduction to algorithms*. 3rd ed. London: MIT Press; 2009. 1292 p.
12. McConnell J.J. *Analysis of algorithms: An active learning approach*. Jones & Bartlett Publishers; 2008. 451 p.
13. Гэри М., Джонсон Д. *Вычислительные машины и труднорешаемые задачи*: пер. с англ. М.: Книга по требованию; 2012. 420 с. ISBN 978-5-458-26100-5
14. Ierusalimschy R. *Programming in Lua*. 3rd ed. PUC-Rio, Brasil, Rio de Janeiro; 2013. 348 p.
15. Fisher J.A. Very long instruction word architectures and the ELI-512. In: *Proceedings of the 10th Annual International Symposium on Computer Architecture* (*ISCA '83*). 1983. P. 140–150. https://doi.org/10.1145/800046.801649
16. Babb R.I. Parallel processing with large-grain data flow techniques. *Computer*. 1984;17(7):55–61. https://doi.org/10.1109/MC.1984.1659186

About the author

Valery M. Bakanov, Dr. Sci. (Eng.), Professor, Department of Hardware, Software and Mathematical Support of Computing Systems, Institute of Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: e881e@mail.ru. Scopus Author ID 6505511355, ResearcherID L-1314-2015, RSCI SPIN-code 8868-4594, https://orcid.org/0000-0002-1650-038X

Об авторе

Баканов Валерий Михайлович, д.т.н., профессор, профессор кафедры КБ-5 «Аппаратное, программное и математическое обеспечение вычислительных систем» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: e881e@mail.ru. Scopus Author ID 6505511355, ResearcherID L-1314-2015, SPIN-код РИНЦ 8868-4594, https://orcid.org/0000-0002-1650-038X

Translated from Russian into English by Vladislav V. Glyanchenko
Edited for English language and spelling by Thomas A. Beavitt

Information systems. Computer sciences. Issues of information security
Информационные системы. Информатика. Проблемы информационной безопасности

UDC 004.49
https://doi.org/10.32362/2500-316X-2022-10-6-20-27

RESEARCH ARTICLE

Genetic programming support vector machine model for a wireless intrusion detection system

Anshita Dhoot [1, @], Alexey N. Nazarov [2], Ilia M. Voronkov [3]

[1] *Moscow Institute of Physics and Technology, Moscow oblast, Dolgoprudny, 141701 Russia*
[2] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[3] *HSE University, Moscow, 123557 Russia*
[@] *Corresponding author, e-mail: anshita.dhoot.23@gmail.com*

Abstract

Objectives. The rapid penetration of wireless communication technologies into the activities of both humans and Internet of Things (IoT) devices along with their widespread use by information consumers represents an epochal phenomenon. However, this is accompanied by the growing intensity of successful information attacks, involving the use of bot attacks via IoT, which, along with network attacks, has reached a critical level. Under such circumstances, there is an increasing need for new technological approaches to developing intrusion detection systems based on the latest achievements of artificial intelligence. The most important requirement for such a system consists in its operation on various unbalanced sets of attack data, which use different intrusion techniques. The synthesis of such an intrusion detection system is a difficult task due to the lack of universal methods for detecting technologically different attacks; moreover, the consistent application of known methods is unacceptably long. The aim of the present work is to eliminate such a scientific gap.

Methods. Using the achievements of artificial intelligence in the fight against attacks, the authors proposed a method based on a combination of the genetic programming support vector machine (GPSVM) model using an unbalanced CICIDS2017 dataset.

Results. The presented technological intrusion detection system architecture offers the possibility to train a dataset for detecting attacks on CICIDS2017 and extracting detection objects. The architecture provides for the separation of the dataset into verifiable and not verifiable elements, with the latter being added to the training set by feedback. By training the model and improving GPSVM training set, better accuracy is ensured. The operability of the new flowchart of the GPSVM model is demonstrated in terms of the entry of input data and output of data after processing using the training set of the GPSVM model. Numerical analysis based on the results of model experiments on selected quality indicators showed an increase in the accuracy of the results as compared to the known SVM method.

Conclusions. Computer experiments have confirmed the methodological correctness of choosing a combination of the GPSVM model using an unbalanced CICIDS2017 dataset to increase the effectiveness of intrusion detection. A procedure for forming a training dataset based on feedback is proposed. The procedure involving the separation of datasets is shown to create conditions for improving the training of the model. The combination of the GPSVM model with an unbalanced CICIDS2017 dataset to collect a sample increases the accuracy of intrusion detection to provide improved intrusion detection performance as compared to the SVM method.

Keywords: cyber security, cyber intrusion detection, rare category detection, IDS dataset, GPSVM

• **Submitted:** 15.02.2022 • **Revised:** 19.04.2022 • **Accepted:** 12.09.2022

For citation: Dhoot A., Nazarov A.N., Voronkov I.M. Genetic programming support vector machine model for a wireless intrusion detection system. *Russ. Technol. J.* 2022;10(6):20–27.https://doi.org/10.32362/2500-316X-2022-10-6-20-27

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

Модель GP SVM для беспроводной системы обнаружения вторжений

А. Дхут [1, @], А.Н. Назаров [2], И.М. Воронков [3]

[1] *Московский физико-технический институт (национальный исследовательский университет), Московская область, г. Долгопрудный, 141701 Россия*
[2] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[3] *Национальный исследовательский университет «Высшая школа экономики», Москва, 123557 Россия*
[@] *Автор для переписки, e-mail: anshita.dhoot.23@gmail.com*

Резюме

Цели. Стремительное проникновение технологий беспроводной связи и устройств Интернета вещей (IoT) в деятельность человека и их повсеместное использование потребителями информации является значимым историческим явлением. Этот процесс сопровождается растущей интенсивностью негативных информационных атак, прежде всего, широким распространением бот-атак через IoT, объем которых наряду с сетевыми атаками достигает критического уровня, и снизить его самостоятельно потребителям контента не представляется возможным. В таких обстоятельствах возрастает потребность в синтезе технологически новой, основанной на новейших достижениях искусственного интеллекта, системы обнаружения вторжений. Важнейшим требованием к системе является ее эффективность при работе на полученных разными способами несбалансированных наборах данных атак, использующих разные технологические приемы вторжения. Синтез такой системы обнаружения вторжений является сложной задачей из-за отсутствия универсальных методов обнаружения технологически разных атак, а последовательное применение известных методов является недопустимо долгим. Ликвидация этого научного пробела и является целью настоящей статьи.

Методы. Используя достижения искусственного интеллекта в борьбе с атаками, авторы предложили способ, основанный на комбинации модели машины опорных векторов генетического программирования (GPSVM) с применением несбалансированного набора данных CICIDS2017.

Результаты. Предложена архитектура системы технологического обнаружения вторжений с возможностью целевого обучения набора данных в интересах обнаружения атак на CICIDS2017 и извлечения объектов обнаружения. Архитектурой предусмотрено разделение набора данных на проверяемые и непроверяемые объекты, которые по результатам обратной связи будут добавлены в обучающий набор. Для того чтобы обеспечить лучшую точность результата, происходит обучение модели и совершенствование обучающего набора GPSVM. Показана работоспособность новой блок-схемы модели GPSVM относительно того, как набор данных вводится в качестве входных данных и выдает выходные данные после обработки с помощью обучающего набора модели GPSVM. Численный анализ результатов модельных экспериментов по выбранным показателям качества показал увеличение точности результатов по сравнению с известным методом SVM.

Выводы. Компьютерные эксперименты подтвердили методическую правильность выбора комбинации модели GPSVM с применением несбалансированного набора данных CICIDS2017 для повышения эффективности обнаружения вторжений. Предложена процедура формирования обучающего набора данных, основанная

на обратной связи. Показано, что применение такой процедуры вместе с разделением наборов данных создает условия для совершенствования обучения модели. Комбинация модели GPSVM с несбалансированным набором данных CICIDS2017 для сбора выборки повышает точность обнаружения вторжений и обеспечивает наилучшую производительность обнаружения вторжений по сравнению с методом SVM.

Ключевые слова: кибербезопасность, обнаружение кибератак, обнаружение редких категорий, набор данных IDS, GPSVM

• **Поступила:** 15.02.2022 • **Доработана:** 19.04.2022 • **Принята к опубликованию:** 12.09.2022

Для цитирования: Dhoot A., Nazarov A.N., Voronkov I.M. Genetic programming support vector machine model for a wireless intrusion detection system. *Russ. Technol. J.* 2022;10(6):20–27.https://doi.org/10.32362/2500-316X-2022-10-6-20-27

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

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

In order to defend a system from malicious behaviors such as attacks and malware, it is necessary to ensure intrusion detection. An intrusion detection system (IDS) represents an essential line of defense to protect complicated networks against increasing activities of intruders. For this reason, improved IDS designs based on wide-ranging and valid datasets for testing and evaluating techniques are proposed [1], along with responses to the challenge of obtaining significant datasets [2].

Some datasets are not shareable because of privacy concerns, while other available datasets do not reflect the latest trends. Moreover, many datasets are anonymized, despite the variety of traffic involving diverse evidence-based attacks. Thus, there are problems concerning a scarcity of definite characteristics, as well as the general inaccessibility of datasets. Hence, a precisely seamless dataset is yet to be comprehended [2–4]. Due to the evolution of malware and consistently changing attack strategies, it is necessary to regularly update standard datasets [2].

Since the year 1999, many frameworks for evaluating the IDS dataset have been proposed [2–9]. As per the latest existing research evaluation of frameworks, namely diversity of attacks, even characteristics, presented protocols, anonymity, wide-ranging interaction, complete capture, comprehensive network configuration, featuring dataset, ample traffic, metadata, heterogeneity, as well as labelling, are critical factors for developing a valid and comprehensive IDS dataset [7, 9].

ATTACK TYPES AND SCENARIOS

The IDS dataset of the Canadian Institute for Cybersecurity (CICIDS2017 dataset[1]), which offers a wide diversity of attacking sources, is intended for intrusion detection purposes, as well as providing generally network security. This dataset identifies six common attack types, which can be executed using required tools and codes [10–13].

- *Brute force attack:* Perhaps the best-known form of attack, used for password cracking as well as discovering hidden pages and available content in web applications. This type of attack is based on trial and error until the intruder succeeds.
- *Denial-of-service* (*DoS*) *Attack*: A kind of attack in which the intruder seeks to make a machine or network resource temporarily unavailable. It accomplishes this by flooding target network/resources/machines with excessive requests to overload the system, preventing some or all authentic requests from being satisfied.
- *Botnet:* An internet-connected device that uses a botnet owner for performing several tasks. This can be used to steal data, provide access for intruders and send spam messages to devices and their connection.
- *Heartbleed Attack:* This is a kind of bug that exists in the OpenSSL cryptography library that is widely used for Transport Layer Security (TLS) protocol implementation, which can be overloaded by sending a heartbeat appeal comprising malformed bulky length field as well as small payload to the susceptible user server to aggravate the target's reaction.
- *DDoS Attacks:* Distributed denial-of-service occurs when there are multiple resources flooded with the large bandwidth or resources of a vulnerable victim that can be targeted easily. In such a situation, an attack can harm a lot of multiple victims that can be compromised resources, for instance botnet; it can be easily targeted to flood system to generate the large traffic network.
- *Information Attack:* Network intrusion from inside usually gets exploited by vulnerable software for instance Adobe Acrobat Reader. After intruding to the software successfully from the backdoor, it

[1] https://www.unb.ca/cic/datasets/ids-2017.html. Accessed November 1, 2021.

conducts discrete attacks on the network of the dupe which is generally known as full-port scan, service inventories by using Nmap, and IP sweep.

- *Web Attack:* A kind of attack that occurs every day to affect individuals, as well as organizations. Structured query language (SQL) injection can be used by an intruder to generate a SQL command string for use to access information held in databases. Cross-site scripting (XSS) is used to inject SQL scripts and brute forcing over HTTP to obtain admin ID-password data.

SAMPLING TECHNIQUES

Several sample techniques are often used in unbalanced class distribution for handling real-world datasets over the network ID and credit-card fraud detection proposed by researchers for handling unbalanced class distribution data to improve classification performance [9]:

- *Over Sampling:* After replicating instances of the minority class, this approach then generates replicates based on minority classes' characteristics that decrease their distinctiveness to reduce the overall class imbalance level.
- *Under Sampling:* Removes the existed instances in the majority for balancing a dataset.
- *Combining Sampling:* Data cleaning method using a combination of sampling techniques to enhance the classification performance of an unbalanced dataset. Includes under- and oversampling, where under sampling is applied prior to oversampling in order to prevent data from being overlapped.

1. PROPOSED WORK

Our contribution: In this paper, we contribute a better result for feature selection using a genetic programming support vector machine (GPSVM) model that covers all the other important criteria to support the system for detecting intrusion attacks. In order to accomplish tests to determine the extracted features which are benign, as well as to analyze dataset for featuring best feature dataset for distinguishing distinct attacks, we used the CICIDS2017 dataset available on the Canadian Institute for Cybersecurity website [14].

This paper expands on earlier methods used as a toolset with which to create tagged datasets that can include data from both host logs and network traffic [15]. In this case, the discussed problem of using unbalanced classes and the approach to batch training of a neural network intrusion classifier with unbalanced classes can be solved by analogy with the results shown in the paper [16].

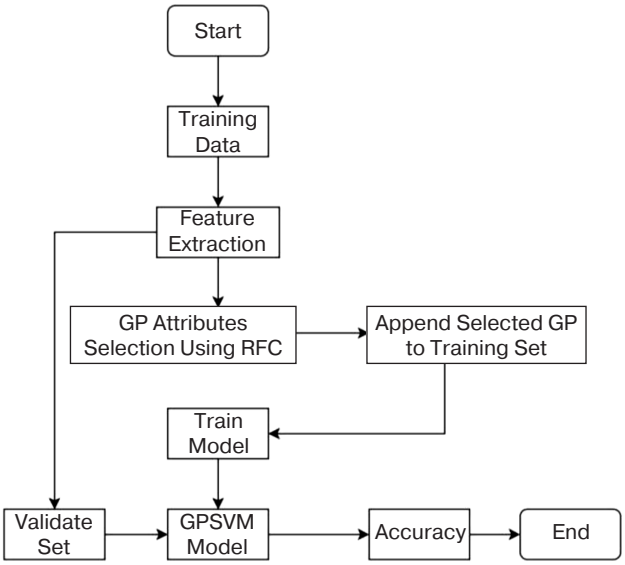


Fig. 1. Flowchart of GPSVM

In our proposed architecture, after carried out training on the CICIDS2017 dataset, features are extracted. For this purpose, datasets are divided into a valid dataset and a dataset that could not be validated, which is sent to genetic programming (GP) attribute selection using request for comments (RFC), after which the selected GP attribute is appended to the training set and sent to the trained model. After that, both datasets go to the GPSVM model to provide better accuracy. Figure 1 shows the flowchart of the GPSVM model as to how the dataset is entered as input to provide an output following its processing via the GPSVM model training set.

2. DATA ANALYSIS

The paper [9] introduced techniques to evaluate random under-sampling (RUS) and synthetic minority over-sampling techniques (SMOTE) and their combination to check the true positive rate of ID results. It shows that RUS gives the superlative performance over the SMOTE that improved results.

While there are a number of available datasets for detecting intrusion attacks, the best dataset as per ID attack is the CICIDS2017, which comprises eight separate files, each containing network activities that have data for five days. After adding eight files to CICIDS2017 dataset, it gives 2830743 instances, 15 classes, and 78 features without any replicated data [9, 14]. However, this dataset is extremely unstable due to the benign class comprising some 80.30% of the dataset. Table 1 shows the CICIDS2017 dataset class distribution with attack labels and the percentage of instances in it.

Below are the three common information that retrieves the evaluation metrics as in this form, precision (P), recall (R), and F-measure (F). Precision (P)

Table 1. CICIDS2017 dataset instances with attack details

No.	Label of attack	Instances	Percentage of total instances
1	Benign attack	2273097	80.3004
2	DoS hulk attack	231073	8.163
3	PortScan attack	158930	5.6144
4	DDoS attack	128027	4.5227
5	DoS Goldeneye attack	10293	0.3636
6	FTP Patator attack	7938	0.2804
7	SSH Patator attack	5897	0.2083
8	DoS Slowloris attack	5796	0.2048
9	DoS Slowhttptest attack	5499	0.1943
10	Bot attack	1966	0.0695
11	Web-Brute force attack	1507	0.0532
12	Web attack XSS	652	0.023
13	Infiltration attack	36	0.0013
14	Web attack-SQL injection attack	21	0.0007
15	Heartbleed attack	11	0.0004
	Total	2830743	100

represents the ratio of accurately classified attacks flow as true positive (TP) in the entire classified flows ($TP + FP$), where FP is false positive:

$$P = \frac{TP}{TP + FP}.$$

Recall (R) is also known as sensitivity, which is a ratio of accurately classified attack flows, i.e., TP, to the generated flows, i.e., ($TP + FN$), where FN is false negative:

$$R = \frac{TP}{TP + FN}.$$

F-measure is the harmonic measurement combination of recall and precision in the single measurement:

$$F = \frac{2}{\frac{1}{P} + \frac{1}{R}}.$$

At the time of execution, training, as well as testing process, is calculated to observe the execution. The weighted average of the three evaluations is considered as comprising a result in the form of (P, R, F) that provides the highest accuracy belonging to the renowned algorithms, i.e., k nearest neighbor (KNN), iterative dichotomizer 3 (ID3), and random forest (RF) methods.

3. RESULTS

In this paper, we have conducted some tests and compared the support vector machine (SVM) model and GPSVM and received better results with the GPSVM model. Although, SVM gives a good result, GPSVM model gives more accuracy in comparison to the SVM model [17]. Table 2 shows the comparison of the confusion matrix in the SVM and GPSVM models, whereas Table 3 shows model accuracy and cross-validation mean score among the SVM and GPSVM models.

Table 2. Confusion matrix

Models	Confusion matrix
SVM	[[62697 88] [166 47663]]
GPSVM	[[62699 86] [73 47756]]

Table 3. Comparison of SVM and GPSVM models

Models	Model accuracy	Cross-validation mean score
SVM	0.99770726422368	0.9974415541592128
GPSVM	0.998562589334081	0.998499276484312

Figure 2 depicts the model accuracy (1) and cross-validation mean scores (2) to compare the SVM and GPSVM models, supporting the conclusion that the GPSVM model works better than any other current alternative. In order to test both the SVM and GPSVM models, we used the CICIDS2017 dataset for making IDS better to track down intrusion attacks and reduce the attack scenario.

Figure 3 shows important features in the GPSVM model. In our proposed model, the model accuracy enhances the IDS from being vulnerable and gives better output compared to other models.

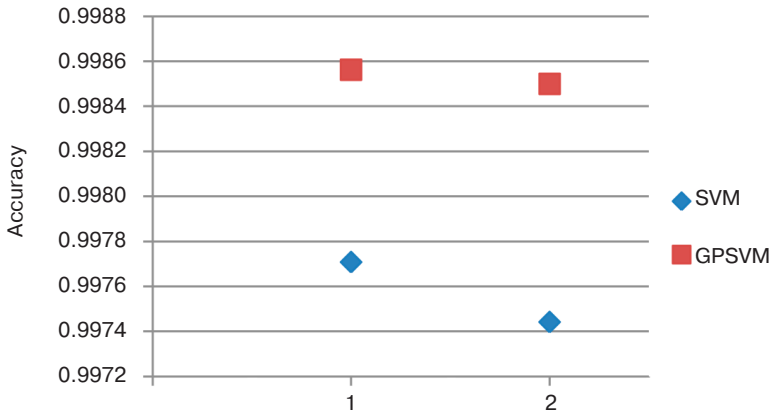


Fig. 2. Comparison of GPSVM and SVM models:
1—model accuracy; 2—cross-validation mean score

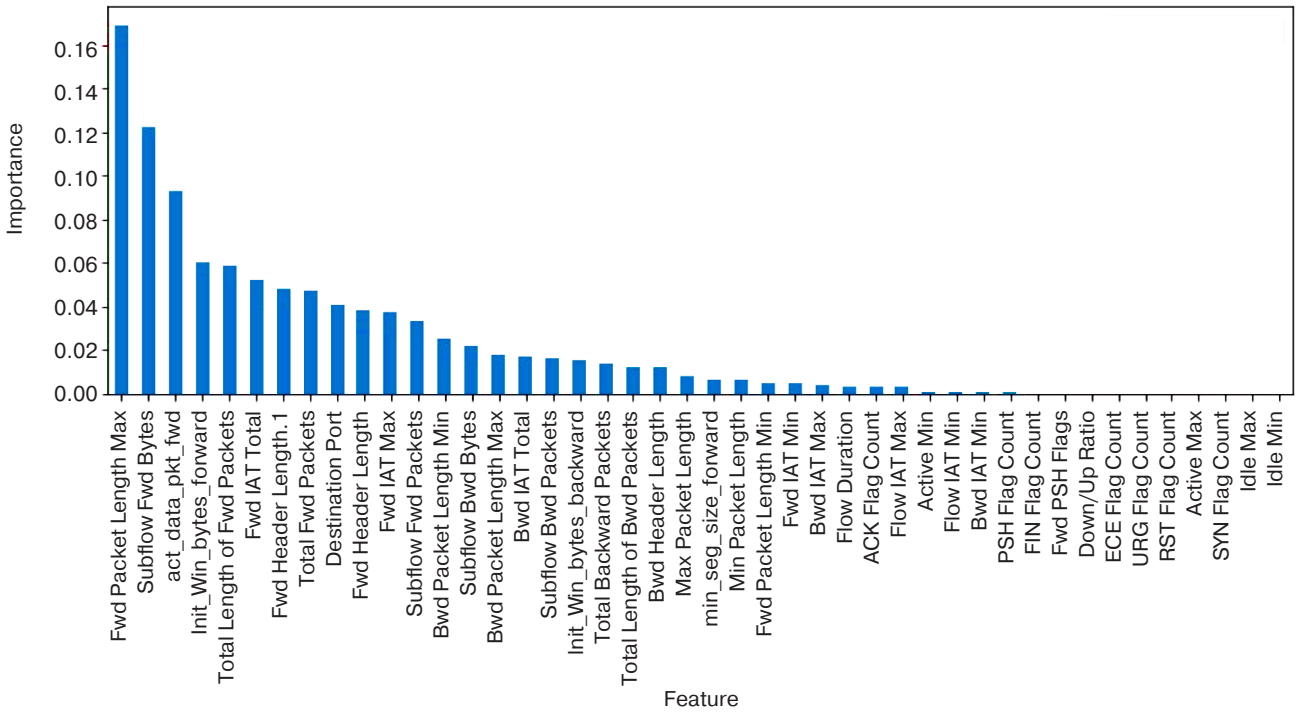


Fig. 3. Important features in the GPSVM model

4. CONCLUSIONS

The presented GPSVM model is concluded to offer an improved approach to tracking intrusion detection that helps to exploit vulnerable systems. This model provides enhanced accuracy for detecting intrusion attacks. In the future, testing this model using a wider variety of resources, along with real-time systems will help to improve the integration of the GPSVM feature extraction technique to develop an improved model version. This will aid IDSs development to improve the reliability of intrusion-detecting machines for various intrusion types. The proposed approach can be used as part of a cloud service for monitoring cyber-attacks using neuro-fuzzy formalism.

ACKNOWLEDGMENTS

The reported was funded by the Russian Foundation for Basic Research, No. 21-57-54002, and Vietnam Academy of Science and Technology, No. QTRU01.14/21-22, 2021.

Authors' contribution

All authors equally contributed to the research work. All authors approved the final text of the manuscript for publication.

REFERENCES

1. Koch R., Golling M., Rodosek G.D. Towards comparability of intrusion detection systems: New data sets. In: *TERENA Networking Conference* (*TNC*). 2014. V. 7.
2. Nehinbe J.O. A critical evaluation of datasets for investigating IDSs and IPSs researches. In: *2011 IEEE 10th International Conference on Cybernetic Intelligent Systems* (*CIS*). IEEE; 2011. P. 92–97. https://doi.org/10.1109/CIS.2011.6169141
3. Shiravi A., Shiravi H., Tavallaee M., Ghorban A.A. Toward developing a systematic approach to generate benchmark datasets for intrusion detection. *Computers & Security.* 2012;31(3):357–374. https://doi.org/10.1016/j.cose.2011.12.012
4. Ghorbani A.A., Lu W., Tavallaee M. Detection approaches. In: *Network Intrusion Detection and Prevention.* Boston, MA: Springer; 2010. P. 27–53. https://doi.org/10.1007/978-0-387-88771-5_2
5. Scott P.D., Wilkins E. Evaluating data mining procedures: techniques for generating artificial data sets. *Inf. Softw. Technol.* 1999;41(9):579–587. https://doi.org/10.1016/S0950-5849(99)00021-X
6. Heidemann J., Papdopoulos C. Uses and challenges for network datasets. In: *2009 Cybersecurity Applications & Technology Conference for Homeland Security.* IEEE; 2009. P. 73–82. https://doi.org/10.1109/CATCH.2009.29
7. Gharib A., Sharafaldin I., Lashkari A.H., Ghorbani A.A. An evaluation framework for intrusion detection dataset. In: *2016 International Conference on Information Science and Security* (*ICISS*). IEEE; 2016. P. 1–6. https://doi.org/10.1109/ICISSEC.2016.7885840
8. Sharafaldin I., Gharib A., Lashkari A.H., Ghorbani A.A. Towards a reliable intrusion detection benchmark dataset. *Software Networking*. 2018;2017(1):177–200. https://doi.org/10.13052/jsn2445-9739.2017.009
9. Ho Y.B., Yap W.S., Khor K.C. The effect of sampling methods on the CICIDS2017 network intrusion data set. In: Kim H., Kim K.J. (Eds.). *IT Convergence and Security*. *Lecture Notes in Electrical Engineering*. Singapore: Springer; 2021. V. 782. P. 33–41. https://doi.org/10.1007/978-981-16-4118-3_4
10. Sharafaldin I., Lashkari A.H., Ghorbani A.A. Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: *Proceedings of the 4th International Conference on Information Systems Security and Privacy* (*ICISSP*). 2018. P. 108–116. https://doi.org/10.5220/0006639801080116
11. Bashir T., Agbata B.C., Ogala E., Obeng-Denteh W. The fuzzy experiment approach for detection and prevention of masquerading attacks in online domain. *East African Sch. J. Eng. Comput. Sci.* 2020;3(10):205–215. https://doi.org/10.36349/easjecs.2020.v03i10.001
12. Fang Y., Zhang C., Huang C., Liu L., Yang Y. Phishing email detection using improved RCNN model with multilevel vectors and attention mechanism. *IEEE Access*. 2019;7:56329–56340. https://doi.org/10.1109/ACCESS.2019.2913705
13. Zhu E., Ju Y., Chen Z., Liu F., Fang X. DTOF-ANN: An artificial neural network masquerading detection model based on decision tree and optimal features. *Appl. Soft Comput.* 2020;95:106505. https://doi.org/10.1016/j.asoc.2020.106505
14. Lashkari A.H., Draper-Gil G., Mamun M.S.I., Ghorbani A.A. Characterization of tor traffic using time-based features. In: *Proceedings of the 3rd International Conference on Information Systems Security and Privacy* (*ICISSP*). 2017, February. P. 253–262. https://doi.org/10.5220/0006105602530262
15. Nazarov A.N., Sychev A.K., Voronkov I.M. The role of datasets when building next generation intrusion detection systems. In: *2019 Wave Electronics and its Application in Information and Telecommunication Systems* (*WECONF*). IEEE; 2019. https://doi.org/10.1109/WECONF.2019.8840124
16. Pantiukhin D., Nazarov A., Voronkov I.M. Intelligent methods for intrusion detection in local area networks. In: Pozin B., Cavalli A.R., Petrenko A. (Eds.). *Actual Problems of System and Software Engineering. Proceedings of the 6th International Conference* (*APSSE 2019*). Moscow; 2019. P. 138–149. URL: http://ceur-ws.org/Vol-2514/paper84.pdf
17. Dhoot A., Zong B., Saeed M.S., Singh K. Security analysis of private intellectual property. In: *2021 International Conference on Engineering Management of Communication and Technology* (*EMCTECH*). IEEE; 2021. https://doi.org/10.1109/EMCTECH53459.2021.9619179

About the authors

Anshita Dhoot, Postgraduate Student, Moscow Institute of Physics and Technology (National Research University) (9, Institutskii per., Moscow oblast, Dolgoprudny, 141701 Russia). E-mail: anshita.dhoot@phystech.edu. https://orcid.org/0000-0002-5024-6194

Alexey N. Nazarov, Dr. Sci. (Eng.), Professor, Department of Corporate Information Systems, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: a.nazarov06@bk.ru. ResearcherID G-3154-2013, Scopus Author ID 7201780424, RSCI SPIN-code 6032-5302, https://orcid.org/0000-0002-0497-0296

Ilia M. Voronkov, Visiting Lecturer, HSE University (11, Pokrovskii bul., Moscow, 109028 Russia); Deputy Head, Center for Neural Network Technologies, International Center for Informatics and Electronics (19, Presnenskii val, Moscow, 123557 Russia). E-mail: ivoronkov@hse.ru. ResearcherID L-6207-2016, Scopus Author ID 24802429000, RSCI SPIN-code 3869-9670, https://orcid.org/0000-0002-1552-5083

Об авторах

Дхут Аншита, аспирант, ФГАОУ ВО «Московский физико-технический институт (национальный исследовательский университет)» (141701, Московская область, г. Долгопрудный, Институтский переулок, д. 9). E-mail: anshita.dhoot@phystech.edu. https://orcid.org/0000-0002-5024-6194

Назаров Алексей Николаевич, д.т.н., профессор, профессор кафедры корпоративных информационных систем Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: a.nazarov06@bk.ru. ResearcherID G-3154-2013, Scopus Author ID 7201780424, SPIN-код РИНЦ 6032-5302, https://orcid.org/0000-0002-0497-0296

Воронков Илья Михайлович, приглашенный преподаватель, ФГАОУ ВО «Национальный исследовательский университет «Высшая школа экономики» (109028, Россия, Москва, Покровский бульвар, д. 11); заместитель начальника, Центр нейросетевых технологий, Международный центр по информатике и электронике (123557, Россия, Москва, ул. Пресненский Вал, д. 19). E-mail: ivoronkov@hse.ru. ResearcherID L-6207-2016, Scopus Author ID 24802429000, SPIN-код РИНЦ 3869-9670, https://orcid.org/0000-0002-1552-5083

The text was submitted by the authors in English
Edited for English language and spelling by Thomas A. Beavitt

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

**Роботизированные комплексы и системы.
Технологии дистанционного зондирования и неразрушающего контроля**

UDC 007.52; 004.89
https://doi.org/10.32362/2500-316X-2022-10-6-28-41

RESEARCH ARTICLE

Prototype multi-agent robotic debris removal system: principles of development and experimental studies

Sergey V. Manko [@], Valery M. Lokhin, Sekou Abdel Kader Diane

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

Abstract

Objectives. The article substantiates the relevance of the creation and the prospects of application of multi-agent robotic systems for elimination of consequences of emergency situations. The purpose of this work was to test the practical feasibility of algorithms for controlling a group of autonomous robots when performing multi-stage missions.

Methods. The theses of the finite automata theory in planning actions of a multi-agent system, methods of automatic control in organizing a goal-directed movement of robots, and methods of computer vision in searching and analyzing debris geometry were used.

Results. The principles of development, architecture, and composition are described for the software and algorithms of a prototype of the multi-agent robotic system created at RTU MIREA as part of integrated research for the creation of tools and methods of group control of robots. The multi-stage task of searching and removing debris in the process of eliminating the consequences of emergency situations is analyzed. A proposed algorithm for planning the actions of robotic agents determines the time sequence of the mission stages. Tasks are allocated among the performing robots according to assessments of their suitability. The autonomous functioning of robotic agents is determined by commands coming from the group control level, as well as an a priori embedded knowledge base with scenario models of appropriate actions. Compensation of local environmental uncertainties in the process of robot movement is based on a comprehensive analysis of visual and navigation information. Along with the main elements of the multi-agent system, the developed infrastructure of hardware and software for visual navigation and wireless communication is described.

Conclusions. The results of the experimental studies demonstrated the efficiency of the developed approaches to the creation of intelligent technologies for group control of autonomous robots on the example of debris search and removal tasks. The feasibility of the multi-agent robotic system is demonstrated by the development and integration of a number of information management and infrastructure subsystems.

Keywords: autonomous robot, intelligent control, group control, multi-robot system, debris removal

• **Submitted:** 11.02.2022 • **Revised:** 01.03.2022 • **Accepted:** 12.09.2022

For citation: Manko S.V., Lokhin V.M., Diane S.A.K. Prototype multi-agent robotic debris removal system: principles of development and experimental studies. *Russ. Technol. J*. 2022;10(6):28–41. https://doi.org/10.32362/2500-316X-2022-10-6-28-41

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

Принципы построения и экспериментальные исследования прототипного образца многоагентной робототехнической системы для разбора завалов

С.В. Манько [@], **В.М. Лохин**, **С.А.К. Диане**

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

Резюме

Цели. В статье обоснованы актуальность создания и перспективы применения многоагентных робототехнических систем для ликвидации последствий чрезвычайных ситуаций. Целью работы является проверка практической реализуемости алгоритмов управления группой автономных роботов при решении многоэтапных прикладных задач.

Методы. Использованы положения теории конечных автоматов в задачах планирования действий многоагентной системы, методы автоматического управления при организации целенаправленного движения роботов-исполнителей, методы машинного зрения в задачах поиска и анализа формы завала.

Результаты. Описаны принципы построения, архитектура и состав программно-алгоритмического обеспечения прототипного образца многоагентной робототехнической системы, разработанной в РТУ МИРЭА в рамках проведения комплексных исследований развития средств и методов группового управления роботами. Рассмотрена многоэтапная задача поиска и устранения завалов. Предложенный алгоритм планирования действий робототехнических агентов определяет временну́ю очередность выполнения этапов данной задачи. Распределение заданий между роботами-исполнителями производится с учетом оценок их пригодности. Автономное функционирование робототехнических агентов определяется командами, поступающими с уровня группового управления, а также априорно заложенной базой знаний со сценарными моделями целесообразных действий. Компенсация неопределенностей внешней среды локального характера в процессе движения роботов базируется на комплексном анализе визуальной и навигационной информации. Наряду с основными элементами многоагентной системы дано описание разработанной инфраструктуры аппаратно-программных средств визуальной навигации и беспроводной связи.

Выводы. Результаты экспериментальных исследований показали эффективность развиваемых подходов к созданию интеллектуальных технологий группового управления автономными роботами на примере задач поиска и разбора завалов. Подтверждена возможность практической реализации многоагентной робототехнической системы на основе разработки и интеграции ряда информационно-управляющих и инфраструктурных подсистем.

Ключевые слова: автономный робот, интеллектуальное управление, групповое управление, многоагентная робототехническая система, разбор завалов

• **Поступила:** 11.02.2022 • **Доработана:** 01.03.2022 • **Принята к опубликованию:** 12.09.2022

Для цитирования: Манько С.В., Лохин В.М., Диане С.А.К. Принципы построения и экспериментальные исследования прототипного образца многоагентной робототехнической системы для разбора завалов. *Russ. Technol. J.* 2022;10(6):28–41. https://doi.org/10.32362/2500-316X-2022-10-6-28-41

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

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Multi-agent robotic systems (MARS) have been attracting steadily increasing interest since the early 21st century owing to the wide prospects for their application in various applied fields. The main advantage of the joint use of autonomous robots as a united group is the significant increase in efficiency gained by combining functionality and resources when organizing the necessary interactions. Although the study of problems associated with the development of technologies for group control of autonomous robots is a priority area worldwide, practical results obtained in the United States and a number of other countries are ahead of the curve [1, 2].

Up until a certain point, programs of similar works in Russia have significantly lagged behind the international level. However, the implementation of a national strategy for scientific and technological development, one of whose main goals is the transition to advanced intellectual and robotic technologies, has provided a sharp intensification of theoretical and experimental studies on this subject. These studies are carried out in a number of research centers, academic institutes, design organizations, and universities [3–5]. As regards the proposed formulations and solutions, the fundamental and applied groundwork formed against this background by the Russian scientific school is not only keeping pace with contemporary global trends, but also is gradually implemented in practical developments related to the creation of intelligent control systems for autonomous robots and multi-agent groups.

In particular, the generalization of the results obtained at RTU MIREA in the Research Project "Methods, Models, and Algorithms for Group Control of Autonomous Robots by the Integrated Application of the Apparatus of the Theory of Finite Automata" made it possible to develop an experimental prototype of MARS for testing technologies for automatic search and analysis of debris during emergency recovery.

PROBLEMS AND PRINCIPLES OF DESIGN OF A MARS FOR AUTOMATIC DEBRIS SEARCH AND REMOVAL

An analysis of natural and man-made emergencies shows that their most characteristic consequences are large-scale destruction and debris, the difficulties of recovery of which are often intensified by such aggravating factors as fires, contamination spots, etc. Under such conditions, when the involvement of human resources may be limited or impossible, the use of autonomous robots and, in particular, MARS, may become the most feasible alternative.

Multi-agent systems for emergency recovery, automatic debris removal, and performing other tasks of similar nature are formed from specialized models of autonomous robots of the corresponding type [6, 7]. Many Russian and foreign manufacturers offer a wide range of multifunctional remote-controlled mobile robots for engineering purposes (with advanced sensor equipment, a replaceable set of tools and attachments) (Fig. 1). Such robots can potentially be used in multi-agent systems for removing debris, provided that the necessary modernization of control tools is made to ensure autonomous operation including the receipt and transmission of data via wireless network communication channels.

The main requirements for the functionality of MARS are related to the need to analyze the assigned applied task, its decomposition into a set of subtasks or composite technological operations, and their subsequent execution in autonomous mode by the joint efforts of individual performers, which coordinate their actions [8–10]. The creation of such systems that fully meet the requirements imposed on them involves the solution of the following key problems:

- organizing a developed human-machine interface, which allows for the prompt formulation of a common applied task;
- organizing the appropriate interaction between individual agents in the interests of performing a common applied task;
- ensuring the autonomy of agents and the system as a whole.

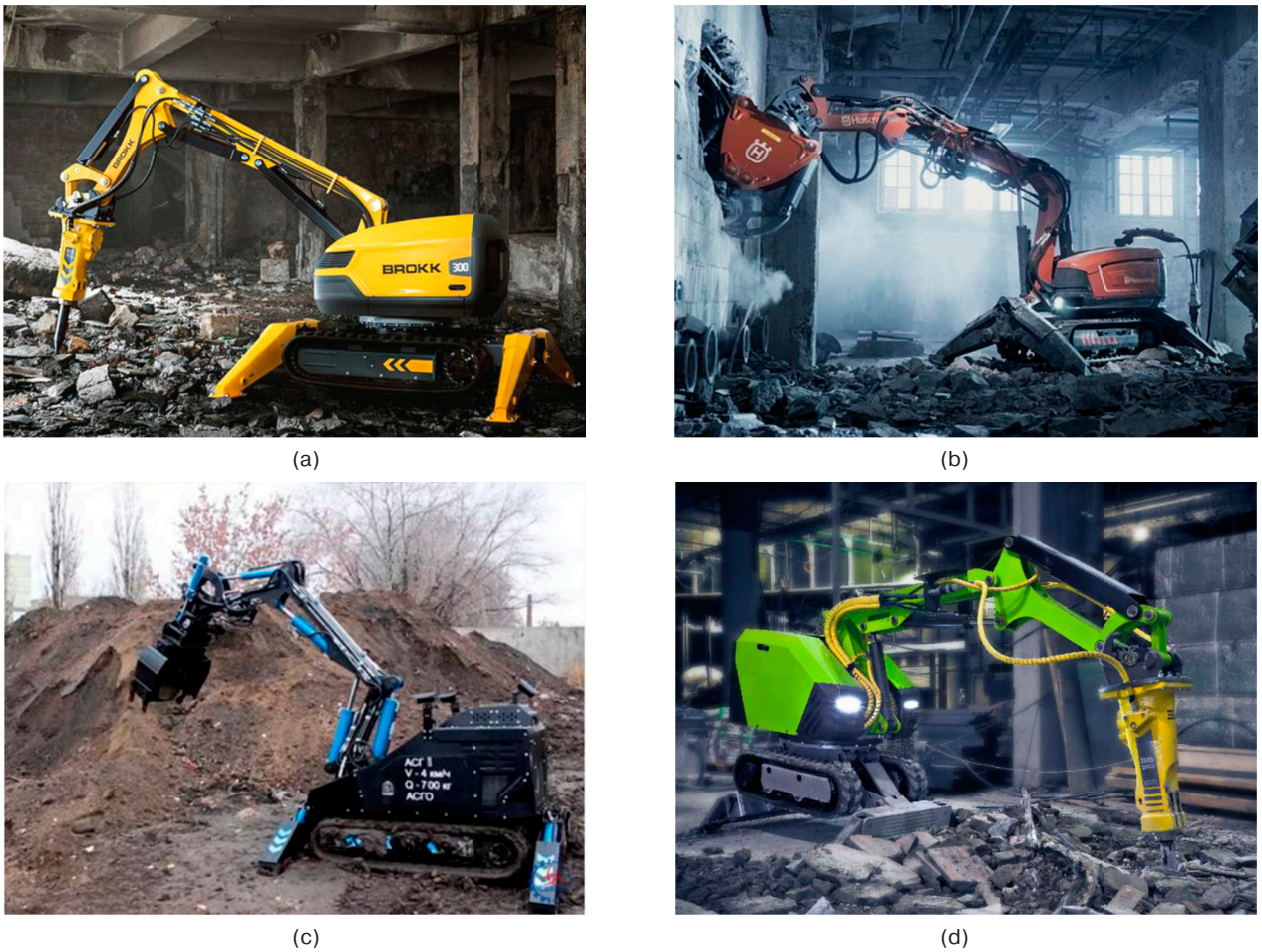
In turn, the problem of organizing the appropriate interactions between active elements of MARS contains two major subproblems:

- planning coordinated actions of agents (based on the analysis of the assigned applied task, the process of its implementation, and available resources) with subsequent formulation and assignment of appropriate tasks;
- ensuring information-logical interaction and compatibility of agents (both at the level of using common formats of presentation of data, message systems, commands, and target designations, and at the level of technical channels for their transmission).

An analysis of the identified problems shows that the integration of potential approaches to their solution within a single system implies the need to build this system in accordance with the following principles:

- commonality of goals for the functioning of agents;
- adequacy of the logical and functional capabilities of agents to the complexity of the tasks being performed;
- common information space of the system;
- mutual information-logical compatibility of agents.

When creating and implementing MARS solutions, the most acute problems are experienced when determining the sequence and methods for the joint



(a) (b) (c) (d)

Fig. 1. Models of specialized teleoperator-controlled engineering robots: (a) Brokk 300 (Brokk, Sweden), (b) Husqvarna 310 (Husqvarna, Sweden), (c) ROIN R-070 (Intekhros, Russia), and (d) Betonolom 2000 (Robotekhniks, Russia)

use of autonomous robots [9, 11] in combination with the development of an adequate set of the software and algorithms of group control with their subsequent testing and debugging.

SEQUENCE AND METHODS FOR THE TARGETED USE OF THE MARS SYSTEM FOR AUTOMATIC DEBRIS SEARCH AND REMOVAL

The effectiveness of the targeted use of MARS is generally determined by the rationality of its actions for performing the required applied task. The corresponding scenario for determining the sequence and methods of using MARS for automatic search and analysis of debris (Fig. 2) involves the phased implementation of the following set of necessary operations:

- movement of existing robots to the designated assembly point in the emergency zone from the places of their prompt delivery or permanent placement;
- reconnaissance of the general situation, assessment of the destruction centers, and determination of the parameters of their location using one or more specialized search robots;
- targeted movement of MARS to the zone of the detected debris;
- automatic layer-by-layer analysis of the debris with the allocation of recognized operations between individual MARS agents.

An analysis of the scenario shows that the implementation of the last stage involves the development of specialized software tools that not only process visual information to determine the parameters of the position of the observed elements of the debris, but also to automatically synthetize a scenario model for dismantling its upper layer.

GENERALIZED ARCHITECTURE OF THE MARS SOFTWARE FOR AUTOMATIC DEBRIS SEARCH AND REMOVAL

The developed suite of software and algorithms, the architecture of which is shown in Fig. 3, provides all the necessary functions for information processing and intelligent control of MARS that are focused on performing operations of automatic search and removal of debris and formed from among autonomous mobile

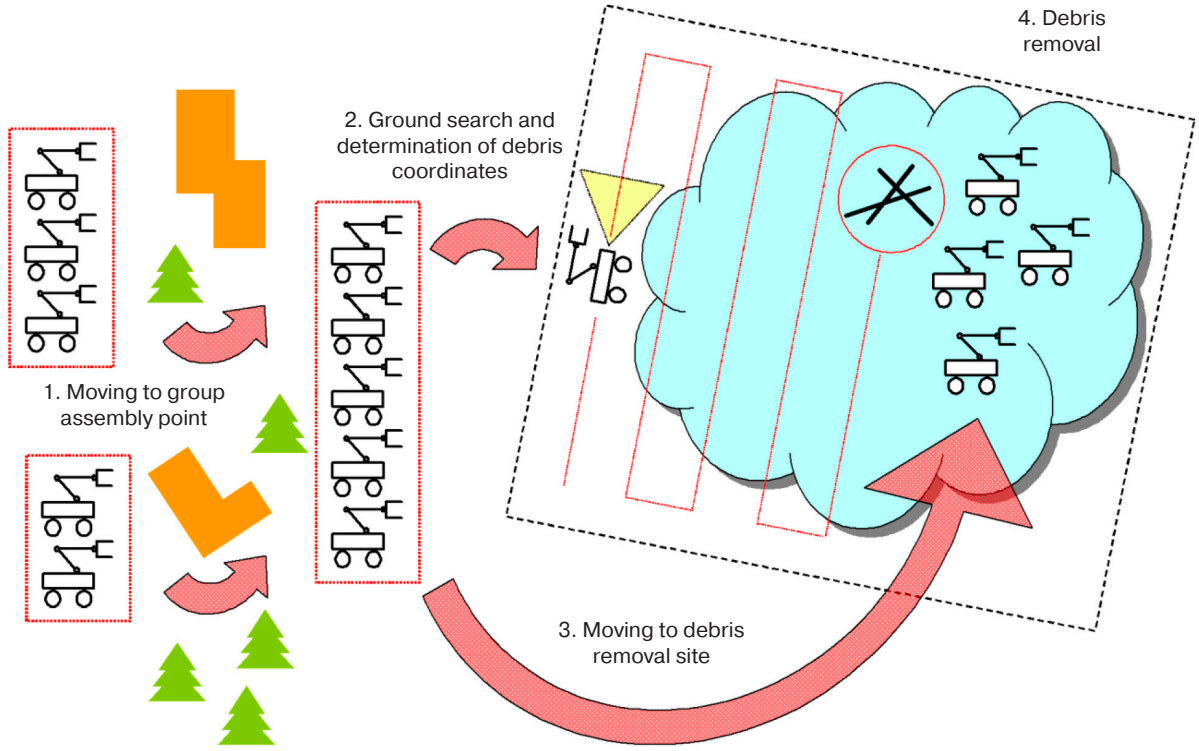


Fig. 2. Generalized scenario of the MARS functioning in performing the tasks of automatic search and analysis of debris in emergency recovery measures

robots (AMR) with an onboard manipulator and advanced sensor equipment. The main structural elements of the suite, which is built on a block-modular principle based on modern knowledge processing technologies, are the human–machine interface subsystem, the subsystem for intelligent planning and task allocation in MARS, as well as the subsystem for intelligent control of autonomous robots.

The software for the intelligent task planning subsystem of MARS includes the following set of main functional modules:

- knowledge base with scenario models for performing applied tasks created according to the description of the operator [9, 10] or in automatic mode (based on information on the structure of the observed debris) [12] using the apparatus of finite automata;
- module for summarizing sensory information for constructing a structural model of the observed debris;
- module for the dynamic construction of scenario models for layer-by-layer removal of detected debris based on the apparatus of finite automata [12];
- module for intelligent planning of stages and operations for testing scenario models of assigned applied tasks [9, 10];
- module for allocating tasks between robots from the united group using algorithms for multi-criteria assessment of the usefulness of potential performers [9].

In turn, the subsystems for intelligent control of autonomous robots of MARS have the following common software composition:

- knowledge base with a priori embedded scenario models for executing individual operations, which are built on the basis of the apparatus of finite automata;
- module for intelligent planning of appropriate actions based on the processing scenario models for performing certain operations [13];
- library of software tools for planning movements of the mobile platform and the manipulator (by the selective use of a number of specialized algorithms, such as A*, potential fields, rapidly exploring random trees (RRT), etc.), and for solving mapping problems (using SLAM algorithms) [14];
- a module for integrating odometric information and data of the external navigation system;
- a module for processing and summarizing sensory information from the onboard video camera and the laser scanning rangefinder for assessing and recognizing the external environment.

Onboard subsystems for controlling the actuating devices of autonomous robots ensure the executing of the set of commands generated by the higher-level subsystem.

The human-machine interface subsystem combines a wide range of software and tools for monitoring the performance and current state of MARS, setting goals

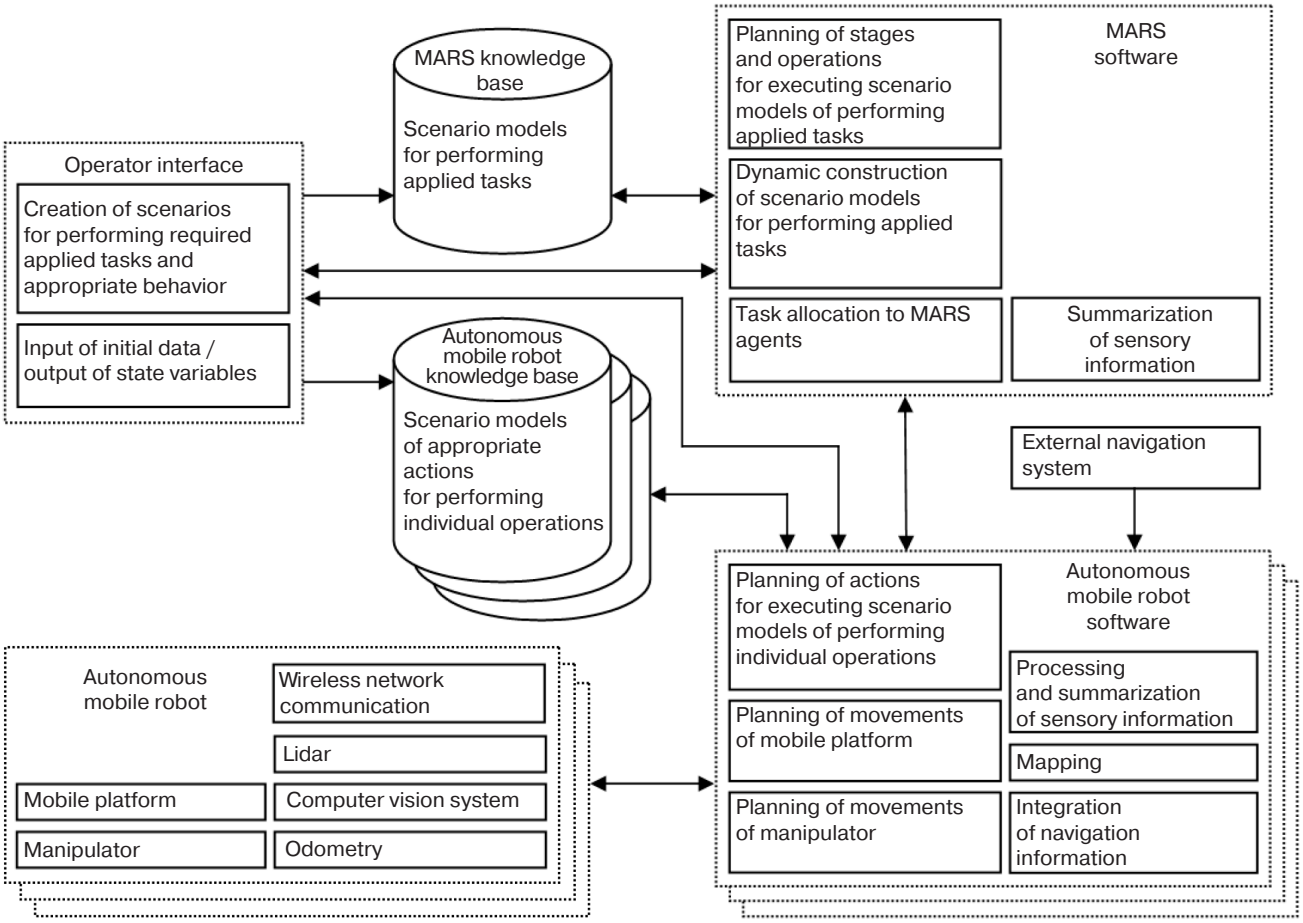


Fig. 3. Generalized architecture of the suite of software and algorithms of a demonstrator of MARS for automatic search and analysis of debris

and objectives for its operation, constructing scenario models for performing tasks and controlling robots, as well as monitoring telemetric information on the progress and results of their operations [15].

DEMONSTRATION MODEL OF THE MARS FOR AUTOMATIC DEBRIS SEARCH AND REMOVAL

The creation of the demonstrator of MARS was aimed at the fundamental testing of intelligent technologies for group control of autonomous robots on the example of automatic search and analysis of debris during field experiments with imitation of the corresponding conditions at a specialized test site. The prototype autonomous robot was represented by the KUKA youBot mobile platform (KUKARoboter, Germany). Designed as a mobile-base manipulator. This model is equipped with a laser scanning rangefinder, a stereo video camera, wireless network communication devices (Fig. 4), as well as software and algorithms for supporting modes of autonomous operation and group control.

The closed rooms of the test site were equipped with a set of 16 Beward BD3670M IP video cameras (NPP Beward, Russia), which were placed uniformly under

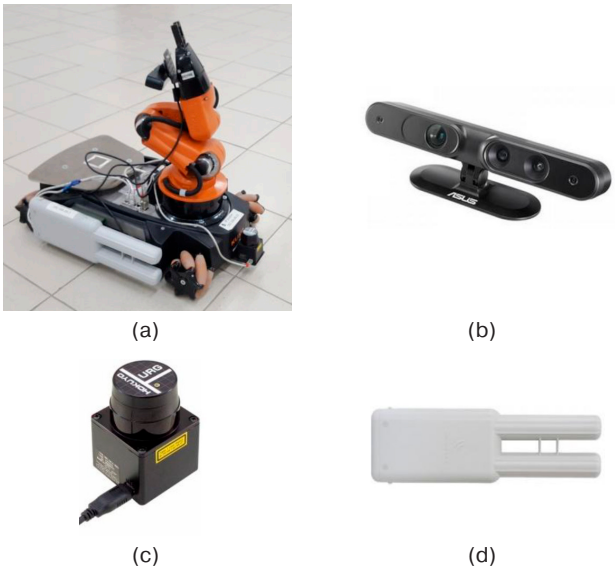


Fig. 4. (a) KUKA youBot mobile robot with an onboard set of information and measurement tools: (a) autonomous transport platform, (b) Asus Xtion PRO live RGBD video camera (AsusTek Computer Inc., China), (c) Hokuyo URG 04LX UG01 laser scanning rangefinder (Hokuyo Automatic Co., Japan), and (d) MikroTik OmniTIK UPA-5HnD wireless access point (MikroTik, Latvia)

the ceiling at a height of 4 m. As well as providing direct determination of the coordinates of observed moving objects within the common integrated navigation field (Fig. 5), these were used for emulating satellite navigation systems. The information interaction of autonomous robots of the united group was performed via wireless network communication channels in accordance with Wi-Fi technology standards using TCP/IP protocols. The unity of the common information space of MARS was ensured by the integration of the network infrastructure of the hardware and software components, related to the wireless communication and the external visual navigation systems (Fig. 6).

The prompt formulation of an applied task and control over the progress of its implementation is carried out using specialized options on the operator interface panel (Fig. 7) on the monitor screen of the central computer.

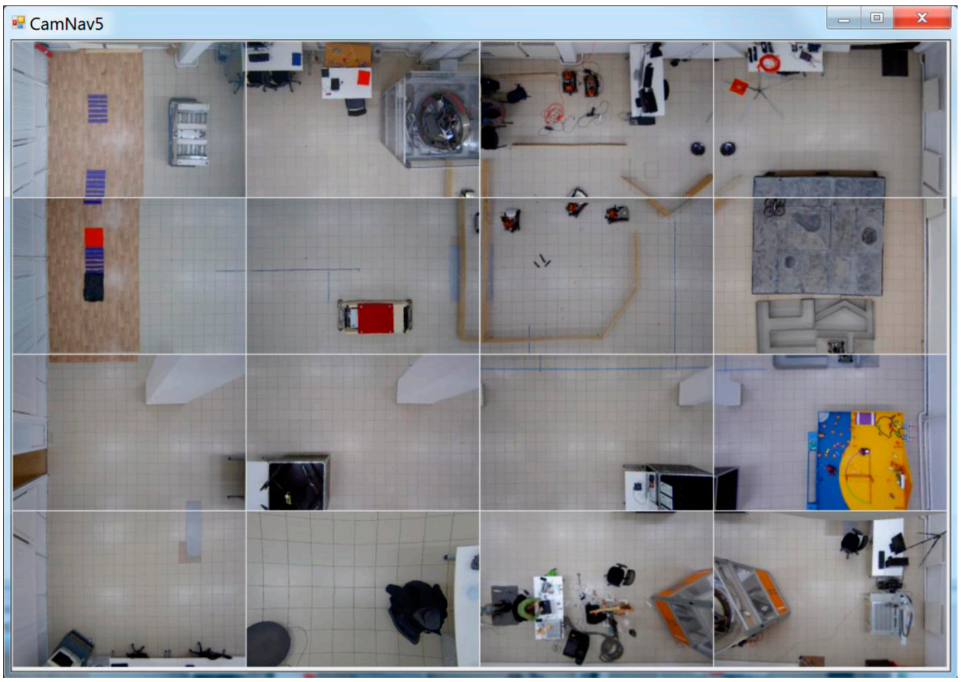


Fig. 5. Integrated field of the visual navigation system

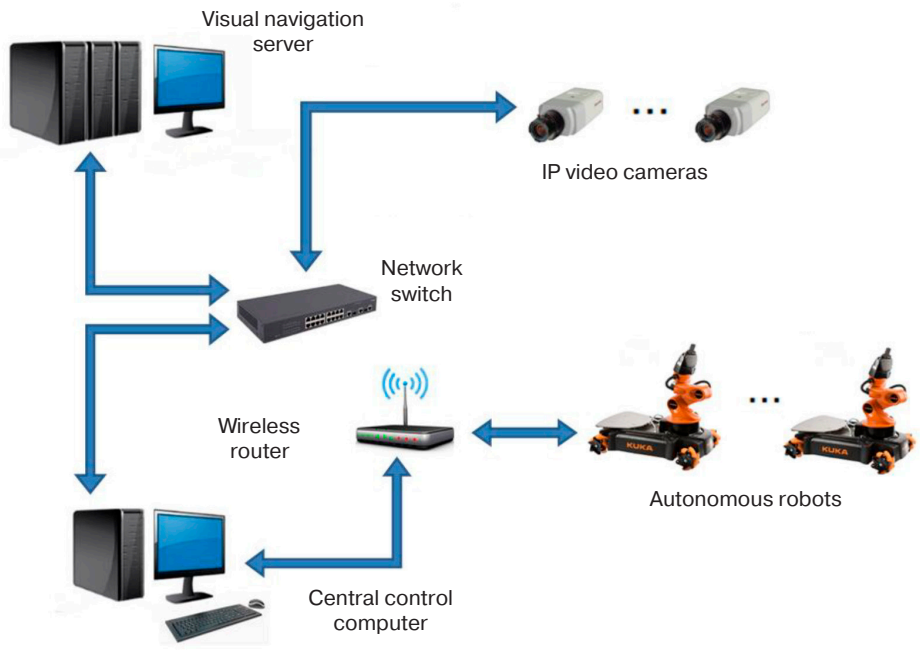


Fig. 6. United network infrastructure of the hardware and software of the visual navigation and the wireless communication systems

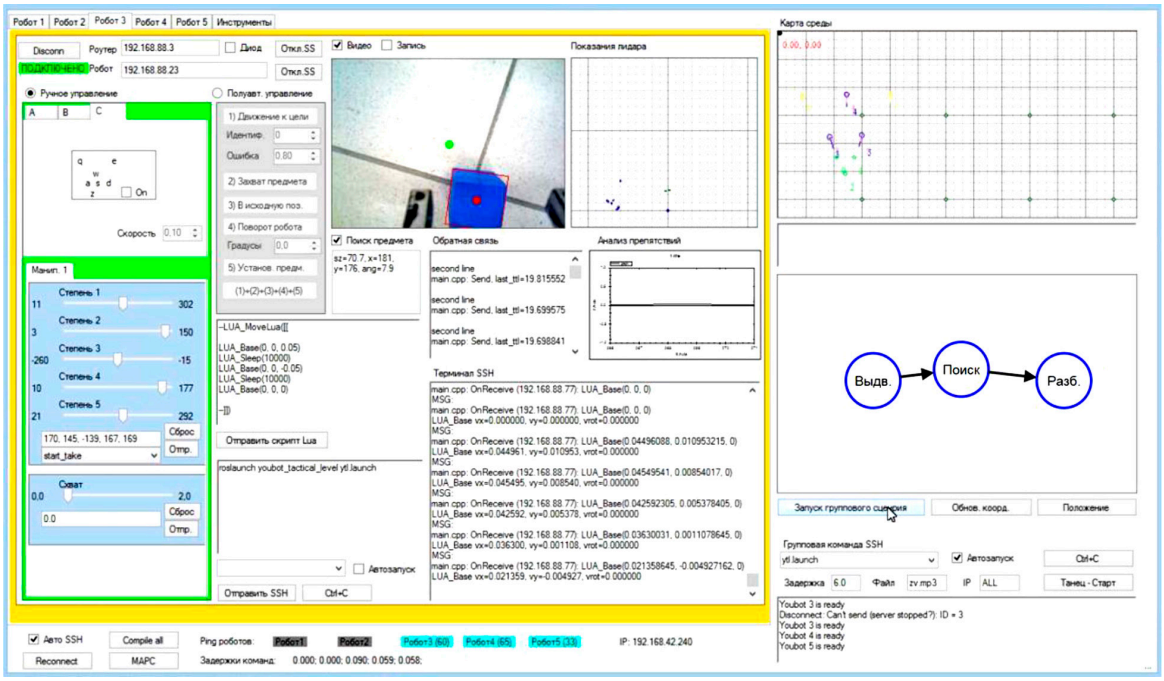


Fig. 7. Interactive operator interface panel of the demonstration model of MARS for automatic search and removal of debris

FIELD TESTING OF INTELLIGENT CONTROL TECHNOLOGIES OF THE MARS FOR AUTOMATIC DEBRIS SEARCH AND REMOVAL

A series of planned experiments for field testing of the software and algorithms for group control of autonomous robots for automatic search and removal of debris was carried out by the example of imitation of consequences of an emergency with conditional imitation of a zone of destruction and debris (Fig. 8) at a specialized test site of the Institute of Artificial Intelligence of RTU MIREA.

The formulation of the task by the operator using interactive tools of the human-machine interface (Fig. 7) involves the construction of a scenario graph, whose nodes and edges represent the composition and sequence of the main stages of the applied mission to be performed. Each of the nodes of the constructed graph is assigned a dataset that determines the parameters of the corresponding stage. If necessary, the assigned data may contain references to the involved algorithms or names of nested scenario graphs.

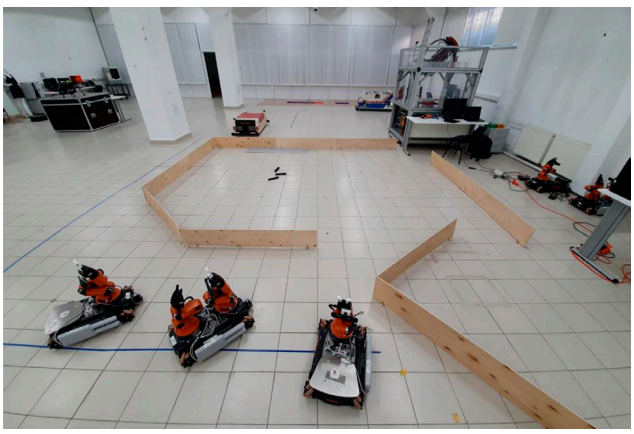


Fig. 8. Stage model for making experiments for automatic search and removal of debris using MARS

The preparation and specification of scenario graphs ensures the automatic construction of a formalized model of the applied task to be solved in the form of an appropriate network of finite automata. This approach ensures the implementation of the control and monitoring functions while planning the behavior and appropriate actions of MARS autonomous robots.

In particular, the formulation of the applied problem for MARS within the ongoing experiments required the construction of a scenario graph with the sequential inclusion of the stages of movement of robots to the assembly point, search, and subsequent removal of debris (Fig. 8).

It is important to note that the parameters for conducting the second stage (ground search) provide for two possible outcomes: a negative search result leads to the completion of the entire mission, whereas the detection of debris causes a transition to the next stage of the task.

In turn, the last stage is specified by a nested scenario graph with cyclic repetition of operations for recognizing the structure of the upper layer of the debris, as well as dynamic model synthesis and planning of dismantling operations up to the last constituent element.

The final stage of setting the task is the initialization of robots, indicating their network addresses and

operating parameters, and entering the coordinates of the assembly place and the zone of subsequent work.

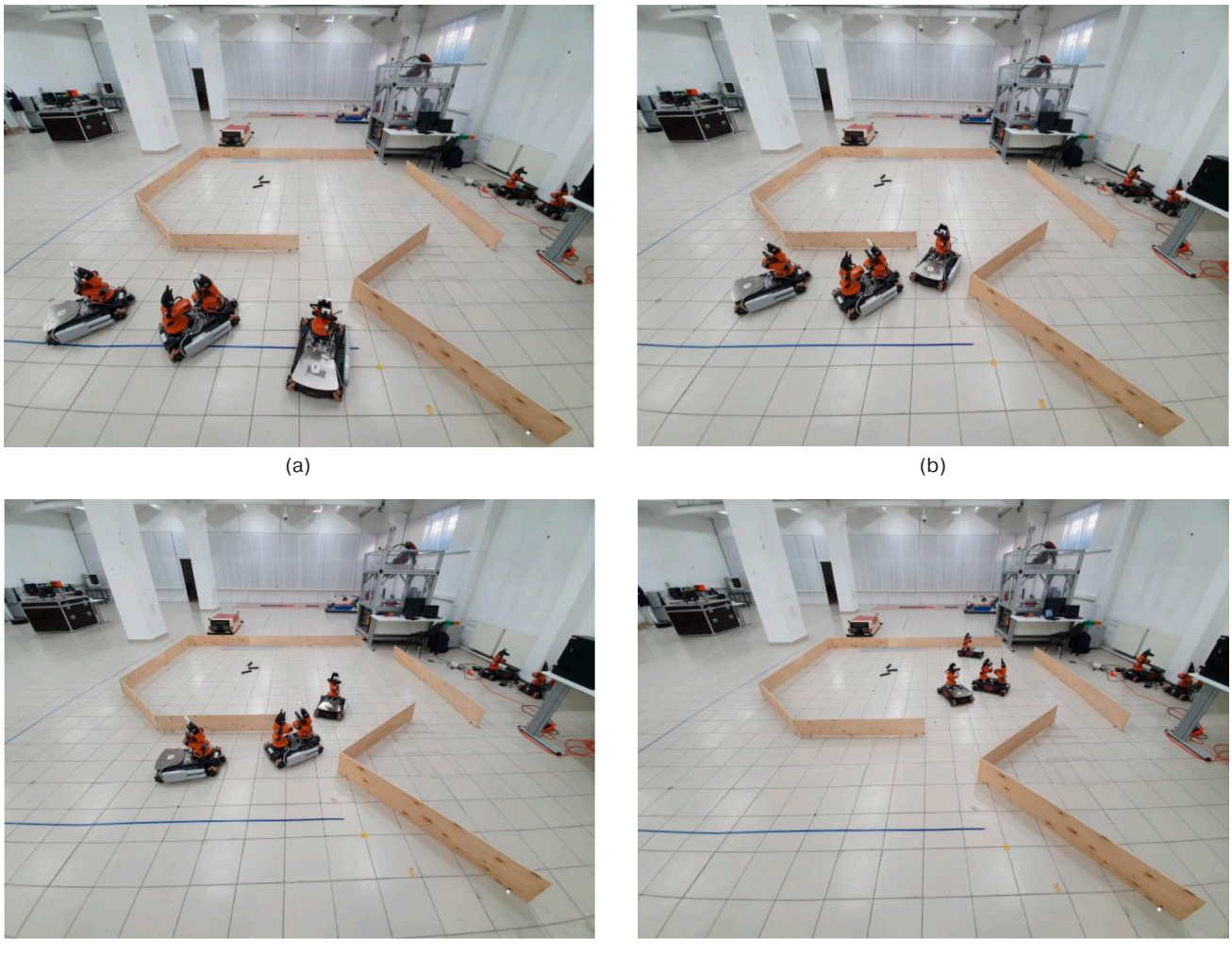
The launch of MARS, which is carried out at the command of the operator, initiates the direct execution of the assigned applied task in automatic mode in accordance with the operationally introduced and a priori embedded scenarios of autonomous and group control.

Figures 9–11 present the main fragments of one of the test-site experiments for testing methods and technologies for group control of autonomous robots of MARS in the co-execution of the common applied task of automatic search and removal of debris.

Figures 9a–9d illustrate in detail the execution of the first stage of the task, which is the advance of autonomous robots to the designated assembly point. While their targeted movement is routed using the A* [16] and RRT [17] algorithms, the movements associated with maintaining convoy formation are planned using the algorithmic implementation of the method of potential fields [18] and the apparatus for processing visual, ranging, and navigational information. The algorithms are chosen based on the scenario for performing the assigned task at the level of the corresponding references established during the specification of the parameters for this stage. The arrival of robots at the designated assembly place, confirmed via wireless network communication channels, serves not only as an actual confirmation of the successful completion of the current stage of the task being performed, but also as a signal for the transition to its next phase. The current status of the robots and the task they perform is displayed on the operator interface panel.

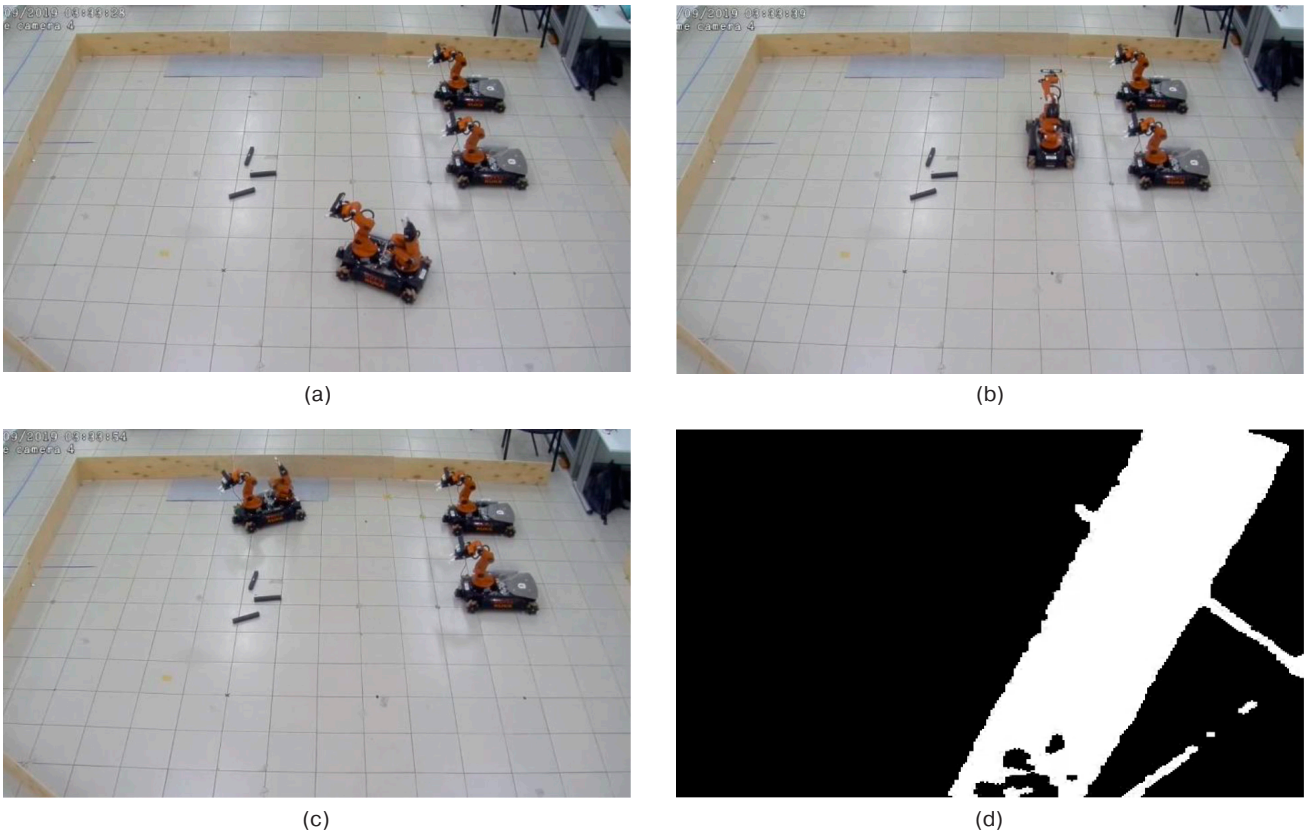
Figure 10 illustrates the second stage of the task, which is the area reconnaissance and the search for debris.

A robot for these purposes is selected according to the criterion of a suitable functional or the smallest serial number (other conditions being equal). The completeness of the inspection of a given area is determined by the scenario of planning and testing search movements using the A* algorithm and available means of processing visual, ranging and navigation information. The sequence of necessary actions is regulated by the scenario for performing the task, as well as the scenario for conducting search operations, which is a priori embedded in the knowledge bases of intelligent control systems for autonomous robots.



(a) (b) (c) (d)

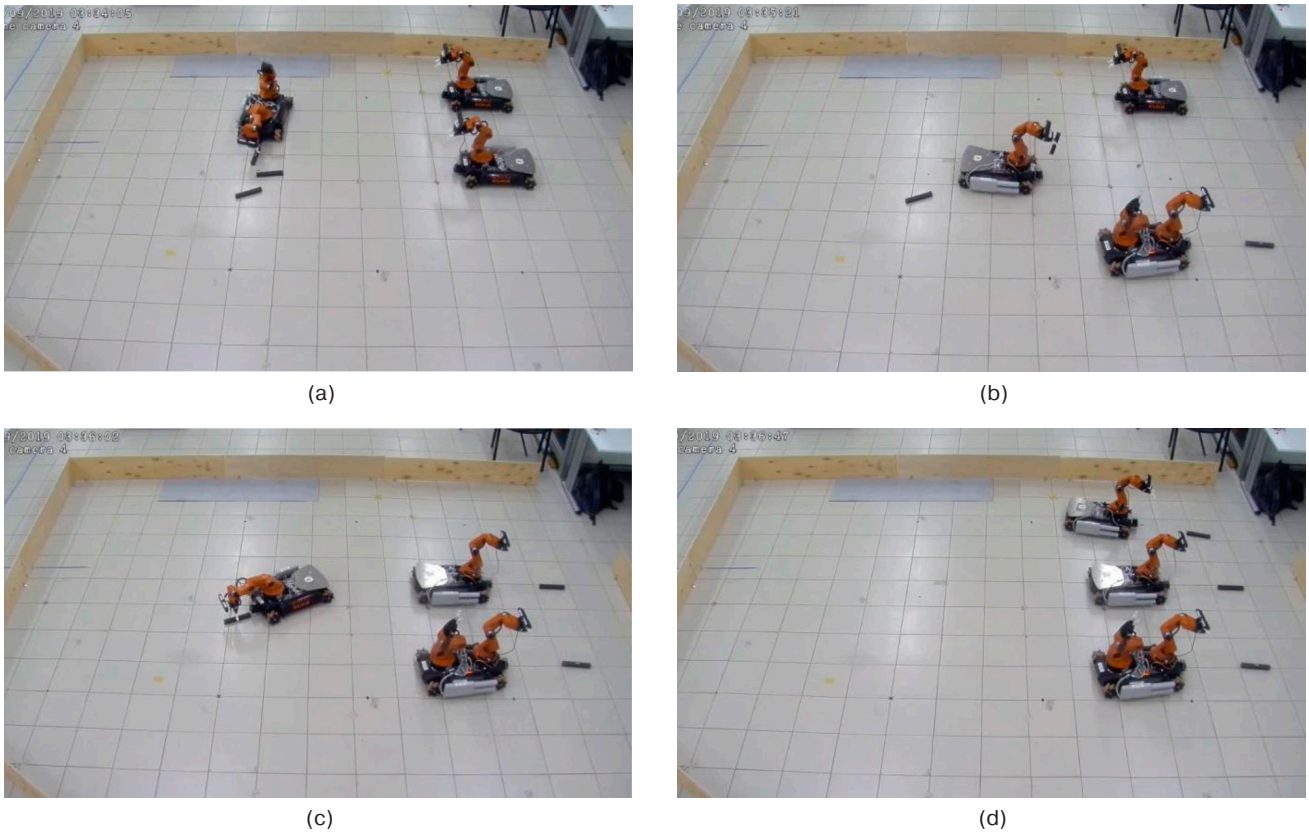
Fig. 9. Movement of a group of autonomous robots maintaining a convoy-type formation configuration toward the designated MARS assembly point



(a) (b)

(c) (d)

Fig. 10. Operation of MARS during the ground search: (a, b) the selection of a robot to search for debris, (c) search movements of the robot during ground search, and (d) the detection of debris based on the results of processing the images from the onboard camera of the robot



(a) (b)

(c) (d)

Fig. 11. Operation of MARS performing debris removal operations

The criterion for detection of debris is the excess of the critical number of destruction elements per unit area, which is found from the results of processing the data from the onboard set of information and measurement tools of the robot.

Following the detection of debris, the third stage of the task involving the removal and evacuation of the found objects can be initiated (Figs. 11a–11d). The nested scenario of its implementation includes the cyclic repetition of image processing procedures, the analysis and restoration of the structure of the upper layer of the debris, and the dynamic construction of a scenario model for dismantling the selected elements.

The construction of such a model based on the use of the apparatus of finite automata makes it possible to plan the layer-by-layer removal of the debris with the prompt issuance of an ordered set of subtasks, indicating the elements to be removed and the parameters of their spatial position.

The tasks are allocated between robots based on assessments of their current occupancy and distance from the place of the corresponding operations. The generated allocations are transmitted via wireless network communication channels to selected addressees from MARS. The execution of the received tasks is regulated by a set of scenario models that are a priori embedded in the onboard knowledge base of the intelligent control system of each of the robots and determine the procedure for carrying out the necessary operations to capture and evacuate the indicated elements of the debris based on the combined use of a number of appropriate algorithms. For example, the local movements of the mobile platform and the manipulator installed on it are planned on the basis of the algorithmic implementation of the RRT method. The targeted movements of the mobile platform are routed using the A* algorithm. Bypassing moving obstacles during the movements of the robot in a dynamically changing scene is controlled by the algorithmic implementation of the method of potential fields. The entire set of problems of planning the movements and controlling the movement of the robot is solved on a real-time basis using the means of processing visual, ranging, and navigation information to take into account the specifics of the environment.

Objective control of the results of testing each component of scenarios of each level on the basis of processing the readings of information and measurement tools of individual robots ensures the transition to the next stage of the task or its completion with all necessary information displayed on the interactive operator interface panel.

The means and methods for group control of autonomous robots, whose coordinated interaction is ensured by the combined use of various information strategies, was practically tested by creating the MARS demonstrator and carrying out its field tests:

- strategies of centralized control in planning the stages of performing a given applied task with the allocation of subtasks to individual robots for performing the necessary operations;
- strategies of autonomous control in planning appropriate actions and coordinating the movements of robots during their joint functioning in performing the assigned tasks.

Note that the principles of constructing the set of hardware, software, and algorithms of the MARS demonstrator allow for its scaling according to the total number of robots in the group. The implementation of such properties makes it possible to increase not only the fault tolerance of the system, but also the degree of its universality in application to various practical problems, the complexity and extent of whose solution may require a prompt change in the required number of involved robots.

In this context, of special interest is the question of the amount of information transmitted via wireless network communication channels during the operation of MARS. As the analysis showed, the intensity of information flows in the data transmission network of the developed MARS configuration with five robots does not exceed a critical level (Table).

Thus, the conducted field experiments convincingly demonstrated the feasibility of MARS solutions focused on performing various applied tasks, including automatic search and removal of debris during emergency recovery.

Table. Intensity of information flows in the wireless data transmission network of MARS

Data	Frequency, Hz	Size of one message, kB	Number of devices	Total transmission rate, kbps	Content
Navigation	5	102	16	65 280	Images from IP cameras
Navigation	5	0.5	1	20	Line with coordinates of detected robots
Navigation	1	0.1	5	4	Lines with coordinates of each robot
Control	1	0.1	5	4	Tactical-level robot control commands
Control	1	0.05	5	2	Confirmation of execution of tactical-level commands
Interface	10	25	5	10 000	Image from onboard video camera
Interface	10	2	5	800	Data from onboard lidar

CONCLUSIONS

The creation of the MARS prototype enabled the integrated testing not only of group control methods and algorithms, but also technologies for processing the necessary knowledge along with human–machine interface tools for the prompt setting of the applied tasks to be performed. The results of experimental studies performed by the example of tasks of automatic search and removal of debris confirmed the efficiency of the developed software and algorithms, the current version of which implements the centralized strategy of group control of autonomous robots constituing the united group. Proposals for further development of the system include supplementing its structure with bulletin board mechanisms to support decentralized strategies for the functioning of autonomous robots at the level of coordinating their behavior plans and interactions.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 16-29-04397) under the Research Project "Methods, Models, and Algorithms for Group Control of Autonomous Robots by the Integrated Application of the Apparatus of the Theory of Finite Automata."

Authors' contributions

S.V. Manko has developed algorithms for planning and distributing tasks between MARS agents; proposed a formalized description of the scenario of the specified applied problem in the form of a network of finite automata.

V.M. Lokhin has developed the MARS architecture and the generalized structure of its autonomous robotic agent; proposed the concept of a full-scale experiment to search for and eliminate the blockage, and controlled the stages of its implementation.

S.A.K. Diane has developed vision algorithms for autonomous robotic agents in the MARS; programmatically implemented key algorithms for planning, distributing tasks, and processing information in MARS.

REFERENCES

1. Yasuda T., Ohkura K. (Eds.). *Multi-robot systems: trends and development*. InTech; 2011. 596 p. ISBN 978-953-307-425-2. https://doi.org/10.5772/544
2. Seenu S.N., Kuppan Chetty R.M., Ramya M.M. Review on state-of-the-art dynamic task allocation strategies for multiple-robot systems. *Industrial Robot*. 2020;47(6): 929–942. https://doi.org/10.1108/IR-04-2020-0073
3. Kalyaev I.A., Gaiduk A.R., Kapustyan S.G. *Modeli i algoritmy kollektivnogo upravleniya v gruppakh robotov* (*Models and algorithms of collective control in groups of robots*). Moscow: Fizmatlit; 2009. 280 p. (in Russ.). ISBN 978-5-9221-1141-6
4. Pshikhopov V.Kh. (Ed.). *Gruppovoe upravlenie podvizhnymi ob"ektami v neopredelennykh sredakh* (*Group control of moving objects in undetermined*). Moscow: Fizmatlit; 2015. 305 p. (in Russ.).
5. Rubtcov V.I., Mashkov K.Yu., Lapshov V.S., Konovalov K.V. Multilevel control system for a group of robots to work in Arctic. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki = Izvestiya TulGU*. 2021;7:53–61 (in Russ.).
6. Datskov I.V., Seroshtanov A.V. Special purpose robotic systems in service of FGKU TsSOOR "Lider." In: *Proc. of The 27th International Scientific-Practical Conference "Prevention. Salvation. Help*." Khimki: Akademiya grazhdanskoi zashchity MChS Rossii; 2017. P. 20–32 (in Russ.).
7. Datskov I.V., Kubeko A.V. Capabilities of TsSOOR "Lider" in application of robotic means. In: *Proc. of The 27th International Scientific-Practical Conference "Prevention. Salvation. Help*." Khimki: Akademiya grazhdanskoi zashchity MChS Rossii; 2017. P. 32–34 (in Russ.).

СПИСОК ЛИТЕРАТУРЫ

1. Yasuda T., Ohkura K. (Eds.). *Multi-robot systems: trends and development*. InTech; 2011. 596 p. ISBN 978-953-307-425-2. https://doi.org/10.5772/544
2. Seenu S.N., Kuppan Chetty R.M., Ramya M.M. Review on state-of-the-art dynamic task allocation strategies for multiple-robot systems. *Industrial Robot*. 2020;47(6): 929–942. https://doi.org/10.1108/IR-04-2020-0073
3. Каляев И.А., Гайдук А.Р., Капустян С.Г. *Модели и алгоритмы коллективного управления в группах роботов*. М.: ФИЗМАТЛИТ; 2009. 280 с. ISBN 978-5-9221-1141-6
4. Пшихопов В.Х. (ред.). *Групповое управление подвижными объектами в неопределенных средах*. М.: ФИЗМАТЛИТ; 2015. 305 с. ISBN 978-5-9221-1674-9
5. Рубцов В.И., Машков К.Ю., Лапшов В.С., Коновалов К.В. Многоуровневая система управления группой роботов для работы в условиях Арктики. *Известия Тульского государственного университета. Технические науки*. 2021;7:53–61.
6. Дацков И.В., Сероштанов А.В. Робототехнические комплексы специального назначения, состоящие на вооружении ФГКУ ЦСООР «Лидер». В сб.: *Материалы XXVII Международной научно-практической конференции «Предупреждение. Спасение. Помощь»*. 16 марта 2017 г. Химки: Академия гражданской защиты МЧС России; 2017. С. 20–32.
7. Дацков И.В., Кубеко А.В. Возможности ЦСООР «Лидер» по применению робототехнических средств. В сб.: *Материалы XXVII Международной научно-практической конференции «Предупреждение. Спасение. Помощь»*. 16 марта 2017 г. Химки: Академия гражданской защиты МЧС России; 2017. С. 32–34.

8. Makarov I.M., Lokhin V.M., Man'ko S.V., Romanov M.P. Multiagent robotic systemś construction principles and design problems. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2012;3:11–16 (in Russ.).
9. Makarov I.M., Lokhin V.M., Man'ko S.V., Romanov M.P., Kryuchenkov E.N., Kucherskii R.V., Khudak Yu.I. Action planning and tasks distribution models and algorithms for multiagent robotic systems. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2012;5:44–50 (in Russ.).
10. Lokhin V.M., Man'ko S.V., Karpov S.A., Margolin I.D. Behavioral mechanisms ensuring network communications in multi-agent robotic systems. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2017;18(12):802–811 (in Russ.). https://doi.org/10.17587/mau.18.802-811
11. Man'ko S.V., Diane S.A.K., Lokhin V.M., Novosel'skii A.K. Group control of robots for debris removal and construction disassembly in the atomic industry. *Ekstremal'naya robototekhnika = Extreme Robotics*. 2017;1(1):302–311 (in Russ.).
12. Diane S., Manko S., Lokhin V. Task planning in robot groups for problems with implicity defined scenarios based on finite-state automata technique. In: *Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements* (*SCM 2017*). 2017. P. 348–351. https://doi.org/10.1109/SCM.2017.7970581
13. Diane S., Manko S., Margolin I., Novosselskiy A. Hierarchical scenarios for behavior planning in autonomous robots. In: *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering* (*EIConRus*). 2019. P. 479–484. https://doi.org/10.1109/EIConRus.2019.8657067
14. Egortsev M.V., Diane S.K., Kaz N.D. Algorithmic support of the system of external observation and routing of autonomous mobile robots. *Russ. Technol. J.* 2021;9(3):15–23 (in Russ.). https://doi.org/10.32362/2500-316X-2021-9-3-15-23
15. Lokhin V.M., Manko S.V., Alexandrova R.I., Romanov M.P., Diane S.A. Man-machine interface for autonomous robots and multi-agent robotic systems. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2016;17(9):606–614 (in Russ.). https://doi.org/10.17587/mau.17.606-614
16. Duchoň F., Babinec A., Kajan M., Beňo P., Florek M., Fico T., Jurišica L. Path planning with modified a star algorithm for a mobile robot. *Procedia Engineering*. 2014;96:59–69. https://doi.org/10.1016/j.proeng.2014.12.098
17. Mohammed H., Jaradat M.A., Romdhane L.B. RRT∗N: An improved rapidly-exploring random tree approach for reduced processing times. In: *2018 11th International Symposium on Mechatronics and its Applications* (*ISMA*). 2018. P. 1–6. https://doi.org/10.1109/ISMA.2018.8330126
18. Platonov A.K., Karpov I.I., Kiril'chenko A.A. *Metod potentsialov v zadache prokladki trassy* (*Method of potentials in the problem of laying a route*). Moscow; 1974. 27 p. (Preprint. Institute of Applied Mathematics, USSR Academy of Sciences; No. 124) (in Russ.).
8. Макаров И.М., Лохин В.М., Манько С.В., Романов М.П. Принципы построения и проблемы разработки мультиагентных робототехнических систем. *Мехатроника, автоматизация, управление*. 2012;3:11–16.
9. Макаров И.М., Лохин В.М., Манько С.В., Романов М.П., Крюченков Е.Н., Кучерский Р.В., Худак Ю.И. Модели и алгоритмы планирования действий и распределения заданий в мультиагентных робототехнических системах. *Мехатроника, автоматизация, управление*. 2012;5:44–50.
10. Лохин В.М., Манько С.В., Карпов С.А., Марголин И.Д. Поведенческие механизмы обеспечения сетевой связи в мультиагентных робототехнических системах. *Мехатроника, автоматизация, управление*. 2017;18(12):802–811. https://doi.org/10.17587/mau.18.802-811
11. Манько С.В., Диане С.А.К., Лохин В.М., Новосельский А.К. Групповое управление роботами в задачах разбора завалов и демонтажа объектов атомной отрасли. *Экстремальная робототехника*. 2017;1(1):302–311.
12. Diane S., Manko S., Lokhin V. Task planning in robot groups for problems with implicitly defined scenarios based on finite-state automata technique. In: *Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements* (*SCM 2017*). 2017. P. 348–351. https://doi.org/10.1109/SCM.2017.7970581
13. Diane S., Manko S., Margolin I., Novosselskiy A. Hierarchical scenarios for behavior planning in autonomous robots. In: *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering* (*EIConRus*). 2019. P. 479–484. https://doi.org/10.1109/EIConRus.2019.8657067
14. Егорцев М.В., Диане С.А.К., Кац Н.Д. Алгоритмическое обеспечение системы внешнего наблюдения и маршрутизации автономных мобильных роботов. *Российский технологический журнал*. 2021;9(3):15–23. https://doi.org/10.32362/2500-316X-2021-9-3-15-23
15. Лохин В.М., Манько С.В., Александрова Р.И., Романов М.П., Диане С.А.К. Принципы построения и программно-алгоритмическое обеспечение человеко-машинного интерфейса для автономных роботов и мультиагентных робототехнических систем. *Мехатроника, автоматизация, управление*. 2016;17(9):606–614. https://doi.org/10.17587/mau.17.606-614
16. Duchoň F., Babinec A., Kajan M., Beňo P., Florek M., Fico T., Jurišica L. Path planning with modified a star algorithm for a mobile robot. *Procedia Engineering*. 2014;96:59–69. https://doi.org/10.1016/j.proeng.2014.12.098
17. Mohammed H., Jaradat M.A., Romdhane L.B. RRT∗N: An improved rapidly-exploring random tree approach for reduced processing times. In: *2018 11th International Symposium on Mechatronics and its Applications* (*ISMA*). 2018. P. 1–6. https://doi.org/10.1109/ISMA.2018.8330126
18. Платонов А.К., Карпов И.И., Кирильченко А.А. *Метод потенциалов в задаче прокладки трассы*. М.; 1974. 27 с. (Препринт. Институт прикладной математики АН СССР; № 124).

About the authors

Sergey V. Manko, Dr. Sci. (Eng.), Professor, Department of Management Problems, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). He is a specialist in the field of intelligent control of autonomous robots and multi-agent robotic systems. Russian Federation State Prize Laureate in the field of education. Member of the Scientific Council of the Russian Academy of Sciences on robotics and mechatronics. E-mail: manko@mirea.ru. Scopus Author ID 55761014700, RSCI SPIN-code 2070-1592, https://orcid.org/0000-0002-6297-8894

Valery M. Lokhin, Dr. Sci. (Eng.), Professor, Department of Management Problems, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). Author of more than 160 scientific and educational works in the field of intelligent information technologies and automatic control. Laureate of the State Prize of the Russian Federation in the field of science and technology. Russian Federation State Prize Laureate in the field of education. Member of the Scientific Council of the Russian Academy of Sciences on robotics and mechatronics. Honored Worker of Science of the Russian Federation. E-mail: kpu-mirea@yandex.ru. Scopus Author ID 6602931640, https://orcid.org/0000-0001-6708-9124

Sekou Abdel Kader Diane, Cand. Sci. (Eng.), Assistant Professor, Department of Management Problems, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: sekoudiane1990@gmail.com. ResearcherID T-5560-2017, Scopus Author ID 57188548666, RSCI SPIN-code 8691-0290, https://orcid.org/0000-0002-8690-6422

Об авторах

Манько Сергей Викторович, д.т.н., профессор кафедры проблем управления Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). Специализируется в области интеллектуального управления автономными роботами и мультиагентными робототехническими системами. Лауреат премии Правительства РФ в области образования. Член научного Совета РАН по робототехнике и мехатронике. E-mail: manko@mirea.ru. Scopus Author ID 55761014700, SPIN-код РИНЦ 2070-1592, https://orcid.org/0000-0002-6297-8894

Лохин Валерий Михайлович, д.т.н., профессор кафедры проблем управления Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). Автор более 160 научных и учебно-методических работ в области интеллектуальных информационных технологий и автоматического управления. Лауреат государственной премии РФ в области науки и техники. Лауреат премии Правительства РФ в области образования. Член научного Совета РАН по робототехнике и мехатронике. Заслуженный деятель науки РФ. E-mail: kpu-mirea@yandex.ru. Scopus Author ID 6602931640, https://orcid.org/0000-0001-6708-9124

Диане Секу Абдель Кадер, к.т.н., доцент кафедры проблем управления Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). Специализируется на вопросах обработки информации и управления в интеллектуальных робототехнических системах. E-mail: sekoudiane1990@gmail.com. ResearcherID T-5560-2017, Scopus Author ID 57188548666, SPIN-код РИНЦ 8691-0290, https://orcid.org/0000-0002-8690-6422

Translated from Russian into English by Vladislav V. Glyanchenko
Edited for English language and spelling by Thomas A. Beavitt

Modern radio engineering and telecommunication systems

Современные радиотехнические и телекоммуникационные системы

UDC 621.391
https://doi.org/10.32362/2500-316X-2022-10-6-42-51

RESEARCH ARTICLE

Comparison of algorithms for multi-objective optimization of radio technical device characteristics

Alexander V. Smirnov @

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

Abstract

Objectives. The selection of a method for solving multi-objective optimization problems has many practical applications in diverse fields. The present work compares the results of applying different methods to the selected classes of problems by solution quality, time consumption, and various other criteria.

Methods. Five problems related to the multi-objective optimization of analog and digital filters, as well as multistep impedance-matching microwave transformers, are considered. One of the compared algorithms comprises the Third Evolution Step of Generalized Differential Evolution (GDE3) population-based algorithm for searching the full approximation of the Pareto set simultaneously, while the other three algorithms minimize the scalar objective function to find only one element of the Pareto set in a single search cycle: these comprise Multistart Pattern Search (MSPS), Multistart Sequential Quadratic Programming (MSSQP) method and Particle Swarm Optimization (PSO) algorithms.

Results. The computer experiments demonstrated the capability of GDE3 to solve all considered problems. MSPS and PSO showed significantly inferior results than to GDE3 for two problems. In one problem, MSSQP could not be used to reach acceptable decisions. In the other problems, MSPS, MSSQP, and PSO reached decisions comparable with GDE3. The time consumption of the MSPS and PSO algorithms was much greater than that of GDE3 and MSSQP.

Conclusions. The GDE3 algorithm may be recommended as a basic method for solving the considered problems. Algorithms minimizing scalar objective function may be used to obtain several elements of the Pareto set. It is necessary to investigate the impact of landscape features of individual quality indices and scalar objective functions on the extreme search process.

Keywords: multi-objective optimization, Pareto optimality, Pareto front, quality index, scalarizing objective function, population-based algorithm

• **Submitted:** 02.05.2022 • **Revised:** 20.05.2022 • **Accepted:** 15.09.2022

For citation: Smirnov A.V. Comparison of algorithms for multi-objective optimization of radio technical device characteristics. *Russ. Technol. J.* 2022;10(6):42–51. https://doi.org/10.32362/2500-316X-2022-10-6-42-51

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

The author declares no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Сравнение алгоритмов многокритериальной оптимизации характеристик радиотехнических устройств

А.В. Смирнов [@]

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

Резюме

Цели. Вопрос о выборе метода решения задачи многокритериальной оптимизации из множества известных методов актуален для многих практических областей. Цель исследования – сравнить результаты применения разных методов на выбранных классах задач по качеству решений, затратам времени и другим критериям.

Методы. В работе сравниваются результаты применения различных алгоритмов при решении пяти задач многокритериальной оптимизации характеристик аналоговых и цифровых фильтров и многоступенчатых согласующих СВЧ-трансформаторов. Исследовались популяционный алгоритм GDE3, осуществляющий поиск одновременно всей аппроксимации множества Парето-оптимальных решений, и три алгоритма, основанные на скаляризации целевой функции, которые в одном цикле поиска находят один элемент указанного множества. Это многократный запуск покоординатного поиска MSPS, многократный запуск алгоритма последовательного квадратичного программирования MSSQP и алгоритм роя частиц PSO.

Результаты. Проведенное исследование показало, что популяционный алгоритм GDE3 позволяет успешно находить множества решений для всех рассмотренных задач. В двух задачах из пяти алгоритмы MSPS и PSO существенно уступили GDE3 как по качеству решений, так и по затратам времени на поиск одного решения. В одной из задач алгоритм MSSQP оказался неработоспособным. В других задачах алгоритмы, основанные на скаляризации, находили решения, не только не уступающие, а в некоторых случаях и превосходящие результаты GDE3. При этом затраты времени на поиск одного решения у MSPS и PSO оказались значительно бóльшими, чем у GDE3 и MSSQP.

Выводы. Алгоритм GDE3 можно рекомендовать как базовый для решения подобных задач. Алгоритмы, основанные на скаляризации, целесообразно применять при поиске небольшого числа элементов множества Парето-оптимальных решений. Необходимо исследовать влияние особенностей рельефов отдельных показателей качества и скалярных целевых функций на процесс поиска решения.

Ключевые слова: многокритериальная оптимизация, Парето-оптимальное решение, фронт Парето, показатель качества, скаляризация целевой функции, популяционный алгоритм

• **Поступила:** 02.05.2022 • **Доработана:** 20.05.2022 • **Принята к опубликованию:** 15.09.2022

Для цитирования: Смирнов А.В. Сравнение алгоритмов многокритериальной оптимизации характеристик радиотехнических устройств. *Russ. Technol. J.* 2022;10(6):42–51. https://doi.org/10.32362/2500-316X-2022-10-6-42-51

Прозрачность финансовой деятельности: Автор не имеет финансовой заинтересованности в представленных материалах или методах.

Автор заявляет об отсутствии конфликта интересов.

INTRODUCTION

Multi-objective optimization (MOO) problems are of considerable interest in radio engineering and other fields of research [1]. Multicriteria optimization is aimed at finding an approximation of the set of Pareto-optimal solutions (POS) [2], which cannot be improved by one of the quality indices (QI) without deteriorating at least one of the other QIs.

Methods for solving MOO problems can be divided into two main classes [3, 4]. The first is based on solving the problem of minimizing the scalar objective function (OF) written in the generalized form [5]:

$$f(\mathbf{x}) = \sum_{j=1}^{M}\left(W_j \cdot \max\left(\frac{\left(Q_j(\mathbf{x}) - Q_{jt}\right)}{Q_{jt}^{\,q_j}}, 0 \right)\right), \quad (1)$$

where $Q_j(\mathbf{x})$ and Q_{jt} are the current and target values of the jth QI; W_j is the weighting factor of the jth QI.

The exponent $q_j = 1$ provided that normalizing the QI deviation from the target value for bringing the summands in (1) to the same range of values is required. The exponent $q_k = 0$ provided that such normalization is not required. Equality (1) covers different scalarization techniques for MOO problem. Taking all $Q_{jt} = 0$ and considering the denominators equal to one, the problem of minimizing the QI weighted sum arises. Another technique is to set target values Q_{jt}, as well as $W_k = 1$ and $W_j >> 1$, where $j = 1, ..., M, j \neq k$. In this case, the problem of minimizing QI Q_k while fixing other QI near the target values arises. Methods based on scalarization allow one POS to be found in a search cycle.

The methods grouped in the second class (MOO population-based algorithms) permit the computation of several elements of approximation of the POS set in a single search cycle using agent population. The classification and description of algorithms of this type are given in [3, 4].

The sets of test functions [6] and quality criteria for approximating the POS set [3] are used to compare different methods for solving the MOO problem. Generally, the results of applying MOO population-based algorithms are evaluated using test functions, e.g., as in [7, 8]. Here, no comparison with the results of scalarization-based methods is performed. This can be justified in the case of test functions whose properties are all known; the quality of found approximations of the POS set can be evaluated objectively. In terms of practical applications, however, where nothing generally is known in advance about functions describing QI, it becomes necessary to solve the "Black Box Optimization" problem for collecting information on QI values while finding.

A determination of which algorithm would give the best approximation may be achieved only by comparing the results of different methods, including both population-based and scalarization-based approaches. Such comparison for a specific class of problems is performed in [9] to support a conclusion about the superiority of population-based algorithms in terms of the quality of obtained solutions. Nevertheless, a different result may be obtained for other types of problems. In [10], the possibility of applying population-based algorithms from the PlatEMO open source [11] running in the *MATLAB* environment to solve the MOO problem of the frequency response of analog electrical filters is investigated. Here, the authors conclude that, in the case of optimization by two QIs, population-based methods provide better solutions than scalarization-based approaches. While, in the case of optimization by three QIs, the opposite result is obtained, this conclusion is reached when not using the most effective algorithm for finding an extremum of the scalar OF in experiments. Different scalarization-based algorithms are compared in [12], albeit without considering population-based approaches.

The present work aims to compare the results of applying MOO algorithms of different classes on examples of several problems of optimizing the characteristics of radio engineering devices. Among the indicators characterizing the compared methods, the highest priority is given to obtaining the best results. In case these criteria are equal, a comparison may be performed in terms of search duration per one POS along with other indices.

PROBLEM FORMULATION AND SELECTION OF OPTIMIZATION ALGORITHMS

Problem 1. MOO of frequency response of analog filters. The QI definitions and methods for calculating them are given in [13] and other works by the author. Below is a list of QIs:

- uneven gain-frequency response in passband *DHp*, dB;
- minimum attenuation in stopband *Hs*, dB;
- exceeding gain-frequency response of a given level in transition band *DHt*;
- uneven delay-frequency response in passband *DTd*, %.

Frequency response is calculated on a dimensionless frequency scale normalized to the upper bound frequency of the low pass filter (LPF) passband. Here, it is necessary to minimize *DHp* and *DTd* together while fulfilling the constraints $DHt \leq 0$, $Hs \geq Hst$.

Problem 2. MOO of simultaneous frequency and time responses of analog filters. The QI definitions and methods for calculating them are given in [13]. In addition to the frequency domain QI mentioned above, the following time domain QIs are introduced:

- maximum voltage (overshoot) of the transient process *Um* normalized to the steady-state value;
- transient rising (front) duration *Tfr*;
- transient-process duration *Tss*.

The last two QIs are calculated on a dimensionless time scale referenced to the normalized frequency scale. Here, it is necessary to jointly minimize *Tss* and maximize *Hs* under the constraints $DHt \leq 0$, $Um \leq Umt$, and $Tfr \leq Tfrt$.

Problem 3. MOO of frequency responses of digital filters. The QI definitions and methods for calculating them are given in [5]. The QI list and formulation of optimization problem is the same as for Problem 1.

Problem 4. MOO of frequency responses of matching multistep microwave transformers (transitions). The following QIs are defined:

- maximum *KPmax* and minimum *KPmin* power transfer coefficients in the matching band;
- uneven power transfer coefficient in matching band *DKP* = *KPmax* – *KPmin*.

The method for calculating these QIs is not given here due to the limited scope of the paper and will be published separately. It is necessary to jointly minimize *DKP* and maximize *KPmin* at a given value of the matching bandwidth *DFM*.

Problem 5. Another MOO problem of frequency response of matching multistep microwave transformers. In addition to QI defined for Problem 4, the relative unevenness of the delay-frequency response in matching band *DTd*, %, is introduced. It is necessary to minimize *DKP* and *DTd* at given values of matching bandwidth *DFM* and maximum power transfer coefficient *KPt*.

We shall now consider the algorithms used for solving the above problems.

The population-based algorithm is implemented using the PlatEMO library mentioned above. The algorithm GDE3 (The Third Evolution Step of Generalized Differential Evolution) showing the best response values according to [10] is selected from 71 algorithms presented in the library. The specified parameters are the size of population *Npop* and the number of calculations of the QI set *Neval*. These values are found experimentally for each problem. For this purpose, search iterations with increasing *Npop* and *Neval* values were performed until the found approximation of the POS set is improved.

Below are the algorithms for finding extrema of scalar OFs.

Multistart Pattern Search (MSPS), which describes the repeated start of stepwise search, is implemented by SOFTD [13] in Problems 1 and 2 and by HODF in Problem 3 [5]. Both programs are written in C++. In Problems 4 and 5, the algorithm is implemented in the *MATLAB* environment.

Particle Swarm Optimization (PSO) is the particle swarm algorithm [3, 4] implemented in *MATLAB* by the *particleswarm*(..) function from the *Global Optimization Toolbox* module. According to the results presented in [12], this algorithm demonstrates its capacity to find global extrema of scalar OFs with complex landscape.

Multistart Sequential Quadratic Programming (MSSQP) is the repeated start of sequential quadratic programming algorithm implemented in *MATLAB* by the *fmincon*(..) function from *Optimization Toolbox* module. Unlike MSPS and PSO algorithms searching for minima of scalar OF of the type (1), the constraints on QI are not considered in MSSQP as penalty terms but are included in the Lagrange function that may be written in the following form:

$$L(\mathbf{x},\boldsymbol{\lambda}) = Q_k(\mathbf{x}) + \sum_{i=1}^{M-1} \lambda_i g_i(\mathbf{x}),$$
$$g_i(\mathbf{x}) = Q_j(\mathbf{x}) - Q_{jt},\ i = 1,...,M-1,\ j = 1,...,M,\ j \neq k, \quad (2)$$

where λ_i are Lagrange multipliers; Q_j are quality indices; Q_{jt} are their target values.

In the strict sense, the algorithm is not intended for finding global extrema of OFs having a complex landscape containing many local extrema. However, as shown in [12], it obtains a good approximation to the global minimum of OF from most starting points evenly distributed in the search space for some problems. At the same time, the search duration turns out to be much shorter than for other algorithms.

The number of starts, *NT,* of the scalarization-based algorithms for finding one POS is selected in each problem so as to find a solution that could not be significantly improved by further *NT* increasing. Other parameters are set equal to default values.

PROBLEM SOLUTIONS

Problem 1. We shall consider experimental results for the analog LPF whose transfer function (TF) contains *NP* = 6 poles and *NZ* = 0 zeros. The lower bound frequency of the stopband on the frequency scale normalized to the upper bound frequency of the passband is *Fs* = 2. The results obtained using the GDE3 algorithm are shown in Fig. 1 in the form of Pareto front approximation graphs for *Hst* = 30 dB and *Hst* = 40 dB. In the first case, *Neval* = $1 \cdot 10^6$, while in the second case, *Neval* = $0.5 \cdot 10^6$. The search duration is 4 and 2 min, respectively. In both cases, *Npop* = 50.

The solutions obtained using algorithms based on scalarizing OFs are also presented in Fig. 1. Finding one solution takes an average of 2 min by MSPS algorithm, 10 s by MSSQP algorithm, and 5 min when using the PSO algorithm. The number of MSPS starts is 3000, while the number of PSO starts is 40; the number of MSSQP starts is 20. From the start results, the best one is selected. Here, it should be noted that about half of MSSQP starts gives the same best result when searching for each POS, with the remainder resulting in unacceptable solution. The results of the other two algorithms are distributed over a wide range of values in most experiments.

The comparison of QI solutions obtained by different algorithms shows that only MSSQP provides benefits for solution quality as compared to GDE3 at *DHp* small values. The other two algorithms based on scalarizing OFs show at best results similar to GDE3, being at a disadvantage to it at small values of *DHp*. In addition, they require significantly more time for finding solutions.

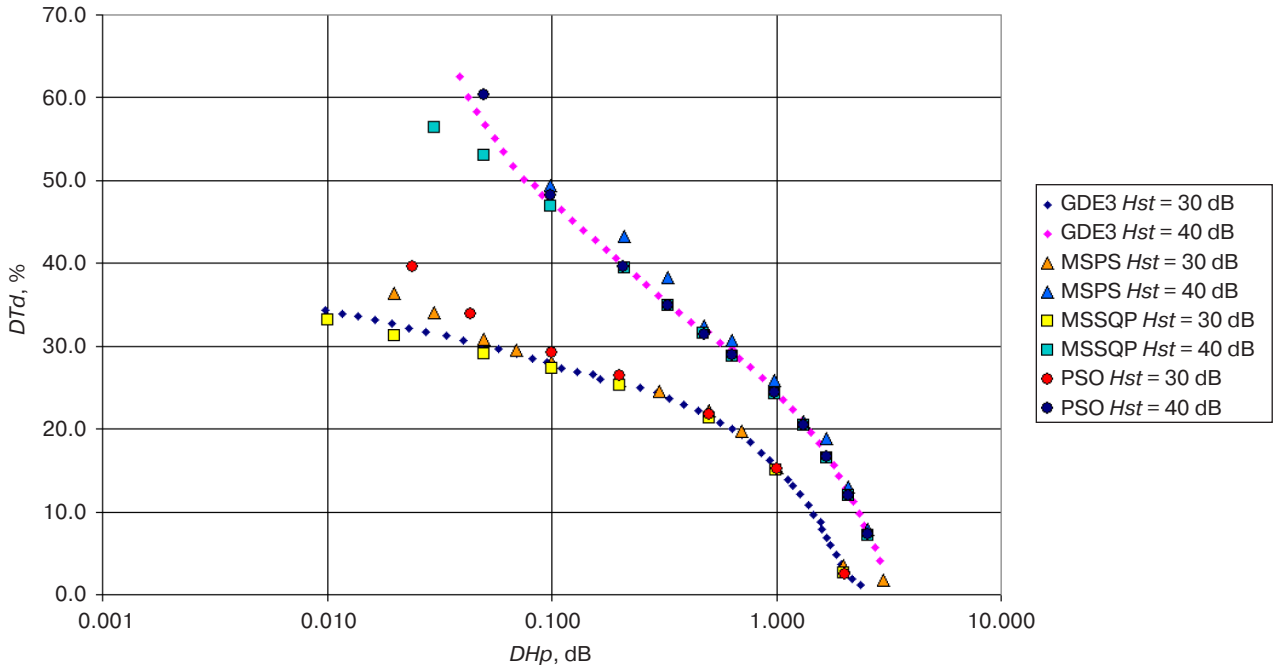


Fig. 1. Problem 1 solution

Problem 2. We shall consider experimental results for the same LPF as in Problem 1. The results obtained using the compared algorithms are shown in Fig. 2. The constraints are set to $Tfr \leq 0.5$, $Um \leq 1.1$, and $DHt \leq 0$ for the first series of experiments, while the constraint on *Tfr* is excluded in the second series.

First, it should be noted that the MSSQP algorithm cannot find admissible solutions, i.e., those satisfying all the restrictions, under the constraint on *Tfr* at all. In the absence of this constraint, the ability of MSSQP to find valid solutions appeared to be noticeably worse than that for GDE3 and PSO. The number of valid solutions in series of $NT = 200$ starts taking 40–60 s is measured in units. In both series of experiments, the MSPS algorithm also shows results significantly inferior to the GDE3 and PSO solutions.

In both series of experiments, the GDE3 and PSO algorithms show similar results in terms of QI. The GDE3 parameters are $Npop = 50$ and $Neval = 200000$, respectively, while finding an approximation of the POS

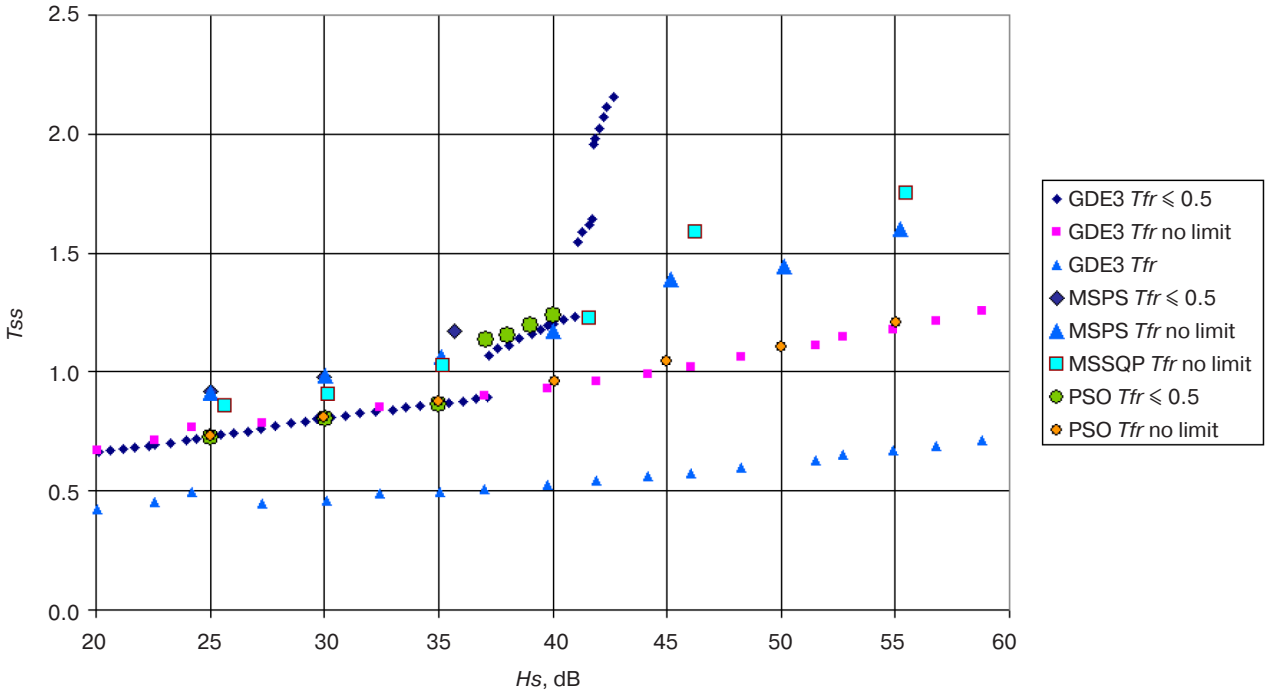


Fig. 2. Problem 2 solution

set takes about 7 min. For PSO, the number of starts is $NT = 40$; the duration of finding one solution is 8–10 min.

The Pareto front turns out to be discontinuous in the presence of a constraint on *Tfr*. To explain the reasons for this effect, the *Tfr* values obtained with no constraint imposed on this QI are plotted on the graph also shown in Fig. 2. As long as *Tfr* is less than constraint $Tfr = 0.5$, approximations of Pareto fronts obtained in the presence and absence of this constraint coincide. The constraint on *Tfr* is not active and does not affect search results. However, if *Tfr* value should exceed the specified threshold, the constraint becomes active, and the transition process becomes oscillatory in order to fulfill it. In this case, the Pareto front discontinuity is due to transitions of the moment when the condition of the transient process completion [5] from one wave to another is satisfied.

Problem 3. Experimental results for LPF with $NP = 4$ for pole and $NZ = 4$ for zero are shown in Fig. 3. The upper bound frequency of the passband on the frequency scale normalized to the sampling frequency is $Fp = 0.1$; the lower bound frequency of the stopband is $Fs = 0.2$. Using the GDE3 population-based algorithm, approximations of the Pareto front for problems of minimizing QI *DHp* and *DTd* at $Npop = 50$, under constraints $Hst = 30$ dB and $Hst = 40$ dB. The numbers of QI calculations are $Neval = 10^6$ and $Neval = 1.5 \cdot 10^6$, the duration is 6 and 9 min, respectively.

Next, points of these approximations are obtained using algorithms based on scalarizing OFs. The MSPS algorithm gives the best results, since the solutions found by it cover the ranges of solutions obtained using GDE3 completely, not only equalling, but even slightly exceeding them in QI terms. In case of $Hst = 30$ dB, the set of MSPS solutions has the least low bound in terms of *DHp* parameters. In all experiments, the number of search starts is $NT = 2000$. The execution time per search ranges from 5 to 8.5 min.

The PSO algorithm for Problem 3 turns out to be worse than MSPS; here, the solutions found are within narrower ranges of QI values taking approximately the same time required for one search.

The MSSQP algorithm also loses to GDE3 and MSPS by the value of the lower bound of the *DHp* range. However, at $HSt = 30$ dB, the algorithm finds solutions within the *DHp* value range of 0.2–0.8 dB with lower *DTd* values compared to other algorithms. These solutions are located in the area of the search space into which other algorithms have not fallen. At the same time, the phase-frequency response of the solutions obtained using MSSQP differ from the phase-frequency response of the other algorithm solutions (Fig. 4) significantly, although the gain-frequency response is similar. At the same time, MSSQP results are close to those obtained by other methods at $Hst = 40$ dB.

Problem 4. We shall consider an example of the problem solution at matching the lines with the ratio of wave impedances $Z_{w2}/Z_{w1} = 10$, frequency matching bandwidth $DFM = 1.2$, and the number of transformer stages $Nst = 4$ and $Nst = 5$ (Fig. 5).

The search by GDE3 algorithm for both values of *Nst* is performed at $Npop = 100$ and $Neval = 1 \cdot 10^6$, taking approximately 2 minutes. Since the results only insignificantly deteriorate with the number of QI calculations, the search duration is decreased by 10 times.

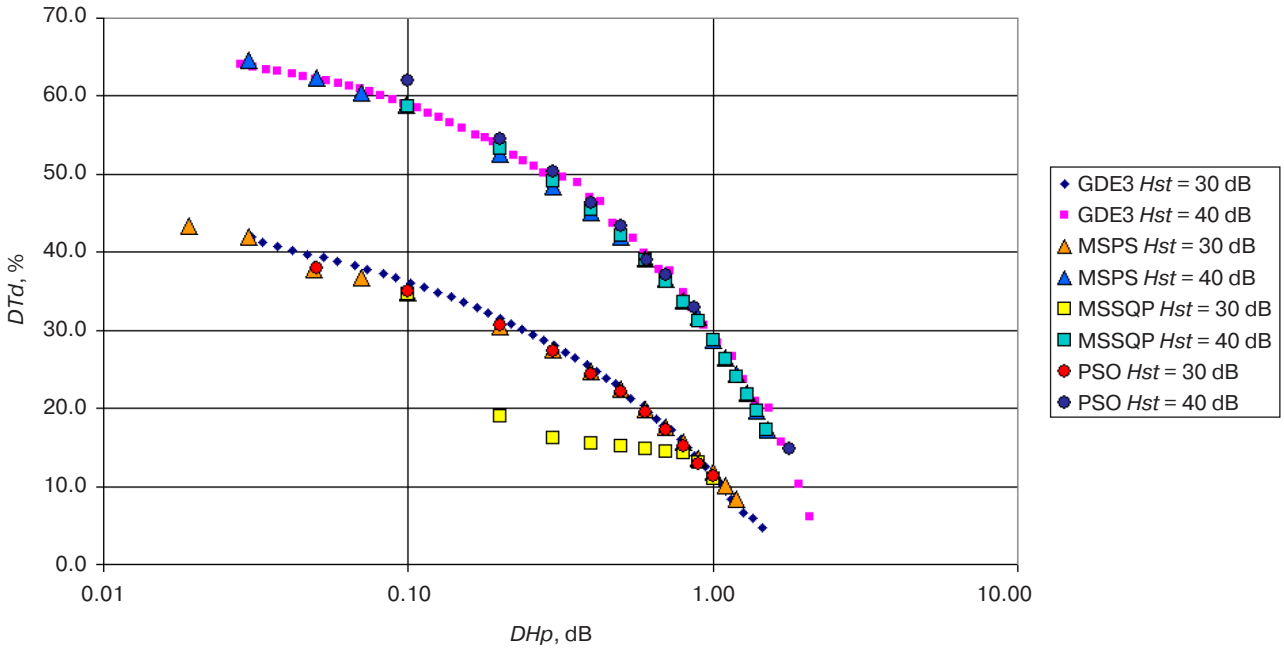


Fig. 3. Problem 3 solution

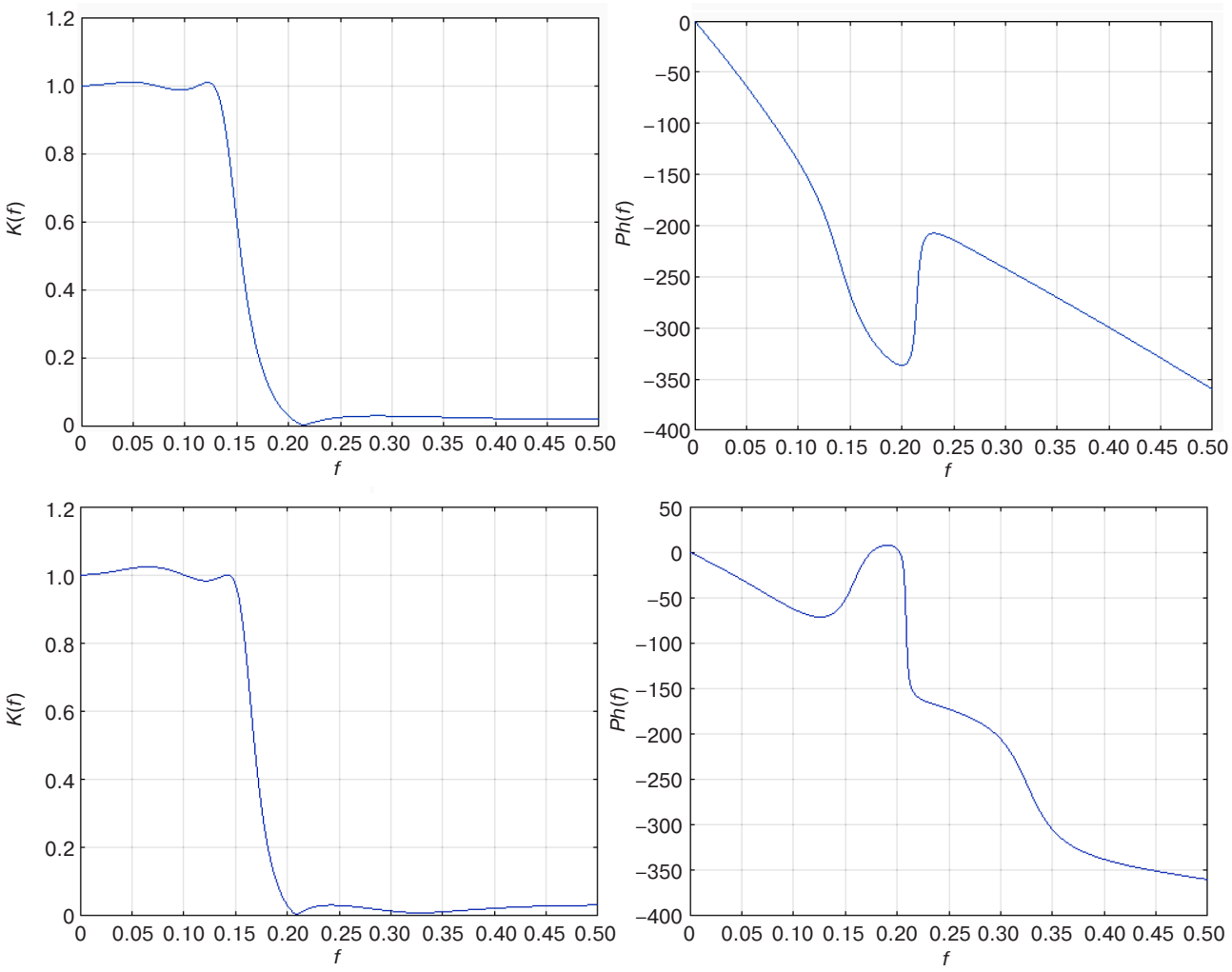


Fig. 4. Comparison of the gain-frequency response $K(f)$ and phase-frequency response $Ph(f)$ of filters found using MSPS algorithm (top) and MSSQP algorithm (bottom) at HSt = 30 dB and $DHp \leqslant 0.2$ dB

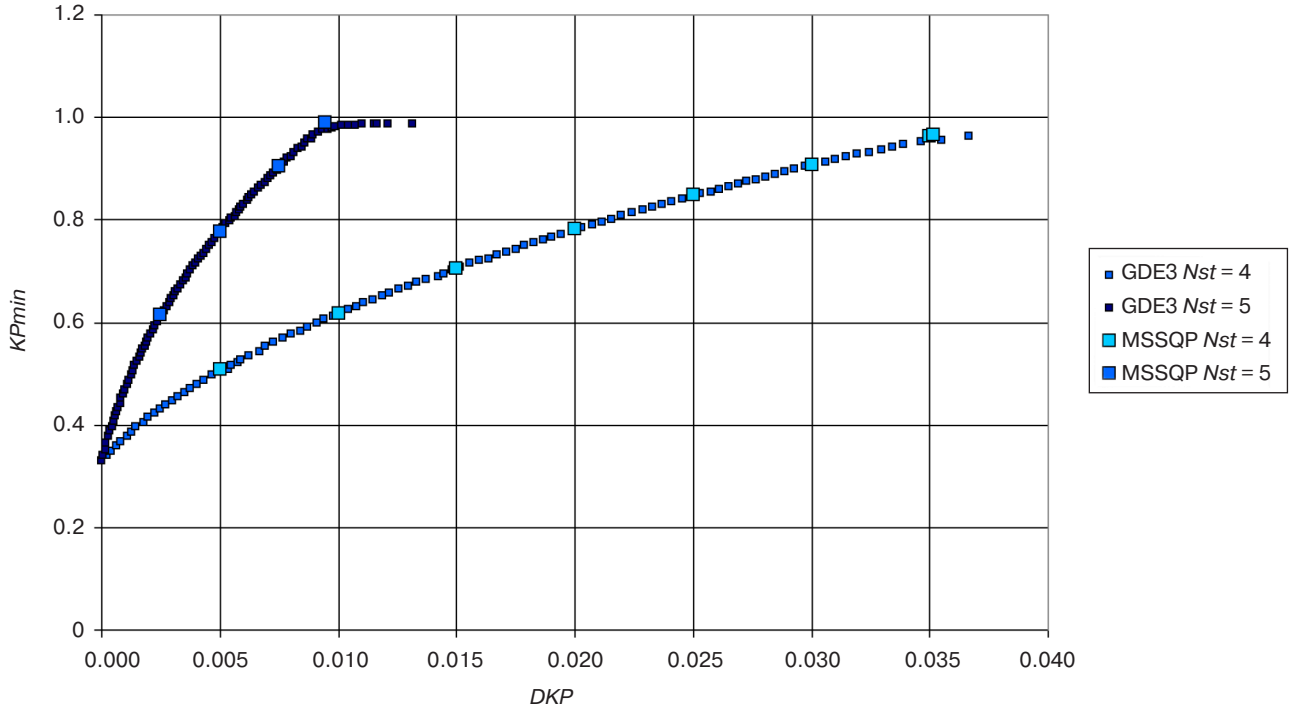


Fig. 5. Problem 4 solution

All three algorithms based on scalarizing OFs show the same results. Therefore, only the solutions obtained with MSSQP falling on Pareto front approximations found by GDE3 algorithm are shown in Fig. 5. However, the methods differ significantly in terms of search duration for a single solution. While MSPS and PSO require 40–50 s, MSSQP requires only 3–4 s. It should be noted that the rightmost points of the series obtained using MSSQP coincide with the results for Chebyshev approximations, while the points of GDE3 series located to their right are not POS.

Problem 5. The problem is solved under conditions Z_{w2}/Z_{w1} = 12, *DFM* = 1, and *KPt* = 1 for the number of stages *Nst* = 3, 4, and 5. Approximations of Pareto fronts obtained by the compared algorithms are shown in Fig. 6.

For GDE3 algorithm, parameters *Npop* = 100 and *Neval* = $1 \cdot 10^5$ are set. No further increase in these parameters has any positive effect. With increasing number of steps *Nst*, the search duration increases within the range from 55 to 71 s. The upper bound of *DKP* values is set to 0.1. At *Nst* = 5, the Pareto front approximation turns out to be discontinuous within the range *DKP* > 0.04. This is due to the *DTd* value varying insignificantly within this range, thus making it difficult to estimate the solution dominances.

For the MSPS algorithm, *NT* = 200. The average search duration per solution for three *Nst* values is 7, 14, and 20 s. In all cases, the found solutions are significantly worse than those obtained using GDE3; moreover, increasing the number of *NT* starts yields no improvement. Since the PSO algorithm gives solutions coincident with MSPS solutions at close durations, its results are not included in Fig. 6.

The MSSQP algorithm demonstrates high efficiency in solving this problem. If *NT* = 10, then 60–100% of starts yield the same result matching GDE3 solutions over the entire range of values. The other starts result in unacceptable solutions with constraint violations. The average search duration for three *Nst* values is 1.6, 5.2, and 7.5 s. At the same time, the algorithm also finds solutions with the given DKP values in the area wherein the Pareto front approximation obtained using GDE3 has turned out to be discontinuous.

CONCLUSIONS

The study demonstrates that the GDE3 MOO population-based algorithm can be used to find a solution for all of the considered problems and can therefore be recommended for use in solving of such types of MOO problems to obtain an approximation of the POS set across a wide range of QIs. Thus, it is reasonable to check the applicability of the MSSQP algorithm to a particular problem, as well as the possibility of obtaining solutions superior in quality to those obtained by GDE3. However, due to the lack of answers to questions why MSSQP algorithm is effective for some problems and unsuitable for others, as well as how it finds solutions for Problem 3 inaccessible to other algorithms, further experimental

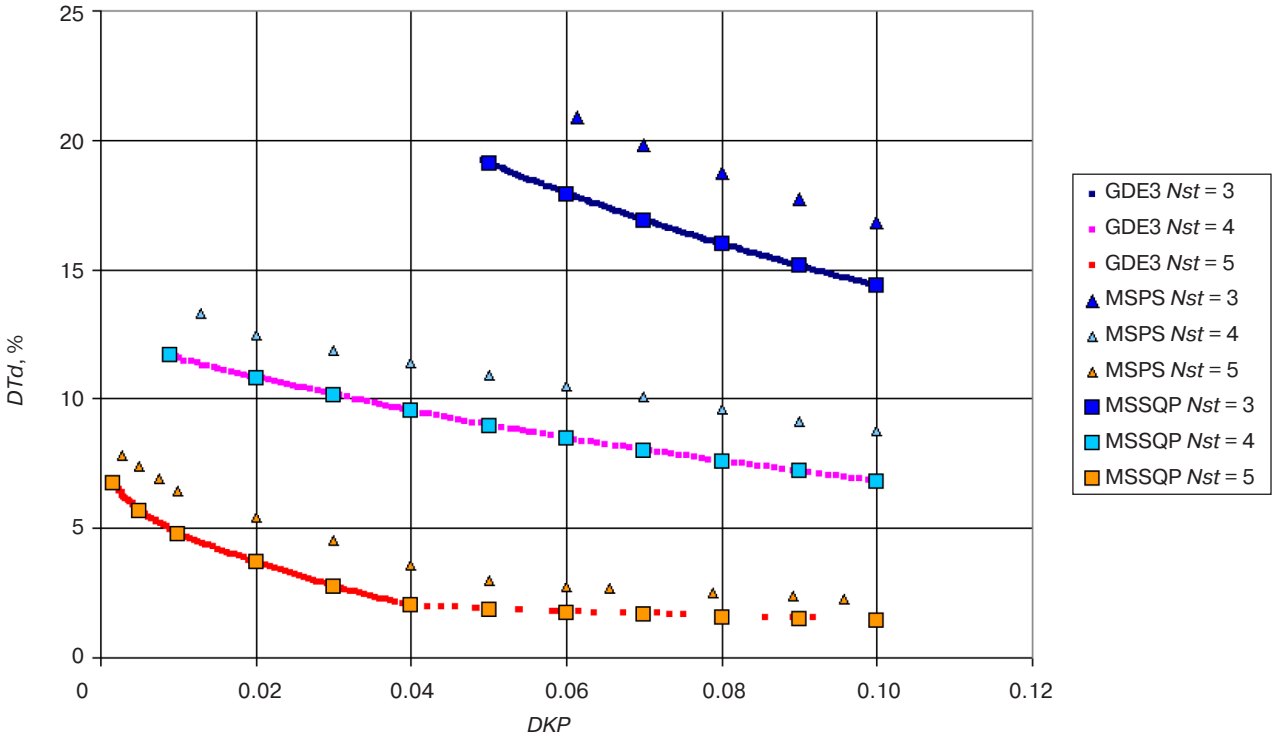


Fig. 6. Problem 5 solution

verification is required. For this it will be necessary to analyze the impact of landscape features of individual QI in MOO problems on the solution finding process. Despite active investigations in recent years, including the use of intelligent technologies [14, 15], in the field continues to be characterized by a lack of sufficiently general results.

The MSSQP algorithm (or, in case it cannot find suitable solutions, other algorithms based on scalarizing OFs) can be recommended for use in cases where it is necessary to find a small number of POS or to provide accurate values for part of QI, which is difficult when using population-based MOO algorithms.

REFERENCES

1. Gutkin L.S. *Optimizatsiya radioelektronnykh ustroistv po sovokupnosti pokazatelei kachestva* (*Optimization of Radio Electronic Devices with Aggregation of Quality Indices*). Moscow: Sovetskoe radio; 1975. 368 p. (in Russ).
2. Chernorutskii I.G. *Metody optimizatsii v teorii upravleniya: uchebnoe posobie* (*Optimization Methods in the Control Theory*). St. Petersburg: Piter; 2004. 256 p. (in Russ).
3. Karpenko A.P. *Sovremennye algoritmy poiskovoi optimizatsii. Algoritmy, vdokhnovlennye prirodoi:* uchebnoe posobie (*Modern Search Optimization Algorithms. Nature-Inspired Optimization Algorithms*). Moscow: Izd. MGTU im. Baumana; 2017. 446 p. (in Russ).
4. Jasbir S.A. *Introduction to optimum design*. 4th edition. Elsevier; 2017. 670 p. https://doi.org/10.1016/C2013-0-15344-5
5. Smirnov A.V. Optimization of digital filters performances simultaneously in frequency and time domains. *Russ. Technol. J*. 2020;8(6):63–77 (in Russ.). https://doi.org/10.32362/2500-316X-2020-8-6-63-77
6. Liu S., Lin Q., Tan K.Ch., Li Q. Benchmark problems for CEC2021 competition on evolutionary transfer multiobjective optimization. Technical Report. 2021. 24 p. https://doi.org/10.48550/arXiv.2110.08033
7. Ray T., Mamun M.M., Singh H.K. A simple evolutionary algorithm for multi-modal multiobjective optimization. 2022. 8 p. Available from URL: https://arxiv.org/pdf/2201.06718.pdf
8. Wang Z., Yao X. An efficient multi-indicator and many-objective optimization algorithm based on two-archive. 2022. 15 p. Available from URL: https://arxiv.org/pdf/2201.05435v1.pdf
9. Chen T., Li M. The weights can be harmful: Pareto search versus weighted search in multi-objective search-based software engineering. 2022. 40 p. Available from URL: https://arxiv.org/pdf/2202.03728.pdf
10. Smirnov A.V. Application of population algorithms in the problems of multiobjective optimization of electrical filters characteristics. *Modelirovanie, optimizatsiya i informatsionnye tekhnologii = Modeling, Optimization and Information Technology.* 2021;9(3) (in Russ.). https://doi.org/10.26102/2310-6018/2021.34.3.015
11. Tian Y., Cheng R., Zhang X., Jin Y. PlatEMO: A MATLAB platform for evolutionary multi-objective optimization. *IEEE Computational Intelligence Magazine*. 2017;12(4):73–87. https://doi.org/10.1109/MCI.2017.2742868

СПИСОК ЛИТЕРАТУРЫ

1. Гуткин Л.С. *Оптимизация радиоэлектронных устройств по совокупности показателей качества*. М.: Советское радио; 1975. 368 с.
2. Черноруцкий И.Г. *Методы оптимизации в теории управления*: учебное пособие. СПб.: Питер; 2004. 256 с.
3. Карпенко А.П. *Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой*: учебное пособие. М.: Изд-во МГТУ им. Н.Э. Баумана; 2017. 446 с.
4. Jasbir S.A. *Introduction to optimum design*. 4th edition. Elsevier; 2017. 670 p. https://doi.org/10.1016/C2013-0-15344-5
5. Смирнов А.В. Оптимизация характеристик цифровых фильтров одновременно в частотной и временной областях. *Российский технологический журнал*. 2020;8(6):63–77. https://doi.org/10.32362/2500-316X-2020-8-6-63-77
6. Liu S., Lin Q., Tan K.Ch., Li Q. Benchmark problems for CEC2021 competition on evolutionary transfer multiobjective optimization. Technical Report. 2021. 24 p. https://doi.org/10.48550/arXiv.2110.08033
7. Ray T., Mamun M.M., Singh H.K. A simple evolutionary algorithm for multi-modal multi-objective optimization. 2022. 8 p. https://arxiv.org/pdf/2201.06718.pdf
8. Wang Z., Yao X. An efficient multi-indicator and many-objective optimization algorithm based on two-archive. 2022. 15 p. URL: https://arxiv.org/pdf/2201.05435v1.pdf
9. Chen T., Li M. The weights can be harmful: Pareto search versus weighted search in multi-objective search-based software engineering. 2022. 40 p. URL: https://arxiv.org/pdf/2202.03728.pdf
10. Смирнов А.В. Применение популяционных алгоритмов в задачах многокритериальной оптимизации характеристик электрических фильтров. *Моделирование, оптимизация и информационные технологии.* 2021;9(3). https://doi.org/10.26102/2310-6018/2021.34.3.015
11. Tian Y., Cheng R., Zhang X., Jin Y. PlatEMO: A MATLAB platform for evolutionary multi-objective optimization. *IEEE Computational Intelligence Magazine*. 2017;12(4):73–87. https://doi.org/10.1109/MCI.2017.2742868
12. Смирнов А.В. Свойства целевых функций и алгоритмов поиска в задачах многокритериальной оптимизации. *Russ. Technol. J.* 2022;10(4):75−85. https://doi.org/10.32362/2500-316X-2022-10-4-75-85

12. Smirnov A.V. Properties of objective functions and search algorithms in multi-objective optimization problems. *Russ. Technol. J*. 2022;10(4):75−85 (in Russ.). https://doi.org/10.32362/2500-316X-2022-10-4-75-85
13. Smirnov A.V. Method of simultaneous optimization of radio devices performance in frequency and time domain. *Russ. Technol. J*. 2018;6(6):13–27 (in Russ.). https://doi.org/10.32362/2500-316X-2018-6-6-13-27
14. Mersmann O., Bischl B., Trautmann H., Preuss M., Weihs C., Rudolf G. Exploratory landscape analysis. In: *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation* (*GECCO'11*). 2011. P. 829–836. https://doi.org/10.1145/2001576.2001690
15. Trajanov R., Dimeski S., Popovski M., Korošec P., Eftimov T. Explainable landscape-aware optimization performance prediction. In: *2021 IEEE Symposium Series on Computational Intelligence* (*SSCI*). 2021. https://doi.org/10.1109/SSCI50451.2021.9660124

13. 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
14. Mersmann O., Bischl B., Trautmann H., Preuss M., Weihs C., Rudolf G. Exploratory landscape analysis. In: *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation* (*GECCO'11*). 2011. P. 829–836. https://doi.org/10.1145/2001576.2001690
15. Trajanov R., Dimeski S., Popovski M., Korošec P., Eftimov T. Explainable landscape-aware optimization performance prediction. In: *2021 IEEE Symposium Series on Computational Intelligence* (*SSCI*). 2021. https://doi.org/10.1109/SSCI50451.2021.9660124

About the author

Alexander V. Smirnov, Cand. Sci. (Eng.), Professor, Department of Telecommunications, Institute of Radioelectronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: av_smirnov@mirea.ru. https://orcid.org/0000-0002-2696-8592

Об авторе

Смирнов Александр Витальевич, к.т.н., доцент, профессор кафедры телекоммуникаций Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: av_smirnov@mirea.ru. https://orcid.org/0000-0002-2696-8592

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Modern radio engineering and telecommunication systems

Современные радиотехнические и телекоммуникационные системы

UDC 621.396.66
https://doi.org/10.32362/2500-316X-2022-10-6-52-59

RESEARCH ARTICLE

Protection of battery-powered devices against accidental swap of power supply connections

Valery P. Babenko,
Vladimir K. Bityukov @

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

Abstract

Objectives. Battery-powered devices (e.g., wireless sensors, pacemakers, watches and other wrist-worn devices, virtual reality glasses, unmanned aerial vehicles, robots, pyrometers, cars, DC/DC converters, etc.) are widely used today. For such devices, it is highly important to ensure safe primary power supply connection, including protection against reverse polarity. The conventional solution to the reverse polarity problem, involving the use of Schottky diodes during system redundancy or increasing power by combining two or more power supplies in the OR-ing circuit due to a large voltage drop, results in significant power losses at high currents, heat dissipation problems, and an increase in the mass and size of the equipment. For this reason, it becomes necessary to develop efficient battery-powered equipment protection against incorrect reverse polarity connection.

Methods. The problem is solved using circuit simulation in the *Electronics Workbench* environment.

Results. When protecting equipment against reverse voltage polarity, it is shown that the minimum level of losses and low voltage drop are provided by "ideal diode" circuit solutions based on discrete components and microcircuits of the "integrated diode" type with external and internal power metal–oxide–semiconductor field-effect transistors (MOSFETs). The circuit simulation of ideal diodes based on *p*- and *n*-channel transistors with superior technical parameters allows the characteristics and voltage and power losses in the protected circuits to be specified along with a presentation of the proposed technical solution simplicity. The contemporary component base of protection devices is discussed in terms of efficiency.

Conclusions. Examples of equipment for protecting against reverse voltage polarity are given along with circuit solutions based on discrete and integrated components. The simulation of the transfer characteristics of protection devices shows the limit for the minimum input voltage value of around 4 V using a MOSFET transistor.

Keywords: protection, battery power, MOSFET, Schottky diode, parasitic diode, ideal diode, connection, voltage, reverse polarity

• **Submitted:** 02.02.2022 • **Revised:** 31.03.2022 • **Accepted:** 05.09.2022

For citation: Babenko V.P., Bityukov V.K. Protection of battery-powered devices against accidental swap of power supply connections. *Russ. Technol. J*. 2022;10(6):52–59. https://doi.org/10.32362/2500-316X-2022-10-6-52-59

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

Защита аппаратуры с батарейным питанием от ошибочного подключения напряжения обратной полярности

В.П. Бабенко,
В.К. Битюков [@]

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

Резюме

Цели. В настоящее время широкое применение находит аппаратура с батарейным питанием (беспроводные датчики, кардиостимуляторы, «умные» браслеты, очки виртуальной реальности, беспилотные летающие аппараты, роботы, пирометры, автомобили, DC/DC преобразователи и др.). Для этих устройств принципиально важным вопросом является безопасное подключение первичных источников электропитания и наличие защиты от напряжения обратной полярности. Традиционное решение проблемы «переполюсовки» (подачи на прибор напряжения питания обратной полярности) с использованием диодов Шоттки при резервировании системы или увеличения мощности путем объединения источников питания по схеме ИЛИ вследствие большого падения напряжения приводит к значительным потерям мощности при больших токах, сложной проблеме теплоотвода и увеличению массогабаритных параметров. Это предопределило реализацию эффективных средств защиты аппаратуры с батарейным питанием от ошибочного подключения напряжения обратной полярности.

Методы. Задача решена с использованием схемотехнического моделирования в среде *Electronics Workbench*.

Результаты. Показано, что минимальный уровень потерь и малое падение напряжения при защите аппаратуры от обратной полярности питающего напряжения обеспечивают схемные решения «идеального диода» на дискретных компонентах и микросхемы типа «интегрального диода» с внешним и внутренним силовым транзистором MOSFET. Схемотехническое моделирование «идеальных диодов» на *p*- и *n*-канальных транзисторах, которые отличаются высокими техническими параметрами, позволило уточнить характеристики, потери напряжения и мощности в защищаемых цепях и показать простоту непосредственно самого технического решения. В статье обсуждены вопросы эффективности и современная элементная база устройств защиты.

Выводы. Приведены примеры элементной базы устройств защиты от «переполюсовки» источников питания, варианты защиты аппаратуры от воздействия напряжения обратной полярности, а также схемотехнические решения на дискретных и интегральных компонентах. Моделирование передаточных характеристик устройств защиты показало ограничение на минимальную величину входных напряжений около 4 В, обусловленную используемым MOSFET транзистором.

Ключевые слова: защита, батарейное электропитание, MOSFET, диод Шоттки, паразитный диод, «идеальный диод», подключение, напряжение, обратная полярность

• **Поступила:** 02.02.2022 • **Доработана:** 31.03.2022 • **Принята к опубликованию:** 05.09.2022

Для цитирования: Бабенко В.П., Битюков В.К. Защита аппаратуры с батарейным питанием от ошибочного подключения напряжения обратной полярности. *Russ. Technol. J.* 2022;10(6):52–59. https://doi.org/10.32362/2500-316X-2022-10-6-52-59

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

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

In order to provide portability, functionality, and usability of electronic equipment (wireless sensors, pacemakers, "smart" bracelets, virtual reality glasses, unmanned aerial vehicles, robots, pyrometers, cars, DC/DC converters, etc.), a self-contained power supply is required [1–6].

Since battery-powered devices may be damaged by improper connection to the primary power supply, it is advisable to provide protection against reverse polarity. Although the battery compartments and contacts of the battery-powered equipment are typically designed in such a way as to exclude the possibility of misplacing the power supply element, however, it is still possible that battery terminals in automotive electronics, mobile electric vehicles, telecommunications servers, storage systems, server infrastructure equipment, and disk batteries may be misconnected. With polarity reversal or transients occurring during switching of inductive loads, serious failures and damage of electronic systems and units may occur. When the batteries of electric vehicles are incorrectly connected, reverse polarity can be dangerous due to possible generation of significant currents of tens to hundreds of amperes for sustained periods of time.

The use of circuit solutions and auxiliary electronic components to provide protection against power supply with wrong polarity in such cases forms the subject of the analysis presented in this paper.

In order to better understand the processes and factors affecting the characteristics of devices for protection against reverse polarity, circuits are simulated in the *Electronics Workbench* (*EWB*) environment [7]. In addition to conventional Simulation Program with Integrated Circuit Emphasis (SPICE) analysis, *EWB* allows the connection of virtual controls and measuring devices to the investigated circuit, which closely approximate real devices. Along with a significant set of methods for analyzing various characteristics of electronic circuits, *EWB* offers an extensive built-in library of analog and digital electronic components, including powerful *n*- and *p*-channel metal–oxide–semiconductor field-effect transistors (MOSFETs) from International Rectifier (USA) and Zetex Semiconductors (United Kingdom), powerful *p-n* junction diodes, and Schottky diodes (Motorola, USA). Considerable experience of using *EWB* in various fields of analog and digital electronics has been accumulated in extensive literature [7, 8]. *EWB* provides convenient use of two simulation methods supported by the program:

- electronic laboratory simulation implying virtual measuring instruments similar to real devices are installed in the electronic circuit while simulation is started by the Activate Simulation switch on the work panel;
- quasi-professional simulation implying the analysis type is set from the Analysis menu, in which window the analysis type and scheme nodes for which the simulation result is viewed are set.

DIODE PROTECTION AGAINST REVERSE VOLTAGE

The simplest methods of protecting equipment against reverse voltage involve using diode VD1 (Fig. 1a) connected in series with load R1 [9].

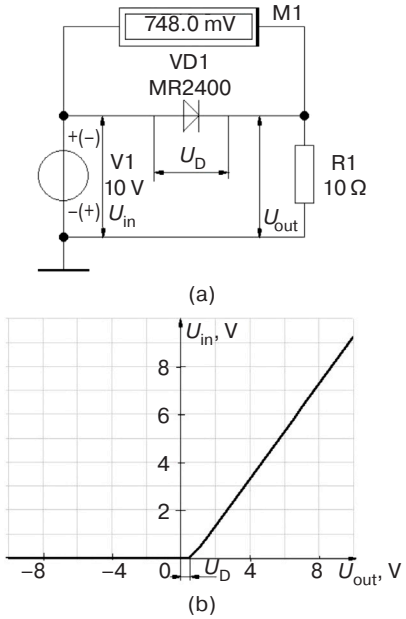


Fig. 1. Diode protection against reverse voltage: (a) circuit, (b) transfer characteristic. MR2400—diode (Infineon Technologies AG, Germany). Here and in following figures, the designations of circuit elements correspond to the designation system adopted by GOST 2.710-81[1]

The transfer characteristic (the dependence of output voltage U_{out} on load R1 on input voltage U_{in}) is shown in Fig. 1b. For convenient estimation of results during simulation in the Analysis/DC Sweep mode, the voltage source V1 determining input voltage U_{in} is varied within the range from −10 V to +10 V. Voltmeter M1 shows voltage drop U_D across diode VD1 in forward direction U_D = 0.748 V with a load current I_1 of about 1 A flowing through it. In such a circuit, a power loss of 0.748 W in the protection diode is extremely undesirable for battery-powered devices. Although Schottky diodes slightly improve the diode protection characteristics at voltage drop U_D = 0.3…0.4 V, even such a small voltage drop may be unacceptable for low supply voltages of

[1] GOST 2.710-81. Unified system for design documentation. Alpha-numerical designations in electrical diagrams. Moscow: Izd. Standartov; 1985 (in Russ.).

the order of 3.3 V typically used for powering modern integrated circuits (ICs) and microcontrollers.

In some cases, where special protection IC against reverse polarity is not available, discrete component circuits using cheap and available MOSFETs having low channel resistance in the open state may be applied. These circuits are often called the "ideal diode" or sometimes "smart diode" for their characteristics[2].

IDEAL DIODE BASED ON *p*-MOSFET

The *p*-channel MOSFET circuit performing the function of protection diode is shown in Fig. 2a.[3]

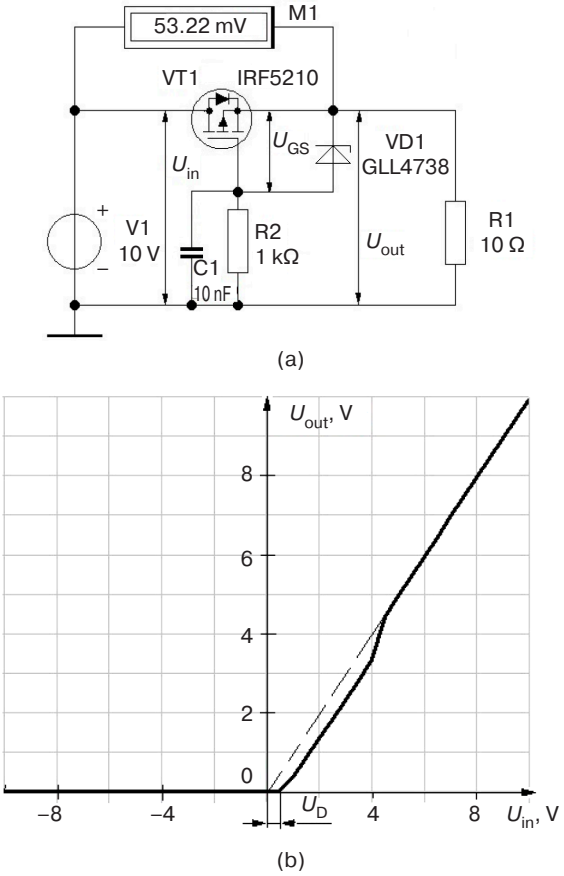


Fig. 2. Protection against reverse polarity based on *p*-MOSFET: (a) circuit, (b) transfer characteristic. GLL4738—zener diode (Vishay, USA)

In the circuit, the drain of transistor VT1 is connected to the positive voltage of the input voltage source V1. Prior to the application of voltage U_{in}, the transistor channel is clamped due to the gate and the source having ground potential along with the voltage U_{GS} between gate and source being zero. When positive input voltage U_{in} is applied, current flows through forward biased parasitic diode (body diode) in the MOSFET system and load R1 resulting in a voltage drop close to U_{in}. The parasitic diode is open until the transistor VT1 channel opens (this occurs at voltage U_{in} about 4.5 V). Due to its low resistance, the open channel shunts the parasitic diode to provide a small voltage drop in it, as can be seen from the transfer characteristic shown in Fig. 2b [10, 11]. Voltmeter M1 measures voltage drop U_{DS} = 53.22 mV across the open transistor. At load current I_l = 1 A, this corresponds to open transistor channel resistance $r_{DS\ ON}$ = 0.053 Ω being close to the value for the IRF5210 transistor $r_{DS\ ON}$ = 0.06 Ω. When reversing the polarity of input voltage U_{in} generated by source V1, transistor VT1 clamps and blocks the load current (second quadrant of the characteristic shown in Fig. 2b).

Zener diode VD1 protects transistor VT1 against exceeding the permissible gate-source voltage. The value of stabilization voltage is selected equal to U_{st} = 9...10 V, so that the voltage is sufficient for reliable opening of transistor VT1. Capacitor C1 smooths out negative voltage surges across the possible load occurring when the input voltage polarity reverses rapidly.

If input voltage U_{in} does not exceed the permissible voltage of the MOSFET switch (typically about 20V), then protection against exceeding the permissible voltage is not required. The circuit may be simplified by excluding VD1, C1, and R2 with the transistor gate being connected to ground.

In normal operation, the minimum input voltage from which the circuit has minimal losses similar to that of an ideal diode occurs with the unlocking transistor voltage (about 4.5 V). At a lower voltage, the transistor channel is clamped while only the parasitic diode remains open, while the voltage drop across the switch is about U_D = 0.7 V (Fig. 2b). An example of the automotive electronic equipment protection against reverse voltage polarity is given in the relevant literature.[4]

The ideal-diode *p*-MOSFET circuit may be attractive due to its simplicity, low voltage drops across the transistor switch under normal operation and current blocking occur in case of reverse polarity. However, other things being equal, *p*-channel MOSFET is inferior to its *n*-channel analogs in terms of such parameters as open channel resistance, maximum current, input capacitance, and cost.

IDEAL DIODE BASED ON *n*-MOSFET

A similar protection circuit against reverse polarity may be implemented on the basis of *n*-MOSFET provided it is connected to the negative terminal of voltage source V1 as shown in Fig. 3a [12].

[2] *Basics of Ideal Diodes*. Application Note. Texas Instruments Incorporated. SLVAE57B – FEBRUARY 2021. 24 p. https://www.ti.com/lit/an/slvae57b/slvae57b.pdf?ts=1639001451460. Accessed January 10, 2022.

[3] https://www.terraelectronica.ru/news/5444. Accessed January 10, 2022 (in Russ.).

[4] https://www.terraelectronica.ru/news/5446. Accessed January 10, 2022.

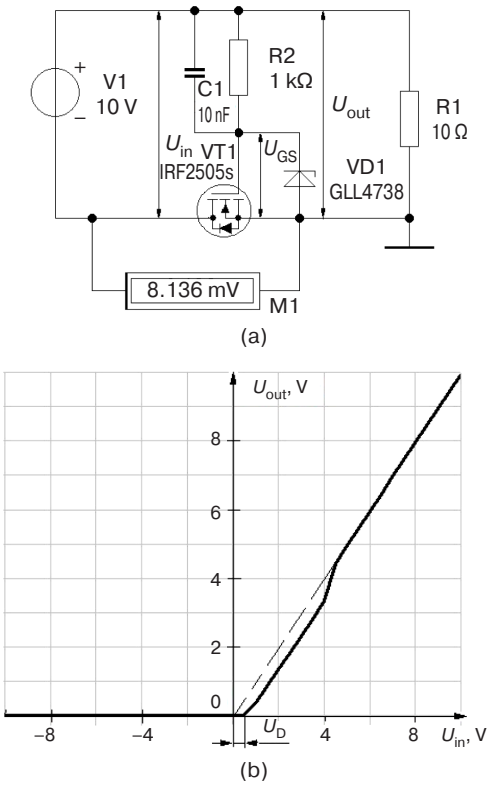


Fig. 3. Protection against reverse polarity based on *n*-MOSFET: (a) circuit, (b) transfer characteristic

As in circuit above, the channel of transistor VT1 is clamped before applying input voltage U_{in}, since U_{GS} = 0. When positive input voltage U_{in} is applied, the parasitic diode of transistor VT1 opens, and current flows through load resistance R1. As can be seen from the transfer characteristic (Fig. 3b), the channel of transistor VT1 becomes conductive at voltage U_{in} about +4.5 V. The voltage drop across the open channel of the transistor becomes minimal at 8.136 mV. At load current about 1 A, this corresponds to the open channel resistance $r_{DS\ ON}$ = 8.136 mΩ being close to the Datasheet[5] data for the IRF2505S transistor $r_{DS\ ON}$ = 0.008 Ω. The scattered power across transistor VT1 is insignificant being about 8 mW.

When the power supply is reverse polarized, transistor VT1 clamps and blocks the reverse current (second quadrant of the transfer characteristic, shown in Fig. 3b). As in the previous circuit, zener diode VD1 protects against exceeding the permissible voltage of transistor VT1. The capacitor C1 ≈ 10 nF is necessary to smooth the negative voltage surge at the output possible of occurring at the moment of reversing polarity of the input voltage and damaging electronic components. Increasing the capacitor C1 capacity, the voltage rise time across the gate voltage may be increased, thus allowing implementing the "soft start" function.

When input voltage U_{in} is less than the voltage permissible for transistor, the overvoltage protection circuit between gate and source may be excluded by eliminating elements VD1, C1, and R2. In this case, the transistor gate is connected to the power supply "plus."

Thus, the *n*-MOSFET circuit provides the most effective protection against reverse polarity. However, it may be someway inconvenient to build a large star ground system due to the protection being included in the ground bus.

INTEGRATED IDEAL DIODE

Electronic component manufacturers offer a fairly wide range of high-efficiency ideal diode solutions. This may be exampled by the LM74610-Q1[6] integrated controller (Texas Instruments, USA) with external *n*-MOSFET VT1 performing the ideal diode functions; its connection circuit is shown in Fig. 4. With the correct (positive) input voltage polarity, the transistor opens to pass current. Here, power dissipation is minimal due to low open channel resistance. In the event of reverse polarity voltage being applied to the circuit input, the controller shuts down the transistor in less than 8 µs. Like a common diode, an ideal diode is connected to the power line via the "Anode" and "Cathode" pins only. The controller has no connection to the common output, thus providing zero self-consumption current. The charge pump circuit with external capacitor V_{cap} is used for controlling the external power transistor.

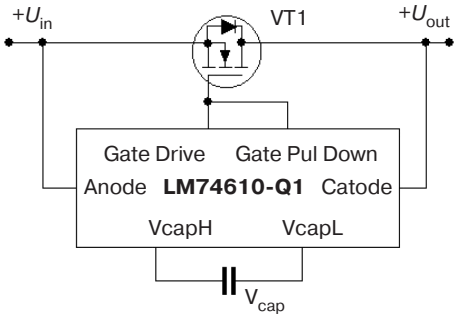


Fig. 4. Connection circuit for the LM74610-Q1 ideal diode

The controller, which is designed for a wide range of automotive applications, is capable of withstanding reverse voltages up to 45 V. Values for current in the circuit and power dissipation are determined by the external transistor characteristics.

[5] Datasheet IRL2505S International Rectifier. https://www.alldatasheet.com/datasheet-pdf/pdf/84685/IRF/IRL2505S.html. Accessed January 10, 2022.

[6] Datasheet LM74610-Q1 Texas Instruments. https://www.alldatasheet.com/datasheet-pdf/pdf/810348/TI1/LM74610-Q1.html. Accessed January 10, 2022.

The LTC4411[7] and LTC4412[8] ideal diodes from Linear Technology (USA) shown in Fig. 5 are oriented towards using *p*-channel MOSFET. However, the LTC4411 model (Fig. 5a) contains a built-in transistor with open state channel resistance $r_{DS\ ON}$ = 0.14 Ω, while the LTC4412 variant has an external transistor, thus extending the range of currents it can control.

In normal operation of ideal diodes, the voltage drop across transistor may be up to 28 mV at a current of no more than 2.6 A with input voltages ranging from +2.6 to +5.5 V. The IC chip has a thermal protection device for blocking the current should the permissible temperature be exceeded. The transistor also switches off when the output voltage exceeds the input voltage with the return current not exceeding 1 μA. The CTL pin can be used to control on/off in response to external commands. The status output STAT indicates the redundant state where power is being supplied from an alternate source, there is voltage across the load, and no current flowing from the main power supply through the ideal diode.

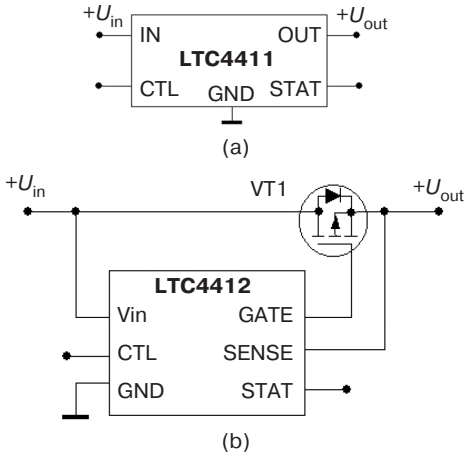


Fig. 5. Ideal diode connection circuit: (a) LTC4411, (b) LTC4412

The LTC4412 controller (Fig. 5b) controls the external *p*-MOSFET (according to established rules, if the transistor is external, then the control circuit is called a controller). The maximum permissible input voltage is up to 40 V, while the switching current is entirely determined by the external transistor characteristics. The SENSE and STAT outputs are used for switching or sharing the load current under operation from multiple power supplies. The gate driver includes an internal voltage clamp for protecting the transistor gate.

Based on the LTC4412, the LTC4412ES[9] power management IC produced by the Linear Technology company comprises an ideal diode with external *p*-channel MOSFET and a built-in Schottky diode to permit efficient OR-ing of multiple power supplies over total load (Fig. 6).

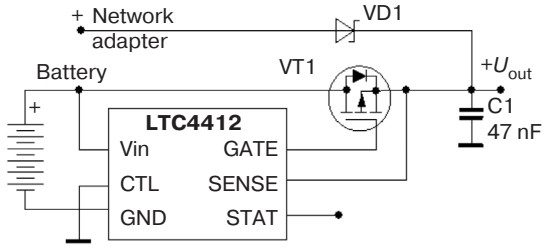


Fig. 6. Redundant power management circuit

In normal operation, the current is brought in contact with the load from the network adapter via the Schottky diode VD1. When external power failure occurs, the current is supplied from the backup battery. The voltage drop across external *p*-channel MOSFET is less than 20 mV.

The IC is designed for use in cellular phones, laptops, digital camcorders, uninterruptible power supplies, powerful USB peripherals, and alternative energy devices. The reverse battery voltage protection circuit based on specialized LTC4359[10] controller (Linear Technology) with external *n*-MOSFET is shown in Fig. 7.

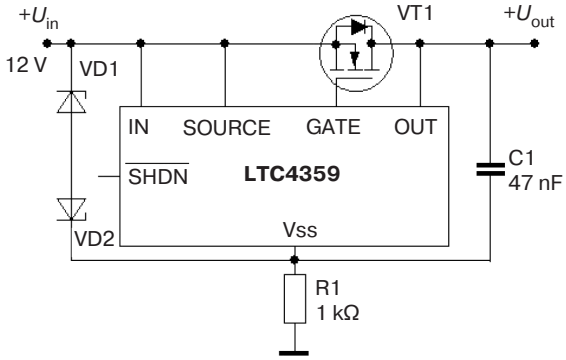


Fig. 7. Automotive equipment protection circuit against the battery reverse polarity

Zener diodes VD1 and VD2 are used in conjunction with resistor R1 to expand the range of possible input voltages. The *n*-channel MOSFET BSC028N06NS[11] (Infineon Technologies AG, Germany) with channel resistance $r_{DS\ ON}$ = 2.8 mΩ and current up to 132 A is

[7] Datasheet LTC4411 Linear Technology. https://www.alldatasheet.com/datasheet-pdf/pdf/94411/LINER/LTC4411.html. Accessed January 10, 2022.

[8] Datasheet LTC4412 Linear Technology. https://www.alldatasheet.com/datasheet-pdf/pdf/82845/LINER/LTC4412.html. Accessed January 10, 2022.

[9] Datasheet LTC4412ES6 Linear Technology. https://www.alldatasheet.com/datasheet-pdf/pdf/82846/LINER/LTC4412ES6.html. Accessed January 10, 2022.

[10] Datasheet LTC4359 Linear Technology. https://www.alldatasheet.com/datasheet-pdf/pdf/1039943/LINER/LTC4359.html. Accessed January 10, 2022.

[11] Datasheet BSC028N06NS Infineon Technologies. https://www.alldatasheet.com/datasheet-pdf/pdf/470560/INFINEON/BSC028N06NS.html. Accessed January 10, 2022.

recommended for use as external VT1 to ensure low heat dissipation, low voltage loss, and small overall dimensions. As well as meeting strict requirements for automotive and telecommunication equipment, the controller allows for redundant power management (SHDN pin).

CONCLUSIONS

The component base of devices providing protection against reverse polarity connection to power supply sources, protection options of devices against reverse voltage polarity has been analyzed along with circuit solutions based on discrete and integrated components. Circuit simulation of ideal diodes based on having high specification *p*- and *n*-channel transistors is used to specify the characteristics and voltage/power losses in protected circuits and demonstrate the simplicity of the technical solution.

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

REFERENCES

1. Guy J. Automatic polarity correction circuit for reverse polarity protection. *RadioLotsman = RadioLocman*. 2021;(7–8):120–121 (in Russ.). Available from URL: https://www.rlocman.ru/shem/schematics.html?di=643745
2. Choudhary V. Transient resistant primary DC/DC converter solutions for automotive power systems. *Silovaya elektronika = Power Electronics*. 2017;3(66): 30–34 (in Russ.). Available from URL: https://power-e.ru/wp-content/uploads/6630.pdf
3. Kazantsev D.P., Shcherbak V.F., Zakamaldin D.A., Kazantsev Yu.E., Pronin A.V., Molodykh S.V. *Electric cardiac pacemaker*: RF Pat. 2531695. Publ. 27.10.2014 (in Russ.).
4. Garelina S.A., Latyshenko K.P., Frunze A.V., Gorbunov R.A. Practical implementation of pyrometers for measuring flame temperature and objects through a flame. *Nauchnye i obrazovatel'nye problemy grazhdanskoi zashchity = Scientific and Educational Tasks of Civil Defence*. 2020;2(45):110–115 (in Russ.).
5. Babenko V.P., Bityukov V.K., Simachkov D.S. DC/DC buck-boost converter with single inductance. *Mikroelektronika = Russian Microelectronics*. 2022;51(1):60–70 (in Russ.). https://doi.org/10.31857/S0544126921060041
6. Oleynik V. Polarity correcting circuit protects battery-powered device. *RadioLotsman = RadioLocman*. 2021; (7–8):122–124 (in Russ.). Available from URL: https://www.rlocman.ru/shem/schematics.html?di=643793
7. Babenko V.P., Bityukov V.K. Simulation of switching of high-power FETS using the Electronics Workbench software. *J. Commun. Technol. Electron*. 2019;64(2): 176–181. https://doi.org/10.1134/S1064226919020025 [Original Russian Text: Babenko V.P., Bityukov V.K. Simulation of switching of high-power FETS using the Electronics Workbench software. *Radiotekhnika i elektronika*. 2019;64(2):199–205 (in Russ.). https://doi.org/10.1134/S0033849419020025]
8. Babenko V.P., Bityukov V.K., Kuznetsov V.V., Simachkov D.S. Simulation of static and dynamic losses in MOSFET keys. *Russ. Technol. J.* 2018;6(1):20–39 (in Russ.). https://doi.org/10.32362/2500-316X-2018-6-1-20-39

СПИСОК ЛИТЕРАТУРЫ

1. Guy J. Схема защиты от переполюсовки с автоматическим исправлением полярности. *РадиоЛоцман*. 2021;(7–8):120–121. URL: https://www.rlocman.ru/shem/schematics.html?di=643745
2. Чоудхари В. Решения устойчивого к переходным процессам первичного DC/DC-преобразователя системы питания автомобильной электроники. *Силовая электроника*. 2017;3(66):30–34. URL: https://power-e.ru/wp-content/uploads/6630.pdf
3. Казанцев Д.П., Щербак В.Ф., Закамалдин Д.А., Казанцев Ю.Е., Пронин А.В., Молодых С.В. *Электрокардиостимулятор*: пат. 2531695 РФ. Заявка № 2012156979/14; заявл. 24.12.2012, опубл. 27.10.2014. Бюл. № 30.
4. Гарелина С.А., Латышенко К.П., Фрунзе А.В., Горбунов Р.А. Практическая реализация пирометров для измерения температуры пламени и объектов сквозь пламя. *Научные и образовательные проблемы гражданской защиты*. 2020;2(45):110–115.
5. Бабенко В.П., Битюков В.К., Симачков Д.С. Понижающе-повышающий DC/DC преобразователь с единственной индуктивностью. *Микроэлектроника*. 2022;51(1):60–70. https://doi.org/10.31857/S0544126921060041
6. Oleynik V. Схема коррекции полярности защищает устройства с батарейным питанием. *РадиоЛоцман*. 2021;(7–8):122–124. URL: https://www.rlocman.ru/shem/schematics.html?di=643793
7. Бабенко В.П., Битюков В.К. Имитационное моделирование процессов переключения силовых полевых транзисторов в программе Electronics Workbench. *Радиотехника и электроника*. 2019;64(2):199–205. https://doi.org/10.1134/S0033849419020025
8. Бабенко В.П., Битюков В.К., Кузнецов В.В., Симачков Д.С. Моделирование статических и динамических потерь в MOSFET-ключах. *Российский технологический журнал*. 2018;6(1):20–39. https://doi.org/10.32362/2500-316X-2018-6-1-20-39

9. Babenko V.P., Bityukov V.K., Simachkov D.S. AC power MOSFET switch. In: *Actual Problems and Perspectives for the Development of Radio Engineering and Infocommunication* (*Radioinfocom 2021*): *Proceedings of the 5th International Scientific and Practical Conference*. Moscow: RTU MIREA; 2021. P. 305–308 (in Russ.).
10. Bityukov V.K., Simachkov D.S., Babenko V.P. *Secondary power sources*. Moscow: Infra-Inzheneriya; 2020. 376 p. (in Russ.).
11. Babenko V.P., Bityukov V.K. Gallium nitride transistors in DC and AC power circuits. In: *Perspective Technology in Means of Information Transfer – PTMIT*-2021: *Proceedings of the 14th International Scientific and Practical Conference*. Vladimir: VlGU; 2021. P. 54–59 (in Russ.).
12. Walker J. FET supplies low-voltage reverse-polarity protection. *RadioLotsman = RadioLocman*. 2019;2:58–59 (in Russ.). Available from URL: https://www.rlocman.ru/shem/schematics.html?di=588375

9. Бабенко В.П., Битюков В.К., Симачков Д.С. Силовой MOSFET-ключ переменного тока. В сб.: *Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем» «Радиоинфоком-2021»*: *Сборник научных статей V Международной научно-практической конференции*. М.: РТУ МИРЭА; 2021. С. 305–308.
10. Битюков В.К., Симачков Д.С., Бабенко В.П. *Источники вторичного электропитания*. М.: Инфра-Инженерия; 2020. 376 с.
11. Бабенко В.П., Битюков В.К. Нитрид-галлиевые транзисторы в силовых цепях постоянного и переменного тока. В сб.: *Перспективные технологии в средствах передачи информации ПТСПИ – 2021*: *Материалы XIV Международной научно-технической конференции*. Владимир: ВлГУ; 2021. С. 54–59.
12. Walker J. MOSFET в низковольтных схемах защиты от обратного напряжения. *РадиоЛоцман*. 2019;2: 58–59. URL: https://www.rlocman.ru/shem/schematics.html?di=588375

About the authors

Valery P. Babenko, Cand. Sci. (Eng.), Associate Professor, Department of Radio Wave Processes and Technology, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: vbabenko16091940@gmail.com. Scopus Author ID 57220534735, RSCI SPIN-code 2815-3866

Vladimir K. Bityukov, Dr. Sci. (Eng.), Professor, Department of Radio Wave Processes and Technology, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: bitukov@mirea.ru. ResearcherID Y-8325-2018, Scopus Author ID 6603797260, RSCI SPIN-code 3834-5360, https://orcid.org/0000-0001-6448-8509

Об авторах

Бабенко Валерий Павлович, к.т.н., доцент, доцент кафедры радиоволновых процессов и технологий Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: vbabenko16091940@gmail.com. Scopus Author ID 57220534735, SPIN-код РИНЦ 2815-3866

Битюков Владимир Ксенофонтович, д.т.н., профессор, профессор кафедры радиоволновых процессов и технологий Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: bitukov@mirea.ru. ResearcherID Y-8325-2018, Scopus Author ID 6603797260, SPIN-код РИНЦ 3834-5360, https://orcid.org/0000-0001-6448-8509

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

UDC 621.385.6
https://doi.org/10.32362/2500-316X-2022-10-6-60-69

RESEARCH ARTICLE

Technology of synthesis and electronic structure of triple barium–strontium–calcium carbonates for cathodes of microwave devices

Vladimir I. Kapustin [1, @], Illarion P. Li [2, @], Natalya E. Kozhevnikova [2], Elvira F. Khudaigulova [3]

[1] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[2] *Pluton, Moscow, 107120 Russia*
[3] *Moscow State University, Moscow, 119991 Russia*
[@] *Corresponding authors, e-mail: kapustin@mirea.ru, i.li@pluton.msk.ru*

Abstract

Objectives. Triple barium–strontium–calcium carbonates of various grades are widely used for depositing oxide coatings on cathodes of electrovacuum devices. Of all types of cathodes used in electrovacuum devices, oxide cathodes are among the most common, due to combining efficiency, durability, operation at relatively low temperatures, and a relatively low cost. The aims of this work were to: create a technology for the synthesis of triple barium–strontium–calcium carbonates with nonequilibrium phase compositions that comprise the triple carbonate proper and a pure barium carbonate phase; develop a quality control procedure for such a carbonate for using it as a component of the cathode material for microwave devices; study how the electronic structure of barium oxide crystallites is affected by doping microimpurities from other phases of the cathode material.

Methods. The study used precision X-ray diffraction analysis and electron spectroscopy.

Results. A technology was developed for the co-precipitation of triple barium–strontium–calcium carbonates from their nitrate salts. Under optimal precipitation conditions, this produces triple carbonate powders with nonequilibrium phase compositions. Electron spectroscopy showed that the parameters of the electronic structure of the crystallites are significantly affected by doping impurities of calcium, strontium, and nickel in barium oxide crystallites formed by heat treatment of triple carbonates.

Conclusions. Calcium and strontium have a synergistic effect on the doping of barium oxide with the two other chemical elements. As well as efficiently controlling the quality of the nonequilibrium phase composition of triple carbonates, which is formed during the synthesis of triple carbonates by the titration method, precision X-ray diffraction analysis can be used to efficiently control the processes of agglomeration of nanoparticles or recrystallization of nanostructured phases formed during the synthesis of triple carbonates.

Keywords: porous-metal cathodes, cathode material, thermionic emission, electronic structure, X-ray diffraction analysis

• **Submitted:** 28.03.2022 • **Revised:** 05.04.2022 • **Accepted:** 07.09.2022

For citation: Kapustin V.I., Li I.P., Kozhevnikova N.E., Khudaigulova E.F. Technology of synthesis and electronic structure of triple barium–strontium–calcium carbonates for cathodes of microwave devices. *Russ. Technol. J.* 2022;10(6):60–69. https://doi.org/10.32362/2500-316X-2022-10-6-60-69

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

Технология синтеза и электронная структура тройных карбонатов бария-стронция-кальция для катодов СВЧ-приборов

В.И. Капустин [1, @], И.П. Ли [2, @], Н.Е. Кожевникова [2], Э.Ф. Худайгулова [3]

[1] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[2] *АО «Плутон», Москва, 107120 Россия*
[3] *Московский государственный университет, Москва, 119991 Россия*
[@] *Авторы для переписки, e-mail: kapustin@mirea.ru, i.li@pluton.msk.ru*

Резюме

Цели. Тройные карбонаты бария-стронция-кальция различных марок широко используются для нанесения оксидных покрытий на катоды электровакуумных приборов. Из всех типов катодов в электровакуумных приборах оксидные катоды являются одними из самых распространенных, т.к. сочетают в себе эффективность, долговечность, работу при относительно небольших температурах и сравнительно невысокую стоимость. Цели работы – разработка технологии синтеза тройных карбонатов бария-стронция-кальция с неравновесным фазовым составом, состоящим из собственно тройного карбоната и фазы чистого карбоната бария, разработка методики контроля качества такого карбоната для применения его в качестве компонента катодного материала для СВЧ-приборов, а также исследование влияния на электронную структуру кристаллитов оксида бария легирующих микропримесей из состава других фаз катодного материала.

Методы. Использована методика прецизионного рентгеноструктурного анализа и методы электронной спектроскопии.

Результаты. Разработана технология совместного осаждения тройных карбонатов бария-стронция-кальция из их азотнокислых солей, которая при выборе оптимального режима осаждения позволяет получать порошки тройного карбоната с неравновесным фазовым составом. Методами электронной спектроскопии показано, что легирующие примеси кальция, стронция, никеля в кристаллитах оксида бария, формирующихся при термообработке тройных карбонатов, существенно влияют на параметры электронной структуры кристаллитов.

Выводы. Совместное влияние кальция и стронция свидетельствует о наличии так называемого синергетического эффекта при легировании оксида бария двумя другими химическими элементами. Методика прецизионного рентгеноструктурного анализа позволяет эффективно контролировать качество неравновесного фазового состава тройных карбонатов, формирующегося при синтезе тройных карбонатов методом титрования и контролировать процессы агломерации наночастиц либо рекристаллизации наноструктурированных фаз, формирующихся при синтезе тройных карбонатов.

Ключевые слова: металлопористые катоды, катодный материал, термоэлектронная эмиссия, электронная структура, рентгеноструктурный анализ

• **Поступила:** 28.03.2022 • **Доработана:** 05.04.2022 • **Принята к опубликованию:** 07.09.2022

Для цитирования: Капустин В.И., Ли И.П., Кожевникова Н.Е., Худайгулова Э.Ф. Технология синтеза и электронная структура тройных карбонатов бария-стронция-кальция для катодов СВЧ-приборов. *Russ. Technol. J.* 2022;10(6):60–69. https://doi.org/10.32362/2500-316X-2022-10-6-60-69

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

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Triple barium–strontium–calcium carbonates have been widely used as components of various types of cathodes for microwave electrovacuum devices – in particular, as the main component of a nickel oxide cathode. The chief characteristics of triple carbonates are determined by their chemical composition and particle size distribution [1, 2]. While there are various technologies for the synthesis of triple carbonates [3–11], the functional efficiency of cathodes using these materials strongly depends on the technology used for their synthesis. Figure 1a shows the equilibrium ternary phase diagram of ternary carbonates [12], and Fig. 2b presents that of ternary oxides [13], which are formed by the decomposition of ternary carbonates during activation of cathodes by heating at a temperature of about 1000°C. The shaded areas in the diagrams indicate compositions of materials used in cathode electronics.

Examples of practically used materials consisting of triple carbonates include aragonite having a rhombohedral crystal structure, comprising a ternary solid carbonate solution. In the case of the thermal decomposition of such a triple carbonate during activation of cathodes and the formation of a ternary solid solution of barium–strontium–calcium oxides, the work function of such a material is known to be higher than that of pure barium oxide [14]. It is noted in the literature that the phase composition of triple carbonates and, hence, their efficiency in the manufacture of cathodes strongly depends on their synthesis conditions [12, 15].

The aims of this work were to create a technology for the synthesis of triple barium–strontium–calcium carbonates with nonequilibrium phase compositions that comprise the triple carbonate proper and a pure barium carbonate phase, as well as developing a quality control procedure for such a carbonate for using it as a component of the cathode material for microwave devices and studying how the electronic structure of barium oxide crystallites is affected by doping microimpurities from other phases of the cathode material.

TRIPLE CARBONATE SYNTHESIS TECHNOLOGY

Triple barium–strontium–calcium carbonates of KTA-1-6 grade were prepared by rapidly adding a 3.5 molal aqueous ammonium carbonate solution stabilized with ammonia to a 1 molal aqueous solution of the corresponding nitrate salts at room temperature. Triple barium–strontium–calcium carbonate of KTA-1-6 grade was obtained by adding a 0.9 molal aqueous solution of the corresponding nitrate salts heated to 45°C to a 3.5 molal aqueous ammonium carbonate solution stabilized with ammonia.

When developing a technology for the synthesis of triple carbonates with nonequilibrium phase compositions, various synthesis conditions were tested. These differed in the rate of addition of the precipitating

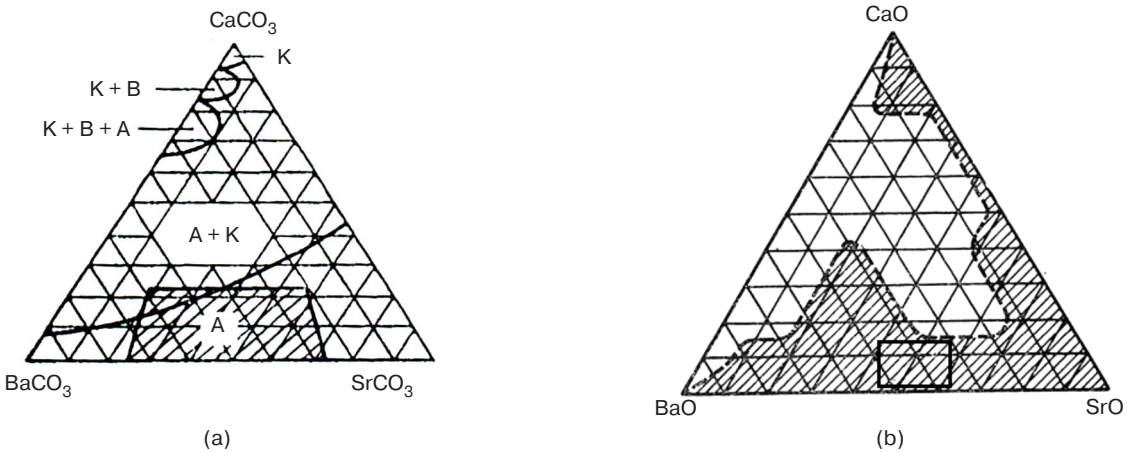
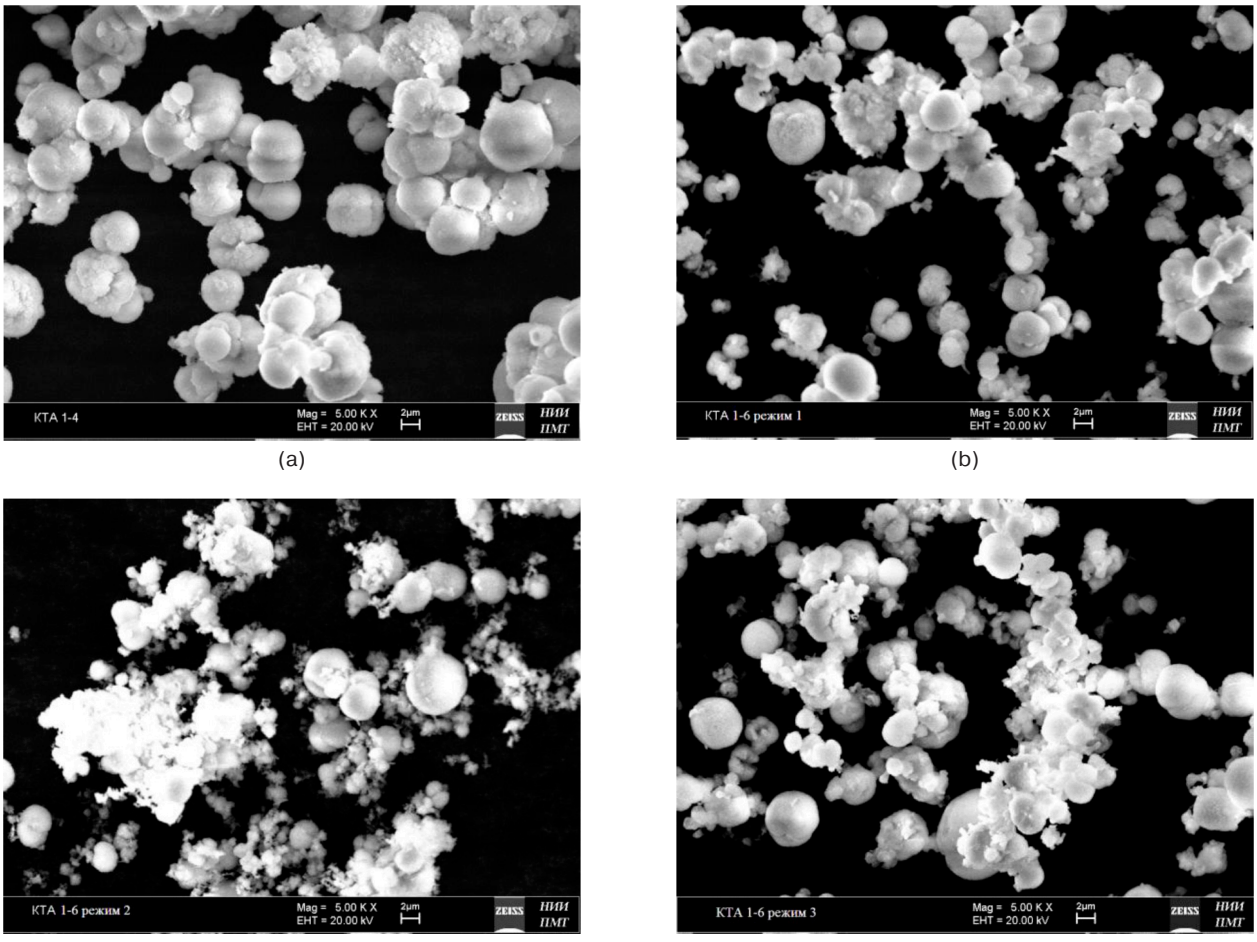


Fig. 1. Phase diagrams of the (a) $BaCO_3$–$SrCO_3$–$CaCO_3$(a) [1] and (b) BaO–CaO–SrO [2] systems

Table 1. Conditions of synthesis of triple barium–strontium–calcium carbonates

Parameter	Variants of synthesis products			
	KTA-1-4	KTA-1-6, conditions 1	KTA-1-6, conditions 2	KTA-1-6, conditions 3
Morphology of product	Fig. 2a	Fig. 2b	Fig. 2c	Fig. 2d
Average grain size, μm (measured with PSKh-10a, LabNauchPribor, Russia)	3.9	1.9	0.9	2.0
Composition $BaCO_3$:$SrCO_3$:$CaCO_3$, %	50:43:6	50:43:6	50:43:6	54:40:5
Temperature of nitrate solution, °C	46	26	30	30
Molar concentration of nitrate solution	0.9	1	1	1



(a) (b) (c) (d)

Fig. 2. Micrographs of particles of triple barium–strontium–calcium carbonate

component to the solution of nitrate salts, as well as the rate of mixing of the reacting solutions.

The structure of the powders of the synthesis products was studied by electron spectroscopy, while the phase composition was studied by X-ray powder diffraction analysis. Table 1 presents the conditions of the synthesis of triple carbonates, while Fig. 2 shows the morphology of particles of the synthesis products.

INVESTIGATION OF THE PHASE COMPOSITION OF SYNTHESIS PRODUCTS

The synthesis products were studied by X-ray powder diffraction analysis. Since the parameters of the crystal structure of germanium are known up to the fifth decimal place, a germanium powder was added to the material to be analyzed at a concentration of 20% to function as an internal standard. Preliminarily, pure barium carbonate

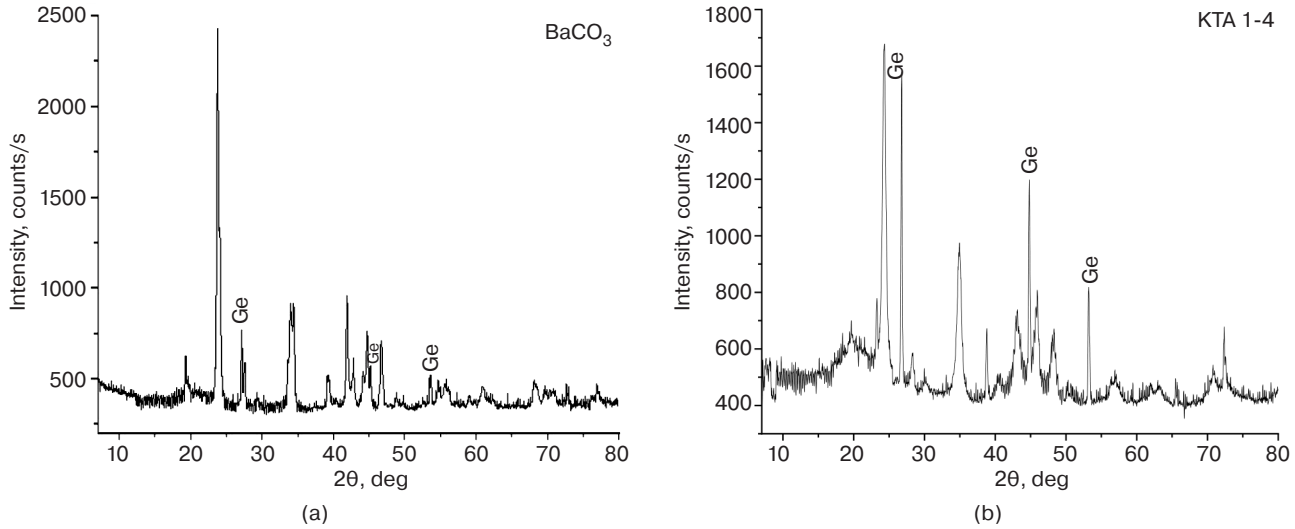


Fig. 3. X-ray powder diffraction patterns of (a) carbonate $BaCO_3$ and (b) triple carbonate of grade KTA-1-4

was studied, the X-ray powder diffraction pattern of which is shown in Fig. 3a. As an example, Fig. 3b presents the X-ray powder diffraction pattern of the synthesized KTA-1-4 grade triple carbonate.

Along with solid solutions of triple carbonates, the analysis of the phase composition of triple carbonate synthesis products demonstrated the presence of pure barium carbonate, which content strongly depends on the synthesis conditions. Table 2 presents the most significant characteristics of the X-ray powder diffraction analysis of the synthesis products of triple carbonates along with the characteristics of pure barium carbonate.

By analyzing the recorded X-ray powder diffraction patterns along with the parameters given in Table 2, it was possible to formulate quality control criteria for triple barium–strontium–calcium carbonates to permit the efficient use of this material as a component of microwave device cathodes. Figure 4 illustrates the selection of quality criteria for triple carbonate using a portion of the X-ray powder diffraction pattern in the 2θ range 22°–28°.

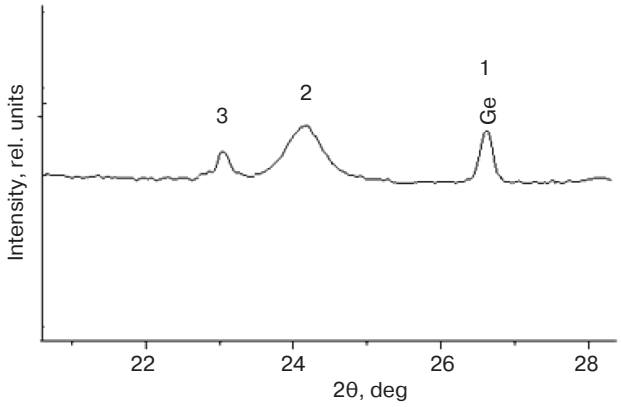


Fig. 4. Selection of quality control parameters for triple carbonates

In Fig. 4, peak 1 is the germanium peak used as an internal standard for determining the positions of other peaks in the X-ray powder diffraction pattern up to the fifth decimal place. Peak 2 characterizes a solid triple carbonate solution phase, while peak 3 corresponds to a phase of pure barium carbonate. To summarize the results of this

Table 2. Quality control parameters for triple carbonates

Parameter	Material				
	$BaCO_3$	KTA-1-4	KTA-1-6, conditions 1	KTA, conditions 2	KTA, conditions 3
a, position of peak of triple carbonate, deg	–	23.318	23.654	24.673	23.644
d, interplanar distance at peak of triple carbonate, Å	–	3.660	3.611	3.608	3.613
δθ, broadening of peak of triple carbonate, deg	–	0.40	0.43	0.46	0.42
a_0, position of peak of barium carbonate, deg	23.778	23.262	23.660	24.22	23.618
Δ*a*, shift of peak of barium carbonate, deg	Standard	−0.516	−0.118	0.442	–0.160
$\delta\theta_0$, broadening of peak of barium carbonate, deg	0.31	0.20	0.06	0.07	0.10
I, barium carbonate to triple carbonate peak area ratio in sample, %	100	7.5	1.1	3.8	2.3

study, physical criteria can be formulated for the quality of triple barium–strontium–calcium carbonate for its use as a component of the cathode materials for microwave devices:

- the position of peak 2 (triple carbonate) represents the composition of the solid solution of triple carbonate;
- the width of peak 2 (triple carbonate) represents the scatter in the composition of triple carbonate;
- the position of peak 3 (pure barium carbonate) represents the level of doping with microimpurities, including calcium and strontium;
- the width of peak 3 (pure barium carbonate) represents the size of crystallites of pure barium carbonate;
- the ratio of the areas of peaks 3 and 2 represents the content of pure barium carbonate in the synthesis product.

INVESTIGATION OF THE ELECTRONIC STRUCTURE OF MATERIALS

Along with triple barium–strontium–calcium carbonate, another component of nickel oxide cathodes is nickel powder having a particle size of 15–25 μm. In addition to functioning as a metal phase to increase the thermal and electrical conductivity of the material, nickel acts as a catalyst for the decomposition of triple carbonate to the corresponding barium–strontium–calcium oxides. In addition, nickel atoms entering the crystal lattice of barium oxide affect the parameters of the electronic structure of the oxide. Obviously, the functions of a catalyst and an alloying element can be performed more efficiently by nickel nanoparticles introduced into the cathode material as an additional component. In this context, we studied the electronic structure of the material of the composition 0.9Ni + 0.1 (KTA-1-6-grade triple carbonate + 0.1Ni (nano)). The nanopowder comprised a nickel powder with a particle size of 80 nm, which was produced by electrical nickel wire explosion by our order at NPK Peredovye poroshkovye tekhnologii (Tomsk, Russia).

The electronic structure of the materials was investigated using a Theta Probe high-resolution electron spectrometer (Thermo Fisher Scientific, USA) equipped with a high-resolution spherical mirror analyzer, an X-ray source, and several types of electron gun. This permits the study of materials by chemical analysis and characteristic electron energy loss spectroscopy.

The initial components of test samples were a KTA-1-6 grade triple carbonate powder (OST 11-OD0.028.002-76[1], Pluton, Russia); a PNK-1L7 grade nickel powder with particle sizes of 15–25 μm (Kolskaya GMK, Russia), and a nano nickel powder (TU 1791-003-36280340-2008, Peredovye poroshkovye tekhnologii,[2] Russia). After mixing the components in a turbula-type mixer (PASM1.000.001, NPK ISTEL, Russia), the samples of the materials were placed into molybdenum vessels (made by hand according to GOST 25442-82[3] from a sheet of MCh grade molybdenum, sheet thickness of 0.22 mm, PO Volfram, Russia) with an alundum-coated molybdenum insert (made by alunding molybdenum sheets according to the internal technology of the enterprise) on nickel plates and sintered in bulk in a vacuum furnace (SShVE-1-2,5/25-I2, NPO TsNIITMASh, Russia) with a gradual temperature rise to 1200°C for 2 h. The materials to be studied were compacted into pellets 1 mm thick and 7.6 mm in diameter on a laboratory press Vaneox 25t (FLUXANA GmbH & Co.KG., Germany) in steel molds at a specific pressing force of P_{sp} ~ 4.5–5 t/cm^2.

Figure 5 shows the electronic spectrum of barium states in the studied sample of the material of the composition 0.9Ni + 0.1 (KTA-1-6-grade triple carbonate + 0.1Ni(nano)), while Table 3 presents the results of the interpretation of this spectrum.

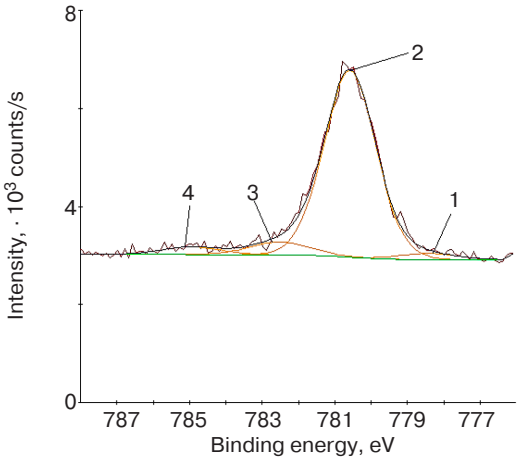


Fig. 5. Electronic spectrum of barium in a sample of the material 0.9Ni + 0.1(KTA + 0.1Ni (nano))

Characteristic electron energy loss spectroscopy was used to study the electron plasma energy losses upon excitation of bulk and surface plasmons in the oxide phases listed in Table 3. Figure 6 shows the characteristic electron energy loss spectrum after its differentiation in order to more clearly distinguish the peaks of characteristic losses.

[1] OST 11-OD0.028.002-76. Electro-vacuum devices. Carbonate of alkaline-earth metals of barium, strontium, calcium. Specifications. http://www.docum.ru/ost.asp?id=262135. Accessed October 28, 2022 (in Russ.).

[2] http://www.nanosized-powders.com. Accessed October 28, 2022 (in Russ.).

[3] GOST 25442-82. Interstate standard. Molibdenum annealed strips for deep drawing. Specifications. Moscow: Izd. Standartov; 2004 (in Russ.).

Table 3. Interpretation of the electronic spectrum of barium in a sample of the material 0.9Ni + 0.1(KTA + 0.1Ni (nano))

No. of peak	Peak energy, eV	Peak intensity, counts/s	Peak width, eV	Barium in compound
1	778.43	124.17	1.96	$Ba(OH)_2 \cdot 2H_2O$
2	780.58	3806.47	1.74	$Ba_{(1-y)}O_{(1-x)}Ni_y$
3	782.51	262.61	1.98	$BaO_{(1-x)}$
4	784.94	159.08	1.95	$Ba_{(1-y)}O_{(1-x)}Ca_y$

Table 4. Parameters of characteristic losses in the oxide phases of the material 0.9Ni + 0.1(KTA + 0.1Ni (nano))

No. of phase	Phase composition	Parameter				
		E_s, eV	E_b, eV	N_s, m^{-3}	N_b, m^{-3}	N_s/N_b
1	$BaO_{(1-x)}$	1.96	2.56	$4.84 \cdot 10^{24}$	$4.15 \cdot 10^{24}$	1.17
2	$Ba_{(1-y-z)}O_{(1-x)}Ca_ySr_z$	3.40	5.42	$1.05 \cdot 10^{25}$	$1.33 \cdot 10^{25}$	0.79
3	$Ba_{(1-y)}O_{(1-x)}Ni_y$	7.81	10.84	$4.76 \cdot 10^{25}$	$4.60 \cdot 10^{25}$	1.03

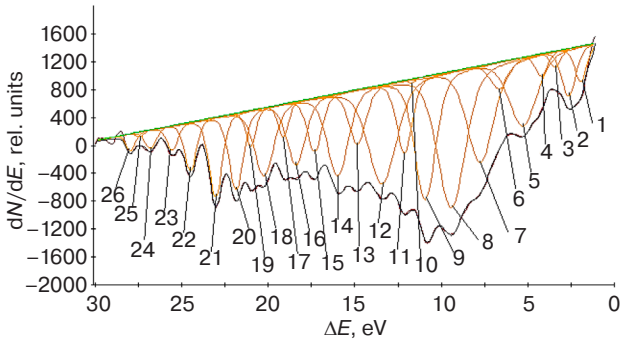


Fig. 6. Characteristic electron energy loss spectrum of the material 0.9Ni + 0.1(KTA + 0.1Ni (nano))

Barium oxide, which is formed by the heat treatment of the material and contains oxygen vacancies, is a donor-type semiconductor, whereas the other carbonate or oxide phases in the cathode material are dielectrics. Therefore, when the surface of the material is bombarded by medium-energy electrons, they undergo characteristic energy losses due to the excitation of bulk and surface plasmons—collective oscillations of oxygen vacancy electrons. Their energies, E_b and E_s, respectively, are found from the well-known expressions

$$\Delta E_b = \sqrt{\frac{e^{*2}\hbar^2 N_b}{\varepsilon\varepsilon_0 m^*}}, \quad (1)$$

$$\Delta E_s = \sqrt{\frac{e^{*2}\hbar^2 N_s}{\varepsilon\varepsilon_0 m^*}}, \quad (2)$$

where e^* is the effective charge of an oxygen vacancy, m^* is the effective mass of oxygen vacancy electrons, ħ is the Planck constant, ε_0 is the vacuum permittivity, $\varepsilon = 3.6$ is the high-frequency permittivity of barium oxide, N_b is the bulk concentration of oxygen vacancies, and N_s is the surface concentration of oxygen vacancies.

The total losses ΔE due to the excitation of plasmon oscillations are written as

$$\Delta E = n_1 E_b + n_2 E_s, \quad (3)$$

where n_1 and n_2 are integers.

Characteristic loss peaks 1–26 in Fig. 6 are well grouped into 9 groups, which (three rows each) correspond to the phases $BaO_{(1-x)}$, $Ba_{(1-y)}O_{(1-x)}Ni_y$, $Ba_{(1-y)}O_{(1-x)}Ca_y$. Table 4 presents summary data on the parameters of the characteristic electron energy losses, along with the calculated values of the volume and surface concentrations of oxygen vacancies in various phases of barium oxide.

CONCLUSIONS

It is shown that the technology of coprecipitation of triple barium–strontium–calcium carbonates from their nitrate salts can be used under optimal precipitation conditions to produce powders of triple carbonates having nonequilibrium phase compositions.

The results of the study of the electronic structure of carbonates containing a nickel nanopowder suggested the following conclusions:

(1) The higher bulk and surface concentrations of oxygen vacancies in the $Ba_{(1-y)}O_{(1-x)}Ni_y$ phase than in the simple $BaO_{(1-x)}$ phase should be accompanied by a lower work function of this phase and, hence, a higher thermal emission current density.

(2) Although the bulk concentration of vacancies is lower in the $Ba_{(1-y-z)}O_{(1-x)}Ca_ySr_z$ phase than in the $Ba_{(1-y)}O_{(1-x)}Ni_y$ phase, the ratio of the bulk and volume concentrations of vacancies is significantly less than unity; i.e., the surface is depleted in oxygen

vacancies, which should be accompanied by a decrease in the upward curvature of the energy bands, i.e., a decrease in the work function of the material.

(3) Calcium and strontium have a synergistic effect on the doping of barium oxide with the two other chemical elements.

The results of the study of the crystal structure of triple carbonates synthesized by the titration method suggest the following conclusions:

- Precision X-ray diffraction analysis efficiently controls the quality of the nonequilibrium phase composition of triple carbonates, which is formed during the synthesis of triple carbonates by the titration method.
- (5) Precision X-ray diffraction analysis can be used to efficiently control the processes of agglomeration of nanoparticles or recrystallization of nanostructured phases formed during the synthesis of triple carbonates.

Authors' contributions

V.I. Kapustin—definition of the research concept; study of the electronic structure of materials; analysis and interpretation of data; writing the initial text of the article.

I.P. Li—investigation of the powder structure of synthesized products via electron spectroscopy; analysis and interpretation of data; participation in writing the initial text of the article.

N.E. Kozhevnikova—development of technology for the synthesis of non-equilibrium phase compositions of triple carbonates; conducting the synthesis of triple carbonates of KTA-1-6, KTA-1-4 grades; preparation of samples of materials to study their electronic structure; participation in writing the initial text of the article.

E.F. Khudaigulova—investigation of the phase composition of powders of synthesized products via X-ray phase analysis; participation in writing the initial text of the article.

All authors have approved the final manuscript for publication.

REFERENCES

1. Nikonov B.P. *Oksidnyi katod* (*Oxide Cathode*). Moscow: Energiya; 1979. 240 p. (in Russ.).
2. Kapustin V.I., Li I.P., Kozhevnikova N.E., Khudaigulova E.F. New criteria for quality control of raw barium-strontium-carbonates calcium. *Elektronnaya tekhnika. Seriya 2: Poluprovodnikovye pribory*. 2021;3(262):94–98 (in Russ.). https://www.elibrary.ru/item.asp?id=46663947
3. Karsakova M.B., Petrova O.B. Processes of crystal formation of triple carbonates for the technology of cathodes of electrovacuum devices. In: *Youth and 21st century 2022.* Proceedings of the 12th International Youth Scientific Conference. In 4 v. Kursk; 2022. V.1. P. 32–36 (in Russ.).
4. Mosin A.D., Pastushkova A.A. Study of the structure of triple carbonates. In: *Scientific and technological progress: information, technology, mechanism the collection.* Collection of articles on results of the International Scientific and Practical Conference. Sterlitamak; 2020. P. 83–87 (in Russ.).
5. Kapustin V.I., Li I.P., Kozhevnikova N.E., Karsakova M.B., Khudaigulova E.F. Synthesis and analysis of nanoscale crystallites of triple barium–strontium–calcium carbonates for cathodes of microwave devices. In: *XI Vserossiskaya nauchno-tekhnicheskaya konferentsiya* "*Elektronika i mikroelektronika SVCh*" (*The 11th All-Russian Scientific and Technical Conference "Microwave Electronics and Microelectronics*." Collection of reports. St. Petersburg. May 30 – June 3, 2022). St. Petersburg: SPbGETU LETI. P. 427–431 (in Russ.). Available from URL: https://mwelectronics.etu.ru/assets/files/2022/427-431.pdf

СПИСОК ЛИТЕРАТУРЫ

1. Никонов Б.П. *Оксидный катод*. М.: Энергия; 1979. 240 с.
2. Капустин В.И., Ли И.П., Кожевникова Н.Е., Худайгулова Э.Ф. Новые критерии контроля качества тройных карбонатов бария-стронция-кальция. *Электронная техника. Серия 2: Полупроводниковые приборы*. 2021;3(262):94–98. https://www.elibrary.ru/item.asp?id=46663947
3. Карсакова М.Б., Петрова О.Б. Процессы кристаллообразования тройных карбонатов для технологии катодов электровакуумных приборов. В сб.: *Молодежь и XXI век – 2022*. Материалы 12-й Международной молодежной научной конференции: в 4-х т. Курск; 2022. Т. 4. С. 32–36.
4. Мосин А.Д., Пастушкова А.А. Исследование структуры тройных карбонатов. В сб.: *Научно-технический прогресс: информация, технологии, механизм.* Сборник статей по итогам Международной научно-практической конференции. Стерлитамак; 2020. С. 83–87.
5. Капустин В.И., Ли И.П., Кожевникова Н.Е., Карсакова М.Б., Худайгулова Э.Ф. Синтез и анализ наноразмерных кристаллитов тройных карбонатов бария-стронция-кальция для катодов СВЧ приборов. В сб.: *XI Всероссийская научно-техническая конференция «Электроника и микроэлектроника СВЧ»*. Сборник докладов. Санкт-Петербург. 30 мая – 3 июня 2022 г. СПб.: СПбГЭТУ «ЛЭТИ». С. 427–431. URL: https://mwelectronics.etu.ru/assets/files/2022/427-431.pdf
6. Александров Е.М., Шофман Г.С., Лубянецкая К.Ф., Фельдман Ф.С. *Способ изготовления эмиссионного материала для оксидных катодов*: пат. 2019878 РФ. Заявл. 26.02.1992; опубл. 15.09.1994.

6. Aleksandrov E.M., Shofman G.S., Lubyanetskaya K.F., Fel'dman F.S. *Method for manufacturing emission material for oxide cathodes*: RF Pat. 2019878. Publ. 15.09.1994 (in Russ.).
7. Artyushenko A.I., Batura Z.E., Krivobok V.I., Nakhodnova A.P. *Method for obtaining fine powders of triple and double carbonates of alkaline earth elements*: RF Pat. 180572. Publ. 19.03.1966 (in Russ.).
8. Dorzhin G.S., Neznaev G.N., Kitaeva T.I. Rusakova Zh.P. *Method for obtaining mixed carbonates of alkaline earth metals*: RF Pat. 520330. Publ. 05.07.1976 (in Russ.).
9. Evstigneeva N.K., Kalinina I.D. *Method for obtaining powders of double or triple carbonates of alkaline earth metals*: RF Pat. 880983. Publ. 15.11.1981 (in Russ.).
10. Smorodinova L.A., Aryanina T.G. *Emission material*: RF Pat. 873301. Publ. 15.10.1981 (in Russ.).
11. Kiselev A.B., Lobova E.V., Nikonov B.P. *Method for manufacturing an active coating of oxide cathodes*: RF Pat. 383109. Publ. 23.05.1973 (in Russ.).
12. Bol'shakov A.F., Abalduev B.V., Popov A.I. Phase inhomogeneity of mixed crystals of alkaline earth metal carbonates and its influence on the properties of an oxide cathode. *Izvestiya AN SSSR. Ser. Neorganicheskiye materialy*. 1977;13(7):1270–1274 (in Russ.).
13. Moizhes B.Ya. *Fizicheskie protsessy v oksidnom katode* (*Physical processes in the oxide cathode*). Moscow: Nauka; 1968. 480 p. (in Russ.).
14. Kapustin V.I., Li I.P. *Teoriya, elektronnaya struktura i fizikokhimiya materialov katodov SVCh priborov: monografiya* (*Theory, electronic structure and physical chemistry of microwave device cathode materials*). Moscow: INFRA-M; 2020. 370 p. (in Russ.). ISBN 978-5-16-015560-9
15. Chistyakova M.A., Podkopaeva N.N., Konikova R.A. Investigation of structural-phase transformations of the oxide coating of cathodes during the formation and service life. *Elektronnaya tekhnika. Seriya 4. Elektrovakuumnye i gazorazryadnye pribory*.1977;4:106–113 (in Russ.).

7. Артюшенко А.И., Батура З.Е., Кривобок В.И., Находнова А.П. *Способ получения мелкодисперсных порошков тройных и двойных карбонатов щелочноземельных элементов*: пат. 180572 РФ. Заявка № 809296/23-4; заявл. 19.12.1962; опубл. 19.03.1966. Бюл. № 8.
8. Доржин Г.С., Незнаев Г.Н., Китаева Т.И., Русакова Ж.П. *Способ получения смешанных карбонатов щелочноземельных металлов*: пат. 520330 РФ. Заявка № 2013007; заявл. 05.04.1974; опубл. 05.07.1976.
9. Евстигнеева Н.К., Калинина И.Д. *Способ получения порошков двойных или тройных карбонатов щелочноземельных металлов*: пат. 880983 РФ. Заявка № 2895040; заявл. 04.01.1980; опубл. 15.11.1981.
10. Смородинова Л.А., Арянина Т.Г. *Эмиссионный материал*: пат. 873301 РФ. Заявка № 2860081; заявл. 27.12.1979; опубл. 15.10.1981.
11. Киселев А.Б., Лобова Э.В., Никонов Б.П. *Способ изготовления активного покрытия оксидных катодов*: пат. 383109 РФ. Заявка № 1608332/26-25; заявл. 04.01.1971; опубл. 23.05.1973.
12. Большаков А.Ф., Абалдуев Б.В., Попов А.И. Фазовая неоднородность смешанных кристаллов карбонатов щелочноземельных металлов и ее влияние на свойства оксидного катода. *Известия АН СССР. Сер. Неорганические материалы*. 1977;13(7):1270–1274.
13. Мойжес Б.Я. *Физические процессы в оксидном катоде*. М.: Наука; 1968. 480 с.
14. Капустин В.И., Ли И.П. *Теория, электронная структура и физикохимия материалов катодов СВЧ приборов: монография*. М.: ИНФРА-М; 2020. 370 с. ISBN 978-5-16-015560-9
15. Чистякова М.А., Подкопаева Н.Н., Коникова Р.А. Исследование структурно-фазовых превращений оксидного покрытия катодов в процессе формирования и срока службы. *Электронная техника. Серия 4. Электровакуумные и газоразрядные приборы*. 1977;4:106–113.

About the authors

Vladimir I. Kapustin, Dr. Sci. (Phys.-Math.), Professor, Chief Specialist, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: kapustin@mirea.ru. Scopus Author ID 7006839899, ResearcherID U-9032-2017, RSCI SPIN-code 5072-5992

Illarion P. Li, Dr. Sci. (Eng.), Director of the Technical Center "Basic Technology of Electric Vacuum Devices," Pluton (11, Nizhnyaya Syromyatnicheskaya ul., Moscow, 105120 Russia). E-mail: i.li@pluton.msk.ru. RSCI SPIN-code 9208-6634

Natalya E. Kozhevnikova, Head of the Laboratory for the Synthesis of Emission-Active Compounds, Pluton (11, Nizhnyaya Syromyatnicheskaya ul., Moscow, 105120 Russia). E-mail: azariia@bk.ru. https://orcid.org/0000-0002-8022-1238

Elvira F. Khudaigulova, Undergraduate Student, Faculty of Physics, M.V. Lomonosov Moscow State University (1-2, Leninskie Gory, Moscow, 119991 Russia). E-mail: hudaygulovaelya@gmail.com. https://orcid.org/0000-0003-2778-0329

Об авторах

Капустин Владимир Иванович, д.ф.-м.н., профессор, главный специалист, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: kapustin@mirea.ru. Scopus Author ID 7006839899, ResearcherID U-9032-2017, SPIN-код РИНЦ 5072-5992

Ли Илларион Павлович, д.т.н., директор технического центра «Базовые технологии ЭВП» АО «Плутон» (105120, Россия, Москва, ул. Нижняя Сыромятническая, д. 11). E-mail: i.li@pluton.msk.ru. SPIN-код РИНЦ 9208-6634

Кожевникова Наталья Евгеньевна, начальник лаборатории синтеза эмиссионно-активных соединений АО «Плутон» (105120, Россия, Москва, ул. Нижняя Сыромятническая, д. 11). E-mail: azariia@bk.ru. https://orcid.org/0000-0002-8022-1238

Худайгулова Эльвира Фанильевна, магистрант, Физический факультет, ФГБОУ ВО «Московский государственный университет им. М.В. Ломоносова» (119991, Россия, Москва, Ленинские горы, д. 1, стр. 2). E-mail: hudaygulovaelya@gmail.com. https://orcid.org/0000-0003-2778-0329

Translated from Russian into English by Vladislav V. Glyanchenko
Edited for English language and spelling by Thomas A. Beavitt

Mathematical modeling

Математическое моделирование

UDC 517.63
https://doi.org/10.32362/2500-316X-2022-10-6-70-77

RESEARCH ARTICLE

Methods and effective algorithms for solving multidimensional integral equations

Alexander B. Samokhin

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

Abstract

Objectives. Integral equations have long been used in mathematical physics to demonstrate existence and uniqueness theorems for solving boundary value problems for differential equations. However, despite integral equations have a number of advantages in comparison with corresponding boundary value problems where boundary conditions are present in the kernels of equations, they are rarely used for obtaining numerical solutions of problems due to the presence of equations with dense matrices that arise that when discretizing integral equations, as opposed to sparse matrices in the case of differential equations. Recently, due to the development of computer technology and methods of computational mathematics, integral equations have been used for the numerical solution of specific problems. In the present work, two methods for numerical solution of two-dimensional and three-dimensional integral equations are proposed for describing several significant classes of problems in mathematical physics.

Methods. The method of collocation on non-uniform and uniform grids is used to discretize integral equations. To obtain a numerical solution of the resulting systems of linear algebraic equations (SLAEs), iterative methods are used. In the case of a uniform grid, an efficient method for multiplying the SLAE matrix by vector is created.

Results. Corresponding SLAEs describing the considered classes of problems are set up. Efficient solution algorithms using fast Fourier transforms are proposed for solving systems of equations obtained using a uniform grid.

Conclusions. While SLAEs using a non-uniform grid can be used to describe complex domain configurations, there are significant constraints on the dimensionality of described systems. When using a uniform grid, the dimensionality of SLAEs can be several orders of magnitude higher; however, in this case, it may be difficult to describe the complex configuration of the domain. Selection of the particular method depends on the specific problem and available computational resources. Thus, SLAEs on a non-uniform grid may be preferable for many two-dimensional problems, while systems on a uniform grid may be preferable for three-dimensional problems.

Keywords: integral equations, collocation method, fast Fourier transform

• **Submitted:** 14.12.2021 • **Revised:** 10.01.2022 • **Accepted:** 15.09.2022

For citation: Samokhin A.B. Methods and effective algorithms for solving multidimensional integral equations. *Russ. Technol. J.* 2022;10(6):70–77. https://doi.org/10.32362/2500-316X-2022-10-6-70-77

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

The author declares no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Методы и эффективные алгоритмы решения многомерных интегральных уравнений

А.Б. Самохин

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

Резюме

Цели. Интегральные уравнения давно и широко используются в математической физике для доказательства теорем существования и единственности решения краевых задач для дифференциальных уравнений. Однако, несмотря на то что интегральные уравнения имеют ряд преимуществ по сравнению с соответствующими краевыми задачами – все краевые условия присутствуют в ядрах уравнений, они практически не использовались для численного решения задач. Это связано с тем, что при дискретизации интегральных уравнений возникают системы уравнений с плотными матрицами, в отличие от разреженных матриц в случае дифференциальных уравнений. В последнее время, в связи с развитием вычислительной техники и методов вычислительной математики, интегральные уравнения начали использоваться при численном решении конкретных задач. В работе предложены два метода численного решения двухмерных и трехмерных интегральных уравнений, описывающих многие важные классы задач математической физики.

Методы. Для дискретизации интегральных уравнений использовался метод коллокации на неравномерной и равномерной сетках. Для численного решения получившихся систем линейных алгебраических уравнений (СЛАУ) используются итерационные методы. Для случая равномерной сетки построен эффективный метод умножения матрицы СЛАУ на вектор.

Результаты. Построены соответствующие СЛАУ, описывающие рассматриваемые классы задач. Для решения систем уравнений, полученных с использованием равномерной сетки, предложены эффективные алгоритмы решения, использующие быстрое дискретное преобразование Фурье.

Выводы. СЛАУ с использованием неравномерной сетки имеют преимущество, связанное с хорошим описанием областей сложной конфигурации, но при этом есть существенные ограничения на размерность СЛАУ. При использовании равномерной сетки размерность СЛАУ может быть на несколько порядков больше, однако в этом случае могут возникать трудности с описанием сложной конфигурации области. Выбор того или иного метода зависит от конкретной задачи и имеющихся вычислительных ресурсов. Для многих двухмерных задач может быть предпочтительнее СЛАУ на неравномерной сетке, а для трехмерных задач – предпочтительнее СЛАУ на равномерной сетке.

Ключевые слова: интегральные уравнения, метод коллокации, быстрое преобразование Фурье

• **Поступила:** 14.12.2021 • **Доработана:** 10.01.2022 • **Принята к опубликованию:** 15.09.2022

Для цитирования: Самохин А.Б. Методы и эффективные алгоритмы решения многомерных интегральных уравнений. *Russ. Technol. J.* 2022;10(6):70–77. https://doi.org/10.32362/2500-316X-2022-10-6-70-77

Прозрачность финансовой деятельности: Автор не имеет финансовой заинтересованности в представленных материалах или методах.

Автор заявляет об отсутствии конфликта интересов.

INTRODUCTION

Let the bounded domain Q be given in Euclidean space E_n, where $n = 2, 3$. This means that Q is a figure on a plane ($n = 2$) or in space ($n = 3$). We shall consider the following integral equation in domain Q:

$$(1+\alpha\eta(x))u(x)+\int_Q \frac{K(x-y)}{R^m}\eta(y)u(y)dy = \\ = u^0(x),\ x\in Q,\ m\le n. \quad (1)$$

Here, $R = |x - y|$; $x = (x_1, \ldots, x_n)$; $y = (y_1, \ldots, y_n)$; α, η, K, u^0 are known functions with $K(x - y)$ being a differentiable coordinate function; u is the unknown function.

Equation (1) describes many practically significant classes of problems. Below are some problems in mathematical physics that may be reduced to equation (1):

- acoustic wave scattering on the inhomogeneous transparent obstacle [1]. In this case, $m < n$, $\alpha = 0$ while other functions included in equation (1) are scalar. Then the equation is the classical Fredholm integral equation of the 2nd kind;
- scattering of electromagnetic waves on the inhomogeneous, anisotropic in general, dielectric body [2, 3]. In this case, $m = n$, and therefore the integral operator in (1) is singular; u and u^0 are vector functions; and η and K are tensor functions. Value α determines the non-integral term of the singular operator and depends on the shape of the singularity and its center. For example, if the singularity is a sphere ($n = 3$) or a circle ($n = 2$), then $\alpha = 1/n$ [4, 5].

Other classes of problems in math physics may also be described using integral equations [6–8].

It is assumed that equation (1) has the only solution in the corresponding function space. It is possible to use only numerical methods for solution (1) when describing real problems. Then equation (1) is approximated by the system of linear algebraic equations (SLAE) with dense matrix using the Galerkin method or collocation method. In this case, the dimensionality N of resulting systems is usually very high ($N >> 1000$) due to the multidimensionality of the equation.

The main efficiency criteria of the numerical algorithm are the number of arithmetic operations T required for obtaining the original problem solution and the amount of computer memory M required for implementing the algorithm. When using the direct Gaussian method for solving SLAE, it is necessary to perform $T \sim N^3$ arithmetic operations and store about $M \sim N^2$ numbers in computer memory. It is clear that the solution for the considered problems requires significant computational resources. For iterative methods, the specified characteristics of algorithms are estimated by the following equations [9]:

$$T \sim LT_A,\ M \sim M_A, \quad (2)$$

where T_A is the number of arithmetic operations required to multiply the SLAE matrix by vector; L is the number of iterations required to obtain the solution with a given accuracy; M_A is the number of different matrix elements.

COLLOCATION METHOD

The collocation method is used for approximating integral equation (1) [3, 10, and 11]. Here, a significant difference between multidimensional problems and one-dimensional problems considered on interval $[a, b]$ should be noted. For such one-dimensional problems, the numerical solution does not cause problems related to describing the domain boundary. For two-dimensional and three-dimensional problems, certain difficulties arise in the discretization of integral equations defined in domains of complex shape.

The domain Q is presented as a union of N_Q cells $\Omega(i)$, $i = 1, \ldots, N_Q$. Nodal points in these cells are selected in their centers defined by the following equations [12]:

$$x_l^c = \frac{\int_\Omega x_l dx}{mes\,\Omega},\ l = 1,\ldots,n, \quad (3)$$

where $dx = dx_1dx_2$ for two-dimensional problems, $dx = dx_1dx_2dx_3$ for three-dimensional problems, and $x^c = (x_1^c,\ldots,x_n^c)$ is the center of the cell Ω with $mes\ \Omega$ being its volume ($n = 3$) or area ($n = 2$).

If the differentiable function of its arguments $f(x)$ is defined in domain Ω, then the following approximation is true:

$$\int_\Omega f(x)dx \approx f(x^c)mes\,\Omega. \quad (4)$$

Here, if $f(x)$ is a linear function of arguments, then expression (4) is exact equality.

If tetrahedrons ($n = 3$) or arbitrary triangles ($n = 2$) are considered as cells, then many complex configurations of domain Q may be described quite accurately. The center of corresponding cells is defined by the following simple equation:

$$x_l^c = \frac{\sum_{k=1}^{n+1} x_l^{(k)}}{n+1},\ l = 1,\ldots,n, \quad (5)$$

where $(x_1^{(k)},\ldots,x_n^{(k)})$ is Cartesian coordinates of the kth vertex of the cell.

We shall approximate integral equation (1) by SLAE of dimension $\sim N_Q$ with respect to the values of the unknown function at node points of domain Q located in centers x^{ci} of cells $\Omega(i)$, $i = 1, ..., N_Q$. The dimensions of cells are selected to provide the desired function weakly changing within a cell. Then corresponding SLAE may be written in the following form [3, 13]:

$$\gamma(i)u(i) + \sum_{j=1}^{N_Q} A(i,j)\eta(j)u(j) = u^0(i),$$
$$i = 1,..,N_Q,\ \gamma(i) = 1 + \alpha(i)\eta(i),$$

$$A(i,j) = \int\limits_{\Omega(j)} \frac{K(x^{ci} - y)}{\left|x^{ci} - y\right|^m} dy,\ i \neq j,\quad A(i,i) = 0,$$

$$u(i) = u(x^{ci}),\ u^0(i) = u^0(x^{ci}),\ \eta(i) = \eta(x^{ci}). \qquad (6)$$

For vector problems, tensor $\alpha(i)$ is determined by the shape of cell $\Omega(i)$ and its center.

Approximation (5) or more accurate numerical integration algorithms may be used for calculating integrals in (6). It should be noted that since nodal points are located in the center of cells, the approximation accuracy of integral operators is $\sim h^2$, where h is the maximum cell diameter (cell diameter is considered as the maximum distance between border points). For relatively small values $N_Q \leq 10\,000$, the system of equations (6) may be solved by direct or iterative methods. Efficient algorithms for solving the system of equations (6) using iterative methods are discussed below.

COLLOCATION METHOD ON UNIFORM GRID

In the kernel of integral equation (1), the term depending on the difference between the Cartesian coordinates of points x and y presents itself. However, this is not used in setting up SLAE (6). Below, we shall create the efficient numerical method for solving equation (1) using uniform grid and discrete Fourier transform (DFT).

We shall write some auxiliary equations using DFT first. Consider complex function $f(n)$ of discrete argument $n = 0, \pm1, \pm2, ...$ Here, it is assumed that $f(n)$ is a periodic function having period N, i.e., $f(n \pm N) = f(n)$ for any n.

The DTF of function $f(n)$ is defined by the following well-known equation:

$$F[f] = f^F(k) = \sum_{n=0}^{N-1} \exp\left(i\frac{2\pi}{N}kn\right) f(n),\ k = \overline{0, N-1}, \qquad (7)$$

where, obviously, the Fourier transform $f^F(k)$ is also a periodic function with period N.

If the Fourier transform $f^F(k)$ is known, then the initial function $f(n)$ may be restored using inverse DFT (IDFT), as follows:

$$F^{-1}[f^F] = f(n) =$$
$$= \frac{1}{N}\sum_{k=0}^{N-1} \exp\left(-i\frac{2\pi}{N}kn\right) f^F(k),\ n = \overline{0, N-1}. \qquad (8)$$

Generally, the number of arithmetic operations $T_F(N)$ required for calculating DFT without calculating functions of the form $\exp\left(i\frac{2\pi}{N}kn\right)$ additionally can be estimated by the following equation:

$$T_F(N) \sim N^2. \qquad (9)$$

When using fast Fourier transform (FFT) algorithms, the number of required arithmetic operations is estimated by the following equation [3]:

$$T_{FF}(N) \sim N\,\mathrm{LOG}(N), \qquad (10)$$

where LOG(N) is integer logarithm, i.e., the sum of all prime divisors of N. If N is a power of two, then $T_{FF}(N) \sim N \log2(N)$.

Let $A(l)$ be a periodic function of a discrete argument with period N. Consider sums of the following forms:

$$v(n) = \sum_{m=0}^{N-1} A(n-m)u(m),\quad n = \overline{0, N-1}. \qquad (11)$$

Sums (11) arise from multiplying circular matrices by vector. We shall apply DFT with period N to both parts of (11). It is easy to show that

$$v^F(k) = A^F(k)u^F(k),\quad k = \overline{0, N-1}. \qquad (12)$$

Using (12) and FFT algorithms, circular matrices may be efficiently multiplied by vector. However, circular matrices rarely appear in real problems. Although many problems, particularly those discussed below, require the calculation of sums of form (11) where the function $A(l)$, $-(N-1) \leq l \leq (N-1)$, is arbitrary within the specified range. Such sums arise from multiplying Toeplitz matrices by vector [14, 15]. The specified function $A(l)$ is defined at integer point $(2N-1)$. We shall predefine the function $A(l)$ by zero at point $l = N$ and extend it to all integer values with period $2N$. The function of discrete argument $u(m)$, $m = 0, ..., N-1$, is defined zero at points $m = N, ..., 2N-1$. We shall consider the sums of the following form:

$$v(n)=\sum_{m=0}^{2N-1}A(n-m)u(m),\quad n=\overline{0,2N-1}. \qquad (13)$$

It follows from the above that function $v(n)$ from (13) coincides with values $v(n)$ from (11) at $n = 0, ..., N - 1$. The following equation may be used for quickly calculating sums (13):

$$v^{\mathrm{F}}(k)=A^{\mathrm{F}}(k)u^{\mathrm{F}}(k),\quad k=\overline{0,2N-1}. \qquad (14)$$

In IDFT, components $v(n)$, $n=\overline{0,N-1}$, are of significance only. Thus, it follows from (10) that the number of arithmetic operations for calculating (11) may be estimated by the following equation:

$$T_{\mathrm{A}}\sim 2N\,\mathrm{LOG}(2N). \qquad (15)$$

In this case, the array with the number of the following elements should be stored in computer memory:

$$M_{\mathrm{A}}\sim 2N. \qquad (16)$$

Therefore, we shall discretize integral equation (1). First, three-dimensional problems are considered. Let us define parallelepiped P with domain Q being inside in the Cartesian rectangular coordinate system. The edges of the parallelepiped are parallel to the coordinate axes, while the edge lengths are equal to $N_l h_l$, l, = 1, 2, 3, where h_l are the grid steps in Cartesian coordinates. Then parallelepiped P can be represented as the union of cells (elementary parallelepipeds) P(p), $p = (p_1, p_2, p_3)$, $p_l = 0, \ldots, N_l - 1$. We shall define domain $\tilde{\mathrm{Q}}$ as the union of N_{Q} cells which centers are located inside domain Q. The nodal points at which function values are determined can be defined in the centers of cells and denoted as $x(p)$, while the function values at these points are denoted as $f(p)$.

Integral equation (1) is approximated, similarly to (6), using SLAE of the following form [5]:

$$\gamma(p)u(p)+\sum_{y(q)\in\mathrm{Q}}A(p-q)\eta(q)u(q)=u^0(p),\ x(p)\in\mathrm{Q},$$

$$A(p-q)=\int_{\Pi(q)}\frac{K(x(p)-y)}{|x(p)-y|^m}dy,\ p\neq q,\quad A(0)=0,$$

$$\gamma(p)=1+\alpha(p)\eta(p). \qquad (17)$$

Since the nodal points are located in the center of cells, the approximation accuracy of the integral operator is $\sim h^2$, $h=\sqrt{h_1^2+h_2^2+h_3^2}$.

It follows from (17) that the main computational costs when multiplying the SLAE matrix by vector (performing one iteration) are associated with calculating sums of the following form:

$$W(p)=\sum_{y(q)\in\mathrm{Q}}A(p-q)V(q),\quad x(p)\in\mathrm{Q}. \qquad (18)$$

For calculating $W(p)$ at nodal points $x(p) \in$ Q, it is necessary to perform $\sim N_{\mathrm{Q}}^2$ arithmetic operations, where N_{Q} is the number of nodal points in domain Q. However, the number of arithmetical operations may be reduced through using fast multiplication technique of Toeplitz matrices by vector discussed above.

We shall complete the definition of function $V(q)$ by zero at points $x(q)$ of parallelepiped P, which do not belong to Q. The following sums are considered:

$$W(p_1,p_2,p_3)=$$
$$=\sum_{q_1=0}^{N_1-1}\sum_{q_2=0}^{N_2-1}\sum_{q_3=0}^{N_3-1}A(p_1-q_1,p_2-q_2,p_3-q_3)V(q_1,q_2,q_3). \qquad (19)$$

Obviously, values $W(p)$ from (18) and (19) coincide at $x(p) \in$ Q. In (19), the matrix function of discrete argument $A(p)$ is defined for values $-(N_1 - 1) \le p_1 \le (N_1 - 1)$, $-(N_2 - 1) \le p_2 \le (N_2 - 1)$, $-(N_3 - 1) \le p_3 \le (N_3 - 1)$.

Let us denote the parallelepiped with sides $2N_1h_1$, $2N_2h_2$, and $2N_3h_3$ by P_2. We extend the matrix function of discrete argument $A(p_1, p_2, p_3)$ to all integer values p_1, p_2, and p_3 assuming it to be periodic for each variable with periods $2N_1$, $2N_2$, and $2N_3$, respectively. Here, we shall complete the definition of function $A(p_1, p_2, p_3)$ by zero at points where it is undefined. Let us further define the function of discrete argument $V(p_1, p_2, p_3)$ as zero at all nodal points P_2 not belonging to P and extend it to all integer values p_1, p_2, and p_3 assuming it to be periodic for each variable with periods $2N_1, 2N_2$, and $2N_3$, respectively.

We shall consider the following equation:

$$W(p_1,p_2,p_3)=$$
$$=\sum_{q_1=0}^{2N_1-1}\sum_{q_2=0}^{2N_2-1}\sum_{q_3=0}^{2N_3-1}A(p_1-q_1,p_2-q_2,p_3-q_3)V(q_1,q_2,q_3). \qquad (20)$$

Given the above, it is clear that function $W(p_1, p_2, p_3)$ from (20) coincides with values $W(p_1, p_2, p_3)$ from (18) at $x(p) \in$ Q. Integer parallelepipeds with the number of discrete arguments on each axis N_1, N_2, and N_3 and $2N_1$, $2N_2$, and $2N_3$ are denoted by P and P_2, respectively. Performing FDT on each variable from both parts of (20), the following equation may be written:

$$W^{\mathrm{F}}(k_1,k_2,k_3)=A^{\mathrm{F}}\left(k_1,k_2,k_3\right)V^{\mathrm{F}}(k_1,k_2,k_3),\ k\in\mathrm{P}_2. \qquad (21)$$

Thus, since transformation of function $A(p_1, p_2, p_3)$ is performed once before the iteration procedure, performing one iteration in solving SLAE (17) requires direct Fourier transformation of function $V(p_1, p_2, p_3)$ on each variable and an inverse transformation of function $W^F(k_1, k_2, k_3)$. The number of arithmetic operations and the amount of memory required for performing one iteration is estimated by the following equations:

$$T_A \sim 10N \, \mathrm{LOG}(N), \, M_A \sim 10N, \, N = N_1 N_2 N_3. \quad (22)$$

For solving two-dimensional problems in a Cartesian coordinate system, domain Q is located inside a defined rectangle. Further reasoning and calculations with obvious modifications repeat the considered case. The values M_A and T_A are evaluated by equations (22) where $N_3 = 1$.

When choosing grid steps and values N_1, N_2, and N_3 (three-dimensional problems) or N_1, N_2 (two-dimensional problems), it is necessary to be guided by the following criteria: first, the desired function varies little within the cells; second, domain $\tilde{Q}$, consisting of cells whose centers are inside Q, is sufficient for describing Q.

When using FFT algorithms, values N comprising multiples of a power of two are generally selected. However, when discretizing integral equations, this often results in significant additional computational costs due to the rather high duty cycle of numbers of a power of two. This may be examplified by the following.

Let $N_1 = N_2 = N_3 = N_0$, i.e., P is a cube. It is assumed that it would be sufficient to take value $N_0 = 150$ for approximating the solution with the required accuracy. The closest powers of two are numbers 128 and 256. Since the value 128 does not satisfy the requirement of approximating the solution, value $N_0 = 256$ should be taken for using conventional FFT. Let $T(N_0)$ be the number of arithmetic operations required to multiply the SLAE matrix by vector depending on values of N_0. Then the following maybe derived from (22):

$$\frac{T(256)}{T(150)} \approx \frac{256^3 \log_2(256)}{150^3 \, LOG(150)} \approx 2.5.$$

The amount of memory for storing the SLAE matrix at $N_0 = 256$ is also several times larger than at $N_0 = 150$. Thus, using FFT for $N_0 = 150$ is much more efficient than using FFT with a power of two.

It should be noted that, for solving SLAE (17) using the considered algorithm, only iterative methods can be used. This is due to iterative algorithms being based on multiplying the SLAE matrix by vector. The number of arithmetic operations and the amount of memory required for solving SLAE (17) are estimated by equations (2) and (22). At the same time, the number of iterations required for obtaining a solution is typically much smaller than the SLAE dimensionality. Thus, it may be possible to numerically solve integral equation (1), which is reduced to high dimension SLAE $N_Q > 10^6$.

CONCLUSIONS

The paper deals with two methods for obtaining a numerical solution of two- and three-dimensional integral equations describing many significant classes of problems in mathematical physics. Here, the collocation method on non-uniform and uniform grids is used for discretizing equations. The corresponding SLAEs are set up. SLAEs using non-uniform grid may have the advantage of describing complex configuration domains well. However, there are significant constraints on the SLAE dimensionality. While the SLAE dimensionality may be several orders of magnitude higher when using a uniform grid, difficulties may arise when describing the domain complex configuration. The selection of one or another method depends on the specific problem and available computational resources. In the author's opinion, SLAE on a non-uniform grid may be preferable for many two-dimensional problems, while SLAE on a uniform grid may be preferable for three-dimensional problems.

ACKNOWLEDGMENTS

The study was supported by the Russian Foundation for Basic Research, grant No. 20-07-00006.

REFERENCES

1. Colton D., Kress R. *Inverse Acoustic and Electromagnetic Scattering Theory*. Applied Mathematical Sciences. book series. (AMS, V. 93). 4th ed. Springer; 2019. 535 p. ISBN 978-3030-303-50-1
2. Samokhin A.B. *Integral'nye uravneniya i iteratsionnye metody v elektromagnitnom rasseyanii* (*Integral Equations and Iteration Methods in Electromagnetic Scattering*). Moscow: Radio i svyaz'; 1998. 160 p. (in Russ.). ISBN 5-256-01405-6
3. Samokhin A.B. *Ob"emnye singulyarnye integral'nye uravneniya elektrodinamiki* (*Volume Singular Integral Equations of Electromagnetics*). Moscow: Tekhnosfera; 2021. 218 p. (in Russ.). ISBN 978-5-94836-618-0
4. Michlin S., Prössdorf S. *Singular Integral Operators*. Berlin-New York: Akademie-Verlag; 1986. 528 p.
5. Samokhin A.B., Samokhina A.S., Shestopalov Yu.V. Discretization methods for three-dimensional singular integral equations of electromagnetism. *Differential Equations*. 2018;54(9):1225–1235. https://doi.org/10.1134/S0012266118090100
6. Vasil'ev E.N. *Vozbuzhdenie tel vrashcheniya* (*Excitation of Bodies of Rotation*). Moscow: Radio i svyaz'; 1987. 270 p. (in Russ.).
7. Dmitriev V.I., Zakharov E.V. *Integral'nye uravneniya v kraevykh zadachakh elektrodinamiki* (*Integral Equations for Boundary Problems of Electromagnetics*). Moscow: MGU; 1987. 165 p. (in Russ.).
8. Il'inskii A.S., Kravtsov V.V., Sveshnikov A.G. *Matematicheskie modeli elektrodinamiki* (*Mathematical Models of Electrodynamics*). Moscow: Vysshaya shkola; 1991. 222 p. (in Russ.).
9. Samokhin A.B., Tyrtyshnikov E.E. Numerical method for solving volume integral equations on a nonuniform grid. *Comput. Math. Math. Phys*. 2021;61(5):847–853. https://doi.org/10.1134/S0965542521050158
   [Original Russian Text: Samokhin A.B., Tyrtyshnikov E.E. Numerical method for solving volume integral equations on a nonuniform grid. *Zhurnal vychislitel'nyi matematiki i matematicheskoi fiziki*. 2021;61(5):878–884 (in Russ.). https://doi.org/10.31857/S0044466921050161]
10. Demidovich B.P., Maron I.A., Shuvalova E.Z. *Chislennye metody analiza. Priblizhenie funktsii, differentsial'nye i integral'nye uravneniya* (*Numerical Methods of Analysis. Approximation, Differential and Integral Equations*). St. Petersburg: Lan'; 2021. 400 p. (in Russ.). ISBN 978-58114-0799-6
11. Kudryashova N.Yu., Grunina T.V. *Granichnye integral'nye uravneniya* (*Boundary Integral Equations*). Penza: PGU; 2018. 72 p. (in Russ.). ISBN 978-5-907-102-47-7.
12. Nurutdinova I.N., Pozharskii D.A. *Integral'noe ischislenie*. *Differentsial'nye uravneniya*. *Ryady* (*Integral Calculus*. *Differential Equations*. *Rows*). Rostov-on-Don: Donskoi GTU; 2021. 96 p. (in Russ.).
13. Samarskii A.A., Gulin A.V. *Chislennye metody* (*Numerical Methods*). Moscow: Nauka; 1989. 432 p. (in Russ). ISBN 5-02-102-013996-3
14. Voevodin V.V., Tyrtyshnikov E.E. *Vychislitel'nye protsessy s teplitsevymi matritsami* (*Computational Processes with Toeplitz Matrices*). Moscow: Nauka; 1987. 319 p. (in Russ.).
15. Tyrtyshnikov E.E. *Metody chislennogo analiza* (*Methods of Numerical Analysis*). Moscow: Akademiya; 2007. 317 p. (in Russ). ISBN 978-5-7695-3925-1

СПИСОК ЛИТЕРАТУРЫ

1. Colton D., Kress R. *Inverse Acoustic and Electromagnetic Scattering Theory*. Applied Mathematical Sciences. Book series. (AMS, V. 93). 4th ed. Springer; 2019. 535 p. ISBN 978-3030-303-50-1
2. Самохин А.Б. *Интегральные уравнения и итерационные методы в электромагнитном рассеянии*. М.: Радио и связь; 1998. 160 с. ISBN 5-256-01405-6
3. Самохин А.Б. *Объемные сингулярные интегральные уравнения электродинамики*. М.: Техносфера; 2021. 218 с. ISBN 978-5-94836-618-0
4. Michlin S., Prössdorf S. *Singular Integral Operators*. Berlin-New York: Akademie-Verlag; 1986. 528 p.
5. Самохин А.Б., Самохина А.С., Шестопалов Ю.В. Методы дискретизации объемных сингулярных интегральных уравнений электромагнетизма. *Дифференциальные уравнения*. 2018;54(9):1251–1261. https://doi.org/10.1134/S0374064118090108
6. Васильев Е.Н. *Возбуждение тел вращения*. М.: Радио и связь; 1987. 270 с.
7. Дмитриев В.И., Захаров Е.В. *Интегральные уравнения в краевых задачах электродинамики*. М.: Изд-во МГУ; 1987. 165 с.
8. Ильинский А.С., Кравцов В.В., Свешников А.Г. *Математические модели электродинамики*. М.: Высшая школа; 1991. 222 с.
9. Самохин А.Б., Тыртышников Е.Е. Численный метод решения объемных интегральных уравнений на неравномерной сетке. *Журн. вычисл. матем. и матем. физ*. 2021;61(5):878–884. https://doi.org/10.31857/S0044466921050161
10. Демидович Б.П., Марон И.А., Шувалова Э.З. *Численные методы анализа. Приближение функций, дифференциальные и интегральные уравнения*. СПб.: Лань; 2021. 400 с. ISBN 978-58114-0799-6
11. Кудряшова Н.Ю., Грунина Т.В. *Граничные интегральные уравнения*. Пенза: Изд-во ПГУ; 2018. 72 с. ISBN 978-5-907-102-47-7
12. Нурутдинова И.Н., Пожарский Д.А. *Интегральное исчисление. Дифференциальные уравнения. Ряды*. Ростов-на-Дону: Донской ГТУ; 2021. 96 с.
13. Самарский А.А., Гулин А.В. *Численные методы*. М.: Наука; 1989. 432 с. ISBN 5-02-102-013996-3
14. Воеводин В.В., Тыртышников Е.Е. *Вычислительные процессы с теплицевыми матрицами*. М.: Наука; 1987. 319 с.
15. Тыртышников Е.Е. *Методы численного анализа*. М.: Академия; 2007. 317 с. ISBN 978-5-7695-3925-1

About the author

Alexander B. Samokhin, Dr. Sci. (Phys.-Math.), Professor, Applied Mathematics Department, Institute of Information Technology, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: absamokhin@yandex.ru. Scopus Author ID 7005200099, RSCI SPIN-code 6302-0596, http://orcid.org/0000-0003-1328-6725

Об авторе

Самохин Александр Борисович, д.ф.-м.н., профессор, Заслуженный деятель науки РФ, профессор кафедры «Прикладная математика» Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: absamokhin@yandex.ru. Scopus Author ID 7005200099, SPIN-код РИНЦ 6302-0596, http://orcid.org/0000-0003-1328-6725

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Philosophical foundations of technology and society

Мировоззренческие основы технологии и общества

UDC 378
https://doi.org/10.32362/2500-316X-2022-10-6-78-90

RESEARCH ARTICLE

Model of formation of digital competences in implementing higher education programs

Olga G. Savka @,
Maria N. Gusarova,
Svetlana V. Sumina,
Yaroslav O. Knyazev,
Denis A. Bezrukov

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

Abstract

Objectives. The paper presents and analyzes a model for the formation and evaluation of digital competencies in students. The model is aimed at the implementation of higher education programs for training specialists not directly working in IT, but whose activities are directly related to using ready-made digital products. Digital competences imply an ability to confidently, effectively, and safely select and apply information and communication technologies in various life practices including researching and critically analyzing information, using digital devices and accessing social network functionality, conducting financial and trading transactions, as well as creating digital content. The formation of such digital competences is one of the results of completing higher education programs.

Methods. The study is based on a model of digital competence formation having the following four interconnected stages: basic digital competences; personal competences (soft skills); professional digital competences; digital culture.

Results. The presented general model for the formation and assessment of student digital competences in higher education programs consists of four interrelated steps, each integral to the process of formation and assessment of the student digital competences, none of which can be excluded without the risk of failing to achieving the specified goals.

Conclusions. The model developed in the paper is based on the existing extensive regulatory framework, as well as existing domestic and foreign practices. Relying on expert community opinion (employers, primarily), it accounts for sector- and region-specific features of universities along with specifics of training areas, as well as comprising a list of optimal organizational and methodological conditions for formation of digital competency.

Keywords: digital competence, digital literacy, digital continuum, model of digital competences, assessment, professional competences, digital culture, information and communication technologies

• **Submitted:** 14.03.2022 • **Revised:** 21.04.2022 • **Accepted:** 13.09.2022

For citation: Savka O.G., Gusarova M.N., Sumina S.V., Knyazev Ya.O., Bezrukov D.A. Model of formation of digital competences in implementing higher education programs. *Russ. Technol. J.* 2022;10(6):78–90. https://doi.org/10.32362/2500-316X-2022-10-6-78-90

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

Модель формирования цифровых компетенций при реализации программ высшего образования

О.Г. Савка [@],
М.Н. Гусарова,
С.В. Сумина,
Я.О. Князев,
Д.А. Безруков

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

Резюме

Цели. Целью работы является анализ модели формирования и оценки цифровых компетенций обучающихся при реализации программ высшего образования, ориентированной на подготовку специалистов, не относящихся непосредственно к ИТ-сфере, но деятельность которых напрямую будет связана с применением цифровых продуктов. Под цифровыми компетенциями индивида следует понимать его способность уверенно, эффективно и безопасно выбирать и применять в различных жизненных практиках информационно-коммуникационные технологии (ИКТ), в т.ч. осуществлять поиск и критическое осмысление информации; использовать цифровые устройства, функционал социальных сетей; выполнять в онлайн-режиме финансовые и торговые операции; производить цифровой контент. Формирование цифровых компетенций у обучающихся – один из результатов освоения программ высшего образования.

Методы. Использована модель формирования цифровых компетенций, включающая четыре связанных между собой этапа: базовые цифровые компетенции, личностные компетенции (soft skills), профессиональные цифровые компетенции, цифровая культура.

Результаты. Разработана базовая (общая) модель формирования и оценки цифровых компетенций обучающихся при реализации программ высшего образования. Она состоит из четырех взаимосвязанных ступеней, каждая из которых является неотъемлемой частью процесса формирования и оценки цифровых компетенций обучающегося и не может быть исключена без риска недостижения поставленных задач.

Выводы. Разработанная авторами модель основана на обширной нормативно-правовой базе, существующих отечественных и зарубежных практиках; опирается на мнение экспертного сообщества (прежде всего, работодателей); учитывает отраслевые и региональные особенности вуза, специфику направлений подготовки; содержит перечень оптимальных организационно-методических условий формирования цифровых компетенций.

Ключевые слова: цифровая компетентность, цифровая грамотность, цифровой континуум, модель цифровых компетенций, ассесмент, профессиональные компетенции, цифровая культура, информационно-коммуникационные технологии

• **Поступила:** 14.03.2022 • **Доработана:** 21.04.2022 • **Принята к опубликованию:** 13.09.2022

Для цитирования: Савка О.Г., Гусарова М.Н., Сумина С.В., Князев Я.О., Безруков Д.А. Модель формирования цифровых компетенций при реализации программ высшего образования. *Russ. Technol. J.* 2022;10(6):78–90. https://doi.org/10.32362/2500-316X-2022-10-6-78-90

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

Авторы заявляют об отсутствии конфликта интересов.

GLOSSARY

Digitalization is the process of spreading and implementing digital technologies in various spheres of society: economy, culture, education, etc.

Digital competence is the ability to confidently, effectively, critically, and safely select and apply information and communication technologies in different spheres of life (information environment, communications, consumption, and the technosphere) based on continuous competence mastering (knowledge, skills, motivation, and responsibility), as well as the readiness to participate in such activities.

Digital continuum is the set of tools, techniques, and forms of education integrating the interaction of all educational process participants both within and outside the educational program, based on the use of digital technologies necessary for forming skills and competences demanded by the current labor market.

INTRODUCTION

Digital competences may be represented as the ability to apply information and communication technologies (ICT) in various spheres of life, namely, in creating and distributing digital content, as well as solving complex problems in the digital environment (e.g., project and business tasks). The rapid development of digital technologies necessitates the continuous updating and expansion of digital competences of citizens in order to effectively build a digital society.

It seems reasonable to clarify the above definition of "digital competence." In studies on this topic, various approaches to defining digital competence have been analyzed[1, 2] [1]. For example, in the analytical report for 2012 devoted to analyzing the identification and transformation of digital competences, A. Ferrari has proposed the following definition: "*Digital competence is the set of knowledge, skills, attitudes, abilities, strategies, and awareness that is required when using ICT and digital media to perform tasks, solve problems, communicate, manage information, behave in an ethical and responsible way, collaborate, create and share content and knowledge for work, leisure, participation, learning, socializing, empowerment, and consumerism*" [2].

According to Ya.V. Dmitrieva and I.A. Alyabin [3], digital competence entails a set of several digital skills and acquired knowledge necessary for continuous application in professional activities. Since digital competence simultaneously "encompasses information management, collaboration, communication and sharing, content and knowledge creation, ethics and responsibility, assessment and problem solving, and technical operations," it represents the sum of all digital competences a person possesses.

In the study carried out by I.A. Volkova and V.S. Petrova [4], it is assumed that digital competences are the key to developing the digital economy not only as an IT platform, but also as a digital product. In their view, the following digitalization stages may be conventionally distinguished: 1) digital inclusion; 2) digital law; 3) schools; 4) further education; 5) higher education; and 6) digital apprenticeship. Accordingly, within each of the listed stages, five areas are identified through which digital competence should pass: 1) digital financial literacy; 2) digital information literacy; 3) digital capture and exchange of cultural artifacts; 4) remote work; and 5) business data and systems integration.

The definition of digital competence proposed by G.U. Soldatova covers a wide range of problems and practices related to using ICT in modern society: "*the ability of a person to confidently, effectively, critically and safely select and apply information and communication technologies in various spheres of life* (*information environment, communication, consumption, and technosphere*) *based on continuous mastering the competences* (*knowledge, skills, motivation, and responsibility*)*, as well as his or her readiness for such activity focusing on the efficient result*" [5].

In the digital competence structure proposed by Soldatova, knowledge, abilities, skills, motivation, and responsibility are included. An important component of the last element is safety. Each of these components is implemented to a greater or lesser extent in various spheres of activity in the Internet environment. Four types of digital competence are identified: *media competence* aimed at finding, comprehending, critically understanding digital information, and creating information objects using digital resources (text, image, and audio/visual); *communicative competence* necessary for various forms of communication (e-mail, chat rooms, blogs, forums, social networks, etc.); *technical competence* forming responsible use of hardware and software tools for solving various problems; and *consumer competence* based on solving everyday tasks and life situations using digital devices and the Internet.

The conceptual framework for defining digital competence based on studying cognitive and metacognitive dimensions related to technologies is proposed by A. Kalvani et al., called the digital competence framework [6]. The authors find digital competence to consist in the convergence of three basic dimensions: technological, cognitive, and ethical. The

[1] Country digital readiness: Research to determine a country's digital readiness and key interventions. https://www.cisco.com/c/dam/assets/csr/pdf/Country-Digital-Readiness-White-Paper-US.pdf. Accessed November 15, 2021.

[2] Djumalieva J., Sleeman C. Which digital skills do you really need? 2018. Nesta. The Innovation Foundation. https://media.nesta.org.uk/documents/Which_digital_skills_do_you_really_need.pdf. Accessed November 15, 2021.

technological dimension involves the ability to create technologies, the cognitive dimension demonstrates the ability to critically evaluate digital text and data, while the ethical dimension entails the competency of consciously interacting effectively with others using various technological platforms.

In many cases, the study of digital competences is based on using the concept of professional competences. For example, the KSAO concept is based on knowledge (K), skills (S), abilities (A), and other characteristics (O) required by professionals for doing their job effectively. The technological acceptance model describes the extent to which citizens have faith in their technological abilities (competences) and how they use government e-services.

The study of digital competence can also be based on sociocognitive and planned behavior theories. According to such psychological theories, an understanding and evaluation of digital competence is based on the concept of self-efficacy. The underlying knowledge providing insight into the cognitive determinants of behavior includes a person's beliefs about his or her ability to be successful; here, the power of control may function as an indicator for predicting a person's intentions [7].

The above suggests that digital competences can be understood as a person's ability to confidently, effectively, and safely select and apply ICT as part of various life practices. These include searching and critically evaluating information; using digital devices and social networking functionalities; conducting online financial and trade transactions; and producing digital content. Such abilities are based on the relevant knowledge, skills, and abilities in the field of digital development acquired in the process of mastering various educational trajectories (secondary, higher, and additional professional education). At the same time, the formation and evaluation of digital competences of higher education (HE) students as specified by the relevant federal state educational standards (FSES) should be reflected in implementing HE programs [8].

METHODS

The developed model for forming digital competences is based on the set of methodological materials (developed by leading educational organizations in the country) presenting various digital competence models (DCMs). Each of the analyzed DCMs forms the appropriate focus and content of actions for implementing it in the educational strategy of the university.

The analysis of domestic DCMs allows suggesting that the professional and personal competences supplemented with variances of basic competences and digital culture (digital ethics) are most often identified by their developers relying on the voluminous body of foreign literature covering this issue. When describing the key characteristics of professional and personal competences, foreign and Russian researchers adhere to general methods and principles. For example, the competence model proposed by researchers from the Russian Presidential Academy of National Economy and Public Administration includes four interrelated blocks: basic digital competences, personal competences, professional competences, and digital culture [9].

In the model, personal competences (soft skills) are represented by the following groups of competences:

1) *result orientation*—implying the ability to simulate different development scenarios and thereby achieve digital development goals in the most effective way);
2) *customer centricity*—implying the ability to build trusting long-term relationships between all interested parties);
3) *communicativeness*—implying selecting the most effective communication strategy and tactics;
4) *emotional intelligence*—implying creating the favorable psychological and emotional climate for teamwork;
5) *creativity*—implying the ability to find a way out of difficult situations using non-standard approaches;
6) *criticality*—implying formation of different scenarios for achieving strategic goals as well as the ability to create concepts and versions of strategies for different time periods.

When all indicators cited by the researchers have been achieved, a person is enabled to successfully participate in implementing a digital transformation strategy and participated in digital development projects.

In our view, it may be also reasonable to consider the block of professional competences (hard skills) in the field of digital development whose formation determines the management of processes, projects, products of digital transformation, and regular solving difficult professional tasks in the digital environment. Six key professional competences are presented in the model. Each group may be characterized by specifying the following most relevant indicators:

1) *digital development management*—implying knowledge, skills and abilities to apply the strategic management tools, techniques, and approaches in digital development management;
2) *organizational culture development*—implying knowledge, skills, and abilities to apply tools and techniques of organizational culture formation and broadcasting as well as organizational change management techniques;

3) *management tools*—implying knowledge, skills, and abilities to apply tools and techniques of the project approach (basics of project management; project management system at the state and organization level; projects, project programs, and project portfolios); and project life cycle management processes (initiation, preparation, implementation, monitoring, control, and completion);
4) *data management and use*—implying knowledge, skills, and abilities to apply data-driven decision-making (DDDM) technologies (DDDM culture and ethics; DDDM integration into business processes of the organization; automatic decision-making systems including artificial intelligence systems; ensuring data security; and data lifecycle management techniques (data model design, data lifecycle stages);
5) *application of digital technologies*—implying knowledge, skills, and abilities to apply end-to-end (E2E) technologies (new production technologies, neurotechnologies and artificial intelligence, wireless communication technologies, components of robotics and sensorics, quantum technologies, distributed registry systems, and virtual and augmented reality technologies); means and methods of information security and cybersecurity;
6) *IT infrastructure development*—implying knowledge, skills, and abilities in the field of technical documentation regulating functioning of information systems and IT products.

In the model, a digital culture comprises the system of values, attitudes, and norms and rules of behavior supported and broadcasted by the digital transformation team.

The competence model proposed by the NAFI multidisciplinary analytical center[3] is no less significant for the study. It identifies the following competences: 1) information literacy (searching for information on various resources and assessing its usefulness and harm); 2) computer literacy (ease of use of any digital devices); 3) media literacy (searching for and filtering news from various sources); 4) communication literacy (use of digital communication, i.e., social networks and messengers); and 5) attitudes towards technological innovation (readiness to quickly adapt to new technologies).

Here, the model proposed by E.A. Khodyreva from Innopolis University is worth mentioning. In the report "Assessing digital competences in educational programs" presented at the IV Annual Meeting of Leaders in the Field of Education, Knowledge, and Digitalization Management held in Sochi (October 2021), she specified the following main DCM blocks: *basic digital competences* (digital skills); *professional competences in digital development including industry-specific ones* (hard skills); *personal competences in digital development* (soft skills). Universal professional competences include organization of project activities, business processes, customer centricity, data culture, application of digital products in industry (professional activity), presentation (visualization), delivery of information using data, and prototyping. Personal competences include focusing on results, leadership, communication, critical thinking, self-development, resource management, strategic thinking, creative thinking, and teamwork. In the report, information literacy, communication literacy, digital content creation, digital security, and problem-solving skills in the digital environment are referred to as basic digital competences.

The theoretical and methodological DCM analysis carried out in the present study revealed a significant segment of common characteristics: universality (the ability to update the model for organizations, departments, and teams of any type and economic sector); scalability (the ability to adjust the set of competences to corporate characteristics and industry focus); the presence of block structure (attention is paid not only to competences, but also to the values underlying digital transformation); the ability to create both own models and separate products (e.g., personal trajectories of professional development in digital transformation, personal digital profiles of digital transformation participants, etc.) based on it.

It is reasonable to apply the principles and methods embedded in DCMs, which have been developed by Russian educational organizations for creating a model of digital competence formation (DCF), to identify a step-by-step process integrated into university educational trajectories. At the present stage, the issue of applying the DCF model in implementing the HE program is yet to be comprehensively studied due to the relevant practices being in the process of formation. This gap may be filled by the DCF model proposed in the paper for areas not related to the ICT professional sphere.

RESULTS

The basic (general) model of formation and assessment of student digital competences in implementing HE programs developed in the paper consists of four interrelated stages, none of which can be excluded without the risk of failing to achieving the defined goals due to being integral to the process of formation and assessment of the digital competences

[3] Digital Literacy of Russians: A 2020 study. https://nafi.ru/analytics/tsifrovaya-gramotnost-rossiyan-issledovanie-2020/. Accessed December 3, 2020 (in Russ.).

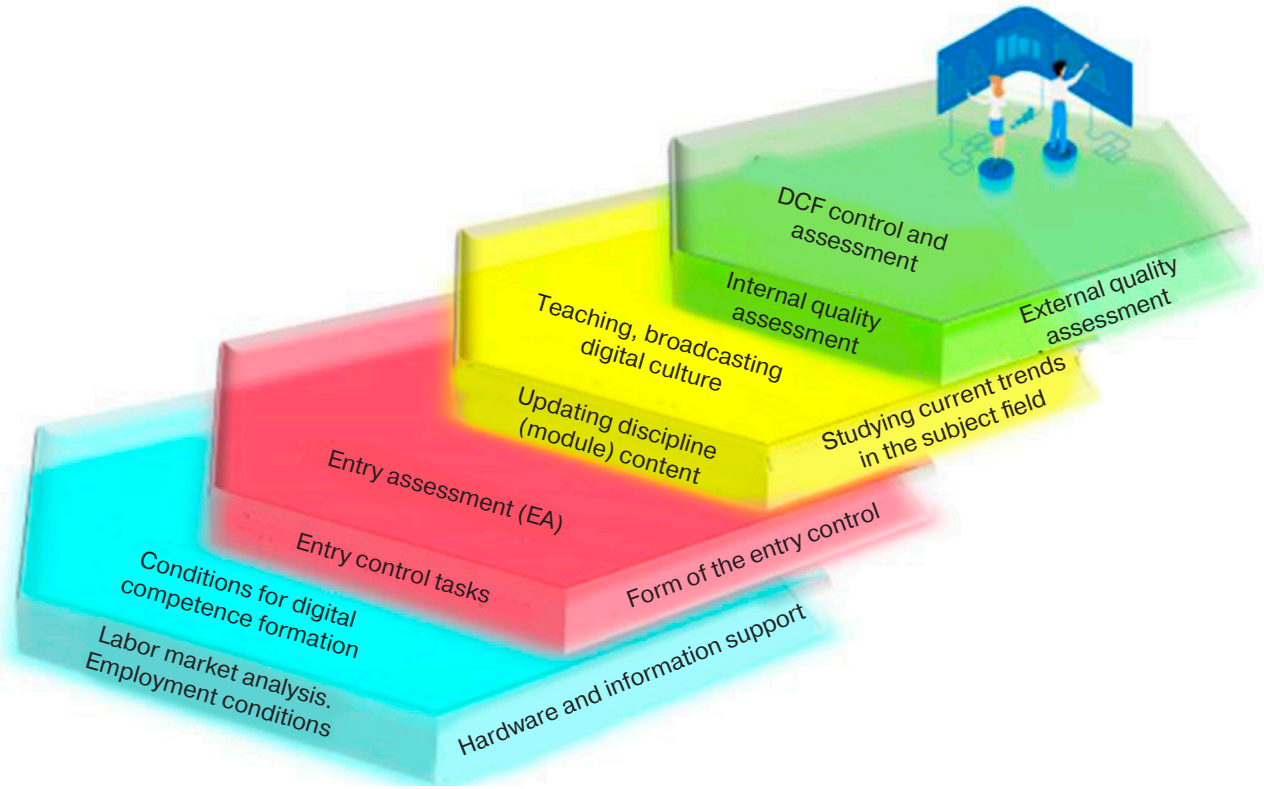


Fig. 1. The basic (general) model of formation and assessment of student digital competences in implementing HE programs

of students (Fig. 1). The model is focused on training professionals for working in areas not directly related to the ICT field, but whose activities are directly related to applying final digital products.

Implementing the first step of the proposed model being ***the conditions for the formation of digital competences*** is a step-by-step process including monitoring the labor market, creating the digital educational continuum, and staffing the application of digital technologies in the educational process.

In our opinion, it is reasonable to involve published sources containing the results of sociological surveys on requirements of employers, not only for qualifications, but also primarily for the digital skills and competences of professionals according to labor market analysis. A review of the relevant literature[4] [10, 11] concludes that the digital economy requires personnel not just oriented towards the digital environment, but also able to design and create it by themselves, as well as to actively apply all its resources and tools in their professional activity. Of paramount importance here is not the amount of information the person receives but the knowledge and skills for effectively applying and managing it, including the use of digital devices, communication applications, and networks. According to the data of the HeadHunter[5] and Superjob[6] companies, which represent the leading national online recruitment agencies, today's employer is more interested in transprofessional competences, which allow the professional not only to interact with digital products and solutions, but literally to immerse him/herself inside the virtual content, modeling and developing it.

The second stage of the first step includes the development and formation of the digital educational continuum as a basic component of organizational and methodological support for DCF in implementing HE programs. In the paper, this is defined as the set of tools, techniques, and forms of education integrating the interaction of all educational process participants, both within and outside the educational program, based on the use of digital technologies necessary for forming skills and competences demanded by the current labor market.

The digital continuum implies much more than the electronic information-educational environment due to representing an effective and safe digital environment for students; its components are educational content, university mobile applications, and electronic interaction

[4] Sorokin D. Participation of the Russian Federation subjects in the implementation of the "Personnel for the Digital Economy" federal project. https://d-russia.ru/wp-content/uploads/2019/09/sorokin.pdf. Accessed December 8, 2021 (in Russ.).

[5] https://hh.ru/. Accessed April 6, 2022 (in Russ.).

[6] https://www.superjob.ru/. Accessed April 6, 2022 (in Russ.).

channels. It is reasonable to assert the creation of a fully digital environment wherein the student is not only the user but also the full participant modeling it, for example, for his/her personal educational trajectories. The digital continuum formation accelerates immersion in the environment of modern digital interactions similar to the environment the graduate finds him/herself after graduation.

An analysis of employer requirements concerning digital skills and competences of graduates identifies a set of hardware and software tools required for setting up the digital continuum[7] [12]. In terms of free software, *1C:Enterprise 8.3*[8] and *Kompas-3D LT*[9] may be recommended, while *EOS*[10], *T-FLEX CAD*[11], *Astra Linux*[12], and *QForm*[13] are examples of licensed software for implementing HE programs. Certainly, this list is not exhaustive. The list of software with regard for the required digital skills in the professional field should be independently determined by the educational organization.

Using software products requires appropriate hardware. The RTU MIREA long-term strategy for increasing the hardware base of the educational process proves that computer classrooms should be equipped with modern multimedia equipment including the projection screen, multimedia projector, plasma panel, camcorder, computers, video conferencing equipment, and audio equipment.

The formation of a digital continuum also implies using modern professional databases, information reference systems, and electronic library systems relevant for corresponding HE programs. It would be advisable to form such a list in consultation with representatives of organizations providing the base for practical training, as well as with potential employers from state organizations and commercial structures.

We believe that the number of current professional databases and information reference systems should include "1C:Enterprise for educational institutions via the Internet,"[14] CLE International[15], COMSOL Multiphysics®[16] software for multiphysical modeling, eLIBRARY.RU[17], etc. Among electronic library systems can be identified, for example, Urait[18], IPR Books[19], EBS Znanium[20], and University Library Online[21].

Staffing of the educational process being the third stage of the first step implies formation of the teaching staff from among teachers capable of applying ICTs in the educational process and integrating digital literacy with their other professional competences. Obviously, a contemporary university teacher—just like professional in any other field—should strive to develop his/her transprofessional skills focused on immersing in the digital environment and work in it. This would allow the teacher to respond flexibly to the development of new teaching models and maintain their competitiveness in the labor market. The educational organization, in its turn, should develop the systematic approach to training and retraining the teaching staff in the field of digital literacy, digital technologies, products, and tools. Practitioners having digital competences can be helpfully involved in the educational process.

The study allows formulating the following requirements for the teacher concerning mastering necessary digital competences: use of the digital environment for organizing the learning process, search and analysis of information, cybersecurity, digital ethics, information management, communication in the digital environment, and continuous self-education.

Implementing the first step in the proposed model is designing the list of required indicators for digital competences (professional and personal).

The entry assessment, comprising the second stage of the proposed model, determines the basic level of student knowledge towards which the DCF educational process is aimed. The entry control determines the necessary requirements as to the content of disciplines (modules) forming digital competences.

The first stage of the second step includes defining the most complete list of basic digital competences and developing the entry control tasks for identifying DCF level among students. Basic digital competences are the basic level of knowledge and skills for using ICT in professional activities and common life practices. It is assumed to be not lower than the level defined by FSES for secondary education.

The entry control content should fully reflect the subject fields related to digital technologies. Thus, the

[7] A set of materials on digital skills. Geneva, 2018:6–7. https://psihdocs.ru/komplekt-materialov-po-cifrovim-navikam.html. Accessed January 18, 2022 (in Russ.).

[8] https://v8.1c.ru/podderzhka-i-obuchenie/uchebnye-versii/distributiv-1s-predpriyatie-8-3-versiya-dlya-obucheniya-programmirovaniyu/. Accessed April 6, 2022 (in Russ.).

[9] https://kompas.ru/kompas-3d-lt/about/. Accessed April 6, 2022 (in Russ.).

[10] https://eos.ru/. Accessed April 6, 2022 (in Russ.).

[11] https://www.tflexcad.ru/. Accessed April 6, 2022 (in Russ.).

[12] https://astralinux.ru/. Accessed April 6, 2022 (in Russ.).

[13] https://qform3d.ru/. Accessed April 6, 2022 (in Russ.).

[14] https://edu.1cfresh.com. Accessed April 6, 2022 (in Russ.).

[15] https://www.cle-international.com. Accessed April 6, 2022 (in Russ.).

[16] https://www.comsol.ru. Accessed April 6, 2022 (in Russ.).

[17] www.elibrary.ru. Accessed April 6, 2022 (in Russ.).

[18] https://urait.ru/. Accessed April 6, 2022 (in Russ.).

[19] https://www.iprbookshop.ru/. Accessed April 6, 2022 (in Russ.).

[20] https://znanium.com/. Accessed April 6, 2022 (in Russ.).

[21] https://biblioclub.ru/. Accessed April 6, 2022 (in Russ.).

following topics of questions/tasks for determining the basic digital competence level of students may be recommended:

- formation and evolution of information society and information technologies;
- basic terms and concepts of digitalization and digital economy;
- the role of information in the modern world;
- information and digital technologies in the media segment and creative activities;
- digital technologies in communication;
- safety engineering when using office equipment;
- information security;
- ethical behavior in the digital environment;
- selecting digital tools and applications required for solving assigned tasks;
- searching the Internet for information;
- analysis and processing of digital information;
- algorithms and algorithmic models for solving professional tasks;
- data storage and processing;
- text editors;
- electronic spreadsheets;
- databases.

At the second stage of the entry assessment (second step), it is necessary to decide on the format of the entry control and its implementation (test, oral questioning, questionnaire, portfolio, essay, and case study). The forms of entry control and using the results are shown in Fig. 2.

Due to each format having its own advantages and disadvantages, it seems reasonable to use the combination of various formats. For example, testing may reveal mastery of terminology and the digital knowledge basics while case study or questionnaire reveals their depth.

The third step of the DCF model is focused on teaching and broadcasting digital culture. This step forms the digital competences required for successful professional activity in the digital economy. Its implementation is based on the achievements of the first and second stages. The results are used to determine the updating of educational disciplines (modules) in DCF terms.

The study suggests that indicators of the required DCF level with allowance for the subject field characteristics, available tools for assessing their formation, and current trends in the digital economy should be relied on when updating working programs of disciplines (modules). The unity of pedagogical requirements for each digital competence may be achieved by developing the unified assessment tool for student progress monitoring and midterm attestation in all educational disciplines.

Digital competences are formed throughout the entire period of study. This is achieved through a combination of various factors. They include the

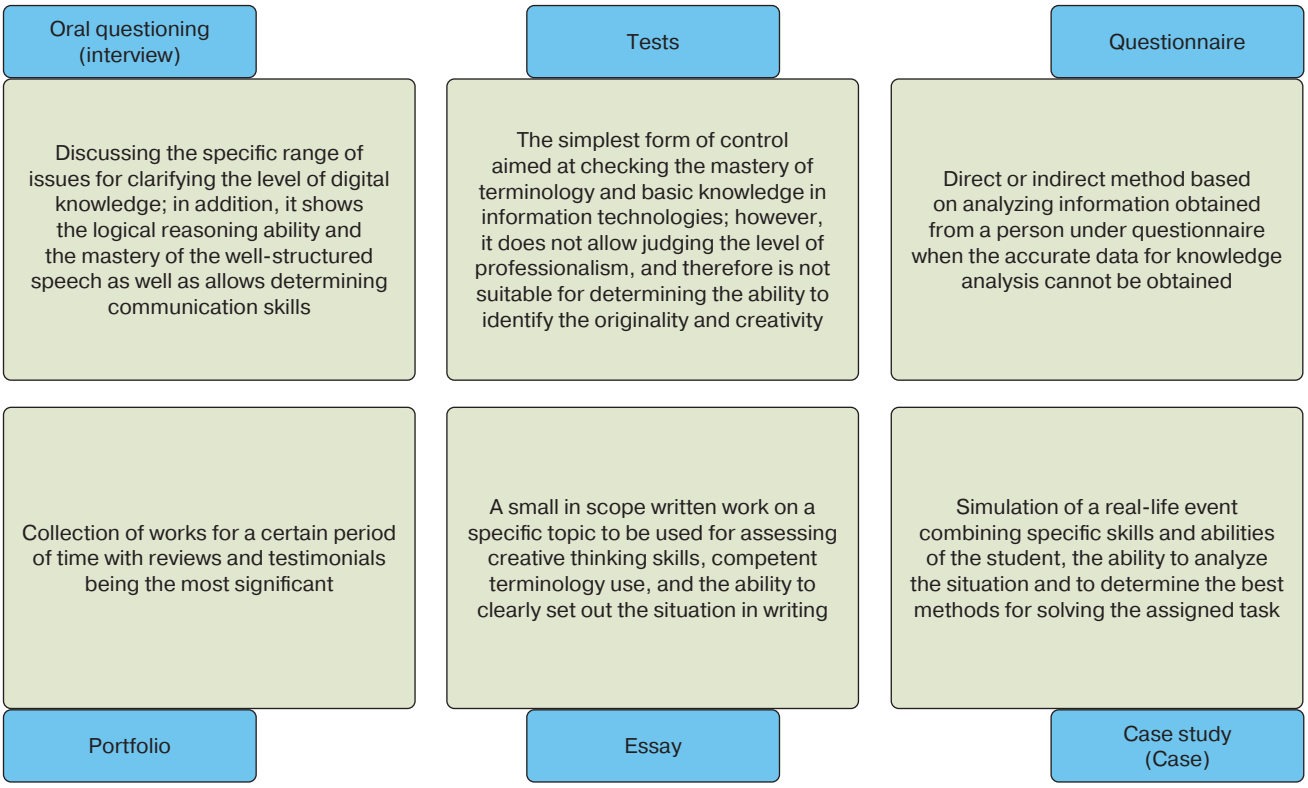


Fig. 2. Forms of entry control and using the results

presence of interdisciplinary links in the educational program and maintaining the continuity of the discipline. Obviously, the required digital competence level can be achieved only by their step-by-step formation at the indicator level for each separate discipline in the presence of interdisciplinary links in the educational program. The significant amount of work in this regard (methodological support, primarily) is carried out by the scientific and methodological council of the university that may involve the potential employer representatives in these activities.

The educational process throughout the updated programs of disciplines (modules) is accompanied by broadcasting digital culture from the digital educational continuum, ensuring harmonious integration of digital competences into the student personal competences. In the process of mastering the educational program, professional digital competences may be transferred into the category of everyday life practices, thus forming the new generation digital personality.

We believe that the student digital competence should be identified among the digital culture key values. In our opinion, it is the readiness for innovative technological challenges that forms the basis for successful digital development and an integral part of the digital culture of future professional having the necessary knowledge and skills, understanding of using ICTs and digital technologies in life and professional activities, as well as respecting digital ethics and safety. Achieving the claimed indicators is the result of the third step aimed at the digital culture formation and broadcasting.

The DCF control and assessment is the fourth step of the proposed model consisting of two stages. The first stage includes the assessment of educational achievements (formation of digital competence indicators) planned for disciplines (modules) and practices within the framework of ongoing and midterm attestation; diagnostics of the DCF level; and final state attestation.

When assessing the achievement of learning outcomes, the study recommends the creation of a register of indicators displaying all DCF indicators of the student.

The student mastery levels are established using appropriate assessment tools; thereby, the educational program should include the set of assessment materials for checking the theoretical and practical knowledge in the field of digital application. In our opinion, an important component of the assessment material system may be cases (problem-situational type of practical tasks requiring comprehensive approach to their solution), since their application is seemed the most effective for solving practical professional tasks requiring implementation of digital competences.

The second stage of the fourth step provides an independent external assessment of the student (future professional) DCF[22] [12]. It includes various implementation formats, namely the following:

- involving potential employers in forming assessment materials; providing the basis for practical training by them; and participation of employer representatives from governmental and non-governmental organizations in the midterm and final state attestation, reviewing the educational program in DCF terms in the professional field;
- systematic questionnaire of employers on their satisfaction with DCF level of graduates conducted by the educational organization (at least once a year). Based on the employer questionnaire results analysis, the educational organization adjusts the educational program;
- questionnaire of graduates (e.g., sending out questionnaires through e-mail) on their professional fulfillment degree in the labor market as well as on the competitiveness and employer demand for digital skills and competences acquired in their training period.

Thus, the educational program may be updated by the educational organization on the basis of the final state attestation results and obtained values for independent assessment of the graduate DCF (professional and personal). Adjusting the educational program (Fig. 3.) is a continuous multi-iterative process.

CONCLUSIONS

Thus, the study concludes by recommending that the basic (general) model for formation and assessment of the student digital competences in implementing HE programs be focused on formation of the fundamentally new personal and professional digital competences. The development of approaches to their formation is determined by new trends in the economy and society development influenced by the rapid spread of digital platforms, ecosystems, E2E technologies, and digital tools. Graduates should acquire the skills, knowledge, and abilities to flexibly adapt to market and employer demands in the digital competence application; easily socialize in the digital society; implement digital projects; use digital tools to identify, access, manage, analyze, assess, and synthesize digital resources; select and apply ICT in everyday life practices and professional activities effectively, critically, and safely.

[22] Methodological recommendations on the use of digital tools for confirming the results of the assessment of the digital economy competencies. https://old.digitalskills.center/upload/iblock/e0d/e0dd8b36038dee3fc37eceddb9bec13e.pdf?_ga=2.111898692.1115568595.1639414150-1681117855.1639414150. Accessed December 2, 2021 (in Russ.).

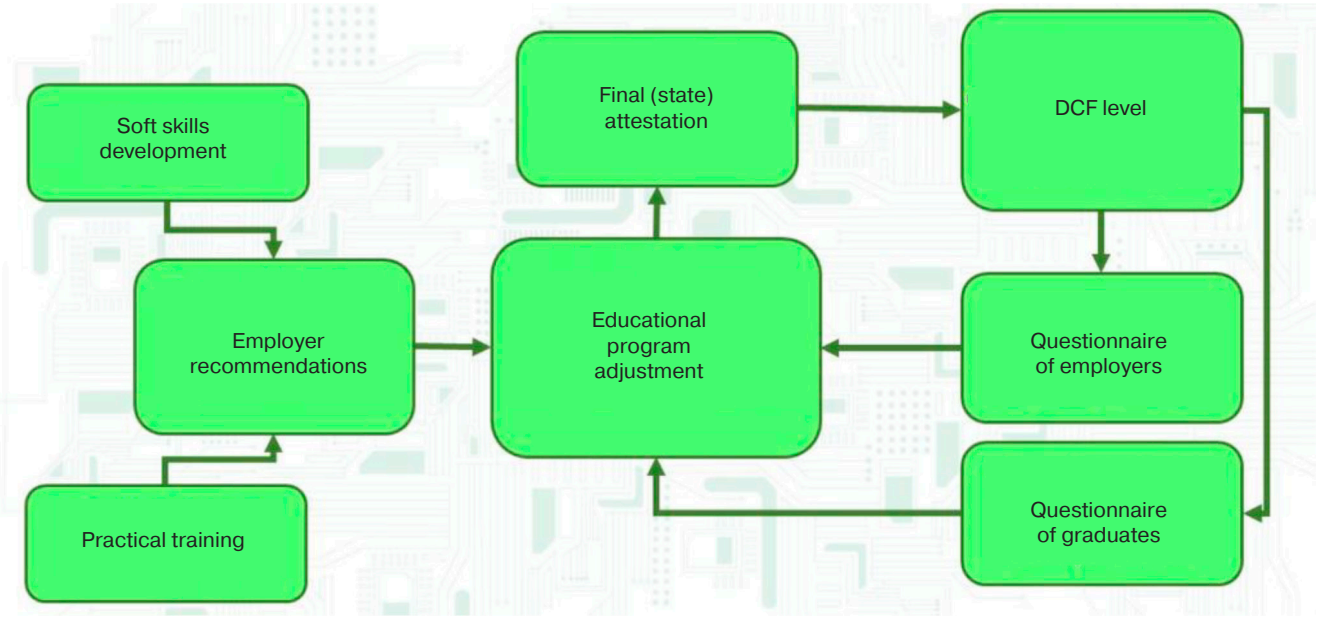


Fig. 3. Educational program adjustment

The presented model is based on the extensive regulatory framework, digital competence models from the leading domestic educational organizations, and the voluminous body of literature revealing the concept of digital competence along with its constituent blocks. It complements and expands the already existing practices of creating, implementing and assessing DCF models in HE programs. The advantages of the model are in terms of its reliance on expert community opinion (employers, primarily), as well as its accounting for sectoral and regional specific features of HE institution, training area specifics, and availability of optimal organizational and methodological conditions for digital competence formation.

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

REFERENCES

1. Sicilia M.-A., García Barriocanal E., Sánchez-Alonso S., et al. Digital skills training in higher education: Insights about the perceptions of different stakeholders. In: *Proceedings 6th International Conference on Technological Ecosystems for Enhancing Multiculturality* (*TEEM*). 2018. P. 781–787. https://doi.org/10.1145/3284179.3284312
2. Ferrari A. *DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe*. JRC Publications Repository. 2013, October 9. https://doi.org/10.2788/52966
3. Dmitrieva Ya.V., Alyabin I.A., Brovko E.I., Dvinina S.Yu., Dem'yanova O.V. University students' digital skills: De Jure vs De Facto. *Universitetskoe upravlenie: praktika i analiz = University Management: Practice and Analysis*. 2021;25(2):59–79 (in Russ.). https://doi.org/10.15826/umpa.2021.02.015
4. Volkova I.A., Petrova V.S. Formation of digital competencies in vocational education. *Vestnik Nizhnevartovskogo gosudarstvennogo universiteta = Bulletin of Nizhnevartovsk State University*. 2019;1:17–24 (in Russ.). Available from URL: https://vestnik.nvsu.ru/2311-1402/article/view/49621
5. Soldatova G.U., Rasskazova E.I. Models of digital competence and online activities of Russian adolescents. *Natsional'nyi psikhologicheskii zhurnal = National Psychological Journal*. 2016;2(22):50–60 (in Russ.). https://doi.org/10.11621/npj.2016.0205
6. Calvani A., Cartelli A., Fini A., Ranieri M. Models and instruments for assessing digital competence at school. *Journal of e-learning and knowledge society.* 2008;4(3):183–193.
7. Dmitrieva N.E., Zhulin A.B., Artamonov R.E., Titov E.A. Assessment of digital readiness level. In: *Assessment of Digital Readiness of the Population of Russia. Report of the HSE University*. Moscow: Izdatel'skii dom Vysshei shkoly ekonomiki; 2021. P. 67–73 (in Russ.). ISBN 978-5-7598-2518-0. Available from URL: https://conf.hse.ru/mirror/pubs/share/464963752.pdf. Accessed 21.01.2022.
8. Kuzmina M.V., et al. (Eds.). *Formation of digital literacy of students: Methodological recommendations for educators in the framework of the implementation of the Federal project "Digital Educational Environment."* Kirov: IRO Kirov. Oblast; 2019. 47 p. (in Russ.). Available from URL: https://kirovipk.ru/wp-content/uploads/2019/12/formirovanie-czifrovoj-gramotnosti-obuchayushhihsya-metodicheskie-rekomendaczii-dlya-rabotnikov-obrazovaniya.pdf. Accessed 21.01.2022.
9. Shklyaruk M.S., Garkusha N.S. (Eds.). *The Competence model of the digital transformation team in the public administration system.* Moscow: RANKhiGS; 2020. 84 p. (in Russ.). Available from URL: https://frpm.ru/wp-content/uploads/2021/09/Competency_Model_CDTO_RANEPA.pdf. Accessed 21.01.2022.
10. Gusev A.A. Labor relations digitalization and its impact on labor productivity and value of companies. *Ekonomika. Nalogi. Pravo = Economics, Taxes & Law*. 2019;12(6):39–47 (in Russ.). Available from URL: http://www.fa.ru/org/div/edition/enp/journals/2019%20%E2%84%966.pdf

СПИСОК ЛИТЕРАТУРЫ

1. Sicilia M.-A., García Barriocanal E., Sánchez-Alonso S., et al. Digital skills training in higher education: Insights about the perceptions of different stakeholders. In: *Proceedings 6th International Conference on Technological Ecosystems for Enhancing Multiculturality* (*TEEM*). 2018. P. 781–787. https://doi.org/10.1145/3284179.3284312
2. Ferrari A. *DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe*. JRC Publications Repository. 2013, October 9. https://doi.org/10.2788/52966
3. Дмитриева Я.В., Алябин И.А., Бровко Е.И., Двинина С.Ю., Демьянова О.В. Развитие цифровых навыков у студентов вузов: де-юре vs де-факто. *Университетское управление: практика и анализ = University Management: Practice and Analysis*. 2021;25(2):59–79. https://doi.org/10.15826/umpa.2021.02.015
4. Волкова И.А., Петрова В.С. Формирование цифровых компетенций в профессиональном образовании. *Вестник Нижневартовского государственного университета*. 2019;1:17–24. URL: https://vestnik.nvsu.ru/2311-1402/article/view/49621
5. Солдатова Г.У., Рассказова Е.И. Модели цифровой компетентности и деятельность российских подростков онлайн. *Национальный психологический журнал*. 2016;2(22):50–60. https://doi.org/10.11621/npj.2016.0205
6. Calvani A., Cartelli A., Fini A., Ranieri M. Models and instruments for assessing digital competence at school. *Journal of e-learning and knowledge society.* 2008;4(3):183–193.
7. Дмитриева Н.Е., Жулин А.Б., Артамонов Р.Е., Титов Э.А. Оценка уровня цифровой готовности. В кн.: *Оценка цифровой готовности населения России. Доклад НИУ ВШЭ*. М.: Издательский дом Высшей школы экономики; 2021. С. 67–73. ISBN 978-5-7598-2518-0. https://conf.hse.ru/mirror/pubs/share/464963752.pdf. Дата обращения: 21.01.2022.
8. Кузьмина М.В. и др. (авт.-сост.). *Формирование цифровой грамотности обучающихся: Методические рекомендации для работников образования в рамках реализации Федерального проекта «Цифровая образовательная среда»*. Киров: ИРО Кировской области; 2019. 47 с. URL: https://kirovipk.ru/wp-content/uploads/2019/12/formirovanie-czifrovoj-gramotnosti-obuchayushhihsya-metodicheskie-rekomendaczii-dlya-rabotnikov-obrazovaniya.pdf. Дата обращения 21.01.2022.
9. Шклярук М.С., Гаркуша Н.С. (ред.). *Модель компетенций команды цифровой трансформации в системе государственного управления*. М.: РАНХиГС; 2020. 84 с. https://frpm.ru/wp-content/uploads/2021/09/Competency_Model_CDTO_RANEPA.pdf. Дата обращения 21.01.2022.
10. Гусев А.А. Цифровизация трудовых отношений и ее влияние на производительность труда и стоимость компаний. *Экономика. Налоги. Право*. 2019;12(6):39–47. URL: http://www.fa.ru/org/div/edition/enp/journals/2019%20%E2%84%966.pdf

11. Kat'kalo V.S., Volkov D.L., Baranov I.N., Zubtsov D.A., Sobolev E.V., Yurchenkov V.I., Starovoitov A.A., Safronov P.A. *Digital Skills Education: Global Challenges and Best Practices. Analytical report for the 3rd International Conference "More than Learning: How to Develop Digital Skills."* Moscow: Korporativnyi universitet Sberbanka; 2018. P. 10–11 (in Russ.). Available from URL: https://sberuniversity.ru/upload/iblock/2f8/Analytical_report_digital_skills_web_demo.pdf. Accessed 21.01.2022.
12. Sukhanova E.A., Frumin I.D. (Eds.). *The quality of education in Russian universities: what we understood during the pandemic*: Analytical report. Tomsk: Izdatel'stvo Tomskogo gosudarstvennogo universiteta; 2021 (in Russ.). Available from URL: http://vital.lib.tsu.ru/vital/access/manager/Repository/koha:000892778. Accessed 21.01.2022.
11. Катькало В.С., Волков Д.Л., Баранов И.Н., Зубцов Д.А., Соболев Е.В., Юрченков В.И., Старовойтов А.А., Сафронов П.А. *Обучение цифровым навыкам: глобальные вызовы и передовые практики. Аналитический отчет к III Международной конференции «Больше чем обучение: как развивать цифровые навыки».* М.: АНО ДПО «Корпоративный университет Сбербанка»; 2018. Р. 10–11. URL: https://sberuniversity.ru/upload/iblock/2f8/Analytical_report_digital_skills_web_demo.pdf. Дата обращения 21.01.2022.
12. Суханова Е.А., Фрумин И.Д. (науч. ред.). *Качество образования в российских университетах: что мы поняли в пандемию*: Аналитический доклад. Томск: Издательство Томского государственного университета; 2021. 46 с. URL: http://vital.lib.tsu.ru/vital/access/manager/Repository/koha:000892778. Дата обращения 21.01.2022.

About the authors

Olga G. Savka, Cand. Sci. (Hist.), Head of the Department of Documentary Studies, History of State and Law, Institute of Management Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: savka@mirea.ru. RSCI SPIN-code 5570-0959, https://orcid.org/0000-0002-0808-1760

Maria N. Gusarova, Dr. Sci. (Hist.), Professor, Department of Documentary Studies, History of State and Law, Institute of Management Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: gusarova@mirea.ru. RSCI SPIN-code 5254-2382, https://orcid.org/0000-0002-4333-7514

Svetlana V. Sumina, Senior Lecturer, Information Protection Department, Institute of Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: sumina@mirea.ru. https://orcid.org/0000-0001-6024-0637

Yaroslav O. Knyazev, Cand. Sci. (Eng.), Assistant Professor, Department of Digital and Additive Technologies, Institute of Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: knyazev_ya@mirea.ru. RSCI SPIN-code 7282-7748, https://orcid.org/0000-0001-9185-6194

Denis A. Bezrukov, Cand. Sci. (Chem.), Assistant Professor, Department of Biotechnology and Industrial Pharmacy, M.V. Lomonosov Institute of Fine Chemical Technologies, MIREA – Russian Technological University (86, Vernadskogo pr., Moscow, 119571 Russia). E-mail: bezrukov@mirea.ru. https://orcid.org/0000-0002-3201-3266

Об авторах

Савка Ольга Геннадьевна, к.и.н., доцент, заведующий кафедрой документоведения, истории государства и права Института технологий управления ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: savka@mirea.ru. SPIN-код РИНЦ 5570-0959, https://orcid.org/0000-0002-0808-1760

Гусарова Мария Николаевна, д.и.н., доцент, профессор кафедры документоведения, истории государства и права Института технологий управления ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: gusarova@mirea.ru. SPIN-код РИНЦ 5254-2382, https://orcid.org/0000-0002-4333-7514

Сумина Светлана Валерьевна, старший преподаватель кафедры КБ-1 «Защита информации» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: sumina@mirea.ru. https://orcid.org/0000-0001-6024-0637

Князев Ярослав Олегович, к.т.н., доцент кафедры цифровых и аддитивных технологий Института перспективных технологий и индустриального программирования ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: knyazev_ya@mirea.ru. SPIN-код РИНЦ 7282-7748, https://orcid.org/0000-0001-9185-6194

Безруков Денис Алексеевич, к.х.н., доцент кафедры биотехнологии и промышленной фармации Института тонких химических технологий им. М.В. Ломоносова ФГБОУ ВО «МИРЭА – Российский технологический университет» (119571, Россия, Москва, пр-т Вернадского, д. 86). E-mail: bezrukov@mirea.ru. https://orcid.org/0000-0002-3201-3266

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

MIREA – Russian Technological University.
78, Vernadskogo pr., Moscow, 119454 Russian Federation.
Publication date November 30, 2022.
Not for sale.

МИРЭА – Российский технологический университет.
119454, РФ, г. Москва, пр-т Вернадского, д. 78.
Дата опубликования 30.11.2022 г.
Не для продажи.


МИРЭА
Российский
технологический
университет