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

11(1) 2023

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
2023, Vol. 11, No. 1

Russian Technological Journal
2023, том 11, № 1

https://www.rtj-mirea.ru

Российский технологический университет

https://doi.org/10.32362/2500-316X ISSN 2500-316X (Online)

Russian Technological Journal 2023, Vol. 11, No. 1

Publication date January 31, 2023.

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 2023, том 11, № 1

Дата опубликования 31 января 2023 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

UDC 004.5
https://doi.org/10.32362/2500-316X-2023-11-1-7-17

RESEARCH ARTICLE

Developing the data management component of an academic discipline program for an educational management information system

Julia V. Starichkova @, Igor E. Rogov, Valeriya S. Tomashevskaya

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

Abstract
Objectives. The need to apply methods and models to support the educational process at universities including the formation and management of academic discipline programs (ADPs) is determined by the growing need for the active implementation of various automation tools including integrated information systems, which arise in response to a number of regulatory and legal factors. Such social factors result in the significant increase in the volume and categories of information circulating within business processes of an educational organization, as well as the expansion of the requirements for ensuring the protection, storage, and transmission of information. In recent years, the Government of the Russian Federation has approved the national "Digital Economy" and "Education" projects (including the Federal Project "Digital Educational Environment") emphasizing the growing role of informatization and digitalization processes in education. In this connection, an obvious discrepancy arises between the theoretical characteristics of information flows existing in educational organizations and the methods of its collection, processing, storage, analysis, and application used in practice. One of the most important conceptual components of the educational process in higher education institutions is the ADP, which organizes the relationship between various components of the educational process: curriculum, competencies, training areas, learning technologies, and methods for conducting the control check of students' knowledge. The labor-intensive and variable nature of ADP development and implementation requires the introduction of information technologies. Thus, the aim of the present work is to analyze the volume and structure of institutional educational programs in order to identify the necessary software requirements.
Methods. The classification of learning management systems according to various criteria, key requirements for academic disciplines, and ADP structure is considered.
Results. An analysis of links between the ADP and key entities of the educational process is presented. The functionality of the self-developed ADP module for implementing at RTU MIREA is aimed at providing interconnection, transparency, and availability of links between academic discipline parameters and its sections.
Conclusions. Introducing the ADP module allows reducing the time spent on developing the program by providing universal templates of academic disciplines, along with the possibility of autofilling the academic discipline parameters and tracking the current status of ADPs, as well as increasing the level of awareness of participants in the educational process.

Keywords: educational process, data management, digitalization, educational environment, information systems to support the educational process, educational programs and standards, academic discipline program

• **Submitted:** 31.05.2022 • **Revised:** 13.07.2022 • **Accepted:** 28.11.2022

For citation: Starichkova Ju.V., Rogov I.E., Tomashevskaya V.S. Developing the data management component of an academic discipline program for an educational management information system. *Russ. Technol. J.* 2023;11(1):7–17. https://doi.org/10.32362/2500-316X-2023-11-1-7-17

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

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Разработка компонента управления данными программы учебной дисциплины для информационной системы управления образовательной средой

Ю.В. Старичкова @, И.Е. Рогов, В.С. Томашевская

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

Резюме

Цели. Потребность в применении методов и моделей поддержки учебного процесса в образовательных организациях высшего образования, в т.ч. формирования и управления программами учебных дисциплин (ПУД), определяется растущей необходимостью активного внедрения средств автоматизации, включая комплексные информационные системы. Это вызвано наличием нормативно-правовых и социальных факторов, приводящих, с одной стороны, к существенному увеличению объемов и категорий информации, циркулирующей в рамках бизнес-процессов образовательной организации, а с другой стороны, к расширению требований, предъявляемых к обеспечению защиты информации, ее хранению и передаче. В 2018–2019 гг. Правительством Российской Федерации утверждены национальные проекты «Цифровая экономика» и «Образование» (в т.ч. федеральный проект «Цифровая образовательная среда»), подчеркивающие растущую роль процессов информатизации и цифровизации в образовании. Следует отметить очевидное несоответствие между характеристиками существующих в образовательных организациях информационных потоков и способами их сбора, обработки, хранения, анализа и применения на практике. Одной из важнейших составляющих образовательного процесса в организациях высшего образования является ПУД, позволяющая организовать взаимосвязи между различными составляющими учебного процесса: учебным планом, компетенциями, направлениями подготовки, технологиями обучения и способами осуществления контрольной проверки знаний обучающихся. Разработка и реализация ПУД является трудоемким и вариативным процессом, который требует внедрения информационных технологий. Цель работы – анализ объема и структуры образовательных программ учреждения для выявления требований к необходимому программному обеспечению.

Методы. Рассмотрена классификация систем управления обучением по различным признакам, ключевые требования к учебным дисциплинам и структура ПУД.

Результаты. Проведен анализ связей ПУД и ключевых сущностей учебного процесса. Рассмотрены функциональные возможности модуля собственной разработки ПУД для внедрения в РТУ МИРЭА, направленные на обеспечение взаимосвязи, прозрачности и доступности связей между параметрами учебной дисциплины и ее разделами.

Выводы. Введение модуля ПУД позволит уменьшить временны́е затраты на разработку программы путем предоставления универсальных шаблонов учебных дисциплин, возможности автозаполнения их параметров, отслеживания текущего статуса ПУД, и повысит уровень информированности участников образовательного процесса.

Ключевые слова: образовательный процесс, управление данными, образовательная среда, информационные системы поддержки учебного процесса, образовательные программы и стандарты, программы учебных дисциплин

• **Поступила:** 31.05.2022 • **Доработана:** 13.07.2022 • **Принята к опубликованию:** 28.11.2022

Для цитирования: Старичкова Ю.В., Рогов И.Е., Томашевская В.С. Разработка компонента управления данными программы учебной дисциплины для информационной системы управления образовательной средой. *Russ. Technol. J.* 2023;11(1):7–17. https://doi.org/10.32362/2500-316X-2023-11-1-7-17

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

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Information technologies are widely used in all spheres of human activities including education. The introduction of information systems (IS) into the educational process contributes to its rational organization and the creation of the unified information space for the educational institution and direct participants in the learning process.

The information educational environment is the systematically organized set of information technologies as well as hardware, software, and methodological support along with electronic educational resources for organizing the educational process wherein a person plays the role of the subject [1, 2]. The digitalization of the following areas is required for the functioning of the information educational environment in a higher educational institution (HEI):

- information and methodological activities,
- planning the learning process and resource provision,
- placement and storage of training materials,
- monitoring,
- distance education [3].

The need to digitize these areas of activities has given rise to the special type of IS being the Learning Management System (LMS). This term implies the use of IS to provide support to the e-learning process in terms management, monitoring, and documenting, as well as providing the learning content and control [4]. This IS support may be distinguished into methodological, administrative, and technological strands. The ISs themselves comprise software or platforms aimed at providing teachers and students with the necessary tools to organize and conduct the learning process including distributing the reference and lecture materials, as well as generating reports.

The functionality of modern LMSs may be complemented by software products such as the Training Management System (TMS) for implementing the learning process under the teacher control and the Learning Record Store (LRS) for tracking and recording user actions. LMSs find application for solving various tasks ranging from IS providing means for online courses up to systems of the corporation level which functionality is regulated by GOST R 52653-2006[1] and GOST R 52655-2006[2] standards.

DOMESTIC AND INTERNATIONAL EXPERIENCE IN IMPLEMENTING IS INTO EDUCATIONAL ACTIVITIES

By the early 1980s, personal computers had already become much more accessible to users, resulting in the increasing use of software in the educational process. The first IS supporting the educational process and its components appeared in 1980. Among the most popular IS, the ToolBook and TenCORE authoring systems were used to create interactive content for learning. The computer managed instruction (CMI) systems FirstClass and TrainingPartner[3], which were among the first IS for automating educational process components, use formal language and specifications to create training courses and integrate them into the distance learning system. The first tools for working with email, forms, and interactive whiteboards were also introduced in FirstClass and TrainingPartner.

A complex of automated management systems for higher education institutions already being developed in Russia at that time would go on to become the first IS to support the educational process. This software developed by the Research Institute of Higher Education Problems of the USSR covered the entire learning process ranging from admission to the university up to the completion of education [5]. According to the centralized implementation process, the system went on to be integrated into more than 50 different educational institutions deemed as having the greatest technical and intellectual potential.

Along with the development and growth of the number of IS supporting the educational process components,

[1] GOST R 52653-2006. *Information and communication technologies in education. Terms and definitions*. Moscow: Standartinform; 2007 (in Russ.).

[2] GOST R 52655-2006. *Information and communication technologies in education. The integrated automated control system of the high professional educational system. General requirements*. Moscow: Standartinform; 2007 (in Russ.).

[3] LMS-Timeline. https://www.rockymountainalchemy.com/cudenver/INTE6750/Emergence/LMS-Timeline.html. Accessed April 15, 2022.

various standards governing the processes of their development and implementation become the subject of active development. An example of this type of standard is the Sharable Content Object Reference Model (SCORM)[4], which focuses on the requirements for the formation of educational literature for distance learning systems. Here, the key idea is to compile electronic resources from shared content objects [6]. SCORM aims at implementing the interoperability of all elements of the education programs with all LMSs and virtual learning environments [7]. In accordance with this standard, educational materials are presented in the form of blocks, allowing them to be included in different courses and disciplines and providing the possibility of their independent use as part of a distance learning system. Since essentially consisting of a set of technical rules focused on providing the reader with knowledge about course design and lesson structure, as well as the description of the principles of interaction with the system, SCORM becomes relevant to programs within academic disciplines (ADPs). This standard puts forward the following three key requirements for the learning program components:

- the presence of manifest file containing the complete description of the course and all its components;
- course metadata represented by the image, video file, or HTML page should be associated with the specific metadata file;
- the presence of the program interaction language to implement communication between the learning organization system and the learning program.

Such educational support systems as Moodle[5], Sakai[6], ATutor[7], ILIAS[8], and others are implemented using this standard.

Today, the most preferable solutions for educational institutions are not off-the-shelf solutions but the systems providing flexible tools to support the learning process, which may be further improved in the future [8, 9].

IS ROLE OF ADP COMPONENT IN SUPPORTING THE EDUCATIONAL PROCESS

Although many LMSs offer similar functionality, individual systems may differ significantly in a number of key parameters, thus complicating the decision-making process for using a particular system in the educational institution. In order to identify different types of LMS, the following characteristics may be used as classifiers:

- license type;
- functionality;
- modularity;
- requirements imposed by the customer to the system;
- the system physical location features [8].

LMSs can also be classified into three types according to the type of license: free, partially paid, and paid. Free systems are characterized by their cost-free distribution. Partially paid systems are characterized by the presence of a minimal free functionality, which may be expanded upon payment of a fee.

Moodle is an example of one of currently most popular free LMSs. The system provides users with tools for both creating courses and monitoring the learning process. The key advantages are its simplicity and ease of use, while offering extensive opportunities for organizing the educational process and monitoring the knowledge of students, as well as its relative user-friendliness to third-party developers allowing their integration into the system [10].

The partially paid system may be exemplified by the eFront[9] system having a wide starting set of tools: glossary, testing, forum, chat, calendar, etc.

The paid systems include Dnevnik.ru[10], Moiuniver[11], YaKlass[12], SharePointLMS[13], BlackBoard[14], Desire2Learn (D2L)[15], and others.

Classifying IS by functionality, two types of LMS may be distinguished:

1) aimed at supporting the learning process upon the whole;
2) aimed at providing educational material and testing the knowledge of students only.

Examples of the first type are Moodle, Sakai, e-University[16], Education Elements[17], Ilias, Odijoo[18], ScormCloud[19], Dnevnik.ru, My University, YaKlass, and eFront.

The systems belonging to the second type may be exemplified by Claroline[20], Dokeos[21], LAMS[22], Learn eXact[23], and Coursera[24].

[4] https://scorm.com/. Accessed April 17, 2022.
[5] https://moodle.org/. Accessed May 15, 2022.
[6] https://www.sakailms.org/. Accessed May 11, 2022.
[7] https://atutor.github.io/. Accessed May 13, 2022.
[8] https://www.ilias.de/. Accessed May 15, 2022.

[9] https://www.efrontlearning.com/. Accessed May 12, 2022.
[10] https://dnevnik.ru/. Accessed May 15, 2022 (in Russ.).
[11] https://moi-univer.ru/. Accessed May 21, 2022 (in Russ.).
[12] https://www.yaklass.ru/. Accessed May 15, 2022 (in Russ.).
[13] https://www.sharepointlms.com/. Accessed May 15, 2022.
[14] https://www.blackboard.com/. Accessed May 15, 2022.
[15] https://www.d2l.com/. Accessed May 15, 2022.
[16] https://dic.academic.ru/dic.nsf/ruwiki/1428198. Accessed May 26, 2022 (in Russ.).
[17] https://www.edelements.com/. Accessed May 26, 2022.
[18] https://rusticisoftware.com/blog/taking-scorm-to-odijoo/. Accessed May 15, 2022.
[19] https://rusticisoftware.com/products/scorm-cloud/?utm_source=google&utm_medium=natural_search. Accessed May 15, 2022.
[20] https://www.claroline.com/. Accessed May 15, 2022.
[21] https://www.dokeos.com/. Accessed May 15, 2022.
[22] https://www.lamsfoundation.org/. Accessed May 15, 2022.
[23] https://www.exactls.com/. Accessed May 15, 2022.
[24] https://www.coursera.org/. Accessed May 15, 2022.

Using the modularity criterion for classifying LMS, the autonomous and modular types of systems may be distinguished. The autonomous LMS is characterized by implementing all tools for the activities in one application. Modular LMSs are independent subsystems.

In terms of requirements, LMSs implemented as an off-the-shelf product or made on order in accordance with the needs of the customer organization may be distinguished.

In terms of physical location, the local, server, and cloud systems may be identified.

The local and server-based LMSs may be exemplified by Moodle, Tandem University[25], Ilias, ATutor, and WebTutor[26]. Cloud LMSs include Coursera, iSpring[27], Ed-modo[28], Odijoo, Scorm Cloud, TalentLMS[29], and Docebo[30].

In addition to the basic requirements for modern LMS such as reliability, convenience, and low cost for the educational process participants, the following additional requirements based on the classification features discussed above may be formulated:

- availability of flexible tools allowing the educational institution to implement the necessary functions within the educational process;
- support of SCORM or Tin Can API standard[31] to migrate content from one IS to another;
- adaptability.

The main trend in the LMS development is the transformation into the Next Generation Digital Learning Environment (NGDLE) being the ecosystem consisting of learning tools and components developed by general standards [11]. For proper functioning, NGDLE requires implementing the following functional features:

- the ability to analyze and evaluate the learning process;
- compatibility and the ability to customize the learning environment;
- simplicity and usability of the tools for both students and academic teaching staff (ATS);
- ensuring interaction and integration between different learning programs.
- The above functional features allow identifying four key dimensions for NGDLE:
- it should be possible to exchange learning content between all components of the system presented in a common format;
- the integration process should be easy and convenient to reduce the time costs and simplify the process of capacity building-up for the users;
- the learning environment should be the main source of data retrieval for the learning process;
- NGDLE should allow creating new interoperability standards in ways being compatible with its other standards to maintain overall consistency.

The NGDLE personalization covers two aspects:

- equipping and configuring the learning environment which is used then for building pathways to complete learning tasks and achieve learning goals;
- adaptive learning in which the automated system provides students with coaching and suggestions tailored to each student needs.

There has been significant recent momentum in adaptive learning, and this should be a feature of the NGDLE landscape. As with other NGDLE functional areas, integrating adaptive learning tools capable of providing data on student to support analytics would be crucial.

Within the context of NGDLE for analytics, two main components may be distinguished:

- learning analytics characterized as the measurement, collection, analysis, and reporting of student data in order to understand and optimize the learning process itself and the environment wherein it occurs;
- integrated planning and advising systems defined as the institutional capacity to make educational progress by creating a common space for all participants in the educational process. This space should contain the information and set of services necessary for a certain level of education.

It should be noted that most of the major LMS platforms have built-in functionality to perform learning process analytics based on data from IS and LMS. Based on this, such modules may be considered as first-generation attempts. Future analytics modules may be placed outside LMS, while their dashboards can be accessible for viewing in LMS or other applications using the Interoperability specification (the protocol describing interaction of learning platforms that is Learning Tools Interoperability, LTI). The results of the transition to NGDLE would be an increase in the amount of stored data; integration of appropriate tools aimed at improving the work quality; the ability to use analytical functions for evaluating the learning process.

SELECTING AN IS MODEL FOR ORGANIZING THE EDUCATIONAL PROCESS AT RTU MIREA

Educational programs offered at RTU MIREA were analyzed in the context of selecting the appropriate IS model (Fig. 1) [12]. The following key factors were identified:

[25] https://tandemservice.ru/products/tandem-university. Accessed May 15, 2022 (in Russ.).

[26] https://webtutor.ru/_wt/main_web. Accessed May 15, 2022 (in Russ.).

[27] https://www.ispring.ru/. Accessed May 15, 2022 (in Russ.).

[28] https://soware.ru/products/edmodo. Accessed May 15, 2022 (in Russ.).

[29] https://www.talentlms.com/. Accessed May 15, 2022.

[30] https://www.docebo.com/. Accessed May 15, 2022.

[31] Tin Can API is a specification of programs in distance learning, which allows training systems to communicate with each other by tracking and recording training classes of all types.

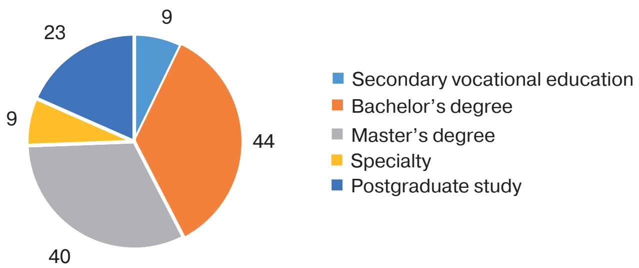


Fig. 1. Number of RTU MIREA educational programs

- territorial—conditioned by the location of classrooms and basic departments;
- quantitative—reflecting the number of students to include those who will only enter the university, i.e., making allowance for the dynamics of their enrollment (Fig. 2);
- implementation—degree of integration of current IS into the institution activities, as well as the level of their interdependence [13, 14].

As a result, it was decided to allocate the ADP management subsystem as a specialized LMS component.

The learning process at RTU MIREA is carried out in accordance with the basic educational programs reflecting its key components [14, 15]. As well as defining the learning process itself, the working ADP reflects the following key components:

- thematic focus of the discipline;
- competencies to be obtained by the student according to his or her achieved results;
- number of hours and order of lectures, practical, and laboratory classes, including independent work carried out by the student
- the classroom hardware and software;
- format and methods for testing and control of the student knowledge;
- recommended literature;
- questions for test or exam, etc. [14, 15].

The process of ADP development, formed in accordance with the needs of the labor market, is regulated by Federal Law No. 273-FZ of December 29 "On Education in the Russian Federation".[32] When creating an ADP, it is necessary first of all to be guided by its purpose. The educational process is aimed at ensuring that a student master a set of key skills and competencies provided by the training area or specialty. There should be a relationship between all types of obtained competencies by determining the place of the academic discipline in the learning process to provide the future specialist with a complete picture of his/her chosen specialty.

The ADP design and implementation process should be guided by three key principles: interconnectedness, transparency, and accessibility.

The first requirement is an attribute of any learning process reflecting the relationship between the various elements of the learning process including curriculum, competency map, educational standards, etc. The second requirement refers to the processes of agreeing and approving the learning program. The third requirement implies the program should be accessible to all participants in the learning process, i.e., including teaching staff and students.

Today, although LMSs are used at almost every university, the main disadvantage is the lack of a "boxed solution" that would reduce the load on the teaching staff by taking over most of the tasks related to creating a working ADP. In future, such a solution should offer the following:

[32] http://pravo.gov.ru/proxy/ips/?docbody=&firstDoc=1&lastDoc=1&nd=102162745. Accessed May 15, 2022 (in Russ.).

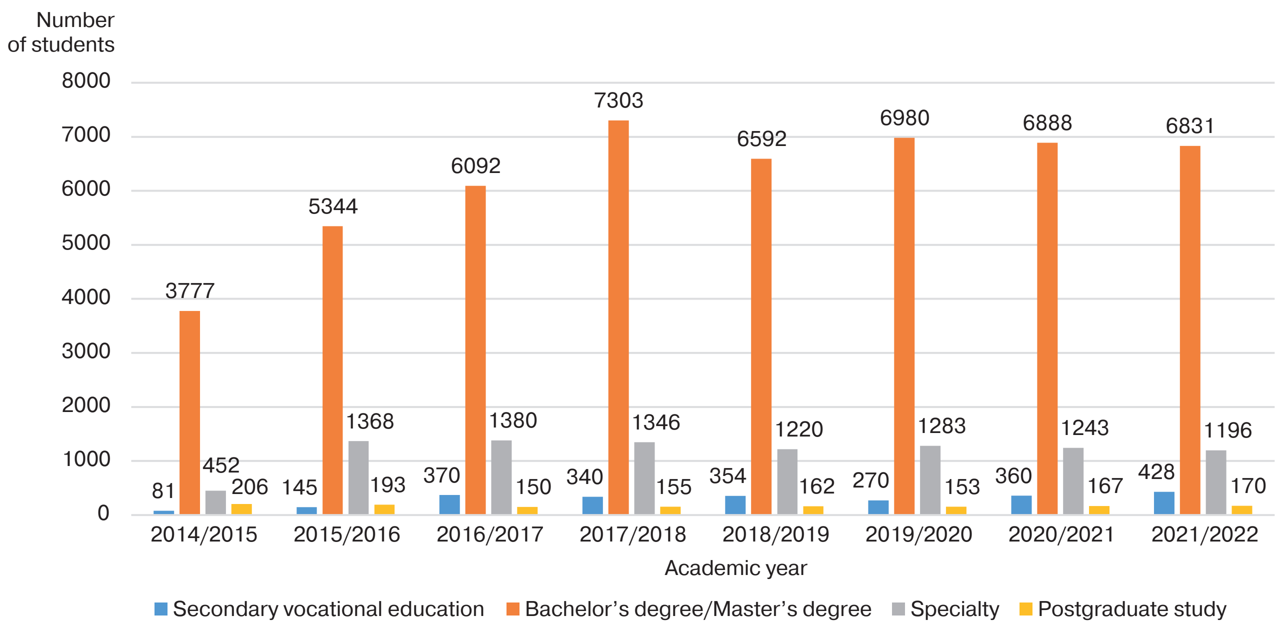


Fig. 2. Dynamics of the number of students enrolled in RTU MIREA

- compensate for the increased teaching staff workload in drafting documents supporting the educational process;
- reduce the document flow;
- accelerate the process of ADP creation;
- form the competency matching matrix of academic disciplines;
- provide the development of work programs with allowance for the sequence of the studied disciplines and the formation of competencies;
- analyze the interrelation of the disciplines involved in the process of the future specialists training;
- synchronize the list of recommended literature for mastering the discipline with the list of literature of the university library;
- provide the possibility of the working ADP individualization;
- reduce the number of mistakes made by teachers;
- automatically identify discrepancies and contradictions;
- track the requirements for the material and technical support of the learning process;
- ensure flexibility to changes in documents governing the development process;
- ensure that working programs comply with the requirements of Federal State Educational Standards for higher education;
- form working programs as separate editions;
- ensure the transparency of the processes of agreeing and approving the ADP formal and substantive parts;
- ensure storage, versioning, and accessibility of all working ADP versions;
- ensure a high level of quality management in developing the working ADP;
- ensure the working ADP availability to all participants in the educational process;
- ensure the collection of statistical data on the current status of the working ADP.

IS KEY ELEMENTS AND THEIR FUNCTIONALITY

When automating the ADP development process, it is necessary to implement the functionality to allowing both the creation and editing of the working ADP and its key elements:

- creating, editing, searching, and using off-the-shelf programs (provided they are available in the database);
- creating and editing templates of working ADP;
- filling the sections of the working ADP based on regulatory requirements;
- adding new documentation;
- ability to integrate with digital library catalogs;
- work with competencies (individually and by creating a common list) implying creation, editing, and automatic generation;
- formation of the teaching load, manually or automatically.

In view of the above, it has been decided to develop the proper module for working with ADP to be further implemented at RTU MIREA. The formalized description of functional requirements for the ADP module may be presented as shown in Fig. 3.

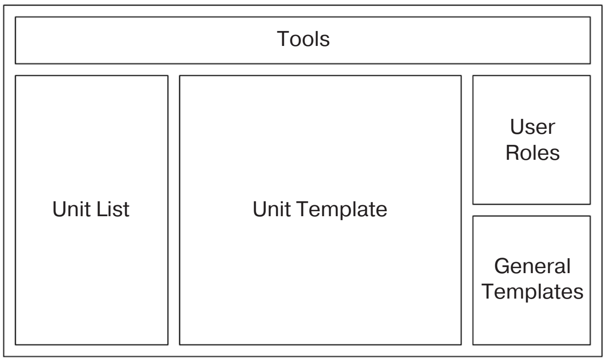


Fig. 3. Basic functionality of the ADP module for forming ADP approval templates

Tools provide the user with various functionalities including navigation between templates and printing. Unit List displays the hierarchical structure of units from Faculty to Department level. Unit Template contains information regarding the current template. User Roles contain links to reference information about ADP developers and those responsible on behalf of the Department and Methodology Department. General templates contain links to departmental programs.

The ADP module aims to solve the following set of tasks:

1. Storage of created and implemented learning programs indicating their versions.
2. Access to information about courses and their key characteristics for ATS.
3. Identification of the ADP place in the curriculum.
4. Ability to set the values assigned to the credits required for the student training areas and the corresponding specialization for each discipline.
5. Formation of the set of student competences to be acquired by him/her at the end of the study discipline.
6. Identifying the interrelationships of various disciplines in the curriculum.
7. ADP status allowing determining its presence or absence.
8. Determination of intersection cases of ADP sections with allowance for the chosen training areas and specialties.

By solving the described problems, three key requirements for the ADP development and implementation process are fulfilled: accessibility, interconnection, and transparency [14–16]. Access to the ADP created using the module is provided to all persons

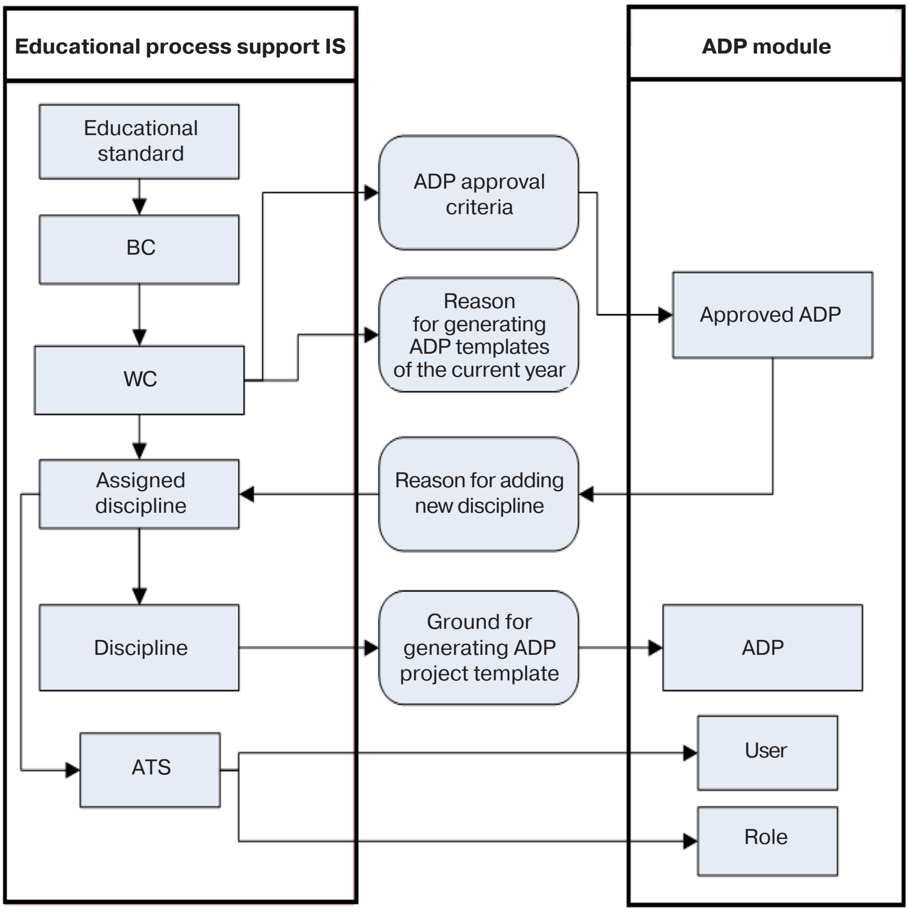


Fig. 4. Relationship between ADP attributes and entities inherent in the learning process. WC is working curriculum. BC is basic curriculum

involved in the educational process thus ensuring openness and transparency in organizing the learning process.

Universal templates for generating standard ADP forms can be divided into two main groups. The first group contains templates included in the work plan and assigned to a particular structural unit and ATS. Its main distinguishing feature is to indicate the presence of the university-wide elective. The second group is intended for templates only being planned to be introduced in the curriculum. Both typified groups of templates created at RTU MIREA in accordance with the classification of academic disciplines are accessible to ATS.

The structural content of the template includes two areas: a functional employee area and ADP area. The first provides the employee with access to their courses, material base, ADP, group chat, and additional tool features. The second area includes the following five tabs providing information between which the user can switch: ADP card, content, disciplines, review, and status log.

By integrating the module with IS, two main features are implemented:

- formation of the academic discipline parameters with the provision of connection between parameters conditioned by the curriculum and the sections of the selected discipline;
- availability of choice between different competences of the corresponding academic discipline for obtaining the necessary information formed on the basis of the data stored in the competence map and specialty direction (Fig. 4).

CONCLUSIONS

The implemention of the ADP module as a part of the learning process support IS at RTU MIREA reduces the time spent by ATS on the ADP development by providing users with various templates along with the possibility to fill them automatically in accordance with the discipline curriculum. The ADP module functionality also ensures the collection of statistical data on the program current status and the availability of all ADP versions, as well as increasing the student awareness level.

Authors' contributions

Ju.V. Starichkova has developed the functionality of the self-developed module "Academic Discipline Programs."

I.E. Rogov has conducted an analysis of the links between the ADP and the key entities of the educational process.

V.S. Tomashevskaya made a classification of learning management systems according to various criteria, key requirements for academic disciplines and the structure of ADPs.

REFERENCES

1. Evdokimova V.E., Kirillova O.A. Information educational environment of the university. *Vestnik Kurganskogo gosudarstvennogo universiteta = Bulletin of Kurgan State University*. 2020;(1–55):81–84 (in Russ.).
2. Ezhova Yu.M. Features of the organization of the educational process in the information and educational environment. *Azimut nauchnykh issledovanii: pedagogika i psikhologiya = Azimuth Scientific Research: Pedagogy and Psychology*. 2021;10(1–34):96–100 (in Russ.). https://doi.org/10.26140/anip-2021-1001-0023
3. Ivanova O.Yu., Kutuzova Z.Yu., Kutuzov A.V. Information and educational environment of higher school: essence and structure. *Kontsept = Concept*. 2020;8:20–29 (in Russ.). Available from URL: https://e-koncept.ru/2020/201054.htm
4. Ponyaeva T.A. Organization of distance learning in higher educational institutions based on LMS systems. *Problemy sovremennogo pedagogicheskogo obrazovaniya = Problems of Modern Pedagogical Education*. 2020;(67–4):330–332 (in Russ.).
5. Vlasenko A.V., Kashirina E.I. Current issues of data management in the context of digital transformation. *Vestnik AGU = The Bulletin of the Adyghe State University*. 2020;3(266):74–79 (in Russ.). Available from URL: http://vestnik.adygnet.ru/files/2020.3/6412/74-79.pdf
6. Solovov A.V. Mathematical modeling of the content of electronic educational resources. *Vestnik Samarskogo gosudarstvennogo aerokosmicheskogo universiteta imeni akademika S.P. Koroleva = Vestnik SGAU*. 2009;4: 245–253 (in Russ.).
7. Stafeev S.K., Sukhorukova M.V., Pashkovskii M.A., et al. Implementation of the open educational standard SCORM in the educational process. *Nauchno-tekhnicheskii Vestnik Sankt-Peterburgskogo universiteta informatsionnykh tekhnologii, mekhaniki i optiki = Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2007;44:10–15 (in Russ.).
8. Popova Yu.B. Classification of learning management systems. *Sistemnyi analiz i prikladnaya informatika = System Analysis and Applied Information Science*. 2016;3:51–58 (in Russ.).
9. Girya I.A. Integration of student knowledge models in the adaptive distance learning environment. *Obrazovatel'nye tekhnologii i obshchestvo = Educational Technology & Society*. 2010;13(4):240–245 (in Russ.).
10. Belko E.S. Experience in designing an electronic training course on mathematical analysis in the Moodle system. *Vestnik Nizhnevartovskogo gosudarstvennogo universiteta = Bulletin of Nizhnevartovsk State University*. 2020;3:4–10 (in Russ.). https://doi.org/10.36906/2311-4444/20-3/01
11. Kluev A. Changes in university management during the pandemic. *Universitetskoe upravlenie: praktika i analiz = University Management: Practice and Analysis*. 2020;24(3):7–12 (in Russ.). Available from URL: https://www.umj.ru/jour/article/view/1242/1060

СПИСОК ЛИТЕРАТУРЫ

1. Евдокимова В.Е., Кириллова О.А. Информационная образовательная среда вуза. *Вестник Курганского государственного университета*. 2020; (1–55):81–84.
2. Ежова Ю.М. Особенности организации образовательного процесса в условиях информационно-образовательной среды. *Азимут научных исследований: педагогика и психология*. 2021;10(1–34):96–100. https://doi.org/10.26140/anip-2021-1001-0023
3. Иванова О.Ю., Кутузова З.Ю., Кутузов А.В. Информационно-образовательная среда вуза: сущность и структура. *Концепт*. 2020;8:20–29. URL: https://e-koncept.ru/2020/201054.htm
4. Поняева Т.А. Организация дистанционного обучения в высших учебных заведениях на основе LMS системы. *Проблемы современного педагогического образования*. 2020;(67–4):330–332.
5. Власенко А.В., Каширина Е.И. Актуальные вопросы управления данными в условиях цифровой трансформации. *Вестник АГУ*. 2020;3(266):74–79. URL: http://vestnik.adygnet.ru/files/2020.3/6412/74-79.pdf
6. Соловов А.В. Математическое моделирование содержания электронных образовательных ресурсов. *Вестник Самарского государственного аэрокосмического университета имени академика С.П. Королёва*. 2009;4:245–253.
7. Стафеев С.К., Сухорукова М.В., Пашковский М.А. и др. Внедрение открытого образовательного стандарта SCORM в учебный процесс. *Научно-технический Вестник Санкт-Петербургского университета информационных технологий, механики и оптики*. 2007;44:10–15.
8. Попова Ю.Б. Классификация автоматизированных систем управления обучением. *Системный анализ и прикладная информатика*. 2016;3:51–58.
9. Гиря И.А. Интеграция моделей знаний ученика в адаптивной среде дистанционного обучения. *Образовательные технологии и общество*. 2010;13(4):240–245.
10. Белько Е.С. Опыт проектирования электронного обучающего курса по математическому анализу в системе Moodle. *Вестник Нижневартовского государственного университета*. 2020;3:4–10. https://doi.org/10.36906/2311-4444/20-3/01
11. Клюев А. Как меняется управление университетами в период пандемии. *Университетское управление: практика и анализ*. 2020;24(3):7–12. URL: https://www.umj.ru/jour/article/view/1242/1060
12. Прокопов Н.И., Иванов С.Ю., Томашевская В.С., Антонюк С.Н., Иванова Д.В. Научный потенциал современного вуза: перспективы подготовки научно-педагогических кадров в аспирантуре. *Общество: социология, психология, педагогика*. 2020;1(69):14–23. URL: https://www.elibrary.ru/item.asp?id=42444059
13. Томашевская В.С., Рогов И.Е., Старичкова Ю.В., Карачунский А.И., Румянцев С.А. Опыт и тенденции развития подготовки специалистов по направлениям математических методов и информационных технологий в области здравоохранения. *Вестник Российского университета дружбы народов. Серия: Информатизация образования*. 2015;2:40–48.

12. Prokopov N.I., Ivanov S.Yu., Tomashevskaya V.S., Antonyuk S.N., Ivanova D.V. Scientific potential of a modern university: prospects for preparation of scientific and pedagogical personnel via postgraduate education. *Obshchestvo: sotsiologiya, psikhologiya, pedagogika = Society: Sociology, Psychology, Pedagogics*. 2020;1(69):14–23 (in Russ.). Available from URL: https://www.elibrary.ru/item.asp?id=42444059
13. Tomashevskaya V.S., Rogov I.E., Starichkova Yu.V., Karachunskii A.I., Rumyantsev S.A. Experience and trends in the development of training specialists in the areas of mathematical methods and information technologies in the field of healthcare. *Vestnik Rossiiskogo universiteta druzhby narodov. Seriya: Informatizatsiya obrazovaniya = RUDN Journal of Informatization in Education*. 2015;2:40–48 (in Russ.).
14. Rogov I.E., Adon'ev A.A., Starichkova Yu.V. Experience in development, trends in the development and implementation of information systems supporting the main educational process. *Sovremennye tekhnologii i IT-obrazovanie = Modern Information Technologies and IT Education*. 2017;13(4):82–90 (in Russ.). https://doi.org/10.25559/SITITO.2017.4.628
15. Shalamkov S.A., Starichkova Yu.V. Experience in the development and implementation of the module automate the process of creating and approving programs of study within the information educational environment of support of the basic educational process. *Vestnik Rossiiskogo universiteta druzhby narodov. Seriya: Informatizatsiya obrazovaniya = RUDN Journal of Informatization in Education*. 2015;4:67–76 (in Russ.).
16. Karpenko A.P., Dobryakov A.A. Modeling support for automated learning systems. Review. *Nauka i obrazovanie = Science & Education*. 2011;7 (in Russ.). Available from URL: http://technomag.edu.ru/doc/193116.html

14. Рогов И.Е., Адоньев А.А., Старичкова Ю.В. Опыт разработки, тенденции развития и внедрения информационных систем поддержки основного образовательного процесса. *Современные технологии и ИТ-образование*. 2017;13(4):82–90. https://doi.org/10.25559/SITITO.2017.4.628
15. Шаламков С.А., Старичкова Ю.В. Опыт разработки и внедрения модуля автоматизации процесса создания и утверждения программ учебных дисциплин в рамках информационной образовательной среды поддержки основного образовательного процесса. *Вестник Российского университета дружбы народов. Серия: Информатизация образования*. 2015;4:67–76.
16. Карпенко А.П., Добряков А.А. Модельное обеспечение автоматизированных обучающих систем. Обзор. *Наука и образование*. 2011;7. URL: http://technomag.edu.ru/doc/193116.html

About the authors

Julia V. Starichkova, Cand. Sci. (Eng.), Head of the Department of Artificial Intelligence Technologies, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: starichkova@mirea.ru. RSCI SPIN-code 3001-6791, https://orcid.org/0000-0003-1804-9761

Igor E. Rogov, Director of the Development Program Directorate, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: 2030@mirea.ru. Scopus Author ID 57220747598, RSCI SPIN-code 5036-3472, https://orcid.org/0000-0001-6095-9502

Valeriya S. Tomashevskaya, Cand. Sci. (Eng.), Assistant Professor, Department of Corporate Information Systems, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: tomashevskaya@mirea.ru. RSCI SPIN-code 9099-0294, https://orcid.org/0000-0002-6460-2866

Об авторах

Старичкова Юлия Викторовна, к.т.н., заведующий кафедрой технологий искусственного интеллекта Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: starichkova@mirea.ru. SPIN-код РИНЦ 3001-6791, https://orcid.org/0000-0003-1804-9761

Рогов Игорь Евгеньевич, директор Дирекции программы развития ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: 2030@mirea.ru. Scopus Author ID 57220747598, SPIN-код РИНЦ 5036-3472, https://orcid.org/0000-0001-6095-9502

Томашевская Валерия Сергеевна, к.т.н., доцент кафедры корпоративных информационных систем Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: tomashevskaya@mirea.ru. SPIN-код РИНЦ 9099-0294, https://orcid.org/0000-0002-6460-2866

Translated from Russian into English by Kirill V. Nazarov
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 004.023
https://doi.org/10.32362/2500-316X-2023-11-1-18-30

RESEARCH ARTICLE

Continuous genetic algorithm for grasping an object of a priori unknown shape by a robotic manipulator

Andrey D. Voronkov @,
Sekou A.K. Diane

MIREA – Russian Technological University, Moscow, 119454 Russia
@ Corresponding author, e-mail: a.voronkov.rtu@yandex.ru

Abstract
Objectives. The problem of providing the interaction of a robotic manipulator with a priori unknown objects in a given workspace is of great interest both to the research community and many industries. By developing a solution to this problem, it will be possible to reduce the time taken for robots to adapt to new environments and objects therein. One of the primary stages of providing the interaction of the robotic manipulator with objects is the search for the target position of the robot gripper based on the onboard sensor subsystem, which can be carried out by a number of methods. Methods associated with machine learning and self-learning technologies may not be suitable for some applications (for example, during rescue operations) when it is necessary to quickly search for the target position of the gripper for an a priori unknown object, about which there is no relevant information in the robot database. Therefore, for this problem, heuristic approaches – for example, genetic algorithms – seem to be applicable. The objectives of this work are to implement a search based on a continuous genetic algorithm for the target position of the robot gripper including collision avoidance and study its performance under virtual simulation.
Methods. A heuristic search algorithm (continuous genetic algorithm) is used. The complex scene analysis algorithm uses classical image processing methods. In order to evaluate the effectiveness of the algorithm, virtual simulation is used.
Results. The possibility of using a continuous genetic algorithm is analyzed in the problem of grasping an object of an a priori unknown shape avoiding collisions with other objects of a static scene. A complex scene analysis algorithm and implementation of a continuous genetic algorithm are presented for finding the target position of the gripper of a Kuka LBR iiwa 7 R800 robotic control system with redundant kinematics. The results of an experimental virtual simulation of the obtained algorithm are presented.
Conclusions. The conducted research demonstrates the effectiveness of the continuous genetic algorithm in obtaining the target position of the gripper of the robotic manipulator under conditions when the static scene represents randomly located objects of various shapes.

Keywords: continuous genetic algorithm, grasping of objects of unknown shape, positioning of gripper, collision avoidance, robotic manipulator

• **Submitted:** 17.03.2022 • **Revised:** 12.04.2022 • **Accepted:** 26.10.2022

For citation: Voronkov A.D., Diane S.A.K. Continuous genetic algorithm for grasping an object of a priori unknown shape by a robotic manipulator. *Russ. Technol. J.* 2023;11(1):18–30. https://doi.org/10.32362/2500-316X-2023-11-1-18-30

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

Непрерывный генетический алгоритм в задаче захвата манипуляционным роботом объекта априорно неизвестной формы

А.Д. Воронков @,
С.А.К. Диане

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

Резюме
Цели. Задача взаимодействия манипуляционного робота с априорно неизвестными объектами рабочей области представляет большой интерес для научного сообщества и множества отраслей. Решение этой задачи позволит сократить время адаптации робота к новым средам и объектам в них. Один из главных этапов взаимодействия манипуляционного робота с объектами сцены – поиск целевого положения захватного устройства на основе бортовой сенсорной подсистемы – может быть осуществлен рядом методов. Методы, связанные с технологиями машинного обучения и самообучения, могут быть неподходящими для некоторых областей применения (например, во время аварийно-спасательных работ), когда требуется быстро осуществить поиск целевого положения захватного устройства для априорно неизвестного объекта, информации о котором нет в базе данных робота. Поэтому для этой задачи представляются применимыми эвристические подходы, например, генетический алгоритм. Целями работы являются реализация поиска целевого положения захватного устройства с избеганием столкновений на основе непрерывного генетического алгоритма и исследование его работоспособности в условиях виртуального моделирования.
Методы. Использован эвристический алгоритм поиска – непрерывный генетический алгоритм. В комплексном алгоритме анализа сцены использованы классические методы обработки изображения. Использовано виртуальное моделирование для оценки эффективности алгоритма.
Результаты. В работе рассмотрена возможность применения непрерывного генетического алгоритма в задаче захвата объекта априорно неизвестной формы с избеганием столкновений с другими объектами статической сцены. Представлен комплексный алгоритм анализа сцены и реализация непрерывного генетического алгоритма для решения задачи поиска целевого положения захватного устройства робота избыточной кинематики Kuka LBR iiwa 7 R800. Проведены эксперименты и приведены результаты виртуального моделирования полученного алгоритма.
Выводы. Проведенное исследование позволяет утверждать, что непрерывный генетический алгоритм достаточно эффективен в задачах поиска целевого положения захватного устройства манипуляционного робота при условиях, когда статическая сцена представляет собой хаотично расположенные объекты разной формы.

Ключевые слова: непрерывный генетический алгоритм, захват объектов неизвестной формы, позиционирование захватного устройства, избегание столкновений, манипуляционный робот

• **Поступила:** 17.03.2022 • **Доработана:** 12.04.2022 • **Принята к опубликованию:** 26.10.2022

Для цитирования: Воронков А.Д., Диане С.А.К. Непрерывный генетический алгоритм в задаче захвата манипуляционным роботом объекта априорно неизвестной формы. *Russ. Technol. J.* 2023;11(1):18–30. https://doi.org/10.32362/2500-316X-2023-11-1-18-30

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

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Topical tasks of modern robotic manipulators typically involve the indoor operation of robots. Under such conditions, a robot may collide with a large number of objects of a priori unknown shape, color, or texture. Images obtained using RGBD cameras constitute the source of a significant amount of initial data used for solving such robotics problems[1]. Among the tasks of manipulation robotics, for which a single RGBD camera is usually sufficient, the most relevant are: cleaning up premises, emergency rescue operations, work with products on a conveyor belt. Such tasks do not require a top-level control system to determine whether objects belong to any particular class. Thus, at the moment, the problem of interaction of robotic manipulator with objects of the workspace, whose shape, class and texture are unknown in advance, is relevant. The solution of this problem would allow robotic manipulators to more effectively interact with dozens of different objects that are found in the environment. Correct positioning of the operating tool of the robotic manipulator plays a key role in the interaction of the robot with the scene.

It can be argued that models, methods and algorithms for grasping and transferring a priori unknown objects based on RGBD images open up wide opportunities in such areas as service robotics and special robotics for emergency rescue operations. These capabilities can be used when the robot is operating in unprepared and uncontrolled real-world environment.

CURRENT STATUS OF THE PROBLEM

The problem of grasping an a priori unknown object by a robotic manipulator is complex due to the need to process data from sensors, undertake scene analysis, and assess the planned grasp in accordance with certain criteria, bearing in mind the need to minimize the time spent on searching for options for the positions of the gripper. Much research is devoted to solving the problem of grasping a priori unknown objects while ensuring collision avoidance [1–4]. The complete solution to this problem includes the execution of a number of subtasks, among which may be included: segmentation of a priori unknown objects in an RGBD image, reconstruction of the shape of objects, determination of the position and orientation of selected a priori unknown objects, as well as synthesis of hypotheses about the optimal gripper pose with collision avoidance.

The authors of [1, 5] resort to approximation of the gripper shape using a mathematical model of a displaced and oriented cylinder, which is subsequently used to synthesize possible gripper positions and select the optimal one. It is noted in the papers that most of the existing approaches that analyze an RGBD image extract a flat surface of a table or floor on which objects are located using RANSAC[2]. By this means, the stage of segmenting the entire scene using a neural network to obtain a set of clusters of points that reflect objects on the surface can be avoided. In addition, the authors distinguish approaches to scene analysis when grasping unknown objects into global, where the selection of the gripper position is performed based on the reconstructed 3D model of the object, and local methods, which rely on the boundaries or segments of objects in the image. In this context, the approach in the present paper is for the most part global, since the search for the target position of the gripper is carried out on the basis of a selected subset of the point cloud. Working with three-dimensional data of a point cloud increases the time for finding a solution, but allows the possibility of checking for collisions of the gripper with scene objects.

Information about the segments of all visible individual objects of a static scene can form the basis for the synthesis of grasping robot configurations. For this, the authors of [6] studied the segmentation of objects without a priori knowledge of their classes. Here, a hierarchical neural network structure is used. With this approach, the neural network is able to segment the visible parts of objects and extract the predicted full segments of objects, including their hidden parts, as well as segment the "blocked" areas of objects in an RGBD image. This hierarchy, comprising the novelty of the approach, lies in the fact that each new unit of information is obtained on the basis of the previous ones. Using such a neural network, a robotic manipulator can obtain a target item from a pile of items. To do this, it removes objects blocking the target object one by one, obtaining the target position of the manipulator using the Contact-GraspNet neural network [2], which generates the position and orientation of the parallel gripper, as long as the target object has "blocked" areas. Thus, the successive grasp and transfer of objects that prevent access to the target object is achieved, followed by the grasping of the target object itself.

Reconstruction of selected objects is also the subject of a large number of studies. The reconstructed object model allows synthesizing many options for the location of the grasping device relative to the object, after which the optimal option is selected. Thus, the problem can be solved using approximation of the found subset of the point cloud with a superquadric model or a primitive-body model and through reconstruction using the

[1] An RGBD camera is a sensor widely used in robotics that provides, in addition to a color image of the environment, depth information, i.e., information about the distance from the camera to the obstacle for each pixel.

[2] Random sample consensus (RANSAC) is an iterative algorithm for estimating the parameters of a mathematical model using random samples from initial data.

machine learning algorithm that analyzes a part of the surface.

In [7], an approach is used based on replacing the selected and processed subset of the point cloud, reflecting a separate object, with a superquadric body. All possible configurations of members of the superquadric family are described by eleven parameters. During the operation of the algorithm, the plane of the table and clusters of points belonging to individual objects are selected. To ensure a robust grasping, two criteria are used: placing the gripper as close as possible to the centroid of the superquadric and placing the touching points in places with the least curvature of the surface. The grasping of an object is carried out by synthesizing the set of possible grasps and selecting the first reachable grasp.

The authors of [8] propose an approach to grasping objects of any shape by representing them as a set of primitive bodies, such as spheres, cylinders, right-angle parallelepipeds and cones. The approach is based on the logic of human behavior when trying to grasp an object. The paper also considers a set of rules that describe the starting positions of the grasping device relative to the primitive body. Thus, summarizing information about the desired location of the gripper can be embedded in the upper-level control system in advance.

In [9], the possibility of reconstructing a scene obtained based on a depth map from one camera position is studied. For the resulting incomplete representation of the scene, the "random forest" algorithm compares each 3D point obtained from the depth map with a prediction of the values of the TSDF function[3] for its neighborhood. An array of such predictions forms the resulting surface. The training sample is formed from a 3D scene model. What makes this work interesting is its lack of reliance on information about the classes of objects that the scene contains. This is of great advantage in this case due to the possibility of applying the solution to reconstruct the shape of a single object.

The continuous genetic algorithm is a classical version of the genetic algorithm, with the exception that the genes of individuals are comprised of real numbers [10]. Thus, the individual of the population is itself a vector of real numbers that contains the solution to the problem. This approach makes it possible to search for a solution in a continuous solution space, which is preferable for a number of problems. A continuous genetic algorithm is especially suited for solving the problem of finding a position and orientation in space that satisfy some criterion.

The authors of [11] propose a method for determining the position and orientation of an object with a non-deformable structure from a black-and-white image of the object and a 3D model of the object known in advance. The initial data consist in a set of key points found on a black and white image of an object detected by the SUSAN[4] algorithm, including a 3D model of the object with a set of key points marked on it. The authors used three Euler angles α, β, γ along with three projections of the transfer vector $\mathbf{T}_x$, $\mathbf{T}_y$, $\mathbf{T}_z$ as genes of an individual's chromosome. The fitness function was calculated as the average value of the distance between each key point recognized by the SUSAN algorithm and the key point of the model closest to it projected onto the image plane. As a result, the genetic algorithm selected such solutions in which the key points of the modeled object coincided with the original ones. The criterion for the end of the algorithm—i.e., the threshold value of the average distance at which the position and orientation were visually found correctly—was determined experimentally. The algorithm scheme used single-point crossing, selection by roulette wheel rotation, and elite generational reproduction—the transition of the best individual of the current generation to the next generation[5].

An example of a local approach in the terminology of [5] is presented in [12]. Here, the authors used a 2D depth image as input. The position and orientation of the robot's grasping device is determined based on a search for the faces of objects. The main disadvantage of this approach is the lack of collision avoidance with other obstacles, while the main the advantage is the high speed of work.

The authors of [3] and [13] also use approaches based on the applications of neural networks. In [3], based on the TSDF representation of the scene, the convolutional neural network calculates the expected grasping quality index, orientation, and width of the grasping device opening in one pass for each scene voxel. To do this, the neural network was trained on reliable data obtained in the simulator. In [13], in order to achieve the segmentation of a priori unknown objects, the neural network generates a feature map that reflects where areas with the same properties are concentrated in the image. Then, based on this map, the clustering algorithm determines the number of clusters in the image

[3] Truncated signed distance function (TSDF) is a function for representing a three-dimensional surface as a voxel grid, each voxel being marked by the distance to the nearest surface.

[4] SUSAN is an algorithm for detecting features in an image that analyzes brightness changes in the local neighborhood of the considered point.

[5] Batishchev D.I., Neimark E.A., Starostin N.V. *Application of genetic algorithms to solving discrete optimization problems.* Teaching and methodological materials for the advanced training program "Information technology and computer modeling in applied mathematics." Nizhny Novgorod: UNN; 2007. 85 p. (in Russ.).

and provides information about the resulting segments. Data on segments of a priori unknown objects can form the basis for further analysis of the scene in order to obtain the target position of the grasping device.

As part of solving the problem of analyzing a complex scene for the interaction of a robotic manipulator with a priori unknown objects of the scene, several training data sets have also been developed, such as the Object Segmentation Database. This dataset is designed to train neural networks in the task of segmenting randomly located unknown objects of various shapes and contains 111 data units, which include the RGB image of the scene, the depth map, and segmentation information. Segmented objects are assigned to only one class comprising "object"; all copies of this class are selected separately on the training images. Each element of this dataset consists of an RGB image, a depth map, and an annotated image with object segments. The scenes presented in the dataset consist of several randomly arranged objects of different shapes and textures.

Thus, the task considered in the present paper is relevant since it enables finding the required position of the grasping device for several epochs of the genetic algorithm via a heuristic approach.

APPLICATION OF A GENETIC ALGORITHM TO SEARCH THE TARGET POSITION OF A GRIPPER IN A COMPLEX SCENE

In the present work, a continuous genetic algorithm was used to find the target position of the gripper in a complex scene with randomly located objects of unknown shape. The general algorithm consists of several stages, shown in Fig. 1. As can be seen, at the beginning of the algorithm, a random scene is generated, consisting of various randomly located objects. Then the Kuka LBR iiwa 7 R800 robotic manipulator (manufactured by KUKA, Germany) moves to a predetermined configuration such that the optical axis of the RGBD camera fixed on the gripper is directed perpendicular to the scene plane. In this position, the RGBD camera takes RGB and depth shots of the scene. The resulting images are processed to produce an RGBD image, fed to the input of a neural network with the U-Net architecture, the output of which is also processed and represents a segment of the object closest to the camera, which is further considered as the target. Based on the depth map, a point cloud of the scene is constructed; this is then divided into two subsets based on information about the segment of the target object. In a subset of the point cloud of the object and obstacles, unobservable areas are eliminated using 2D Delaunay triangulation. Then, using a continuous genetic algorithm, the position of the gripper is determined in order to grasp the target object. To reach the target point, the robot first moves to the pre-grip position, then iteratively moves the gripper towards the object until the target position is reached. The grasped object is then transferred to the target area for the objects. These steps are discussed in more detail below.

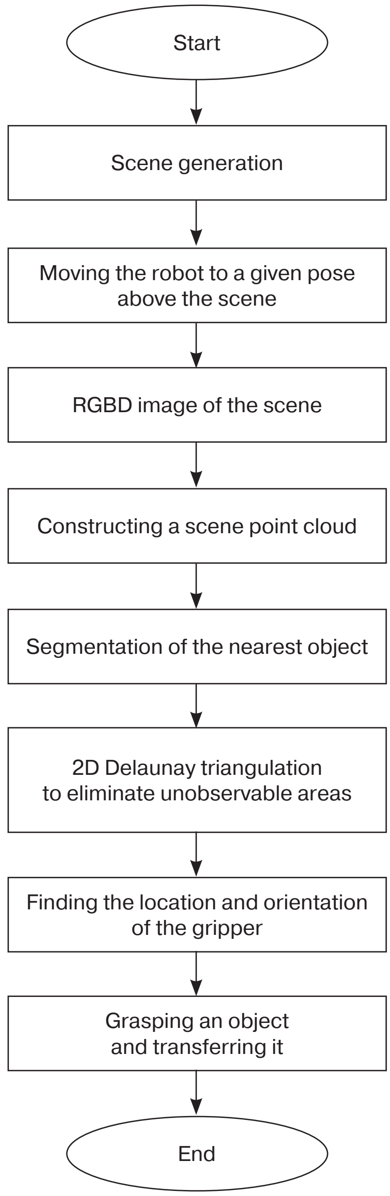


Fig. 1. General diagram of the algorithm

The initial data comprises an RGBD image from a camera attached to the robot's gripper. The result of the algorithm is a vector of generalized coordinates that describes the angles of the robot's drives at which the grasping of the target object becomes possible.

To create a chaotic scene, 20 objects, comprising bricks, stones, beams, etc., were modeled. The objects had a uniform texture and different shapes, including asymmetric, as well as the corresponding mass parameters. In the process of generating the scene, the

(a) (b)

Fig. 2. Creating a chaotic scene:
(a) models of individual scene objects; (b) generated random scene

algorithm randomly selected an object, added it to the scene, applied rotation and displacement to it, after which the algorithm waited 10 s for the object to fall and stop moving in the simulator. Thus, the presence in the desired area of the scene of a given number of randomly selected and randomly located objects was achieved. The modeled objects and generated random scene are shown in Fig. 2.

Following the creation of the random scene, the robot was moved to an initial position specified by a predefined vector of generalized coordinates. On the basis of RGB and depth maps of the scene, a point cloud was constructed and data were generated to be delivered to the input of a neural network with the U-Net architecture [14]. The neural network was preliminarily trained to segment the nearest object in an RGBD image using a simulator-collected and annotated training set. Here, it was necessary to process the data prior to delivery to the input of the neural network separately from the data received from the output of the neural network. The corresponding processing stages are shown in Fig. 3. As can be seen from the figure, preprocessing consisted in finding a point in the depth map with a minimum distance value, cropping the image section with

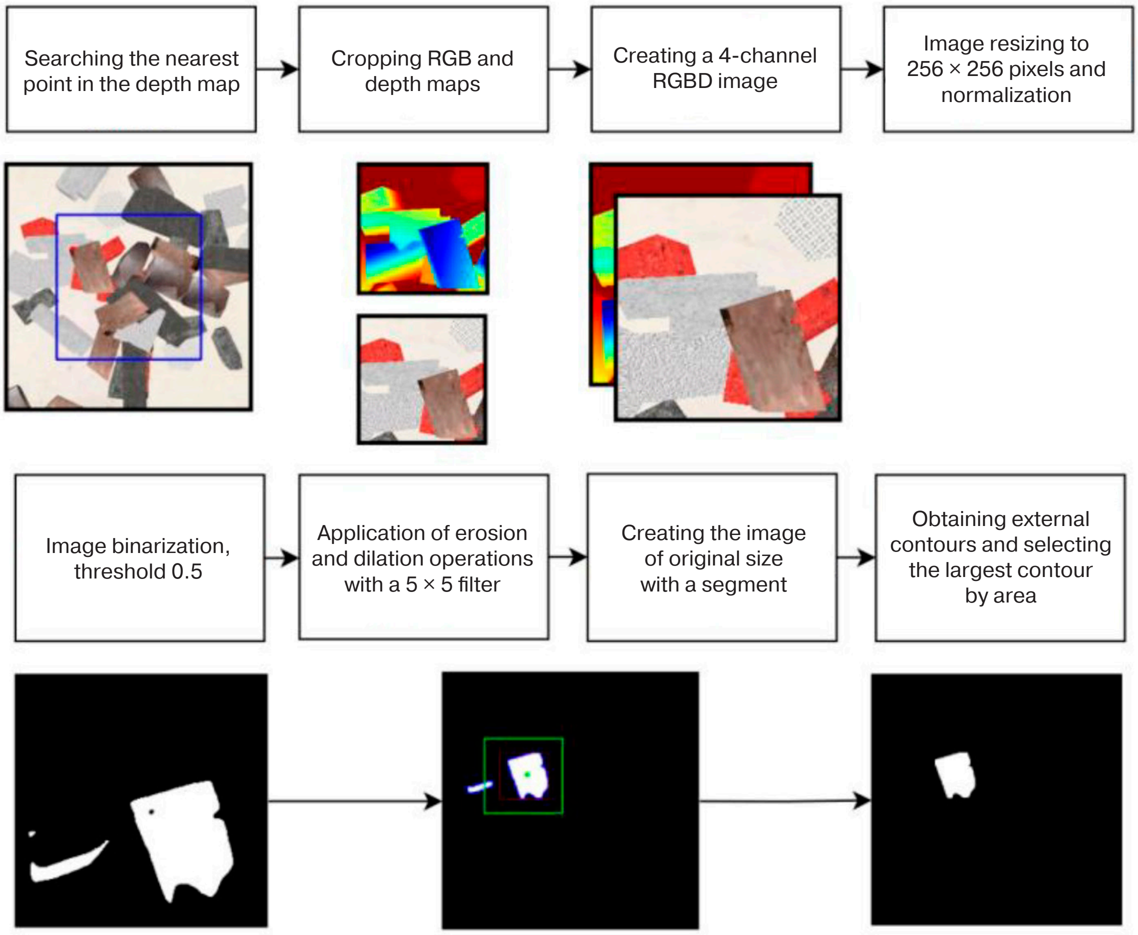


Fig. 3. Preliminary and final data processing for a segmenting neural network with U-Net architecture

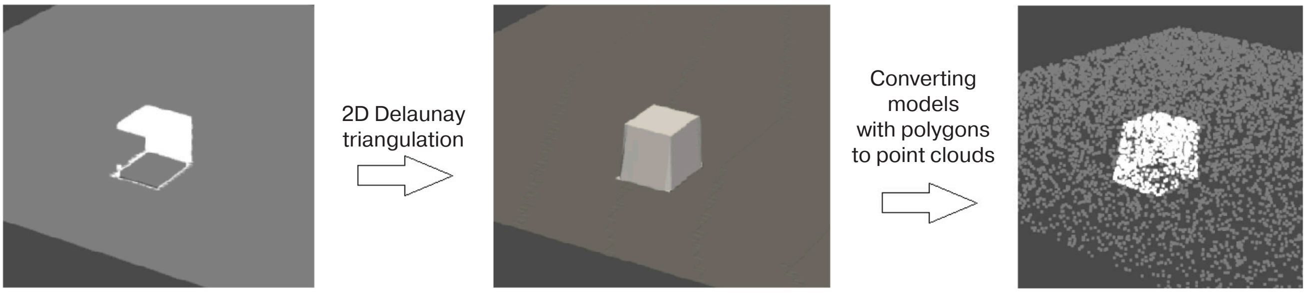
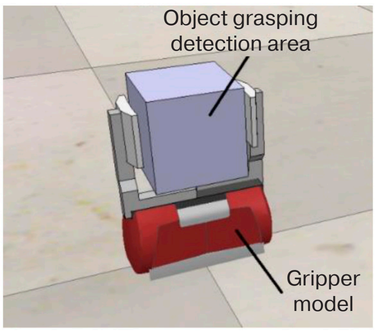


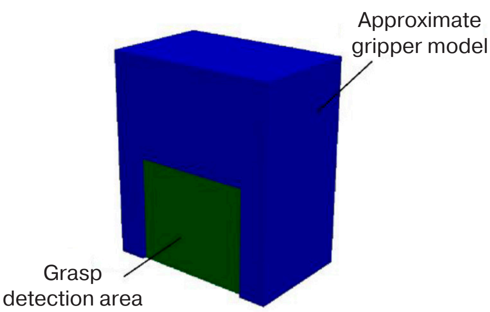
Fig. 4. Transformation of the point cloud in order to eliminate unobservable areas

this point in the center, forming a 4-channel RGBD image from RGB and depth maps, scaling to the size of the neural network input layer 256 × 256, and normalization. Post-processing was required to eliminate the false segments detected by the neural network. To do this, the output of the neural network with a threshold $t = 0.5$ was transformed into a binary mask, with morphological dilation and erosion operators used to eliminate small erroneously segmented areas, and the segment with the largest area being taken as the true segment.

To increase the speed of the algorithm, only a part of the point cloud was constructed near the selected segment. Due to the depth map obtained from a single camera position being the only source of data for constructing a point cloud, there were unobservable areas in the constructed cloud. This could prevent the genetic algorithm from correctly assessing the fitness of an individual and searching for the correct position of the gripper. To solve this problem, the point cloud was divided into two subsets based on the selected segment:



(a)



(b)

Fig. 5. Gripping device models: (a) original model; (b) approximated model

one subset of points belonging to the obstacle and another subset belonging to the target object. Further, using two-dimensional Delaunay triangulation, the polygons between the points were completed. The resulting models with polygons were then sent to the input of an algorithm that converted the 3D model into a point cloud. The resulting 2 point clouds, which did not contain unobservable areas, were consequently suitable for the operation of a continuous genetic algorithm (Fig. 4).

The position and orientation of the working tool of the robotic manipulator relative to the base coordinate system are uniquely described by a six-parameter vector: three projections of the transfer vector and three Tait-Bryan angles describing successive rotations of the object in the *ZYX* order; these six real numbers make up the chromosome of an individual. For the operation of the genetic algorithm, a 3D model of the gripper was used. The detection area was set within which the target object for grasping was to be located. To reduce the time spent on processing the 3D model of the gripper, it was approximated by a low-poly model having the same geometry as the original model (Fig. 5).

To calculate the fitness of an individual, the calculation of the number of points of the target object and the obstacle located inside the gripper model and the grasp detection area is used. To do this, we used an algorithm based on the ray tracing method, which determines the fact that a point is located inside a closed three-dimensional surface. In order for the resulting position of the working body to be oriented vertically, the desired intervals for the angles of rotation relative to the *X* and *Y* axes were set: $-20° < \alpha_g < 20°$, $-20° < \beta_g < 20°$. When calculating the fitness function, going beyond the boundaries of these intervals was taken into account. The fitness of an individual was calculated by the formula:

$$F(p)=\begin{cases}1000+x, \\ \text{if } O_{\text{inside gr}}>0 \text{ or } I_{\text{inside gr}}>0, \\ 500+x, \\ \text{if } O_{\text{inside gr}}=0 \text{ and } I_{\text{inside gr}}=0 \text{ and } I_{\text{inside vol}}=0, \\ \dfrac{1}{I_{\text{inside vol}}}+x, \\ \text{if } I_{\text{inside vol}}>0 \text{ and } O_{\text{inside gr}}=0 \text{ and } I_{\text{inside gr}}=0,\end{cases}$$

where x is the sum of the absolute values of the excess of the selected angles α and β over the given intervals; $O_{\text{inside gr}}$ is the number of points in the obstacle point cloud inside the gripper model; $I_{\text{inside vol}}$ is the number of object points inside the detection area; $I_{\text{inside gr}}$ is the number of object points inside the gripper model.

The genetic crossover and mutation operators used are shown in Fig. 6. As can be seen from the figure, the crossover operator is an elementwise weighted sum of the genes of the two original parental chromosomes, while the weight coefficient is a random number. The mutation operator, designed to introduce random changes in the chromosome of an individual, is presented in three variations: introducing a random change in each gene of the original chromosome with a probability of 50%, introducing a change in a randomly selected gene of each logical subgroup of genes, and initiating a change in one randomly selected gene. The choice of the mutation operator is made with equal probability.

The following scheme was used in the implemented continuous genetic algorithm:

- construction of the initial population near the central point of the segmented object;
- population size: 34 individuals;
- 16 individuals are selected for crossover: 10 best and 6 random individuals;
- pairs for crossover are random, with each pair of individuals interbreeding twice, giving four offsprings;
- offsprings undergo mutation with a probability of 30%, while the type of mutation is chosen with equal probability;
- elite generational reproduction: the 2 most adapted individuals of the previous generation are copied to the next generation and do not undergo mutation;
- criterion for terminating the algorithm's work: reaching a given number of epochs or reaching a certain value of the fitness function.

The following conditions were chosen as criteria for stopping the search for a solution by the genetic algorithm:

- the number of object points located inside the grasping detection area is more than 40% of their total number;
- more than 50 epochs have passed.

The criteria for stopping the search were chosen based on the analysis of the simulation results of the robotic manipulator. During a series of runs of the algorithm with different target objects, it was noticed that the threshold value of the proportion of object points located in the grasping detection area of the total number of points, equal to 40%, in most cases ensures successful grasp of the object. At the same time, the selected ultimate number of epochs of the algorithm, equal to 50, provides a balance between the speed of calculation and the probability of finding a solution; this is despite the fact that in most cases the search for the position of the gripper to successfully grasp the object took much less than 50 epochs.

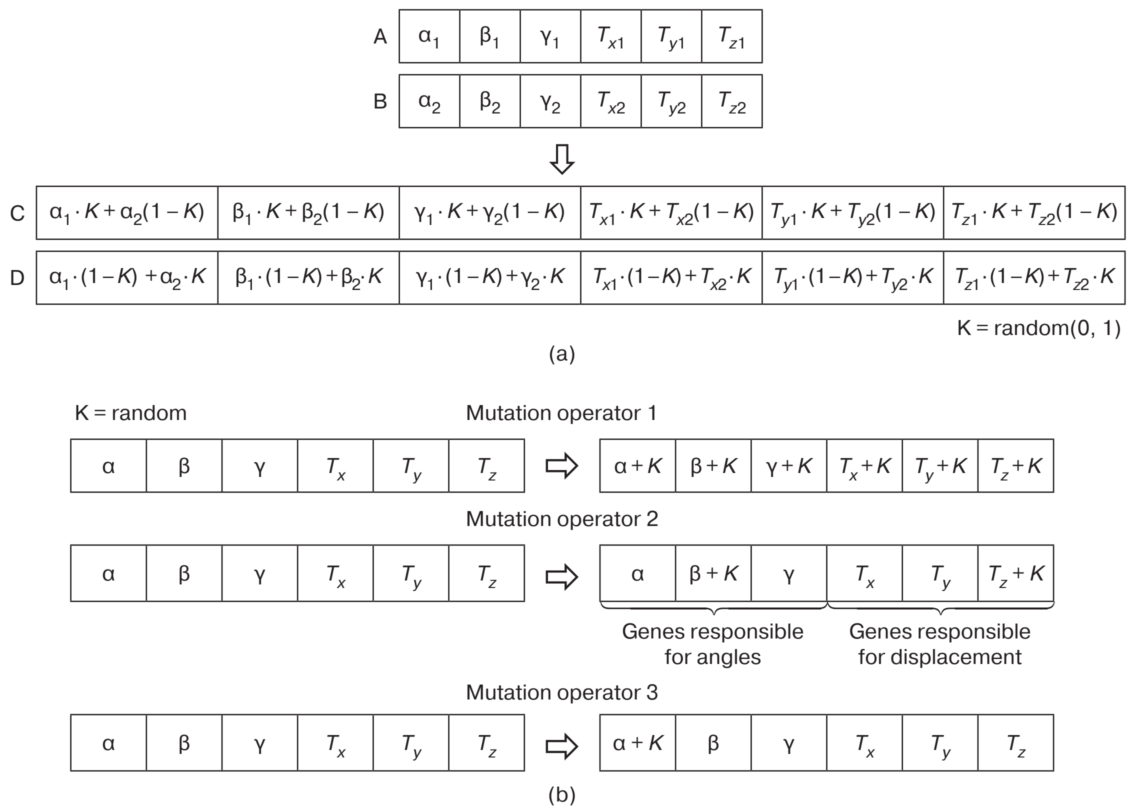


Fig. 6. Genetic operators: (a) crossover operator; (b) set of mutation operators

The target position and orientation of the gripper are determined by the individual with the highest degree of fitness (minimal $F(p)$ value) at the end of the genetic algorithm. Based on these data, the pre-grasp position is calculated by shifting the obtained position along the OZ axis by -0.1 m. Then the drive angles for ten intermediate positions of the gripper between the pre-grasp position and the target position are calculated based on solving the inverse kinematics problem. To grasp the object, the robot approaches the gripper by sequentially setting the drive angles for each intermediate position. In this way, approximation along the vector is achieved to avoid collisions with other objects in the scene. The robot with the grasped object then moves to a position above the target area using a predetermined generalized coordinate vector and the gripper opens. The least squares procedure of the SciPy Python library was used to solve the inverse problem of kinematics. This procedure implements the nonlinear least squares method, in which the search for the minimum of the objective function of several variables is carried out:

$$F(\mathbf{Q}) = \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{4}\left(A_{ij}^{08}(\mathbf{Q}) - A_{ij}^{08\_goal}\right)^2,$$

where $\mathbf{A}^{08}(\mathbf{Q})$ is a homogeneous transformation matrix describing the transition from the world coordinate system to the coordinate system fixed on the gripper, calculated based on the generalized coordinate vector $\mathbf{Q}$ found by the numerical method and representing a matrix of 1 column and 7 rows, the elements of which are the angles of the joints of a robotic manipulator; $\mathbf{A}^{08\_goal}$ is the target uniform transformation matrix computed using the given position and orientation.

Thus, it is possible to determine the assumptions made during the development of the algorithm. Here, it is assumed that the scene objects have a uniform texture and are not deformable, that one of the dimensions of each scene object is less than the maximum opening width of the grippers, that all objects are in the working area of the robot, and that the approximate gripper model has the maximum opening width.

EXPERIMENTAL EVALUATION OF THE METHOD

In order to simulate the interaction of a robotic manipulator with objects of a complex visual scene, a virtual scene was developed in the CoppeliaSim simulator. Its original form is shown in Fig. 7. The scene comprises a model of the Kuka LBR iiwa 7 R800 robotic manipulator with a 512×512-pixel RGBD camera mounted on a gripper and an ambient lighting module designed to make the RGB image uniform by eliminating shadows. There is a source zone for scene generation and a target zone.

In order to study the performance of the continuous genetic algorithm when searching for the target position of the grasping device, five experimental runs of the algorithm were carried out on the generated scene. The process of grasping one object is shown in Fig. 8. Convergence graphs of the continuous genetic algorithm are shown in Fig. 9. As can be seen from the graphs, the algorithm converged to a satisfactory solution in less than 31 epochs in all experimental runs. The initial fitness of the best individual in most cases was high due to individuals of the initial population being created in the vicinity of the central point of the segmented object, along with a maximal width of the opening of the grasping device in the approximated model. This increased the probability that a part of the point cloud

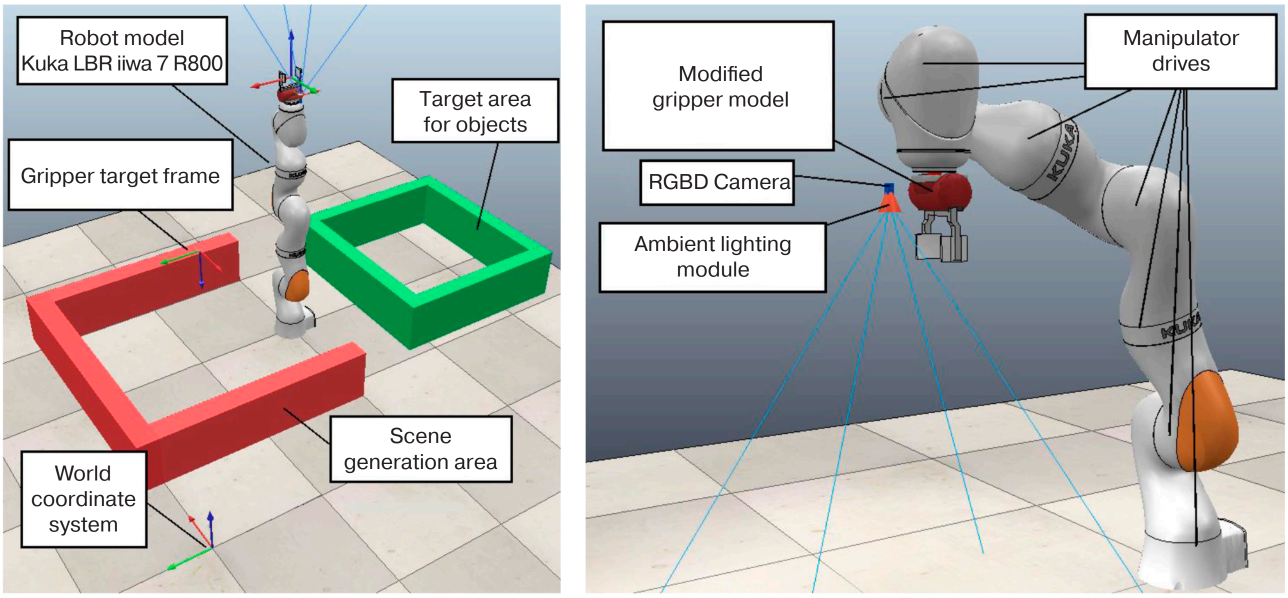


Fig. 7. Scene for simulation: (a) composition of the scene; (b) robot model

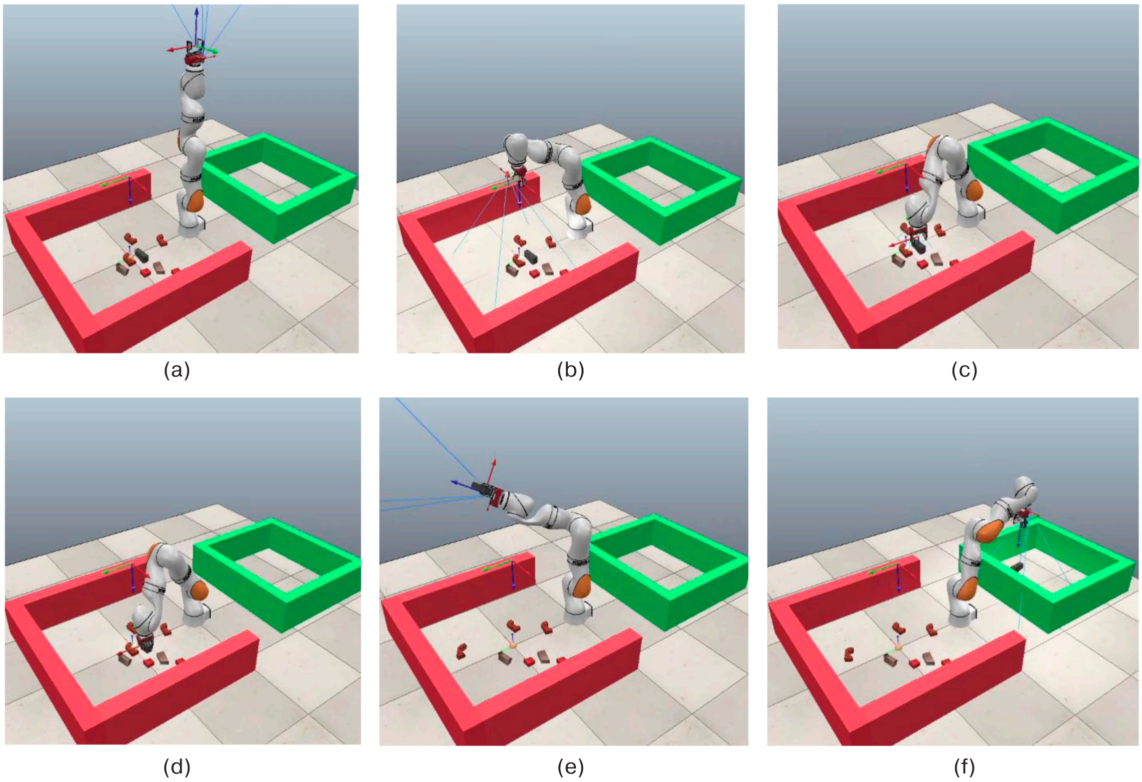


Fig. 8. The process of grasping one object: (a) the configuration of the robot at the start of the program; (b) robot configuration for obtaining RGB and depth maps of the scene; (c) pre-grasp configuration; (d) last intermediate configuration; (e) move to end position; (f) configuration for dropping the object to the target area

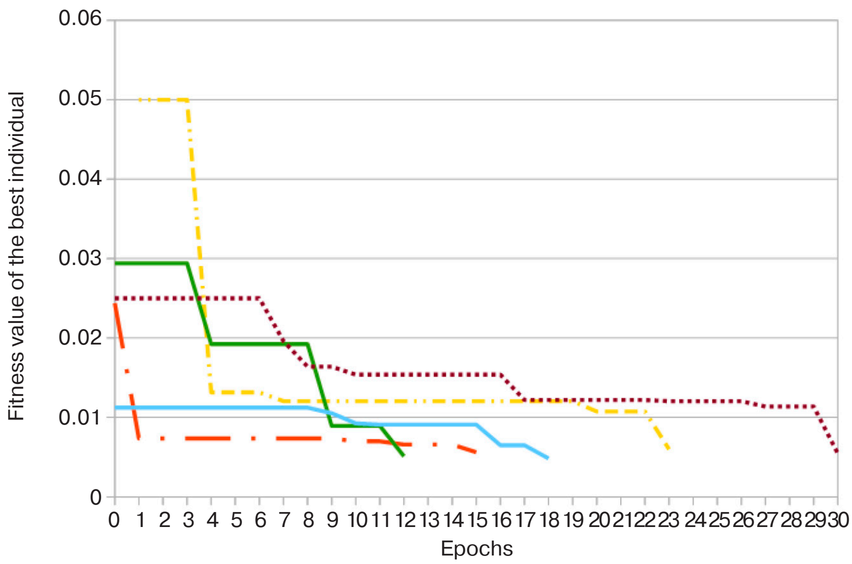


Fig. 9. Charts of convergence

of the object in the created individual was in the grasping detection area, and the gripper did not collide with the object. Since in this case the genetic algorithm needs more iterations to find the position of the gripper without colliding with the object's point cloud, an increase in the convergence time of the algorithm is associated with an increase in the size of the target object. The average convergence time was 1.9 s on a computer equipped with an AMD Ryzen 5 3500U processor; the GPU was not involved.

CONCLUSIONS

On the basis of the developed software for implementing the analysis of the RGBD image of the scene, the genetic algorithm is confirmed as suitable for solving the problem of obtaining the target position of the grasping device. The advantages of the developed solution are the ability to work using a point cloud obtained from one position of the camera, as well as the capability to eliminate unobservable areas. An important role is played by genetic operators and the implementation of the fitness function since their calculation directly affects the average time of finding a solution, which should therefore be minimized. Therefore, measures were taken to increase the speed of the algorithm: use of a subset of the original point cloud and approximation of the gripper with a primitive low-poly model.

Despite the success of the resulting solution, there are opportunities for its further improvement. In particular, the study conducted in [5] clearly indicates that fixing a camera on the end of the robot offers an advantage over a camera fixed above the stage, since it then becomes possible to generate a more detailed point cloud of the scene taken from several positions along a predetermined scanning path. Generating a point cloud based on RGBD images from several camera positions allows it to be constructed more informatively to avoid excessive use of Delaunay triangulation to complete the cloud. Also, in a number of studies [4, 15] parallelism of actions is used; this involves the use of calculations to obtain the target gripper position during the movement of the robot to some intermediate position above the scene. Finally, it is of interest to study the optimal scheme of the genetic algorithm in terms of the number of individuals in the population for obtaining the most efficient crossover and mutation operator.

The novelty of the results obtained lies in the study of the applicability of a continuous genetic algorithm in the problem of positioning the grasping tool of a robotic manipulator. The results of the study can be used to solve the problems of positioning a robotic manipulator in an environment with a priori unknown objects. Examples of such tasks include cleaning premises, performing warehouse operations, and clearing rubble.

Authors' contributions

A.D. Voronkov participated in the development of the concept of using a continuous genetic algorithm in the task of positioning the gripper of the manipulator; developed algorithmic and software implementing a continuous genetic algorithm and a complex algorithm for scene analysis; participated in the drafting of the text and making edits to the article.

S.A.K. Diane participated in the development of the concept of the application of the continuous genetic algorithm; proposed the scenario of the final experiment; provided scientific guidance; participated in the drafting of the text and making edits to the article.

REFERENCES

1. Lei Q., Wisse M. Fast grasping of unknown objects using cylinder searching on a single point cloud. In: *Ninth International Conference on Machine Vision* (*ICMV 2016*). 2016. V. 10341. https://doi.org/10.1117/12.2268422
2. Sundermeyer M., Mousavian A., Triebel R., Fox D. *Contact-GraspNet: efficient 6-DoF grasp generation in cluttered scenes*. arXiv [Preprint]. 2021. 7 p. Available from URL: https://arxiv.org/abs/2103.14127
3. Breyer M., Chung J., Ott L., Siegwart R., Nieto J. Volumetric grasping network: Real-time 6 DOF grasp detection in clutter. In: *4th Conference on Robot Learning* (*CoRL 2020*). P. 1602–1611. Available from URL: https://proceedings.mlr.press/v155/breyer21a/breyer21a.pdf
4. Lippiello V., Ruggiero F., Siciliano B., Luigi V. Visual grasp planning for unknown objects using a multifingered robotic hand. *IEEE/ASME Transactions on Mechatronics*. 2013;18(3):1050–1059. https://doi.org/10.1109/TMECH.2012.2195500
5. Lei Q., Meijer J., Wisse M. A survey of unknown object grasping and our fast grasping algorithm-C shape grasping. In: *2017 3rd International Conference on Control, Automation and Robotics* (*ICCAR*). P. 150–157. https://doi.org/10.1109/ICCAR.2017.7942677
6. Back S., Lee J., Kim T., Noh S., Kang R., Bak S., Lee K. *Unseen object amodal instance segmentation via hierarchical occlusion modeling*. arXiv [Preprint]. 2021. 8 p. Available from URL: https://arxiv.org/abs/2109.11103
7. Abhijit M., Federico T., Perez-Gracia A. Grasping unknown objects in clutter by superquadric representation. In: *2018 Second IEEE International Conference on Robotic Computing* (*IRC*). P. 292–299. https://doi.org/10.1109/IRC.2018.00062
8. Miller A., Knoop S., Christensen H., Allen P. Automatic grasp planning using shape primitives. In: *Proceedings of the 2003 IEEE International Conference on Robotics and Automation*. P. 1824–1829. https://doi.org/10.1109/ROBOT.2003.1241860
9. Firman M., Aodha O., Julier S., Gabriel J. Structured prediction of unobserved voxels from a single depth image. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). P. 5431–5440. https://doi.org/10.1109/CVPR.2016.586
10. Gridin V.N., Solodovnikov V.I. A continuous genetic data preprocessing algorithm searching coefficients of the approximating function. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh = New Information Technologies in Automated Systems*. 2018;21:302–306 (in Russ.).
11. Rossi C., Abderrahim M., Diaz J. EvoPose: A model-based pose estimation algorithm with correspondences determination. *2005 IEEE International Conference Mechatronics and Automation*. V. 3. P. 1551–1556. https://doi.org/10.1109/ICMA.2005.1626786
12. Jabalameli A., Behal A. From single 2D depth image to gripper 6D pose estimation: A fast and robust algorithm for grabbing objects in cluttered scenes. *Robotics*. 2019;8(3):63. https://doi.org/10.3390/robotics8030063
13. Xiang Y., Xie C., Mousavian A., Fox D. Learning RGB-D feature embeddings for unseen object instance segmentation. In: *4th Conference on Robot Learning* (*CoRL 2020*). P. 461–470. Available from URL: https://proceedings.mlr.press/v155/xiang21a/xiang21a.pdf
14. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2015. P. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

15. Dune C., Marchand E., Collowet C., Leroux C. Active rough shape estimation of unknown objects. In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems.* P. 3622–3627. https://doi.org/10.1109/IROS.2008.4651005

About the authors

Andrey D. Voronkov, Postgraduate Student, Department of Control Problems, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: a.voronkov.rtu@yandex.ru. https://orcid.org/0000-0003-4688-9346

Sekou A.K. Diane, Cand. Sci. (Eng.), Assistant Professor, Department of Control Problems, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). Scopus Author ID 57188548666, ResearcherID T-5560-2017, https://orcid.org/0000-0002-8690-6422

Об авторах

Воронков Андрей Дадашевич, аспирант кафедры проблем управления Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: a.voronkov.rtu@yandex.ru. https://orcid.org/0000-0003-4688-9346

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). Scopus Author ID 57188548666, ResearcherID T-5560-2017, https://orcid.org/0000-0002-8690-6422

Translated from Russian into English by Evgenii I. Shklovskii
Edited for English language and spelling by Thomas A. Beavitt

Modern radio engineering and telecommunication systems

Современные радиотехнические и телекоммуникационные системы

UDC 681.883.45
https://doi.org/10.32362/2500-316X-2023-11-1-31-40

RESEARCH ARTICLE

Hydroacoustic communication channel capacity

Valery E. Denisov @

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

Abstract

Objectives. Capacity, describing the maximum rate of information transmission, is an important characteristic of any communication channel. The main purpose of this work is to determine the capacity of a hydroacoustic communication channel with constrained average intensity of the transmitted signal. An additional aim consists in finding the optimal spectrum of a transmitted signal and calculate its boundary frequencies. A model of a single-path channel was considered, which is characteristic of the deep sea with the receiver or transmitter placed at a sufficient depth.

Methods. Concepts of applied hydroacoustics, the theory of random processes, and information theory were used.

Results. An expression for gain in a hydroacoustic communication channel has been obtained. A novel expression derived for the spectral level of sea noise caused by sea surface waves is based on piecewise linear approximation of the curves of the spectral levels of noise obtained from four sources: turbulence, shipping, sea waves, and the thermal noise of the sea. Dependencies of the hydroacoustic channel capacity on communication distance, intensity of the transmitted signal, and sea state, are characterized. The definition of the optimal spectrum itself is determined along with the lower and upper boundary frequencies of the optimal spectrum of the transmitted signal. The dependence of the bandwidth usage on the intensity of the input signal at various communication distances has been investigated.

Conclusions. On the basis of the Francois–Garrison attenuation coefficient, channel capacity was correlated with the parameters of the marine environment: temperature, salinity, and pH in the study area. At a given intensity of the input signal, channel capacity was shown to decrease significantly with increasing distance and sea wave intensity. It is also shown that the width of the optimal spectrum decreases with increasing distance. Sea wave noise was noted to affect significantly the shape of the optimal spectrum and its boundary frequencies. The possibility of cases where bandwidth usage increases with increasing distance at a given input signal intensity cannot be ruled out.

Keywords: hydroacoustic communication channel, sea noise, spectral intensity of sea noise, signal intensity, bandwidth, optimal spectrum, channel bandwidth utilization

• **Submitted:** 28.04.2022 • **Revised:** 21.06.2022 • **Accepted:** 10.11.2022

For citation: Denisov V.E. Hydroacoustic communication channel capacity. *Russ. Technol. J.* 2023;11(1):31–40. https://doi.org/10.32362/2500-316X-2023-11-1-31-40

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

Пропускная способность гидроакустического канала связи

В.Е. Денисов @

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

Резюме
Цели. Пропускная способность является важной характеристикой любого канала связи, так как определяет предельную скорость передачи информации в нем. Основная цель работы – определение пропускной способности гидроакустического канала связи при ограничении на среднюю интенсивность передаваемого сигнала. Дополнительной задачей являлось определение оптимального спектра передаваемого сигнала и расчет граничных частот этого спектра. Была рассмотрена модель однолучевого канала, характерная для глубокого моря, когда приемник или передатчик расположены на достаточной глубине.
Методы. Использованы положения прикладной гидроакустики, теории случайных процессов и теории информации.
Результаты. Получено выражение для коэффициента передачи гидроакустического канала связи и новое выражение для спектрального уровня шумов моря, обусловленных волнением поверхности моря. На основе кусочно-линейной аппроксимации кривых спектральных уровней шумов турбулентности, судоходства, волнения моря и теплового шума моря определена спектральная плотность интенсивности шума моря. Получены зависимости пропускной способности гидроакустического канала от дальности связи, интенсивности передаваемого сигнала и состояния поверхности моря. Определены нижняя и верхняя частоты оптимального спектра передаваемого сигнала и оптимальный спектр. Исследована зависимость коэффициента использования полосы частот от интенсивности входного сигнала для разных значений дальности связи.
Выводы. Использование коэффициента затухания Франсуа – Гаррисона позволило связать пропускную способность канала с параметрами морской среды: температурой, соленостью, значением водородного показателя в исследуемом районе. При заданной интенсивности входного сигнала пропускная способность существенно уменьшается с ростом дальности и усилением волнения моря. Показано, что с ростом расстояния ширина оптимального спектра уменьшается. Отмечается значительное влияние шума от волнения моря на форму оптимального спектра и значения его граничных частот. Было установлено, что возможны случаи увеличения коэффициента использования полосы частот ростом дальности при заданной интенсивности входного сигнала.

Ключевые слова: гидроакустический канал связи, шумы моря, спектральная интенсивность шумов моря, интенсивность сигнала, пропускная способность, оптимальный спектр, коэффициент использования полосы частот канала

• **Поступила:** 28.04.2022 • **Доработана:** 21.06.2022 • **Принята к опубликованию:** 10.11.2022

Для цитирования: Денисов В.Е. Пропускная способность гидроакустического канала связи. *Russ. Technol. J.* 2023;11(1):31–40. https://doi.org/10.32362/2500-316X-2023-11-1-31-40

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

Автор заявляет об отсутствии конфликта интересов.

INTRODUCTION

When developing digital hydroacoustic communication systems, the capacity of a given hydroacoustic communication channel (HACC), representing the maximum rate of information transmission at a given sea noise and given constraints on the transmitted signal, is of particular interest. Given knowledge of HACC capacity, it is always possible to estimate the efficiency of the communication system under development. HACC capacity has been estimated by a number of national and international researchers under various conditions. Works [1–4] are particularly noteworthy among works in which capacity determination techniques are described. The closest approach to that taken the present work is the one given in Stojanovic's study [1], where the capacity of a single-path HACC was obtained using the frequency dependencies of the attenuation coefficient and sea noise. Assuming nonspherical spreading, as it should be in the case with a point source, the Urick attenuation coefficient [5], which is notionally independent of the parameters of the marine environment, was used. Here, the so-called practical spreading approach was used, in which the sound intensity taken as inversely proportional to the distance to the power of 1.5. However, as was noted [5], if the properties of the marine environment are insufficiently known, it is recommended to apply the spherical law instead. The HACC capacity was also calculated [1] using a special approximating formula for the spectral noise density [6], which gives somewhat overestimated values in comparison with the Urick curves [6]. However, Stojanovic [1] considered only the case of calm sea conditions. In the present work, the capacity of a single-path HACC was calculated using the frequency dependence of the attenuation coefficient described by the Francois–Garrison formula. Considered the most accurate at the present time, this formula is based around the parameters of the marine environment. In addition, a simple and illustrative piecewise linear approximation of the spectral level of sea noise was used. The dependence of the channel capacity on the intensity of sea surface waves was additionally investigated. Considering research carried out in recent years, the works [7–9] should be noted, in which various hydroacoustic communication systems were analyzed and their corresponding information transmission rates estimated.

1. HYDROACOUSTIC CHANNEL GAIN

In the emission of a harmonic signal, the instantaneous pressure $p(t, R)$ in a boundless homogeneous absorbing marine environment at distance R from a point source can be represented as follows [10–12]:

$$p(t,R)=\sqrt{2}P(R)\cos\left[\omega\left(t-\frac{R}{c}\right)\right], \tag{1}$$

where

$$P(R)=P_0\frac{R_0}{R}\exp\left[-\alpha_e\left(R-R_0\right)\right] \tag{2}$$

is the root-mean-square pressure, Pa; $\omega = 2\pi f$ is the angular velocity of the source vibrations, rad/s; c = 1500 m/s is the speed of sound in seawater; R is distance, m; α_e is the attenuation coefficient in the marine environment, Np/m; P_0 is the root-mean-square pressure at reference distance R_0 (typically, R_0 = 1 m).

From expressions (1) and (2), the complex gain of the HACC can be expressed:

$$H(j\omega)=\frac{R_0}{R}\exp\left[-\alpha_e\left(R-R_0\right)-j\omega\frac{R}{c}\right]= \\ =H(\omega)\exp[j\vartheta(\omega)],$$

where $H(\omega)$ is the frequency response and $\upsilon(\omega)$ is the phase response. They are found as follows:

$$H(\omega)=\left|H(j\omega)\right|=\frac{R_0}{R}\exp\left[-\alpha_e\left(R-R_0\right)\right],\ \upsilon(\omega)=-\omega R/c.$$

In practice, the attenuation coefficient is typically expressed in dB/km, and the distance R is expressed in kilometers. Then, at R_0 = 1 m, the formula for the frequency response takes the form

$$H(f)=\frac{1}{R_{\text{km}}\cdot 10^3}\exp(-0.115\alpha R). \tag{3}$$

As follows from expression (3), at a given communication distance R, the properties of the frequency response are completely determined by the frequency response of the attenuation coefficient. Therefore, let us dwell in more detail on the attenuation coefficient of seawater. α has to be calculated using various semi-empirical formulas obtained on the basis of theoretical and experimental data. One of such formulas for calculating α is the Francois–Garrison formula [13, 14], which was derived by analyzing and generalizing the known results of measurements in the ocean, including measurements in the Arctic. The sound attenuation coefficient α in the frequency range from 200 Hz to 1 MHz to a depth of 5 km can be represented as follows [13, 14]:

$$\alpha = \alpha_1 + \alpha_2 + \alpha_3,\ \text{dB/km},$$

where α_1 is the absorption coefficient due to ionic relaxation of boron compounds (borates), α_2 is the absorption coefficient due to ionic relaxation of magnesium sulfate ($MgSO_4$), and α_3 is the absorption coefficient due to the shear viscosity of fresh water and the structural relaxation of fresh water molecules. These quantities are expressed as follows:

$$\alpha_1=\frac{A_1P_1f_1f^2}{f_1^2+f^2},\ \alpha_2=\frac{A_2P_2f_2f^2}{f_2^2+f^2},\ \alpha_3=A_3P_3f^2.$$

In the above formulas, the frequency f is expressed in kHz, and A_i (i = 1, 2, 3), P_i (i = 1, 2, 3), f_1, and f_2 are written as follows:

$$A_1=\frac{8.86}{c}\cdot 10^{0.78\text{pH}-5},\ (\text{dB/km})\cdot\text{kHz}^{-1};$$

$$P_1=1, f_1=2.8(S/35)^{0.5}\cdot 10^{4-1245/\Theta},\ \text{kHz};$$

$$A_2 = 21.44\frac{S}{c}(1+0.025T),\ (\text{dB/km})\cdot\text{kHz}^{-1};$$

$$P_2 = 1 - 1.37\cdot10^{-4}z + 6.2\cdot10^{-9}z^2;$$

$$f_2 = \frac{8.17\cdot10^{8-1990/\Theta}}{1+0.0018(S-35)},\ \text{kHz};$$

$$P_3 = 1 - 3.83\cdot10^{-5}z + 4.9\cdot10^{-10}z^2;$$

$$A_3 = \begin{cases} 4.937\cdot10^{-4} - 2.59\cdot10^{-5}T + \\ + 9.11\cdot10^{-7}T^2 - 1.50\cdot10^{-8}T^3, \\ T \le 20°\text{C}, \\ \\ 3.964\cdot10^{-4} - 1.146\cdot10^{-5}T + \\ + 1.45\cdot10^{-7}T^2 - 6.50\cdot10^{-10}T^3, \\ T > 20°\text{C}, \end{cases} \quad (\text{dB/km})\cdot\text{kHz}^{-2}.$$

Here, c is the speed of sound in seawater, m/s, which is approximately calculated as follows:

$$c \approx 1412 + 3.21T + 1.19S + 0.0167z;$$

T is the water temperature, °C; $\Theta = 273 + T$ is the water temperature expressed in kelvins, K; S is salinity, ‰; z is depth, m; and pH is the pH of the water in the area of measurement α. The pH of the ocean water ranges from 7.5 to 8.4; the pH of the water in a certain area is available in the literature[1].

At frequencies below 200 Hz, the Francois–Garrison equation may be invalid. The results of attenuation measurements at frequencies below 200 Hz were analyzed by Guilysse and Sabathe [15], who argued that the attenuation of sound at these frequencies is caused by the scattering of sound by inhomogeneities of the marine environment. The attenuation coefficient α_4 in this frequency range id independent of frequency, but varies with latitude from 0.004 dB/km in polar waters to 0.0002 dB/km in the tropics. Stojanovic [1] used the value α_4 = 0.003 dB/km. The Francois–Garrison equation is applicable to all ocean conditions in the frequency range from 200 Hz to 1 MHz. Using this equation, the attenuation coefficient can be calculated with an accuracy of 10%. To obtain an expression for the sound attenuation coefficient in the frequency range from 10 Hz to 1 MHz, it is necessary to add the attenuation coefficient α_4 to the Francois–Garrison attenuation coefficient. Further, it is assumed that α_4 = 0.003 dB/km. Therefore, the sound attenuation coefficient at frequency f has the form: $\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$.

To calculate the attenuation coefficient from the Francois–Garrison formula, it is necessary to know the water temperature T, salinity S, pH, and depth z along the entire sound propagation path.

Urick [5] proposed a simpler formula for the sound attenuation coefficient at a water temperature of T = 4°C, which is characteristic of great depths in the deep sea:

$$\alpha_\text{U} = \frac{0.11f^2}{1+f^2} + \frac{44f^2}{4100+f^2} + 3\cdot10^{-4}f^2.$$

Here, α_U is the attenuation coefficient, dB/km; and f is frequency, kHz.

The Urick formula is also based on the results of experimental studies, and its components are similar to those of the Francois–Garrison formula. However, Urick [5] provided no information on the accuracy of the approximation of the known expressions for the attenuation coefficient by this formula. Moreover, the frequency range in which this formula is valid was not indicated. It was of interest to compare the Francois–Garrison and Urick attenuation coefficients. Figure 1a illustrates the frequency dependence of the attenuation coefficient calculated by the Francois–Garrison formula. Since, on a logarithmic scale, the Urick curve differs little from the Francois–Garrison curve, the graphs of the Francois–Garrison and Urick attenuation coefficients and their differences were plotted on a linear scale (Fig. 1b).

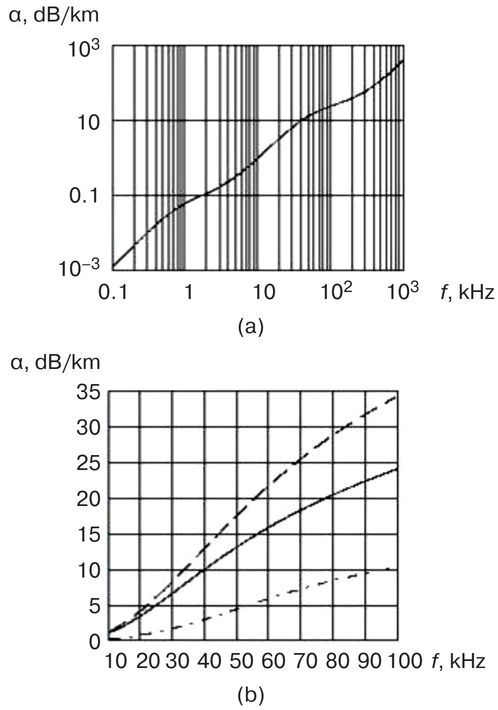


Fig. 1. Attenuation coefficient α versus frequency f in the frequency ranges (a) 0.1–10^3 kHz (Francois–Garrison formula) and (b) 10–100 kHz (the solid and dashed curves were calculated by the Francois–Garrison and Urick formulas, respectively; and the dashed-and-dotted curve represents the difference $\alpha_\text{U}(f) - \alpha(f)$)

[1] *Atlas okeanov* (*Atlas of the Oceans*): in 3 v. V. 1. *Tikhii okean* (*Pacific Ocean*). 323 p. V. 2. *Atlanticheskii i Indiiskii okeany* (*Atlantic and Indian Oceans*). 334 p. V. 3. *Severnyi Ledovityi okean* (*Arctic Ocean*). Leningrad: MO USSR; 1974–1980. 188 p. (in Russ.).

Figure 1b shows that the discrepancies between the attenuation coefficients at high frequencies can be significant. The relative error $\delta_\alpha = (\alpha - \alpha_U)/\alpha$ of the approximation of the Francois–Garrison formula by the Urick formula in the frequency range from 0.1 kHz to 1 MHz at a temperature of $T = 4°C$ and a seawater salinity of $S = 35‰$ is $\delta_\alpha = -0.26$ at $z = 100$ m and $\delta_\alpha = -0.42$ at $z = 1000$ m. At $S = 30‰$, $|\delta_\alpha|$ increases by half. Such error values can hardly be considered acceptable under these conditions. Therefore, the HACC capacity should be calculated using the Francois–Garrison formula.

Further, let us the express frequency f in hertz. For this purpose, the factor 10^{-3} should be added to the formulas for the coefficients A_1 and A_2; the factor 10^{-6}, to the formula for A_3; and the factor 10^3, to the formulas for the frequencies f_1 and f_2.

2. NOISES OF THE MARINE ENVIRONMENT

The interference considered in this work is the noise that is characteristic of the deep sea. Its main sources in this case are ocean turbulence, long-distance shipping, sea surface waves, and thermal noise generated by the thermal motion of water molecules [5]. At short time intervals and moderate depths, the ambient sea noise can be described by a stationary Gaussian process and characterized by spectral intensity density. Due to the variability of the sources, the ambient noise is characterized by the average spectral intensity density. The graphs of the average spectral intensities of turbulence noise, shipping noise, noise from sea surface waves, and thermal noise in the frequency range 1–10^5 Hz were provided by Urick [5, Fig. 7.5]. Using a piecewise linear approximation of these curves, the following analytical expressions for the spectral noise levels were obtained.

The spectral level of turbulence noise, $N_{tu}(f)$, dB relative to 1 μPa:

$$N_{tu}(f) = (107 - 30\log f)\,[1(f-1) - 1(f-50)].$$

The spectral level of shipping noise, $N_{sh}(f)$, dB relative to 1 μPa:

$$N_{sh}(f,k) = (44 + 9k + 20\log f)\,[1(f-1) - 1(f-10)] + (64 + 9k)[1(f-10) - 1(f-100)] + (144 + 9k - 40\log f)\,1(f-100),$$

where f is frequency, Hz; $1(f)$ is the unit step function; and $k = 0$, 1, and 2 are in the cases of low-, medium-, and high-intensity shipping, respectively.

The spectral level of sea wave noise, $N_w(f)$, dB relative to 1 μPa:

$$N_w(f,b) = (N_1 + 20\log f - 46)[1(f-1) - 1(f-200)] + N_1[1(f-200) - 1(f-1000)] + (N_1 + 60 - 20\log f)\,1(f-1000), \quad (4)$$

$$N_1 = N_1(b) = N_w(f,b)|_{f=1000\text{ Hz}} = 44.9 + 13.4\ln(1+b).$$

In the last formulas, b is a numerical quantity, which characterizes the sea state, measured as wave forces, and takes values in the range [0, 6]. Formula (4) was obtained by the least squares approximation of the sequence of the values $N_w(1000, b)$ (b = 0, 0.5, 1, 2, 3, 4, 6) of the family of curves $N_w(f, b)$ presented on a large scale in [16]. The relative approximation error does not exceed 4%.

The spectral level of thermal noise, $N_{th}(f)$, dB relative to 1 μPa:

$$N_{th}(f) = (-75 + 20\log f)\,1(f - 10000). \quad (5)$$

After the determination of the spectral levels of the noises, their spectral intensity densities can be found.

For turbulence noise,

$$J_{tu}(f) = J_0 10^{0.1 N_{tu}(f)} = J_0 10^{10.7} f^{-3}[1(f-1) - 1(f-50)],\ \text{W/(cm}^2\cdot\text{Hz)},$$

where $J_0 = 0.667 \cdot 10^{-22}$ W/(cm^2·Hz) is the reference value of the spectral intensity, which corresponds to a root-mean-square pressure of 1 μPa.

For shipping noise,

$$J_{sh}(f,k) = J_0 10^{0.1 N_{sh}(f)} = J_0 10^{4.4+0.9k} f^2 [1(f-1) - 1(f-10)] + J_0 10^{6.4+0.9k}[1(f-10) - 1(f-100)] + J_0 10^{14.4+0.9k} f^{-4} 1(f-100),\ \text{W/(cm}^2\cdot\text{Hz)}.$$

For noise from sea surface waves,

$$J_w(f,b) = J_0 10^{0.1 N_w(f,b)} = J_0 10^{0.1N_1 - 4.6} f^2 [1(f-1) - 1(f-200)] + J_0 10^{0.1N_1}[1(f-200) - 1(f-1000)] + J_0 10^{0.1N_1 + 6} f^{-2} 1(f-1000),\ \text{W/(cm}^2\cdot\text{Hz)}.$$

For thermal noise,

$$J_{\text{th}}(f) = J_0 10^{0.1 N_{\text{th}}(f)} = \\ = J_0 10^{-7.5} f^2 \, 1(f - 10000), \ \text{W/(cm}^2\cdot\text{Hz)}. \tag{6}$$

Expressions (5) and (6) are valid at a water temperature of T = 4°C. At other water temperatures, the spectral density of thermal noise is calculated from the formula

$$J_{\text{th}}(f) = 4\pi k_B \Theta c^{-2} f^2 1(f - 10000), \ \text{W/(cm}^2\cdot\text{Hz)},$$

where $k_B = 1.38 \cdot 10^{-23}$ W/(Hz·K) is the Boltzmann constant and $c = 1.5 \cdot 10^{-5}$ cm/s is the average speed of sound in seawater.

Figure 2 presents the graphs of the spectral intensity density of the total noise of the marine environment,

$$J_{\text{n}}(f) = J_{\text{tu}}(f) + J_{\text{sh}}(f) + J_{\text{w}}(f) + J_{\text{th}}(f),$$

for the case of medium-intensity shipping in various sea states. The solid, dotted, dashed, and dashed-and-dotted lines are plotted at b = 0 (calm sea), 1, 3, and 6, respectively.

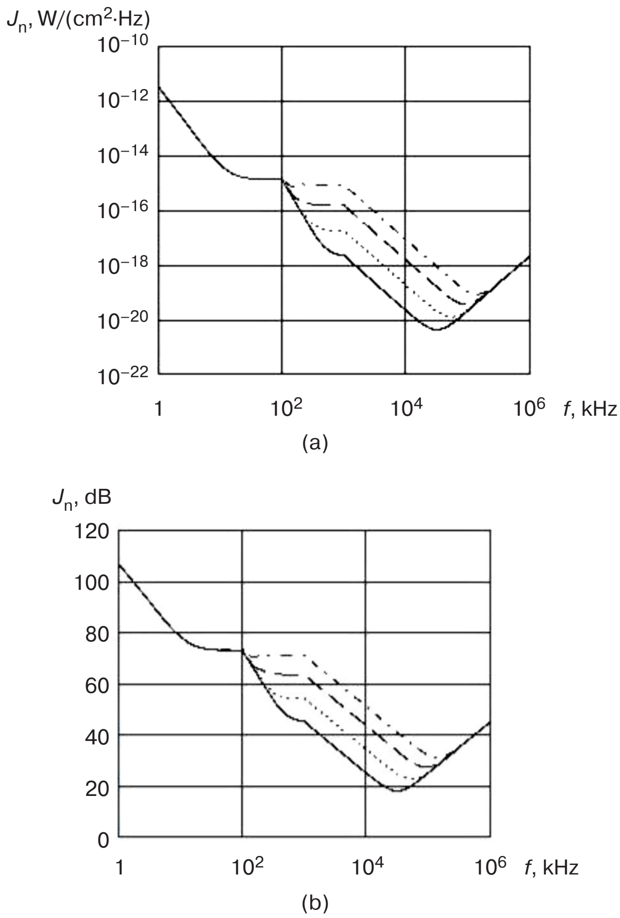


Fig. 2. Spectral intensity density of the total noise of the marine environment in (a) physical units and (b) decibels relative to $J_0 = 0.667 \cdot 10^{-22}$ W/(cm²·Hz)

3. HACC CAPACITY

The HACC under investigation can be represented as a low-pass filter with the complex gain $H(j\omega)$, to the input of which signal (pressure) $p_1(t)$ is applied, and at the output of which signal $p_2(t)$ is received and summed with nonwhite Gaussian sea noise $p_n(t)$. Thus, the process at the HACC has the form $p(t) = p_2(t) + p_n(t)$. Let us consider the signal $p_1(t)$ the realization of a certain constrained-intensity stationary ergodic random process. Let the intensity of the input signal at a distance of 1 m from the source be I_S. Let us determine the HACC capacity under constraints on the bandwidth (by the filter) and on the signal intensity. The capacity of a HACC-like channel under constraints on frequency and average power was found using information theory [17], and it was shown that the capacity is reached if the input signal is a stationary Gaussian process. In terms of this work, the formula for the HACC capacity takes the form

$$C = \int_{f_{\text{lower}}(B)}^{f_{\text{upper}}(B)} \log \left[\frac{H^2(f) B}{J_{\text{n}}(f)} \right] df. \tag{7}$$

Here, $f_{\text{lower}}(B)$ and $f_{\text{upper}}(B)$ is the lower and upper frequencies, respectively, of the bandwidth, for which

$$\frac{J_{\text{n}}(f)}{H^2(f)} \le B, \tag{8}$$

B is the solution of the equation

$$\int_{f_{\text{lower}}(B)}^{f_{\text{upper}}(B)} \left[B - \frac{J_{\text{n}}(f)}{H^2(f)} \right] df = I_S, \tag{9}$$

and $H(f) = |H(j2\pi f)|$ is the the frequency response of the HACC.

The spectral intensity density $J_S(f)$ of the random input signal at which the channel capacity is reached is found by the formula

$$J_S(f) = \begin{cases} B - J_{\text{n}}(f)/H^2(f), & f \in [f_{\text{lower}}(B), f_{\text{upper}}(B)], \\ 0, & f \notin [f_{\text{lower}}(B), f_{\text{upper}}(B)]. \end{cases} \tag{10}$$

Thus, the HACC capacity is determined by two parametric equalities (7) and (9). The parameter is the quantity B, which should satisfy inequality (8). By simultaneously solving equations (7) and (9) under constraint (8), the C and B values can be found, and then from formula (10), the spectrum of the optimal input signal can be determined. This problem can be solved only by numerical methods on a computer.

Figure 3a illustrates the calculated dependence of the capacity $C(R)$ on the communication distance at three values of the input signal intensity for the case of calm sea ($b = 0$) and medium-intensity shipping ($k = 1$).

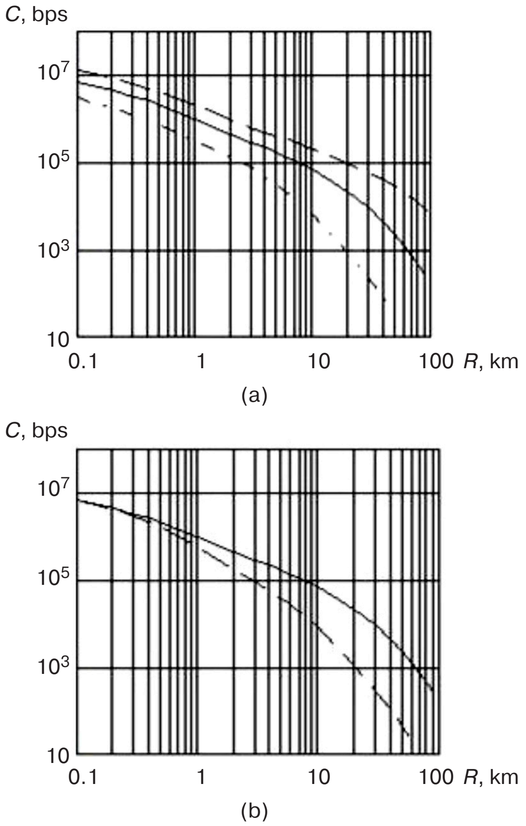


Fig. 3. HACC capacity versus communication distance at $k = 1$: (a) $b = 0$ and $I_S = 10^{-7}$ W/cm^2 (dashed-and-dotted line), $I_S = 10^{-5}$ W/cm^2 (solid line), and $I_S = 10^{-3}$ W/cm^2 (dashed line) and (b) $I_S = 10^{-5}$ W/cm^2 and $b = 3$ (dashed line) and $b = 0$ (solid line)

Figure 3a shows that the capacity decreases with increasing communication distance, and this decrease is the faster, the lower the intensity of the input signal. Figure 3a can be used to estimate the order of magnitude of the capacity under the given communication conditions.

Figure 3b presents the dependences $C(R)$ at $I_S = 10^{-5}$ W/cm^2 and various b. It can be seen that the noise from sea surface waves can significantly decrease the capacity. For example, in calm sea ($b = 0$) at $R = 10$ km, $C = 7 \cdot 10^4$ bps, and if $b = 3$, then $C = 8 \cdot 10^3$ bps.

Calculations showed that, at a constant intensity of the input signal, with increasing distance R, the width of the spectrum of the optimal signal decreases due to an increase in the lower boundary frequency f_{lower} of the spectrum and a decrease in the upper frequency f_{upper}. Figure 4 illustrates the distance dependences of these frequencies at $k = 1$ and $b = 0$.

As an example, Fig. 5 presents two spectra of the optimal signal at constant intensity of the input signal, $k = 1$, $b = 0$, $I_S = 10^{-5}$ W/cm^2, and two values of the distance R.

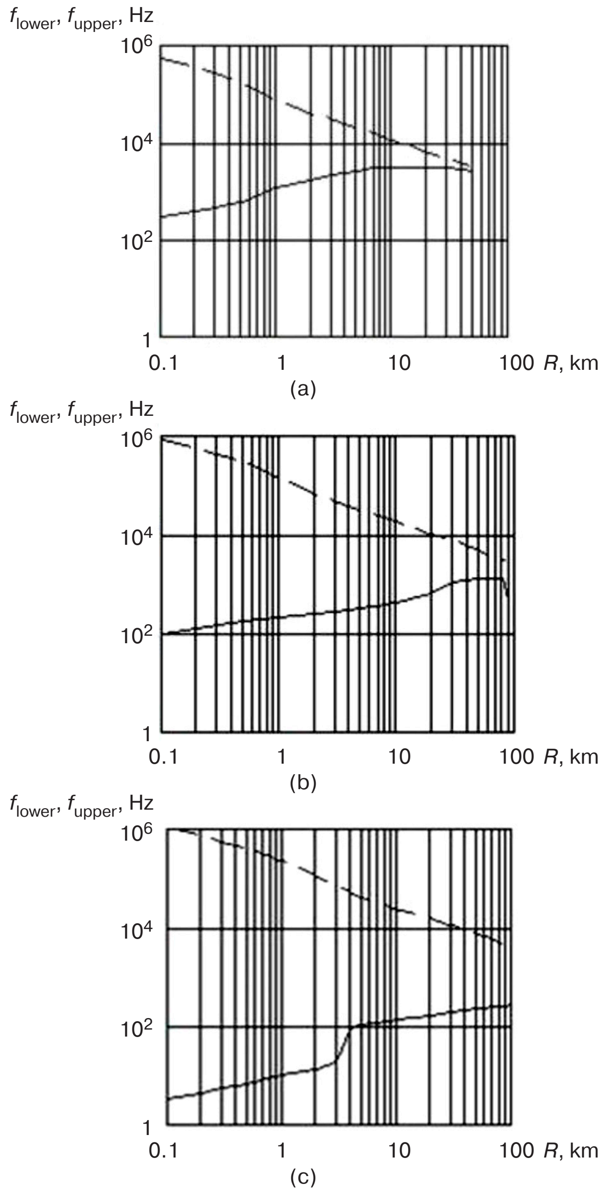


Fig. 4. Dependences of the lower boundary frequency f_{lower} (solid line) and the upper boundary frequency f_{upper} (dashed line) of the optimal spectrum of the input signal on the communication distance at I_S of (a) 10^{-7}, (b) 10^{-5}, and (c) 10^{-3} W/cm^2

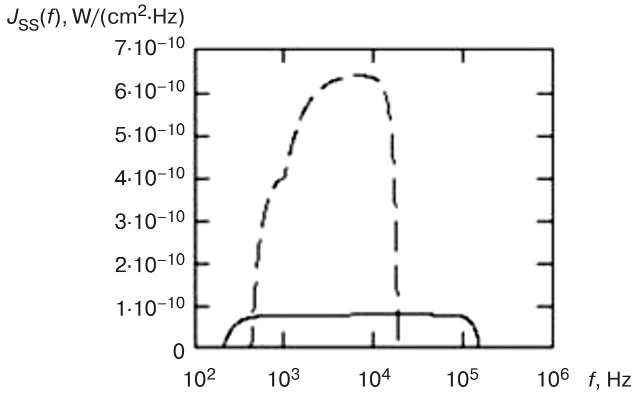


Fig. 5. Optimal spectra of the input signal at various communication distances of $R = 1$ km (solid line) and $R = 10$ km (dashed line)

As can be seen from Fig. 5, with increasing communication distance at constant intensity of the input signal, the optimal spectrum narrows, but its values increase. The spectrum of the optimal input signal is always limited in frequency.

Figure 6 shows the optimal spectra at $k = 1$, $I_S = 10^{-5}$ W/cm^2, and $R = 10$ km in the absence of sea waves ($b = 0$) and in the case of sea waves at $b = 3$. As follows from Fig. 6, the noise from sea waves leads to a decrease in the width of the optimal spectrum, a change in its shape, and an increase in its intensity.

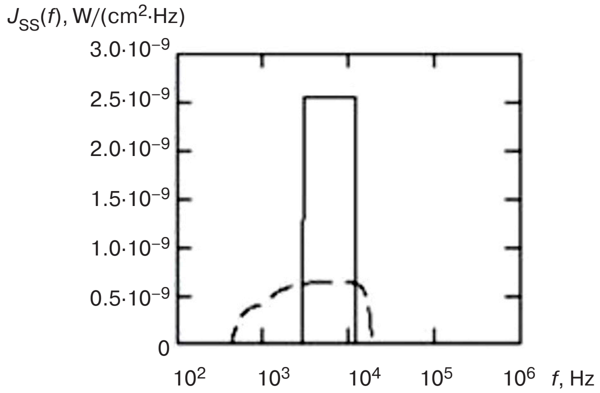


Fig. 6. Optimal spectra of the input signal at various sea wave intensities of $b = 3$ (solid line) an $b = 0$ (dashed line)

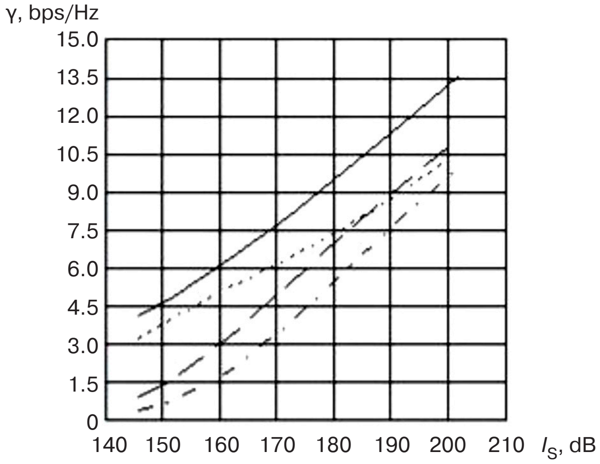


Fig. 7. Bandwidth usage versus input signal intensity at various communication distances of $R = 0.2$ km (solid line), $R = 1$ km (dotted line), $R = 5$ km (dashed line), and $R = 10$ km (dashed-and-dotted line)

As follows from the analysis, the HACC is a constrained-bandwidth channel. Therefore, from a practical point of view, it was of great interest to evaluate the bandwidth usage of the HACC. Figure 7 shows the results of calculations of the dependence of the bandwidth usage $\gamma = C/(f_{upper} - f_{lower})$ on the intensity of the input signal (in decibels relative to 1 μPa) at four communication distances. It can be seen from Fig. 7 that, regardless of the distance, γ increases with increasing input signal intensity. An interesting feature of these dependences is that the curves at $R = 1$ and 5 km intersect at a certain intensity value (in the vicinity of 190 dB). Therefore, there may be cases where the bandwidth usage at a longer distance is greater than that at a shorter distance at the same input signal intensity. From the data in Fig. 7, one can estimate the order of magnitude of the coefficient γ under the specified communication conditions. For example, the maximum value of γ is approximately 13 bps/Hz and occurs at $I_S = 200$ dB and $R = 0.2$ km.

CONCLUSIONS

The main purpose of this work was to determine the capacity of a HACC with constrained average intensity of the transmitted signal. A model of a single-path channel was considered, which is characteristic of the deep sea with the receiver or transmitter placed at a sufficient depth. In this case, the sound propagation to medium distances occurs along "stable" trajectories, which are unaffected by the sea surface and the bottom. Such trajectories exist for vertical and near-vertical (10°–15°) channels.

The comparison of the numerical values of the Francois–Garrison and Urick attenuation coefficients showed that Urick attenuation coefficient can differ markedly from the Francois–Garrison attenuation coefficient. Moreover, the Francois–Garrison attenuation coefficient takes into account the dependence of the sound attenuation on the seawater parameters. Therefore, in this work, the Francois–Garrison attenuation coefficient was used. Using it, the complex gain, the frequency response, and the phase response of the HACC were obtained.

A new expression was derived for the spectral level of sea noise caused by sea surface waves. The expression for the spectral density of the sea noise intensity used to calculate the channel capacity was obtained by subjecting Urick's curves of the spectral levels of noises from turbulence, shipping, sea waves, and thermal noise to a piecewise linear approximation. The Francois–Garrison attenuation coefficient was used to relate the HACC capacity with the parameters of the marine environment: temperature, salinity, and pH of the study area.

The capacity was calculated for the cases of medium-intensity shipping in calm sea and in Beaufort force 3 waves. It was found that HACC capacity is significantly reduced by noise from sea surface waves. The dependencies of channel capacity on communication distance and the intensity of the transmitted signal were calculated. At a given intensity of the input signal, channel capacity was observed to significantly decrease with increasing distance.

The lower and upper boundary frequencies of the optimal spectrum of the input signal were determined along with the optimal spectrum itself. It was shown that, with increasing communication distance, the width of the optimal spectrum also decreases. A significant effect of noise from sea waves on the shape of the optimal spectrum and the values of its boundary frequencies can be noted.

The dependence of bandwidth usage on the intensity of the input signal at various communication distances was characterized. Under some conditions, there may be cases where bandwidth usage increases with increasing distance at a given input signal intensity.

REFERENCES

1. Stojanovic M. On the relationship between capacity and distance in an underwater acoustic communication channel. *ACM SIGMOBILE Mobile Computing and Communications Review.* 2007;11(4):34–43. https://doi.org/10.1145/1347364.1347373
2. Lucani D.E., Stojanovic M., Medard M. On the relationship between transmission power and capacity of an underwater acoustic communication channel. In: *Proceedings of OCEANS Conference*. 2008. 6 p. https://doi.org/10.1109/OCEANSKOBE.2008.4531073
3. Stojanovic M., Preisig J. Underwater acoustic communication channels: propagation models and statistical characterization. *IEEE Communications Magazine*. 2009;47(1):84–89.
4. Matarneh A.M. Thorough investigation of BER simulation of DPSK in underwater acoustic channel. *JJEE*. 2016;2(2):160–171.
5. Urik R.J. *Osnovy gidroakustiki* (*Principles of Underwater Sound*). Leningrad: Sudostroenie; 1978. 445 p. (in Russ.). [Urik R.J. *Principles of Underwater Sound*. NY: McGraw-Hill, Book Company; 1975. 445 p.]
6. Coates R. *Underwater Acoustic Systems*. New York: Wiley; 1990. 188 p.
7. Sivers M.A., Basov V.V. The use of MIMO-OFDM technology in systems of the hydrolocation, hydroacoustic communication, navigation and management. In: *Wave Electronics and Infocommunication Systems*: *Proceedings of the 23rd International Scientific Conference*. In 2 v. V. 2. St. Petersburg; 2020. P. 177–183 (in Russ.).
8. Eroshenko S.V., Krants V.Z., Sechin V.V. On the use of parallel channels to increase information rate in hydroacoustic communication system with compound signals. *Gidroakustika = Hydroacoustics.* 2020;42(2): 21–28 (in Russ.).
9. Rodionov A.Yu., Unru P.P., Kulik S.Yu., Golov A.A. Application of multi-frequency signals with constant envelope in underwater communication systems. *Podvodnye issledovaniya i robototekhnika = Underwater Investigations and Robotics*. 2019;3(29):30–38 (in Russ.). https://doi.org/10.25808/24094609.2019.29.3.004
10. Klei K., Medvin G. *Akusticheskaya okeanografiya* (*Acoustical Oceanography*): trans. from Eng. Moscow: Mir; 1980. 580 p. (in Russ.). [Clay C.S., Medwin H. *Acoustical Oceanography*. New York: Wiley; 1977. 544 p.]
11. Brekhovskikh L.M., Lysanov Yu.P. *Teoreticheskie osnovy akustiki okeana* (*Theoretical Foundations of Ocean Acoustics*). Leningrad: Gidrometeoizdat; 1982. 264 p. (in Russ.).
12. Andreeva I.B. *Fizicheskie osnovy rasprostraneniya zvuka v okeane* (*Physical Foundations of Sound Propagation in the Ocean*). Leningrad: Gidrometeoizdat; 1975. 190 p. (in Russ.).

СПИСОК ЛИТЕРАТУРЫ

1. Stojanovic M. On the relationship between capacity and distance in an underwater acoustic communication channel. *ACM SIGMOBILE Mobile Computing and Communications Review.* 2007;11(4):34–43. https://doi.org/10.1145/1347364.1347373
2. Lucani D.E., Stojanovic M., Medard M. On the relationship between transmission power and capacity of an underwater acoustic communication channel. In: *Proceedings of OCEANS Conference*. 2008. 6 p. https://doi.org/10.1109/OCEANSKOBE.2008.4531073
3. Stojanovic M., Preisig J. Underwater acoustic communication channels: propagation models and statistical characterization. *IEEE Communications Magazine*. 2009;47(1):84–89.
4. Matarneh A.M. Thorough investigation of BER simulation of DPSK in underwater acoustic channel. *JJEE*. 2016;2(2):160–171.
5. Урик Р.Дж. *Основы гидроакустики*: пер. с англ. Л.: Судостроение; 1978. 445 с.
6. Coates R. *Underwater Acoustic Systems*. New York: Wiley; 1990. 188 p.
7. Сиверс М.А., Басов В.В. Использование технологии MIMO-OFDM в системах гидролокации, гидроакустической связи, навигации и управления. В сб.: *Волновая электроника и инфокоммуникационные системы: Материалы XXIII международной научной конференции*: в 2 ч. Часть 2. Санкт-Петербург; 2020. С. 177–183.
8. Ерошенко С.В., Кранц В.З., Сечин В.В. Об использовании параллельных каналов для повышения скорости передачи информации в системе гидроакустической связи со сложными сигналами. *Гидроакустика.* 2020;42(2):23–28.
9. Родионов А.Ю., Унру П.П., Кулик С.Ю., Голов А.А. Оценки применения многочастотных сигналов с постоянной огибающей в гидроакустических системах связи. *Подводные исследования и робототехника*. 2019;3(29):30–38. https://doi.org/10.25808/24094609.2019.29.3.004
10. Клей К., Медвин Г. *Акустическая океанография*: пер. с англ. М.: Мир; 1980. 580 с.
11. Бреховских Л.М., Лысанов Ю.П. *Теоретические основы акустики океана*. Л.: Гидрометеоиздат; 1982. 264 с.
12. Андреева И.Б. *Физические основы распространения звука в океане*. Л.: Гидрометеоиздат; 1975. 190 с.
13. Francois R.E., Garrison G.R. Sound absorption based on ocean measurements: Part I: Pure water and magnesium sulfate contributions. *J. Acoust. Soc. Am*. 1982;72(3): 896–907. https://doi.org/10.1121/1.388170
14. Francois R.E., Garrison G.R. Sound absorption based on ocean measurements: Part II: Boric acid contribution and equation for total absorption. *J. Acoust. Soc. Am.* 1982;72(6):1879–1890. https://doi.org/10.1121/1.388673

13. Francois R.E., Garrison G.R. Sound absorption based on ocean measurements: Part I: Pure water and magnesium sulfate contributions. *J. Acoust. Soc. Am*. 1982;72(3): 896–907. https://doi.org/10.1121/1.388170
14. Francois R.E., Garrison G.R. Sound absorption based on ocean measurements: Part II: Boric acid contribution and equation for total absorption. *J. Acoust. Soc. Am.* 1982;72(6):1879–1890. https://doi.org/10.1121/1.388673
15. Kibblevhite A.C., Hampton L.D. A review of deep ocean sound attenuation data at very low frequencies. *J. Acoust. Soc. Am*. 1980;67(1):147–157. https://doi.org/10.1121/1.384473
16. Guies L., Sabate P. *Osnovy akustiki morya* (*Fundamentals of acoustics of the sea*): trans. from Franc. Leningrad: Gidrometeoizdat; 1967. 212 p. (in Russ.).
17. Galager R. *Teoriya informatsii i nadezhnaya svyaz'* (*Information Theory and Reliable Communication*): trans. from Eng. Moscow: Sovetskoe Radio; 1974. 720 p. (in Russ.).
[Gallager R.G. *Information Theory and Reliable Communication*. NY: J. Wiley; 1968. 603 p.]

15. Kibblevhite A.C., Hampton L.D. A review of deep ocean sound attenuation data at very low frequencies. *J. Acoust. Soc. Am*. 1980;67(1):147–157. https://doi.org/10.1121/1.384473
16. Гийес Л., Сабате П. *Основы акустики моря*: пер. с фр. Л.: Гидрометеоиздат; 1967. 212 с.
17. Галагер Р. *Теория информации и надежная связь*: пер. с англ. М.: Сов. Радио; 1974. 720 с.

About the author

Valery E. Denisov, Senior Lecturer, Department of Radio Wave Processes and Technologies, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78 Vernadskogo pr., Moscow, 119454 Russia). E-mail: denisov@mirea.ru. RCSI SPIN-code 1190-9380, https://orcid.org/0000-0002-5824-8741

Об авторе

Денисов Валерий Евгеньевич, старший преподаватель кафедры радиоволновых процессов и технологий Института радиоэлектроники и информатики ФБГОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр–т Вернадского, д. 78). E-mail: denisov@mirea.ru. SPIN-код РИНЦ 1190-9380, https://orcid.org/0000-0002-5824-8741

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

UDC 621.391.072
https://doi.org/10.32362/2500-316X-2023-11-1-41-50

RESEARCH ARTICLE

Optimal reception of multiple phase shift keying and quadrature amplitude modulation signals with non-coherent processing of harmonic interference

Gennady V. Kulikov [1, @],
Trung T. Do [1],
Andrey A. Lelyukh [2],
Van D. Nguyen [3]

[1] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[2] *Moscow Research Institute of Radio Communications, Moscow, 109029 Russia*
[3] *Le Quy Don Technical University, Hanoi, Vietnam*
[@] *Corresponding author, e-mail: kulikov@mirea.ru*

Abstract
Objectives. Analysis of the reception noise immunity of multiple phase shift keying (M-PSK) and quadrature amplitude modulation (M-QAM) signals has demonstrated a significant reduction in the quality of reception of discrete information due to the presence of various types of non-fluctuating interference in a radio communication channel including targeted harmonic interference. Therefore, the development of algorithms for compensating the influence of such forms of interference is an urgent task. While various methods for combatting this kind of interference, these vary in terms of their effectiveness. The aim of the present work is to synthesize and analyze the optimal algorithm for the reception of M-PSK and M-QAM signals with incoherent processing of harmonic interference.
Methods. Various statistical radio engineering and computer simulation methods were used in accordance with optimal signal reception theory.
Results. Synthesis and analysis of the optimal algorithm for receiving M-PSK and M-QAM signals with incoherent processing of harmonic interference were carried out. In addition to calculating the correlation integrals in the receiver, it is necessary to form weight coefficients, whose value depends on the correlation of the interference oscillation (extracted from the received mixture) with a sample of the interference stored in the receiver. The dependences of the bit error probability on the signal-to-noise ratio, interference detuning, and inaccuracy in setting the frequency and level of the interference sample in the receiver were obtained. It is shown that the higher the gain in the noise immunity of reception, the greater the intensity of the harmonic interference.
Conclusions. The synthesized receiver circuit effectively compensates for harmonic interference. However, the efficiency of its operation depends on the detuning of the harmonic interference relative to the center frequency of the spectrum of the useful signal. The scheme for incoherent processing of harmonic interference remains operational even with small (within ±10%) inaccuracies in setting the frequency and the level of the interference copy in the receiver.

Keywords: multiple phase shift keying, quadrature amplitude modulation, harmonic interference, optimal reception, noise immunity

• **Submitted:** 02.03.2022 • **Revised:** 05.05.2022 • **Accepted:** 24.10.2022

For citation: Kulikov G.V., Do T.T., Lelyukh A.A., Nguyen V.D. Optimal reception of multiple phase shift keying and quadrature amplitude modulation signals with non-coherent processing of harmonic interference. *Russ. Technol. J.* 2023;11(1):41–50. https://doi.org/10.32362/2500-316X-2023-11-1-41-50

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

Оптимальный прием многопозиционных сигналов М-ФМ и М-КАМ с некогерентной обработкой гармонической помехи

Г.В. Куликов [1, @],
Ч.Т. До [1],
А.А. Лелюх [2],
В.З. Нгуен [3]

[1] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[2] *Московский научно-исследовательский институт радиосвязи, Москва, 109029 Россия*
[3] *Вьетнамский государственный технический университет им. Ле Куй Дона, Ханой, Вьетнам*
[@] *Автор для переписки, e-mail: kulikov@mirea.ru*

Резюме

Цели. Проведенный в работах многих исследователей анализ помехоустойчивости приема многопозиционных сигналов с фазовой (М-ФМ) и квадратурной амплитудной модуляцией (М-КАМ) показал, что качество приема дискретной информации существенно снижается при наличии кроме шумовой помехи еще и нефлуктуационных помех разных видов в канале радиосвязи. Одной из наиболее опасных является прицельная гармоническая помеха, поэтому разработка алгоритмов компенсации влияния такой помехи является актуальной задачей. Существуют различные методы борьбы с такого рода мешающими воздействиями, обладающие большей или меньшей эффективностью. Целью настоящей работы является синтез и анализ оптимального алгоритма приема многопозиционных сигналов М-ФМ и М-КАМ с некогерентной обработкой гармонической помехи.

Методы. Использованы методы статистической радиотехники, теории оптимального приема сигналов и компьютерное моделирование.

Результаты. Выполнен синтез и анализ оптимального алгоритма приема многопозиционных сигналов М-ФМ и М-КАМ с некогерентной обработкой гармонической помехи. Показано, что кроме вычисления корреляционных интегралов в приемнике необходимо формировать весовые коэффициенты, величина которых зависит от степени корреляции помехового колебания, выделенного из принимаемой смеси, с копией помехи, хранящейся в приемнике. Получены зависимости вероятности битовой ошибки от отношения сигнал/шум, расстройки помехи и неточности установки частоты и уровня копии помехи в приемнике. Показано, что выигрыш в помехоустойчивости приема тем выше, чем больше интенсивность гармонической помехи.

Выводы. Синтезированная схема приемника позволяет достаточно эффективно бороться с гармонической помехой. Эффективность ее работы зависит от расстройки гармонической помехи относительно центральной частоты спектра полезного сигнала. Схема некогерентной обработки гармонической помехи сохраняет работоспособность и при небольших (в пределах ±10%) неточностях установки частоты и уровня копии помехи в приемнике.

Ключевые слова: многопозиционная фазовая модуляция, квадратурная амплитудная модуляция, гармоническая помеха, оптимальный прием, помехоустойчивость

• **Поступила:** 02.03.2022 • **Доработана:** 05.05.2022 • **Принята к опубликованию:** 24.10.2022

Для цитирования: Куликов Г.В., До Ч.Т., Лелюх А.А., Нгуен В.З. Оптимальный прием многопозиционных сигналов М-ФМ и М-КАМ с некогерентной обработкой гармонической помехи. *Russ. Technol. J.* 2023;11(1):41–50. https://doi.org/10.32362/2500-316X-2023-11-1-41-50

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

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Analysis of the noise immunity of receiving multiple phase shift keying (M-PSK) and quadrature amplitude modulation (M-QAM) signals demonstrates a significant deterioration in the quality of receiving discrete information (up to complete destruction of communication) due to the presence in addition to noise of various kinds non-fluctuating interference in the communication channel [1–10]. This is especially true in cases where such interference is signal-like, for example, harmonic [5–7].

There are various methods of dealing with this kind of interference, which are more or less effective, for example, those developed in [1, 11–14]. Their technical implementation can be very complex, as in the case of the synthesis of optimal algorithms [1], or simpler, as in [12, 14], but less efficient. The purpose of this work is to synthesize and analyze the optimal algorithm for receiving M-PSK and M-QAM signals with incoherent processing of harmonic interference.

1. ALGORITHM FOR RECEIVING OF M-PSK AND M-QAM SIGNALS WITH INCOHERENT PROCESSING OF HARMONIC INTERFERENCE

Consider the optimal reception of M-PSK and M-QAM signals against the background of white Gaussian noise $n(t)$ with one-sided power spectral density N_0 and harmonic interference

$$s_{int}(t, \varphi_{int}) = \mu A_0 \cos(\omega_{int} t + \varphi_{int}), \tag{1}$$

with relative intensity μ, frequency $\omega_{int} = 2\pi f_{int}$, close to the useful signal frequency, and random initial phase φ_{int}. At the same time, we assume that the initial phase of the useful signal is known, and the distribution of the random variable φ_{int} is uniform in the range $(-\pi, \pi]$. In this case, we can speak of incoherent noise processing.

The signal entering the input of the receiver,

$$x(t) = s_i(t) + s_{int}(t, \varphi_{int}) + n(t),$$

is a mixture of interference, noise, and M-PSK signal

$$s_i(t) = A_0 \cos(\omega_0 t + \varphi_i),$$
$$\varphi_i = \frac{i2\pi}{M}, t \in (0, T_s], i = 0, 1, ..., M-1, \tag{2}$$

or M-QAM signal

$$s_i(t) = A_{av}(I_i \cos \omega_0 t - Q_i \sin \omega_0 t),$$
$$t \in (0, T_s], i = 0, 1, \ldots, M-1, \tag{2'}$$

where $A_0 = A_{av} = \sqrt{2E_s / T_s}$ is the signal amplitude; $E_s = kE_b$ is the energy of the channel symbol; E_b is the energy per one bit of information; $k = \log_2 M$; I_i and Q_i are the coefficients that determine the amplitudes of the quadrature signal components; ω_0 is the carrier frequency. In the case of M-QAM signal, the energies and amplitude should be averaged over the ensemble of signals.

For the sake of simplicity, let us denote the sum of the signal and interference as

$$s_{sig,int}(t, C_i, \varphi_{int}) = s_i(t) + s_{int}(t, \varphi_{int}). \tag{3}$$

Let us specify the a posteriori probability of this process, and, consequently, the joint a posteriori probability of the channel symbol C_i and phase φ_{int} as follows [15, 16]:

$$p_{ps}[s_{sig,int}(t, C_i, \varphi_{int})] = p_{ps}(C_i, \varphi_{int}) = K p_{pr}(C_i) p_{pr}(\varphi_{int}) \times$$
$$\times \exp[-\frac{1}{N_0}\int_0^{T_s} s_{sig,int}{}^2(t, C_i, \varphi_{int})dt + \frac{2}{N_0}\int_0^{T_s} x(t) s_{sig,int}(t, C_i, \varphi_{int})dt].$$

Here, the symbol K denotes the normalization coefficient, which takes into account all components that do not contain information about the useful signal and non-fluctuating noise, while $p_{pr}(C_i)$ is the a priori probability of the channel symbol C_i, and $p_{pr}(\varphi_{int}) = 1/2\pi$.

$$\int_0^{T_s} s_{sig,int}{}^2(t, C_i, \varphi_{int})dt =$$
$$= \int_0^{T_s} s_i{}^2(t, C_i)dt + \int_0^{T_s} s_{int}{}^2(t, \varphi_{int})dt +$$
$$+2\int_0^{T_s} s_i(t, C_i)s_{int}(t, \varphi_{int})dt =$$
$$= E_s + E_{int} + 2R_i(\varphi_{int}),$$

where E_{int} is the energy of interference packets of duration T_s; $R_i(\varphi_{int})$ is the correlation coefficient between the signal and interference.

Hence,

$$p_{ps}(C_i, \varphi_{int}) = Kp_{pr}(C_i)\frac{1}{2\pi}\times$$
$$\times\exp[-\frac{1}{N_0}(E_s + E_{int} + 2R_i(\varphi_{int}))]\times$$
$$\times\exp[\frac{2}{N_0}\int_0^{T_s} x(t)s_{sig,int}(t, C_i, \varphi_{int})dt] =$$
$$= K_1\frac{1}{2\pi}\exp[-\frac{2}{N_0}R_i(\varphi_{int})]\times$$
$$\times\exp[\frac{2}{N_0}\int_0^{T_s} x(t)s_{sig,int}(t, C_i, \varphi_{int})dt].$$

The coefficient K_1 includes terms that do not depend on the value of the symbol C_i or the phase φ_{int}.

To obtain the a posteriori probability of a discrete symbol C_i, it is necessary to average the value $p_{ps}(C_i, \varphi_{int})$ over all possible values of the phase φ_{int}. Then we obtain:

$$p_{ps}(C_i) = K_1\frac{1}{2\pi}\times$$
$$\times\int_{-\pi}^{\pi}(\exp[-\frac{2}{N_0}R_i(\varphi_{int})]\exp[\frac{2}{N_0}\int_0^{T_s} x(t)s_{sig,int}(t, C_i, \varphi_{int})dt])d\varphi_{int}.$$

Let us introduce the notation for the integrals that describe the degree of correlation between the received process $x(t)$ and signal copies for different values of the channel symbols C_i:

$$J_0 = \frac{2}{N_0}\int_0^{T_s} x(t)s_i(t, C_i = 0)dt,$$
$$\ldots \quad (4)$$
$$J_{M-1} = \frac{2}{N_0}\int_0^{T_s} x(t)s_i(t, C_i = M-1)dt.$$

Assuming that there is a copy of the interference $s_{int}(t, \varphi_{int})$ in the receiver, then the following integrals, in fact, determine the degree of correlation between it and the received interference sample, formed by subtracting the signal copies from the received process $x(t)$:

$$a_0 = \frac{2}{N_0}\int_0^{T_s}[x(t) - s_i(t, C_i = 0)]s_{int}(t, \varphi_{int})dt,$$
$$\ldots \quad (5)$$
$$a_{M-1} = \frac{2}{N_0}\int_0^{T_s}[x(t) - s_i(t, C_i = M-1)]s_{int}(t, \varphi_{int}).$$

Then, taking into account (1), (3)–(5), a decision-making algorithm for the value of the channel symbol C_i can be written using the modified Bessel functions $I_0(\cdot)$ as follows:

$$C_i \Rightarrow \max\left\{p_{ps}(C_i)\right\} =$$
$$= \max\left\{\exp(J_i)\frac{1}{2\pi}\int_{-\pi}^{\pi}\exp(a_i)d\varphi_{int}\right\} =$$
$$= \max\left\{\exp(J_i)I_0(U_i)\right\},$$
$$p_{ps}(C_i = 0) > \left\{p_{ps}(C_j \neq 0)\Big|_{j\neq i}\right\},\ i, j = 0, \ldots, M-1. \quad (6)$$

The last expression can be written differently:

$$\exp(J_i)I_0(U_i) > \{\exp(J_j)I_0(U_j)|_{j\neq i}\} \quad (7)$$

or

$$J_i + \ln(I_0(U_i)) > \{J_j + \ln(I_0(U_j))|_{j\neq i}\},\ i, j = 0, \ldots, M-1.$$

In the case of M-QAM, when comparing the values to find the maximum $\max\{\cdot\}$, the decision thresholds must not be considered as zero, but as equal to the half-difference of the energies of the corresponding signal packets.

The arguments of the Bessel functions are formed using the quadrature components of the quantities described by expression (5), for example:

$$U_i = \sqrt{X_i^2 + Y_i^2},$$

$$X_i = \frac{2\mu A_0}{N_0}\int_0^{T_s}[x(t) - s_i(t, C_i)]\cos\omega_{\text{int}} t dt,$$
$$Y_i = \frac{2\mu A_0}{N_0}\int_0^{T_s}[x(t) - s_i(t, C_i)]\sin\omega_{\text{int}} t dt. \quad (8)$$

The decision rule for a channel symbol in the presence of harmonic interference with a random initial phase in addition to noise in the communication channel basically coincides with the decision rule for the receiver of M-PSK and M-QAM signals against the background of only white Gaussian noise. However, in addition to calculating the correlation integrals, it is also necessary to form weight coefficients for these integrals (in the form of Bessel functions), the value of which depends on the degree of correlation between the interference oscillation extracted from the received mixture $x(t)$, as well as the sample copy of the interference stored in the receiver. The corresponding block diagram of the receiver shown in Fig. 1 is denoted as follows: 1 – 90° phase shifter; 2 – integrator; 3 – exponent calculation unit; 4 – squaring unit; 5 – modulus calculation unit; 6 – weight coefficients formation unit; 7 – maximum selection unit; 8 – circuit channel symbol estimation; 9 – scheme for generating weight coefficients; 10 – noise copy generator; 11 – unit for generating reference oscillations.

As an example, Fig. 1 shows one additional channel; the rest are constructed in a similar fashion.

2. SIMULATION RESULTS

To evaluate the efficiency of the scheme of incoherent harmonic interference processing, the simulation of reception was performed for 2-PSK, 4-PSK (4-QAM), 8-PSK, and 16-QAM signals. The bit error probability P_{eb} was determined depending on the interference parameters and the receiver settings.

A. Dependence of the bit error probability on the interference detuning

Modeling of optimal receivers for M-PSK and M-QAM signals showed the efficiency of the incoherent interference processing circuit to depend on the detuning Δf_{int} of the interference frequency relative to the center frequency of the useful signal spectrum. Figure 2 illustrates the dependence of the bit error probability P_{eb} on the interference detuning $\Delta f_{\text{int}} T_s$ at $\mu = 0.5$ and the

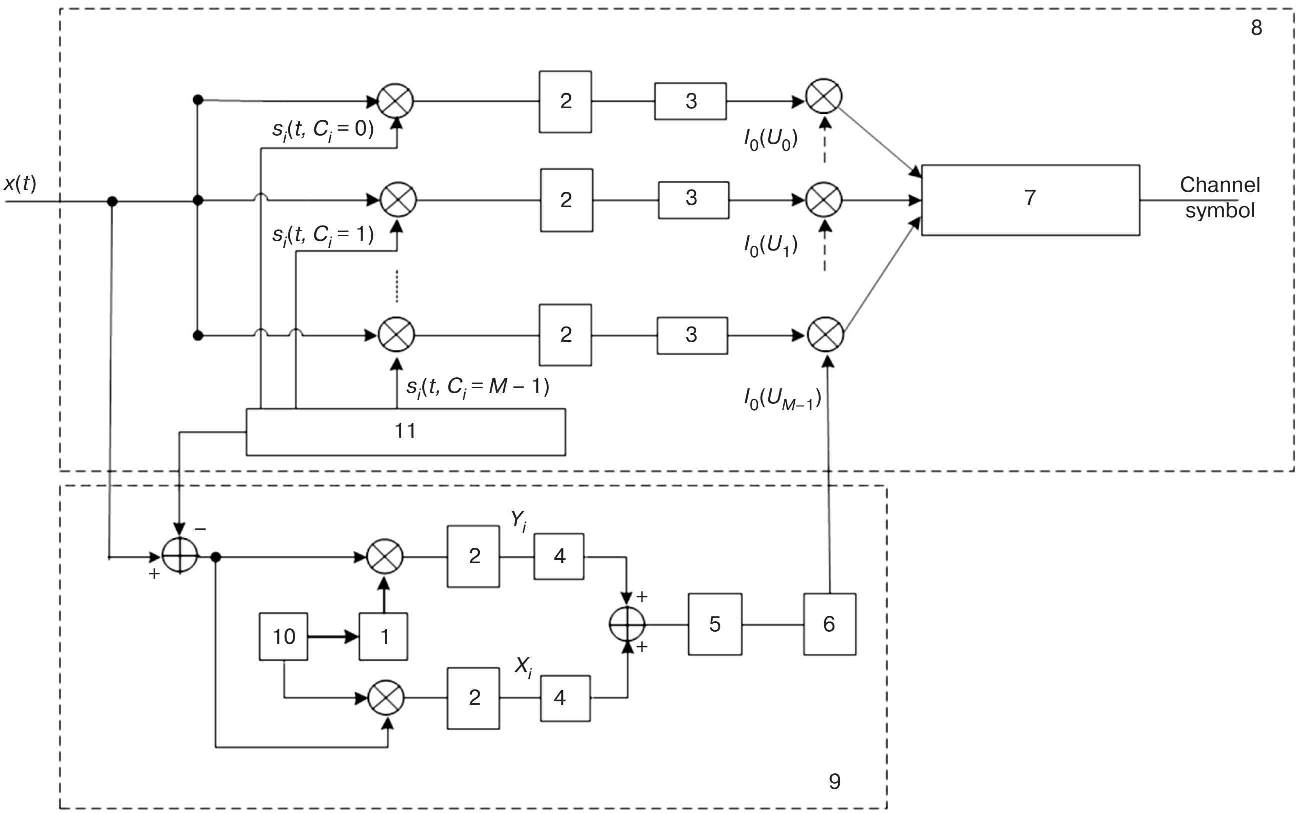


Fig. 1. Block diagram of the optimal receiver of M-PSK and M-QAM signals with incoherent processing of harmonic interference

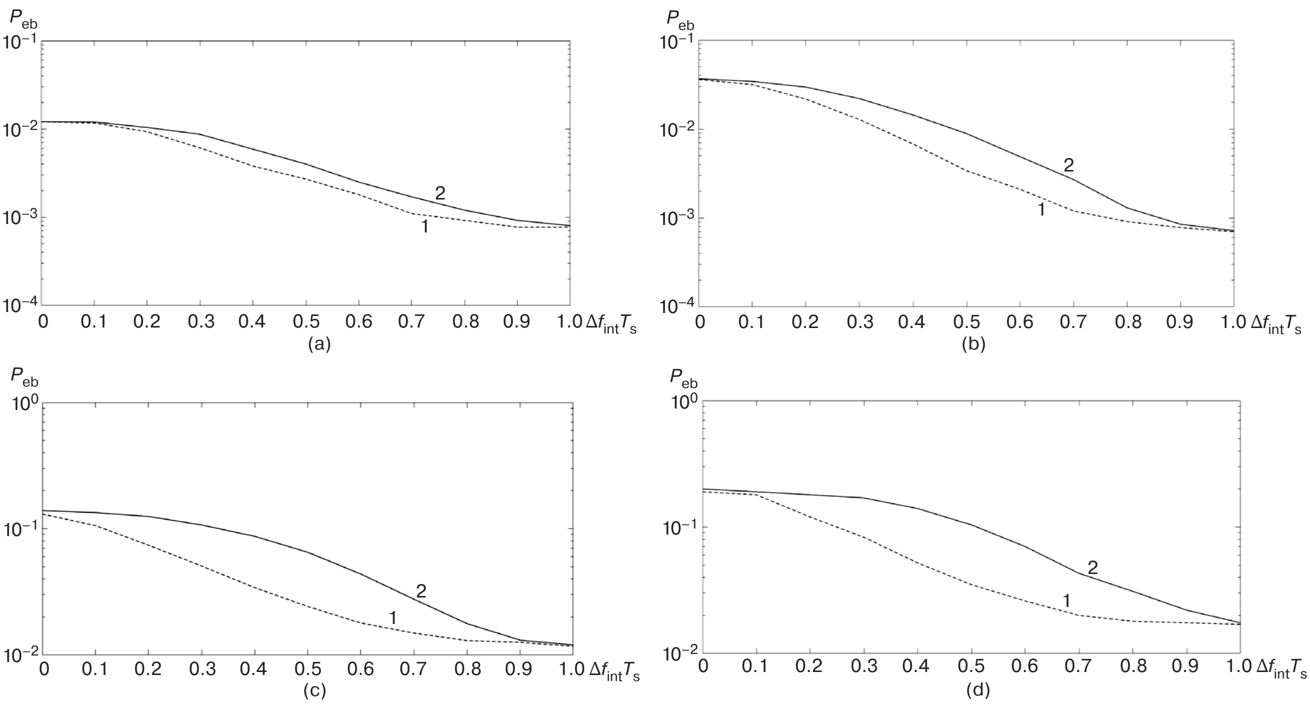


Fig. 2. Dependences of the bit error probability on the interference detuning with (1) and without (2) interference processing: (a) 2-PSK, (b) 4-PSK (4-QAM), (c) 8-PSK, and (d) 16-QAM

signal-to-noise ratio *SNR* = 7 dB. Here, it can be seen that the greatest gain in noise immunity is observed at the detuning $\Delta f_{\text{int}}T_{\text{s}} = 0.5$. Thus, for the 2-PSK signal, the error probability decreases by 1.3 times, for 4-PSK (4-QAM)—by 2.5 times, for 8-PSK—by 2.7 times, and for 16-QAM—by 2.8 times. With targeted interference ($\Delta f_{\text{int}}T_{\text{s}} = 0$), the scheme turns out to be inefficient, while for $\Delta f_{\text{int}}T_{\text{s}} = 1$, there is no gain due to the frequency of the interference coinciding with the frequency of the first zero of the signal spectrum; hence, it can be seen that such interference does not affect reception noise immunity.

Let us now carry out a more detailed analysis of the case when the interference frequency does not coincide with the center frequency of the useful signal spectrum, and when, as follows from Fig. 2, the non-coherent interference processing scheme provides a gain in noise immunity.

B. Dependence of the bit error probability on the signal-to-noise ratio

Figure 3 shows the dependence of the bit error probability P_{eb} on the signal-to-noise ratio SNR (dB) at $\Delta f_{\text{int}}T_{\text{s}} = 0.5$. The dashed lines correspond to the case when the incoherent interference processing circuit is switched on. For comparison, solid lines show the curves obtained without interference processing.

The positive effect of such processing is clearly visible: the probability of bit error is significantly reduced. This is especially noticeable at high interference intensities $\mu = 0.5$ and 0.9. Thus, for $\mu = 0.5$ and $P_{\text{eb}} = 10^{-2}$ switching on the scheme of incoherent interference processing results in an energy gain of no more than 0.5 dB, for 4-PSK (4-QAM)—about 2 dB, and for 8-PSK and 16-QAM—more than 5 dB.

Comparison of the efficiency of the synthesized algorithm with the adaptive algorithm applied in [12, 17] showed that the optimal signal and interference processing provides the best noise immunity parameters over the entire range of interference level changes.

C. Influence of inaccurate setting of the frequency of the copy of the interference in the receiver

Algorithm (7) assumes that the interference frequency in the receiver in known as a sample. Naturally, this assumption is often not fulfilled. Let us consider how the deviation of the frequency of the copy of the interference $\Delta f_{\text{int.rec.}}$ installed in the receiver affects the value of the error probability P_{eb}. Figure 4 shows graphs illustrating this dependence for different signals at SNR = 7 dB. Here, the abscissa shows the value $\Delta f_{\text{int.rec.}}T_{\text{s}}$ normalized to the symbol transmission rate of the useful signal.

Naturally, the interference processing is the most effective when the frequency of the interference $\Delta f_{\text{int.rec.}}T_{\text{s}} = 0$ is precisely known in the receiver. Small inaccuracies in the setting of the frequency of the interference copy $\Delta f_{\text{int.rec.}}T_{\text{s}} = \pm 0.1$ in the receiver slightly reduce the noise immunity of signal reception; although this decrease is stronger, the greater the intensity of the interference, at $\mu = 0.9$ and detuning $\Delta f_{\text{int.rec.}}T_{\text{s}} = \pm 0.5$; for 2-PSK, the error probability already increases by a factor of 5, while for 8-PSK, it decreases by a factor of 7–10.

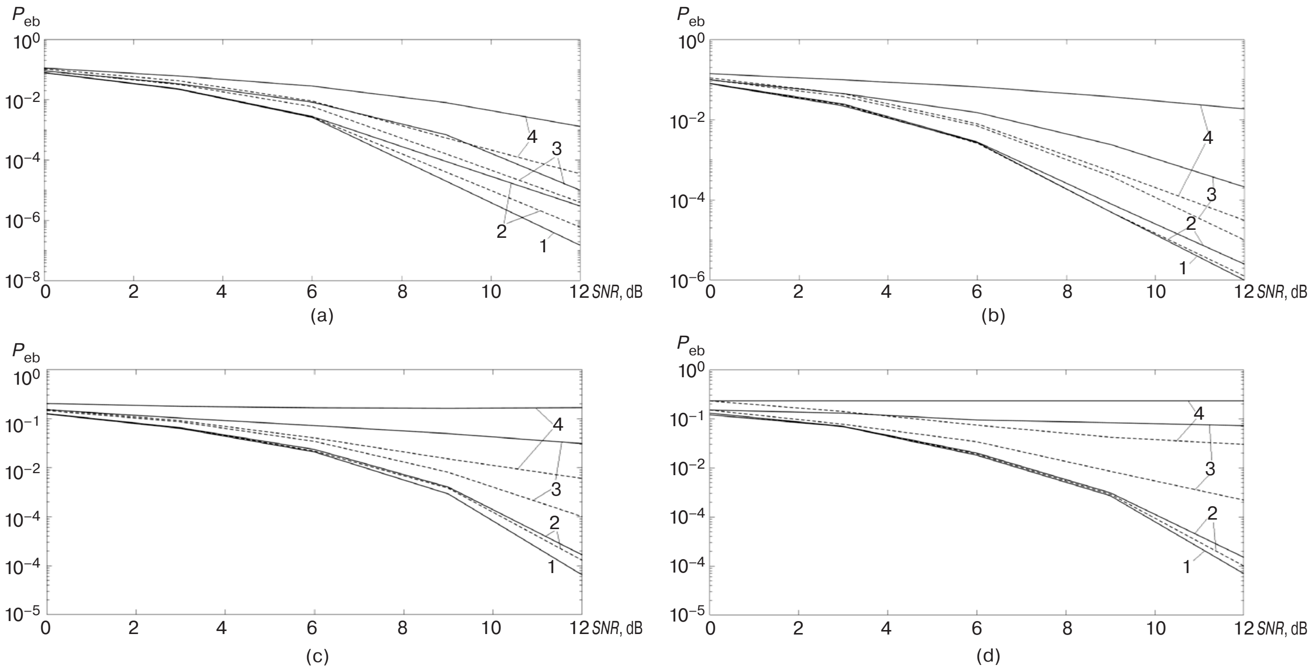


Fig. 3. Dependences of the bit error probability on the signal-to-noise ratio at μ = 0 (1), μ = 0.1 (2), μ = 0.5 (3), μ = 0.9 (4): (a) 2-PSK, (b) 4-PSK (4-QAM), (c) 8-PSK, and (d) 16-QAM

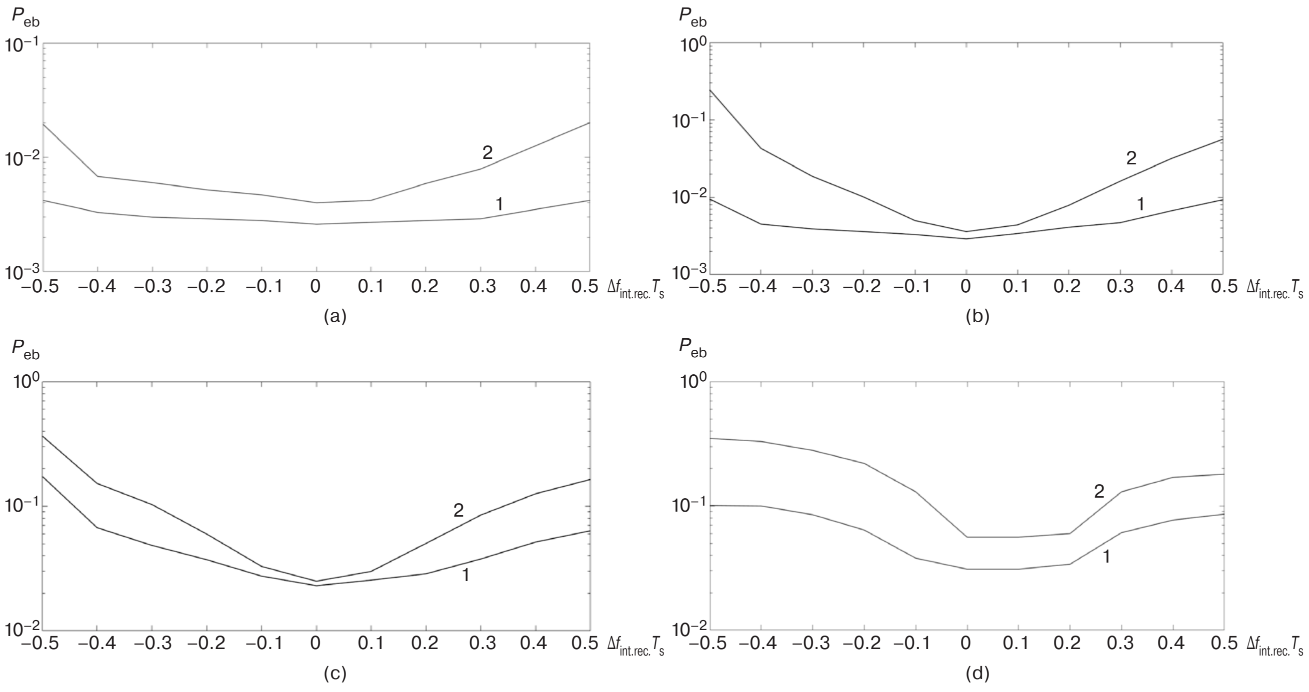


Fig. 4. Influence of the inaccuracy of setting the frequency of the interference copy in the receiver at μ = 0.5 (1), μ = 0.9 (2): (a) 2-PSK, (b) 4-PSK (4-QAM), (c) 8-PSK, and (d) 16-QAM

D. Influence of inaccurate setting of the level of the interference copy in the receiver

Algorithm (7) assumes that the interference level in the receiver is known. Let us consider how the deviation of the level of the copy of the interference $\Delta\mu_{int.rec.}$, installed in the receiver, affects the value of the error probability P_{eb}. Figure 5 shows graphs illustrating this dependence for different signals at SNR = 7 dB and nominal interference intensity μ = 0.9.

From Fig. 5, the value of the error probability can be seen to be minimal at the intensity deviation $\Delta\mu_{int.rec.} = 0$. While small inaccuracies in setting the interference copy level in the receiver $\Delta\mu_{int.rec.} = \pm 0.1$ reduce the noise immunity of signal reception slightly, almost the same (2–3 times) increase in bit error probability is observed for all signals at $\Delta\mu_{int.rec.} = \pm 0.4$.

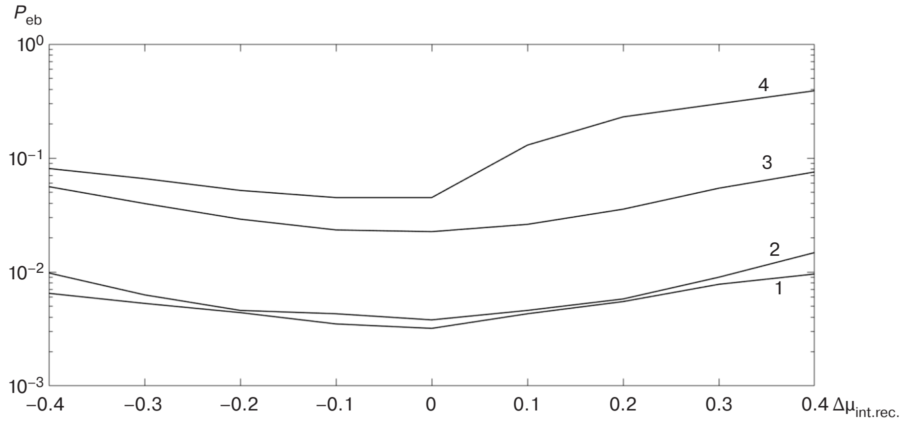


Fig. 5. Influence of the inaccuracy of setting the level of the interference copy in the receiver for 2-PSK (1), 4-PSK (4-QAM) (2), 8-PSK (3), and 16-QAM (4)

CONCLUSIONS

From the analysis of the data obtained, the following conclusions can be drawn:

1. The noise immunity of reception M-PSK and M-QAM signals is improved by optimal non-coherent processing of harmonic interference using an additional threshold correction circuit in the receiver's decision-making unit.
2. The efficiency of the circuit depends on the detuning of the harmonic interference frequency relative to the center frequency of the useful signal spectrum. The greatest gain in noise immunity is observed at the detuning $\Delta f_{int}T_s = 0.5$ (energy gain from 0.5 dB for 2-PSK to more than 5 dB for 8-PSK and 16-QAM at $\mu = 0.5$).
3. The higher the gain in the noise immunity of reception, the greater the intensity of the harmonic interference.
4. The scheme of non-coherent processing of harmonic interference remains operational even with small inaccuracies in the setting of the frequency ($\Delta f_{int.rec.}T_s = \pm 0.1$) and level ($\Delta\mu_{int.rec.} = \pm 0.1$) of the interference copy in the receiver.

Authors' contributions

G.V. Kulikov—the research idea, consultations on the issues of conducting all stages of the study.

T.T. Do—synthesis and analysis of the M-PSK signal receiver.

A.A. Lelyukh—synthesis and analysis of the M-QAM signal receiver.

V.D. Nguyen—computer simulation.

REFERENCES

1. Savvateev Yu.I., Nazarov O.V. (Eds.). *Pomekhozashchishchennost' priema diskretnykh signalov* (*Noise immunity of reception of discrete signals*). Moscow: Radiotekhnika; 2015. 584 p. (in Russ.). ISBN 978-5-93108-094-9
2. Petrov A.V., Beloborodov D.A. Impact of phase-shift keyed noise signal on data transmission channel with multiple phase-shift keying. *Spetsial'naya tekhnika = Special Technique*. 2016;3:2–10 (in Russ.).
3. Buchinskii D.I., Voznyuk V.V., Fomin A.V. Research of noise stability of the receiver with M-PSK modulation under the interference with different structure. *Trudy Voenno-kosmicheskoi akademii imeni A.F. Mozhaiskogo = Proceedings of the Mozhaisky Military Space Academy*. 2019;671:120–127 (in Russ.).
4. Petrov A.V. Symbol error rate in channel, using quadrature amplitude modulation, under the impact of phase-shift keyed jamming. *Informatsionno-izmeritel'nye i upravlyayushchie sistemy = Information-measuring and Control Systems*. 2018;(5):47–53 (in Russ.).

СПИСОК ЛИТЕРАТУРЫ

1. Савватеев Ю.И., Назаров О.В. (ред.). *Помехозащищенность приема дискретных сигналов.* М.: Радиотехника; 2015. 584 с. ISBN 978-5-93108-094-9
2. Петров А.В., Белобородов Д.А. Воздействие фазоманипулированной помехи на канал передачи данных с многопозиционной фазовой манипуляцией. *Специальная техника*. 2016;3:2–10.
3. Бучинский Д.И., Вознюк В.В., Фомин А.В. Исследование помехоустойчивости приемника сигналов с многопозиционной фазовой манипуляцией к воздействию помех с различной структурой. *Труды Военно-космической академии имени А.Ф. Можайского*. 2019;671:120–127.
4. Петров А.В. Вероятность ошибочного приема символа в канале с квадратурной амплитудной манипуляцией под воздействием манипулированной по фазе помехи. *Информационно-измерительные и управляющие системы*. 2018;5:47–53.
5. Ложкин К.Ю., Петров А.В., Прожеторко С.С. Аналитические зависимости средней вероятности искажения бита М-КАМ сигнала на фоне гармонической или фазоманипулированной помех. *Электромагнитные волны и электронные системы*. 2018;23(5):32–41.

5. Lozhkin K.Yu., Petrov A.V., Prozhetorko S.S. Analytical dependences of bit distortion average probability M-QAM of a signal against harmonic or PSK interference. *Elektromagnitnye volny i elektronnye sistemy = Electromagnetic Waves and Electronic Systems*. 2018;23(5):32–41 (in Russ.).
6. Kulikov G.V., Nguyen V.D., Nesterov A.V., Lelyukh A.A. Noise immunity of reception of signals with multiple phase shift keying in the presence of harmonic interference. *Naukoemkie tekhnologii = Science Intensive Technologies.* 2018;19(11):32–38 (in Russ.)
7. Kulikov G.V., Nesterov A.V., Lelyukh A.A. Interference immunity of reception of signals with quadrature amplitude shift keying in the presence of harmonic interference. *Zhurnal Radioelektroniki = Journal of Radio Electronics*. 2018;11:2 (in Russ.). https://doi.org/10.30898/1684-1719.2018.11.9
8. Lozhkin K.Yu., Petrov A.V., Mironov V.A., Mikhalev V.V., Prozhetorko S.S. Analytical dependences of bit distortion average probability M-QAM of a signal against harmonic or PSK jamming subject to fading. *Radiotekhnika = Radioengineering*. 2020;84(4,8):27–35 (in Russ.).
9. Kulikov G.V., Lelyukh A.A., Batalov E.V., Kuzelenkov P.I. Noise immunity of reception of signals with quadrature amplitude modulation in the presence of interference phase-shift keying. *Zhurnal Radioelektroniki = J. Radio Electronics* 2019;7 (in Russ.). https://doi.org/10.30898/1684-1719.2019.7.10
10. Nandi M. Symbol error probablity of coherent PSK system in the presence of two path interference. *Int. J. Phys. Appl.* (*IJPA*). 2013;5(2):133–137. Available from URL: http://www.irphouse.com/ijpa/ijpav5n2_10.pdf
11. Widrow B., Stearns S.D. *Adaptive signal processing*. Prentice-Hall; 1985. 474 p.
12. Kulikov G.V., Do T.T. Efficiency of the phase algorithm of adaptive filtering for receiving signals with multi-position phase shift keying. *Zhurnal Radioelektroniki = Journal of Radio Electronics*. 2020;4:1 (in Russ.). https://doi.org/10.30898/1684-1719.2020.4.9
13. Prilepskii A.V., Prilepskii V.V., Kamenskii S.A. Optimization of the spectra of complex phase-shift keyed signals when received against the background of narrow-band interference and noise. *Teoriya i tekhnika radiosvyazi = Radio Communication Theory and technology*. 2005;1:44–47 (in Russ.).
14. Poborchaya N.E., Smerdova E.O. Variational algorithm for compensating QAM signal distortion against the background of quasi-deterministic band interference and additive white noise. *Sistemy sinkhronizatsii, formirovaniya i obrabotki signalov = Systems of Signal Synchronization, Generation and Processing*. 2014;5(4):141–147 (in Russ.).
15. Tikhonov V.I. Optimal'nyi priem signalov (*Optimal Signal Reception*). Moscow: Radio i svyaz'; 1983. 320 p. (in Russ.).
16. Tikhonov V.I. Statisticheskaya radiotekhnika (*Statistical Radio Engineering*): 2nd ed. Moscow: Radio i svyaz'; 1982. 624 p. (in Russ.).

6. Куликов Г.В., Нгуен Ван Зунг, Нестеров А.В., Лелюх А.А. Помехоустойчивость приема сигналов с многопозиционной фазовой манипуляцией в присутствии гармонической помехи. *Наукоемкие технологии*. 2018;19(11):32–38.
7. Куликов Г.В., Нестеров А.В., Лелюх А.А. Помехоустойчивость приема сигналов с квадратурной амплитудной манипуляцией в присутствии гармонической помехи. *Журнал радиоэлектроники*. 2018;11:2. https://doi.org/10.30898/1684-1719.2018.11.9
8. Ложкин К.Ю., Петров А.В., Миронов В.А., Михалёв В.В., Прожеторко С.С. Аналитические зависимости средней вероятности искажения бита М-КАМ-сигнала на фоне гармонической или фазоманипулированной помех с учетом замираний. *Радиотехника.* 2020;84(4–8):27–35.
9. Куликов Г.В., Лелюх А.А., Баталов Е.В., Кузеленков П.И. Помехоустойчивость приема сигналов с квадратурной амплитудной манипуляцией в присутствии фазоманипулированной помехи. *Журнал радиоэлектроники.* 2019;7. https://doi.org/10.30898/1684-1719.2019.7.10
10. Nandi M. Symbol error probablity of coherent PSK system in the presence of two path interference. *Int. J. Phys. Appl.* (*IJPA*). 2013;5(2):133–137. URL: http://www.irphouse.com/ijpa/ijpav5n2_10.pdf
11. Widrow B., Stearns S.D. *Adaptive signal processing*. Prentice-Hall; 1985. 474 p.
12. Куликов Г.В., До Чунг Тиен. Эффективность фазового алгоритма адаптивной фильтрации при приеме сигналов с многопозиционной фазовой манипуляцией. *Журнал радиоэлектроники*. 2020;4:1. https://doi.org/10.30898/1684-1719.2020.4.9
13. Прилепский А.В., Прилепский В.В., Каменский С.А. Оптимизация спектров сложных фазоманипулированных сигналов при приеме на фоне узкополосных помех и шума. *Теория и техника радиосвязи*. 2005;1:44–47.
14. Поборчая Н.Е., Смердова Е.О. Вариационный алгоритм компенсации искажений КАМ сигнала на фоне квазидетерминированной полосовой помехи и аддитивного белого шума. *Системы синхронизации, формирования и обработки сигналов*. 2014;5(4):141–147.
15. Тихонов В.И. *Оптимальный прием сигналов*. М.: Радио и связь; 1983. 320 с.
16. Тихонов В.И. *Статистическая радиотехника*. 2-е изд., перераб. и доп. М.: Радио и связь; 1982. 624 с.
17. Куликов Г.В., Лелюх А.А., Граченко Е.Н. Адаптивная фильтрация гармонической помехи при приеме сигналов с квадратурной амплитудной модуляцией. *Журнал радиоэлектроники*. 2021;8. https://doi.org/10.30898/1684-1719.2021.8.13

17. Kulikov G.V., Lelyukh A.A., Grachenko E.N. Adaptive filtering of harmonic interference when receiving signals with quadrature amplitude modulation. *Zhurnal Radioelektroniki = Journal of Radio Electronics*. 2021;8 (in Russ.). https://doi.org/10.30898/1684-1719.2021.8.13

About the authors

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

Trung T. Do, Postgraduate Student, Department of Radio Electronic Systems and Complexes, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: dotrungtien1993@gmail.com. http://orcid.org/0000-0002-9596-5440

Andrey A. Lelyukh, Deputy Head of the Technical Center of Special Equipment of JSC "Moscow Research Institute of Radio Communications" (32, Nizhegorodskaya ul., Moscow, 109029 Russia). E-mail: a.lel@mail.ru. RSCI SPIN-code 1021-5094.

Van D. Nguyen, Assistant, Le Quy Don Technical University, (236 Hoang Quoc Viet, Hanoi, Vietnam). E-mail: codonmaimai92@gmail.com. https://orcid.org/0000-0002-2810-1204

Об авторах

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

До Чунг Тиен, аспирант кафедры радиоэлектронных систем и комплексов Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: dotrungtien1993@gmail.com. http://orcid.org/0000-0002-9596-5440

Лелюх Андрей Александрович, заместитель начальника технического центра специальной аппаратуры АО «Московский научно-исследовательский институт радиосвязи» (109029, Россия, Москва, Нижегородская ул., д. 32). E-mail: a.lel@mail.ru. SPIN-код РИНЦ 1021-5094.

Нгуен Ван Зунг, ассистент, Вьетнамский государственный технический университет им. Ле Куй Дона (Вьетнам, Ханой, ул. Хоанг Куок Вьет, 236). E-mail: codonmaimai92@gmail.com. https://orcid.org/0000-0002-2810-1204

Translated from Russian into English by Evgenii I. Shklovskii
Edited for English language and spelling by Thomas A. Beavitt

UDC 621.396.49
https://doi.org/10.32362/2500-316X-2023-11-1-51-59

RESEARCH ARTICLE

Analysis of the effectiveness of methods for ensuring the reliability of a communication satellite transponder

Tatyana E. Gelfman @,
Alexey P. Pirkhavka,
Vladimir O. Skripachev

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

Abstract
Objectives. Since the launch of satellite communication systems in practical use, approaches towards enhancing their operational quality and durability have been developing in the direction of increased reliability of airborne transponders. This is mainly achieved by increasing redundancy and using components with a lower failure rate. In this regard, the creation of new technologies and new materials is a particularly promising direction. However, since durability testing of complex systems can take several years, the problem of ensuring an effective combination of redundancy methods and elements having a reduced failure rate remains challenging. The purpose of the work is to analyze the effectiveness of methods for ensuring the reliability of a communication satellite transponder based on a proposed methodology for determining the durability index using a mathematical model of the probability of failure-free operation.
Methods. In order to describe the complex structure of a satellite communication system transponder, a logical-probabilistic method is used, in which the dependence of the system reliability indicators on the reliability indicators of the transponder elements is formulated as a logical function of operability. Mathematical models of system reliability are created on this basis including for redundant systems. Graphs and analytical methods are used to compare different systems.
Results. The influence of various methods for ensuring the redundancy of transponder devices and the use of more reliable components on the reliability and durability indicators is considered. A gamma-percentage resource-based technique for determining the durability indicator based on the constructed mathematical models of the probability of failure-free operation is presented along with a comparative analysis of measures to increase the gamma-percentage resource of the transponder.
Conclusions. The presented method for determining the durability index using a mathematical model of the probability of no-failure operation can be used to determine the time interval within which redundancy increases the probability of no-failure operation as compared with a decrease in the failure rate of elements. On this basis, the most effective combination of redundancy methods and approaches for reducing the failure rate of elements can be identified.

Keywords: reliability, satellite communication, airborne transponder, redundancy, gamma-percentage resource

• **Submitted:** 08.07.2022 • **Revised:** 29.09.2022 • **Accepted:** 08.11.2022

For citation: Gelfman T.E., Pirkhavka A.P., Skripachev V.O. Analysis of the effectiveness of methods for ensuring the reliability of a communication satellite transponder. *Russ. Technol. J.* 2023;11(1):51–59. https://doi.org/10.32362/2500-316X-2023-11-1-51-59

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

Анализ эффективности методов обеспечения надежности ретранслятора спутника связи

Т.Э. Гельфман @,
А.П. Пирхавка,
В.О. Скрипачев

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

Резюме

Цели. Повышение качества и долговечности работы спутниковых систем связи с момента начала их практического использования развивается в направлении увеличения надежности бортовых ретрансляторов. Это достигается применением методов резервирования и использованием элементной базы с меньшей интенсивностью отказов. Наряду с резервированием большие перспективы в этом плане открывает создание новых технологий и материалов. Весьма актуальной является проблема эффективного сочетания методов резервирования и способов уменьшения интенсивности отказов элементов. Испытания на долговечность сложных систем могут продолжаться несколько лет. Цель работы – анализ эффективности методов обеспечения надежности ретранслятора спутниковой связи на основании предложенной методики определения показателя долговечности по математической модели вероятности безотказной работы.

Методы. Для описания структуры сложной системы, в данном случае бортового ретранслятора системы спутниковой связи, используется логико-вероятностный метод, в котором зависимость показателей надежности системы от показателей надежности элементов формулируется в виде логической функции работоспособности. Эта функция дает возможность создавать различные математические модели надежности систем и построить их логические схемы надежности, в т.ч. для резервированных систем. Для сравнения различных систем применяются графоаналитические методы.

Результаты. Рассмотрено влияние различных методов резервирования устройств ретранслятора и применения более надежной элементной базы на показатели безотказности и долговечности. Представлена методика определения показателя долговечности – гамма-процентного ресурса – по построенным математическим моделям вероятности безотказной работы. Проведен сравнительный анализ мероприятий по увеличению гамма-процентного ресурса ретранслятора.

Выводы. Рассмотренная методика определения показателя долговечности по математической модели вероятности безотказной работы позволяет определить интервал времени, в пределах которого резервирование дает выигрыш по вероятности безотказной работы по сравнению с уменьшением интенсивности отказов элементов. Таким образом возможно обеспечить эффективное сочетание методов резервирования и способов уменьшения интенсивности отказов элементов.

Ключевые слова: надежность, спутниковая связь, бортовой ретранслятор, резервирование, гамма-процентный ресурс

• **Поступила:** 08.07.2022 • **Доработана:** 29.09.2022 • **Принята к опубликованию:** 08.11.2022

Для цитирования: Гельфман Т.Э., Пирхавка А.П., Скрипачев В.О. Анализ эффективности методов обеспечения надежности ретранслятора спутника связи. *Russ. Technol. J.* 2023;11(1):51–59. https://doi.org/10.32362/2500-316X-2023-11-1-51-59

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

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The reliability of satellite communication systems is determined by their characteristics: efficiency, durability, readiness, and risk [1]. To ensure high reliability indicators, it is necessary that the satellite system meets the requirements for a variety of criteria, such as the probability of failure-free operation, mean time to failure, gamma-percentage resource, and operational availability factor [2]. Enhancement in the reliability of satellite communication systems since the beginning of their practical use has been evolving in the direction of increasing the probability of failure-free operation of airborne transponders [3–7]. The lifetime of modern satellite transponders reaches 15 years [4, 8], which is ensured by the use of redundancy methods and the use of components with a lower failure rate. To achieve the required values of reliability indicators, different redundancy methods are used [1, 9–11]. For example, it was shown in [1] that with separate redundancy by replacement with an unloaded reserve, a greater gain in mean time to failure and gamma-percentage resource is provided, and with constant redundancy, a greater gain in the probability of failure-free operation is attained. In addition, the implementation of permanent redundancy is less expensive. The use of redundancy leads to the complication of transponders and, consequently, to an increase in energy consumption, weight, size, and cost indicators of systems. Therefore, it is often necessary to look for the optimal solution that allows one to obtain a given reliability indicator at minimal cost, or maximum reliability for given quality indicators [8, 12–14]. For example, in [15], the effectiveness of optimal redundancy by replacement was evaluated and it was shown that an unloaded redundancy, compared to a loaded one, provides a greater gain in the probability of failure-free operation for any operating time and in average time to failure. In addition, the cost of a transponder with a loaded redundancy is higher, because of increased weight, size, and energy indicators.

Methods for reducing the failure rate of elements are based on new technologies and design principles [3, 8], as well as load redundancy associated with facilitating electrical, thermal, mechanical, and other operating modes of elements [1, 9]. Increased reliability of the elements is characterized by patterns that have pronounced “saturation” areas, determined by the fact that, following an initial period of a sufficiently effective impact on reliability, no further actions or material investments in the development and manufacturing of elements have a significant impact on the increase in reliability. This is due to the achievement of the physical limitations inherent in each class of the components. However, with new generations of component technologies, it is possible to sharply increase reliability levels.

Obviously, the choice of methods for ensuring the reliability of satellite transponders significantly affects their weight, size, energy, and cost indicators.

The aim of the present work is to analyze the effectiveness of combining the methods of redundancy and reducing the failure rate of components based on the determination of the durability indicator in terms of gamma-percentage resource, using a mathematical model of the probability of failure-free operation.

MATHEMATICAL MODELS OF THE PROBABILITY OF NO-FAULT OPERATION

Let the reliability logic diagram of the transponder have the form shown in Fig. 1. The transponder consists of three sections.

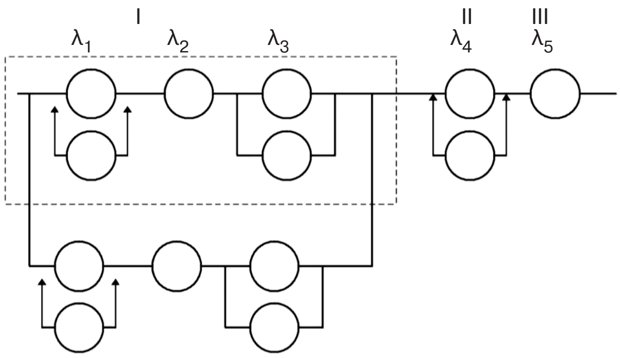


Fig. 1. Logic diagram of the reliability of the transponder

The probability of failure-free operation of the entire transponder P_{trp} is equal to the product of the probabilities of failure-free operation of each section of the device:

$$P_{trp} = P_I P_{II} P_{III}. \quad (1)$$

Section I (highlighted in Fig. 1 by a dashed line), comprising the main receiver, consists of three elements connected in series, the first having a failure rate λ_1 is redundant by substitution (unloaded reserve), while the second element with a failure rate λ_2 is not redundant and the third element with a failure rate λ_3 has permanent redundancy.

The probability of failure-free operation of the main receiver is equal to the product of the probabilities of failure-free operation of each of these three elements, taking into account their redundancy.

$$P_{rec} = P_1 P_2 P_3 \quad (2)$$

For the first element redundant by the substitution,

$$P_1 = e^{-\lambda_1 t}\left(1+\lambda_1 t\right). \quad (3)$$

The second element is not redundant,

$$P_2 = e^{-\lambda_2 t}. \quad (4)$$

For the third element with permanent redundancy,

$$P_3 = 1-\left(1-e^{-\lambda_3 t}\right)^2. \quad (5)$$

Substituting (3)–(5) into (2), we obtain:

$$P_{rec} = e^{-\lambda_1 t}\left(1+\lambda_1 t\right)e^{-\lambda_2 t}\left(1-\left(1-e^{-\lambda_3 t}\right)^2\right).$$

When the receiver is duplicated, the probability of failure-free operation of the first section of the transponder in Fig. 1 is defined by the expression:

$$P_I = 1 - (1 - P_{rec})^2. \quad (6)$$

Section II consists of one transmitter with a failure rate λ_4, redundant by substitution (unloaded reserve). The probability of failure-free operation of this section is determined by the expression:

$$P_{II} = e^{-\lambda_4 t}\left(1+\lambda_4 t\right). \quad (7)$$

Section III consists of one non-redundant element—the antenna-feeder device of the transponder. The probability of failure-free operation of this section:

$$P_{III} = e^{-\lambda_5 t}, \quad (8)$$

where λ_5 is the failure rate of the antenna-feeder device.

Substituting (6)–(8) into (1), we obtain the expression for the probability of failure-free operation of the transponder of the communication satellite for the case of the duplicated receiver and the redundant-by-replacing transmitter, as well as when the first element is redundant by replacing, and when the third element of the receiver has permanent redundancy:

$$P_{trp} = \left(1-\left(1-P_{rec}\right)^2\right)e^{-\lambda_4 t}\left(1+\lambda_4 t\right)e^{-\lambda_5 t}.$$

Without duplication of the receiver, as well as without redundancy of the first and third elements of the receiver, the probability of failure-free operation of the transponder is described by the formula:

$$P_{trp} = e^{-\lambda_1 t}e^{-\lambda_2 t}e^{-\lambda_3 t}e^{-\lambda_4 t}\left(1+\lambda_4 t\right)e^{-\lambda_5 t}.$$

For the case of redundancy by replacing the first element and permanent redundancy of the third element of the receiver, but without duplication of the receiver, the probability of failure-free operation of the transponder is described by the formula:

$$P_{trp} = e^{-\lambda_1 t}\left(1+\lambda_1 t\right)e^{-\lambda_2 t}\left(1-\left(1-e^{-\lambda_3 t}\right)^2\right)e^{-\lambda_4 t}\left(1+\lambda_4 t\right)e^{-\lambda_5 t}.$$

EFFECT OF REDUNDANCY ON PROBABILITY OF FAILURE-FREE OPERATION

Let us consider the effect of redundancy methods on the probability of failure-free operation and the durability of a transponder when there is a change in the failure rate of the receiver, transmitter, and antenna-feeder device. As an indicator of durability, we use the gamma-percentage resource, which is determined from the graphs of the dependence of the probability of failure-free operation of the transponder on time at $P_{trp} = 0.9$.

Graphs of the dependence of the probability of failure-free operation of the transponder on time for three redundancy methods are shown in Fig. 2–4 for different failure rates of elements (cascades).

Method 1. Duplication of the receiver and redundancy by replacing the transmitter, as well as redundancy by replacing the first element and permanent redundancy of the third element of the receiver. Graphs of the dependence of the probability of failure-free operation on time are presented in Fig. 2. The parameters of the graphs and the results of calculating the gamma-percentage resource are shown in Table 1.

As expected, the gamma-percentage resource increases with a decrease in the element failure rate. So,

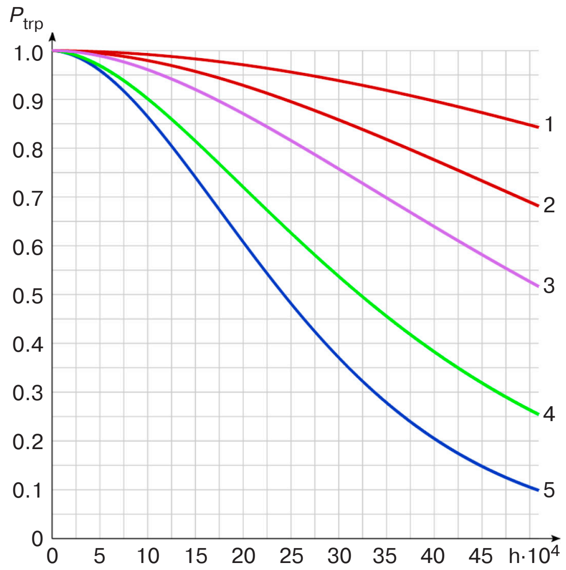


Fig. 2. Dependencies of the probability of failure-free operation on time for the first redundancy method

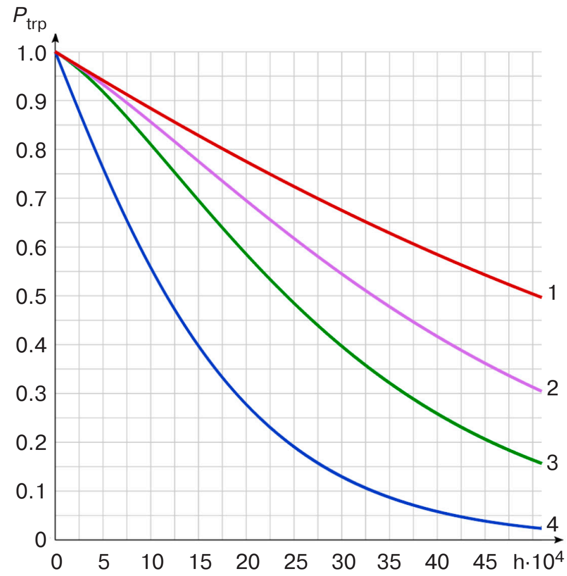


Fig. 3. Dependencies of the transponder failure-free operation probability on time for the second redundancy method

Table 1. Graphs parameters and calculation results (Fig. 2)

Curve No.	Failure rate of elements of the transponder, 10^{-8}/h					Gamma percentage resource, h·10^4
	λ_1	λ_2	λ_3	λ_4	λ_5	
1	20	50	50	100	1	38.8
2	20	50	50	200	1	24.0
3	20	50	50	300	1	16.9
4	20	50	50	500	1	9.8
5	100	200	200	500	1	8.2

Table 2. Graphs parameters and calculation results (Fig. 3)

Curve No.	Failure rate of elements of the transponder, 10^{-8}/h					Gamma percentage resource, h·10^4
	λ_1	λ_2	λ_3	λ_4	λ_5	
1	20	50	50	100	1	8.4
2	20	50	50	300	1	7.1
3	20	50	50	500	1	5.7
4	100	20	20	500	1	2.1

for example, with a decrease in the transmitter failure rate by 2, 3, and 5 times, the gamma-percentage resource increases by 1.6, 2.3, and 4 times, respectively.

Method 2. Redundancy by replacing the transmitter, without redundancy of the receiver and its elements. Graphs of dependence of the probability of failure-free operation on time are presented in Fig. 3. Table 2 shows the parameters of the graphs and the results of calculating the gamma-percentage resource.

Analysis of the graphs in Figs. 2 and 3 along with the data of Tables 1 and 2 shows that, without redundancy of the receiver and with the same failure rates of the elements, the gamma-percentage resource decreases by more than 4 times. However, with more reliable components without redundancy of the receiver, a less significant decrease in the gamma-percentage resource is observed, which does not exceed 1.7 times in the cases under consideration.

Method 3. There is no general duplication of the receiver; redundancy of the transmitter, as well as the first and third elements of the receiver is implemented. Graphs of the dependence of the probability of failure-free operation on time are shown in Fig. 4. The results of the calculation of the gamma-percentage resource and the parameters of the graphs are presented in Table 3.

As follows from the graphs in Fig. 4 and data of Table 3, the antenna-feeder device having the highest reliability has practically no effect on the gamma-percentage resource: in the case under consideration, with an increase in the failure rate λ_5 by 10 times, the gamma-percentage resource decreases by less than 5%.

Let us consider the redundancy options of various methods, in which the values of the gamma-percentage resource vary slightly. We will evaluate the efficiency of the transponder redundancy methods for the above cases in terms of the probability of failure-free operation by determining the efficiency factor of the redundancy method K_{eff} defined as the ratio of the probabilities of failure-free operation for the variants with the same values of the gamma-percentage resource. Figure 5 shows the time dependence of the redundancy efficiency factor of the compared options. Table 4 shows the parameters of

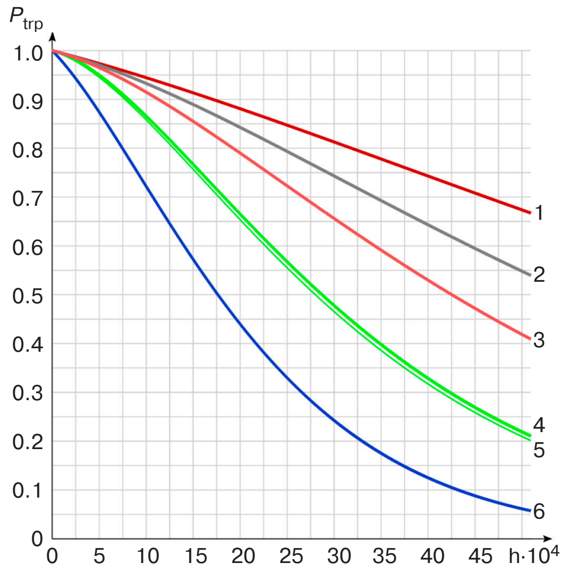


Fig. 4. Dependencies of the transponder failure-free operation probability on time for the third redundancy method

Table 3. Graph parameters and calculation results (Fig. 4)

Curve No.	Failure rate of elements of the transponder, 10^{-8}/h					Gamma percentage resource, h·10^4
	λ_1	λ_2	λ_3	λ_4	λ_5	
1	20	50	50	100	1	16.9
2	20	50	50	200	1	13.6
3	20	50	50	300	1	11.1
4	20	50	50	500	1	8.0
5	20	50	50	500	10	7.6
6	100	200	200	500	1	4.0

the graphs together with the results of determining the effectiveness of redundancy methods.

From the graphs in Fig. 5, an important conclusion can be drawn: there is a threshold value of time up to which redundancy gives a gain in the probability of failure-free operation of the transponder as compared to a decrease in the failure rate of elements. For example, the first redundancy method provides a gain compared to the third method in the time interval from 0 to $1.69 \cdot 10^5$ h even with a 3-fold increase in the transmitter failure rate, as well as an advantage over the second method in the time interval from 0 to $8.7 \cdot 10^4$ h; this applies even in the case of an increase in the failure rates of the first and fourth elements by 5 times, or the second and third elements—by 4 times.

The third redundancy method provides a gain in comparison with the second method in the time interval from 0 to $7.1 \cdot 10^4$ h even with a 5-fold increase in the transmitter failure rate.

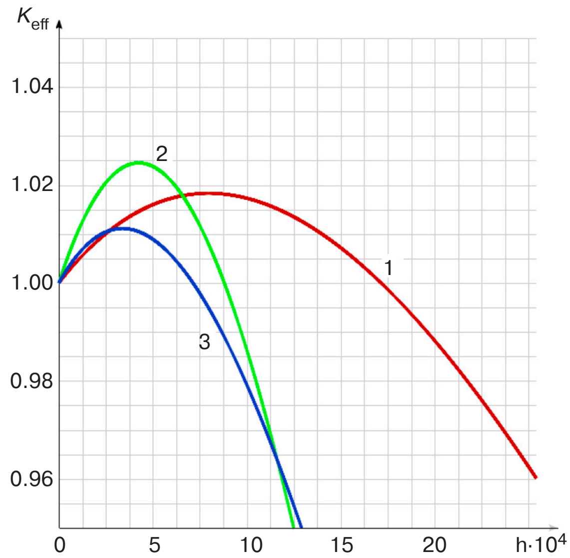


Fig. 5. Redundancy efficiency ratio

Table 4. Graph parameters (Fig. 5) and obtained effectiveness of redundancy methods

Curve No.	Compared redundancy methods	Failure rate of elements of the transponder, 10^{-8}/h					Threshold time h·10^4	Maximum value of K_{eff}
		λ_1	λ_2	λ_3	λ_4	λ_5		
1	1	20	50	50	300	1	16.9	1.018
	3	20	50	50	100	1		
2	1	100	200	200	500	1	8.7	1.025
	2	20	50	50	100	1		
3	3	20	50	50	500	1	7.1	1.011
	2	20	50	50	100	1		

When the threshold value of time is exceeded, there is no gain from redundancy; in this case, the factor that determines the probability of failure-free operation is the reliability of the transponder elements.

CONCLUSIONS

The gamma-percentage transponder resource increases with a decrease in the element failure rate. Although this durability indicator decreases in the absence of redundancy of the receiver and its elements, this decrease is not so significant with more reliable elements. By using the proposed method for calculating the gamma-percentage resource using mathematical models of the probability of failure-free operation, it becomes possible to determine the time interval within which redundancy provides a gain in the probability of failure-free operation in comparison with an increase in the reliability of elements. Thus, it is possible to provide an effective combination of redundancy methods and approaches to reducing the failure rate of elements.

ACKNOWLEDGMENTS

The work was supported by the Foundation for Assistance to Small Innovative Enterprises (FASIE), Contract No. 94С2/МОЛ/73887, dated March 25, 2022.

Authors' contributions

T.E. Gelfman—the research idea, scientific editing.

A.P. Pirkhavka—conducting research, writing and editing the text of the article.

V.O. Skripachev—analysis of literature, preparation of graphic materials, editing of the text of the article.

REFERENCES

1. Polovko A.M., Gurov S.V. *Osnovy teorii nadezhnosti* (*Fundamentals of Reliability Theory*). St. Petersburg: BKHV-Petersburg; 2008. 704 p. (in Russ.). ISBN 978-5-94157-541-1
2. Gelfman T.E., Pirkhavka A.P. The operational readiness factor of satellite communication networks. *Russ. Technol. J.* 2022;10(1):35–40 (in Russ.). https://doi.org/10.32362/2500-316X-2022-10-1-35-40
3. Sevastyanov N.N., Andreev A.I. *Osnovy upravleniya nadezhnost'yu kosmicheskikh apparatov s dlitel'nymi srokami ehkspluatatsii* (*Fundamentals of Reliability Management of Spacecraft with Long Service Life*). Tomsk: TGU; 2015. 266 p. (in Russ.). ISBN 978-5-94621-460-5
4. Maral G., Bousquet M., Sun Zh. *Satellite Communications Systems: Systems, Techniques and Technology.* 6th ed. Wiley & Sons Ltd; 2020. 792 p. ISBN 978-1-119-38208-9
5. Verma A.S., Jaiswal A.K., Kumar M., Nigam G., Srivastava S.K. Measurement of reliability and availability of satellite communication links: Progress and challenges. In: *2013 International Conference on Intelligent Systems and Signal Processing* (*ISSP*). 2013:268–271. https://doi.org/10.1109/ISSP.2013.6526916
6. Elbert B.R. *Introduction to Satellite Communication*. 3rd ed. Artech House Publishers; 2008. 447 p. ISBN 978-1-59693-210-4
7. Jung S., Choi J.P. End-to-end reliability of satellite communication network systems. *IEEE Systems Journal*. 2021;15(1):791–801. https://doi.org/10.1109/jsyst.2020.2980760
8. Mehmet N., Selman D., Hasan H.E., Cenk S. Reliability and cost focused optimization approach for a communication satellite payload redundancy allocation problem. *International Journal of Electrical, Electronic and Communication Sciences*. 2018;11.0(5). https://doi.org/10.5281/zenodo.1316576
9. Antonov A.V., Nikulin M.S. *Statisticheskie modeli v teorii nadezhnosti* (*Statistical Models in Reliability Theory*). Moscow: Abris; 2012. 390 p. (in Russ.). ISBN 978-5-4372-0009-4
10. Gelfman T.E., Gnuchev O.E., Lobyntsev R.Yu. Reservation methods satellite communication systems. In: *Actual Problems and Prospects of Development of Radioengineering and Infocommunication Systems. Proceedings of the III International Scientific and Practical Conference*. Moscow: Technological University (MIREA); 2017. P. 366–374 (in Russ.).

СПИСОК ЛИТЕРАТУРЫ

1. Половко А.М., Гуров С.В. *Основы теории надежности*. СПб.: БХВ-Петербург; 2008. 704 с. ISBN 978-5-94157-541-1
2. Гельфман Т.Э., Пирхавка А.П. Коэффициент оперативной готовности спутниковых сетей связи. *Russ. Technol. J.* 2022;10(1):35–40. https://doi.org/10.32362/2500-316X-2022-10-1-35-40
3. Севастьянов Н.Н., Андреев А.И. *Основы управления надежностью космических аппаратов с длительными сроками эксплуатации*. Томск: ТГУ; 2015. 266 с. ISBN 978-5-94621-460-5
4. Maral G., Bousquet M., Sun Zh. *Satellite Communications Systems: Systems, Techniques and Technology.* 6th ed. Wiley & Sons Ltd; 2020. 792 p. ISBN 978-1-119-38208-9
5. Verma A.S., Jaiswal A.K., Kumar M., Nigam G., Srivastava S.K. Measurement of reliability and availability of satellite communication links: Progress and challenges. In: *2013 International Conference on Intelligent Systems and Signal Processing* (*ISSP*). 2013:268–271. https://doi.org/10.1109/ISSP.2013.6526916
6. Elbert B.R. *Introduction to Satellite Communication*. 3rd ed. Artech House Publishers; 2008. 447 p. ISBN 978-1-59693-210-4
7. Jung S., Choi J.P. End-to-end reliability of satellite communication network systems. *IEEE Systems Journal*. 2021;15(1):791–801. https://doi.org/10.1109/jsyst.2020.2980760
8. Mehmet N., Selman D., Hasan H.E., Cenk S. Reliability and cost focused optimization approach for a communication satellite payload redundancy allocation problem. *International Journal of Electrical, Electronic and Communication Sciences*. 2018;11.0(5). https://doi.org/10.5281/zenodo.1316576
9. Антонов А.В., Никулин М.С. *Статистические модели в теории надежности*. М.: Абрис; 2012. 390 с. ISBN 978-5-4372-0009-4
10. Гельфман Т.Э., Гнучев О.Е., Лобынцев Р.Ю. Методы резервирования спутниковых систем связи. В сб.: *Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем «Радиоинфоком – 2017». Сб. научных трудов III Международной научно-практической конференции*. Московский технологический университет (МИРЭА). 2017. С. 366–374.
11. Ямпурин Н.П., Баранова А.В. *Основы надежности электронных средств*. М.: Академия; 2010. 240 с. ISBN 978-5-7695-5908-2

11. Yampurin N.P., Baranova A.V. *Osnovy nadezhnosti elektronnykh sredstv* (*Fundamentals of Reliability of Electronic Means*). Moscow: Akademiya; 2010. 240 p. (in Russ.). ISBN 978-5-7695-5908-2
12. Bouwmeester J., Menicucci A., Gill E.K.A. Improving CubeSat reliability: Subsystem redundancy or improved testing? *Reliab. Eng. Syst. Saf.* 2022;220:108288. https://doi.org/10.1016/j.ress.2021.108288
13. Castet J.F., Saleh J.H. Satellite and satellite subsystems reliability: Statistical data analysis and modeling. *Reliab. Eng. Syst. Saf*. 2009;94(11):1718–1728. https://doi.org/10.1016/j.ress.2009.05.004
14. Kislaev A.G., Khropov A.N. Optimal redundancy as a method for improving the reliability of space communication systems. In: *Fundamental Problems of Radioengineering and Device Construction*. *Proceedings of the International Scientific and Technical Conference "INTERMATIC 2010."* 2010;10(3):109–112 (in Russ.).
15. Gelfman T.E., Pirhavka A.P. Optimal redundancy of a communications satellite transponder. In: *Proceedings of the International conference "Radio Electronic Devices and Systems for Infocommunication Technologies," REDS 2015*. Moscow: 2015. Issue LXX. P. 217–220 (in Russ.).

12. Bouwmeester J., Menicucci A., Gill E.K.A. Improving CubeSat reliability: Subsystem redundancy or improved testing? *Reliab. Eng. Syst. Saf.* 2022;220:108288. https://doi.org/10.1016/j.ress.2021.108288
13. Castet J.F., Saleh J.H. Satellite and satellite subsystems reliability: Statistical data analysis and modeling. *Reliab. Eng. Syst. Saf*. 2009;94(11):1718–1728. https://doi.org/10.1016/j.ress.2009.05.004
14. Кислаев А.Г., Хропов А.Н. Оптимальное резервирование, как метод повышения надежности систем космической связи. В сб.: *Фундаментальные проблемы радиоэлектронного приборостроения*. *Материалы Международной научно-технической конференции «INTERMATIC – 2010»*. 2010;10(1–3):109–112.
15. Гельфман Т.Э., Пирхавка А.П. Оптимальное резервирование ретранслятора спутника связи. В сб.: *Труды международной конференции «Радиоэлектронные устройства и системы для инфокоммуникационных технологий», РЭУС-2015*. М.: 2015. Вып. LXX. С. 217–220.

About the authors

Tatyana E. Gelfman, Associate Professor, Department of Radio Electronic Systems and Complexes, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: gelfman@mirea.ru. https://orcid.org/0000-0002-8302-6999

Alexey P. Pirkhavka, Cand. Sci. (Eng.), Associate Professor, Department of Radio Electronic Systems and Complexes, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: pirkhavka@mirea.ru. https://orcid.org/0000-0002-2460-7507

Vladimir O. Skripachev, Cand. Sci. (Eng.), Head of the Department of Radiomonitoring and Remote Sensing of Earth, Associate Professor, Department of Radio Electronic Systems and Complexes, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: skripachev@mirea.ru. Scopus Author ID 35105951000, ResearcherID M-9770-2015, RSCI SPIN-code 8322-7260, https://orcid.org/0000-0003-3538-5943

Об авторах

Гельфман Татьяна Элевна, доцент кафедры радиоэлектронных систем и комплексов Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: gelfman@mirea.ru. https://orcid.org/0000-0002-8302-6999

Пирхавка Алексей Петрович, к.т.н., доцент кафедры радиоэлектронных систем и комплексов Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: pirkhavka@mirea.ru. https://orcid.org/0000-0002-2460-7507

Скрипачев Владимир Олегович, к.т.н., начальник отдела радиомониторинга и дистанционного зондирования Земли, доцент кафедры радиоэлектронных систем и комплексов Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: skripachev@mirea.ru. Scopus Author ID 35105951000, ResearcherID M-9770-2015, SPIN-код РИНЦ 8322-7260, https://orcid.org/0000-0003-3538-5943

Translated from Russian into English by Evgenii I. Shklovskii
Edited for English language and spelling by Thomas A. Beavitt

Russian Technological Journal. 2023;11(1):60–69 ISSN 2500-316X (Online)

Mathematical modeling

Математическое моделирование

UDC 004.021: 65.012.26
https://doi.org/10.32362/2500-316X-2023-11-1-60-69

RESEARCH ARTICLE

Algorithm for finding subcritical paths on network diagrams

Mikhail A. Anfyorov @

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

Abstract
Objectives. Network diagrams are used as an information support element in planning and project management processes for structuring planned work and calculating project efficiency characteristics. In order to optimize and balance resources used in projects, it becomes necessary to locate in these models not only the critical path of the maximum weighted length, but also the subcritical paths closest to it having a shorter length in relation to it. The aim of the work is to synthesize and analyze an algorithm for finding *k*-shortest paths between the input and output network vertices, on which basis the above-mentioned subcritical paths can be identified.
Methods. The provisions of graph theory and group theory, as well as the method of dynamic programming, were used.
Results. An algorithm for finding *k*-shortest paths in contourless directed graphs having a strict order relation was developed. Abstract elements were defined according to group theory in graphs as *p*-contours, between which a multilevel structure of relations for implementing the necessary search of paths was then established. For substantiating the efficiency of the constructed algorithm, the validity of the main provisions was demonstrated as follows: firstly, the multilevel system of relations is exhaustive; secondly, there is no loss in the final solution during the operation of the algorithm; thirdly, the paths obtained as a result of the work of the algorithm satisfy the main required relation between them. Numerically, the algorithm was implemented by the dynamic programming method extended by means of an additional functional relationship, implying the presence of suboptimal policies.
Conclusions. The conducted runs of computational experiments confirmed the operability and efficiency of the software-implemented algorithm. The performed analysis demonstrated the good convergence characteristics of the proposed algorithm as compared with other algorithms of this class applied to network diagrams. On this basis, it can be recommended for practical use in project management information systems.

Keywords: project management, network diagram, critical path, algorithm, computational experiment

• **Submitted:** 27.03.2022 • **Revised:** 06.06.2022 • **Accepted:** 24.10.2022

For citation: Anfyorov M.A. Algorithm for finding subcritical paths on network diagrams. *Russ. Technol. J.* 2023;11(1):60–69. https://doi.org/10.32362/2500-316X-2023-11-1-60-69

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

Алгоритм поиска подкритических путей на сетевых графиках

М.А. Анфёров @

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

Резюме
Цели. Информационная поддержка процессов планирования и управления проектами использует в качестве модели сетевые графики, помогающие в формировании структуры планируемых работ и расчете характеристик эффективности проекта. С целью оптимизации и выравнивания ресурсов, используемых в проектах, возникает необходимость нахождения на этих моделях не только критического пути максимальной взвешенной длины, но и ближайших к нему подкритических путей с меньшей по отношению к нему длиной. Цель работы – синтез и анализ алгоритма поиска *k*-кратчайших путей между вершинами входа и выхода сети, позволяющего идентифицировать вышеназванные подкритические пути.
Методы. Использованы положения теории графов и теории групп, а также метод динамического программирования.
Результаты. Разработан алгоритм поиска *k*-кратчайших путей на ориентированных графах без контуров с отношением строгого порядка. С использованием теории групп на графах были определены абстрактные элементы – *p*-контуры, между которыми была установлена многоуровневая структура отношений, позволившая реализовывать необходимый поиск путей. В рамках обоснования работоспособности построенного алгоритма доказана справедливость основных положений: во-первых, многоуровневая система отношений является исчерпывающей; во-вторых, не происходит потерь в окончательном решении в процессе работы алгоритма; в-третьих, пути, найденные в результате работы алгоритма, удовлетворяют основному требуемому соотношению между ними. Численно алгоритм реализован методом динамического программирования, который был расширен за счет использования дополнительного функционального соотношения, предполагающего наличие подоптимальных политик.
Выводы. Проведенная серия вычислительных экспериментов подтвердила работоспособность и эффективность программно реализованного алгоритма. Выполненный анализ показал хорошие характеристики сходимости предложенного алгоритма в сравнении с алгоритмами данного класса, применяемыми к сетевым графикам. Это позволяет рекомендовать его к практическому использованию в информационных системах управления проектами.

Ключевые слова: управление проектами, сетевой график, критический путь, алгоритм, вычислительный эксперимент

• **Поступила:** 27.03.2022 • **Доработана:** 06.06.2022 • **Принята к опубликованию:** 24.10.2022

Для цитирования: Анфёров М.А. Алгоритм поиска подкритических путей на сетевых графиках. *Russ. Technol. J.* 2023;11(1):60–69. https://doi.org/10.32362/2500-316X-2023-11-1-60-69

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

Автор заявляет об отсутствии конфликта интересов.

INTRODUCTION

Established project management methodologies use a contourless network-oriented graph as the main model for displaying the structure of the mutual dependence of project work stages and calculating the time characteristics of these stages, as well as those of the project as a whole [1–3]. In the terminology of network planning and management, such a model is referred to as a network diagram, which may be constructed according to one of two principles. In the first case, typical for the Project Evaluation and Review Technique (PERT) methodology, the work in the model display uses graph arcs to represent the “event-work” principle, while the second uses graph vertices according to the “work-relationship” principle. However, both of these approaches involve the definition of a critical path on the network diagram, having a maximum weighted length and connecting a hanging vertex to a dead end, which may correspond to the beginning or end of the project. In this case, the length is calculated based on the value of the critical work execution time displayed by the active elements of the path (arcs or vertices). The use of a critical path is an integral part of network planning and project management methodology, as reflected in studies at the stage of its theoretical development (an extensive bibliography of this period is presented in [4]) and subsequent development [5–8].

Actual projects are implemented in various areas under conditions of limited material, labor and financial resources. This implies the development of project management theory in the direction of optimizing the structure of the project according to the criterion of its minimum cost under particular resource constraints [7–12]. In this case, the question arises of the need to search the network diagram not only for critical, but also for subcritical paths in order to implement the above-mentioned optimization and leveling of resources. If the critical path Ω_0 can be defined as

$$\Omega_0 : L(\Omega_0) = \max_{\Omega_j \in \mathbf{\Omega}} \{L(\Omega_j)\}, \tag{1}$$

where $L(\Omega_0)$ is the length of the critical path, and $\mathbf{\Omega}$ is the set of all full paths on the network diagram, then the subcritical paths will form an ordered set of full paths of the graph $\{\Omega_1, \Omega_2, ...\}$, characterized as follows:

$$L(\Omega_0) \geq L(\Omega_1);\ L(\Omega_j) \geq L(\Omega_{j+1}). \tag{2}$$

To find the critical path, it is sufficient to use any algorithm to find the shortest weighted path between the vertices that represent the beginning and end points of the project on the network diagram after changing the weights of its active elements (time to complete the work) to negative. Such algorithms can be based on dynamic programming [13, 14] or heuristic methods [15]. When obtaining subcritical paths, algorithms for finding k-shortest paths between graph vertices either use dynamic programming directly [16–22] or with repeated application of the shortest path search [23–25], or other contemporary approaches [26–28], as well as those focused on more complex network designs [29–32].

The present paper describes an algorithm for finding k-shortest paths on network graphs based on the selection of elements of a higher order and structuring relations between them.

STRUCTURING RELATIONSHIPS ON GRAPHS

The performed studies are focused on the use of network graphs of the "works-connections" type, using as a model network directed graphs **G**(**X**, **U**) (**X** is a set of vertices, **U** is a set of arcs) without contours with a strict order relation, and with division into layers [33] (Fig. 1).

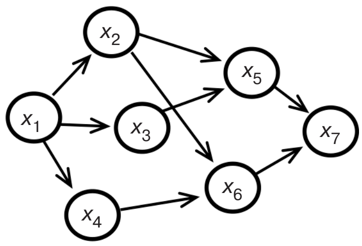


Fig. 1. Network diagram

Let us construct a free Abelian group **P**(**U**) over the generating set **U** of all arcs of the graph **G** = (**X**, **U**) as follows. As elements of **P**(**U**) we will consider the set of formal linear combinations of elements from **U** with integer coefficients in the form of

$$p_j = \sum_{i=1}^{n} (\gamma_i \cdot u_i),\ p_j \in \mathbf{P}(\mathbf{U});\ u_i \in \mathbf{U};\ \gamma_i = 0, \pm 1, \ldots,$$

where n is the number of graph arcs.

As a binary additive operation, we define the sum of elements from the set **P**(**U**) by the formula:

$$p_j = \sum_{i=1}^{n} (\gamma_i \cdot u_i) + \sum_{i=1}^{n} (\gamma'_i \cdot u_i) = \sum_{i=1}^{n} (\gamma_i + \gamma'_i) \cdot u_i.$$

Similarly, we construct a group **H**(**X**) over the set **X** of all vertices of the graph, defining its elements as $h_j = \sum_{i=1}^{m} (\gamma_i \cdot x_i),\ h_j \in \mathbf{H}(\mathbf{X});\ x_i \in \mathbf{X};\ \gamma_i = 0, \pm 1, \ldots,$ where m is the number of vertices. In what follows, when writing elements of the groups **P**(**U**) and **H**(**X**), we will omit their constituent elements that have zero coefficients.

Definition 1. A differential d of a group **P**(**U**) is called a homomorphism d: **P**(**U**) → **H**(**X**) defined as follows:

1) if $u_t = (x_i, x_j)$, then $du_t = x_j - x_i$;
2) if $p_q = \Sigma\gamma_t u_t$, then $dp_q = \Sigma\gamma_t du_t$

for $p_q \in$ **P**(**U**); $u_t \in$ **U**; $x_i, x_j \in$ **X**.

Definition 2. An element $r \in$ **P**(**U**) is called a p-contour if $dr = 0$. The subgroup **R** = *Ker d* ⊂ **P**(**U**) is called a subgroup of p-contours.

Lemma 1. The sum of several p-contours is a p-contour.

Proof. First, the sum Σr_i is an element of the group **P(U)** as a result of addition of its elements. Second, the distributivity of the mapping d according to *Definition 1* allows us to write down $d\Sigma r_i = \Sigma dr_i = 0$. The assertion of the lemma is proved.

Using the generally accepted concept of a path on a graph as a connected finite sequence of arcs, we will denote it as

$$\Omega = \{u_1, u_2, \ldots, u_w\}. \quad (3)$$

Moreover, if $u = (x_1, x_2)$, то $x_1 = u^-$, $x_2 = u^+$. We will determine the length of this path through the above numerical weighted estimates of the vertices $\varepsilon(x)$ by the expression

$$L(\Omega) = \varepsilon(u_1^-) + \sum_{i=1}^{w} \varepsilon(u_1^+). \quad (4)$$

If the path connects the input vertex x' with the output vertex x'' of the network, then we will call such a path complete.

We define the mapping φ as follows:

$$\varphi(\{\gamma_1 u_1, \gamma_2 u_2, \ldots, \gamma_n u_n\}) = \sum_{i=1}^{n} \gamma_i u_i, \gamma_i > 0, \quad (5)$$

and also, the converse to the above:

$$\varphi'\left(\sum_{i=1}^{n} \gamma_i u_i\right) = \{|\gamma_1| u_1, |\gamma_2| u_2, \ldots, |\gamma_n| u_n\}. \quad (6)$$

Definition 3. The image of a path $\Omega \subset \mathbf{U}$ on the graph $\mathbf{G} = (\mathbf{X}, \mathbf{U})$ is called a mapping $\varphi(\Omega)$.

Lemma 2. The differential of the image of any complete path on a network graph is defined as $d\varphi(\Omega) = x'' - x'$.

Proof. Since the full path Ω on the graphs under consideration is a simple path that does not have multiple arcs, then its image, taking into account (3), can be written as $\varphi(\Omega) = u_1 + u_2 + \ldots + u_w$. In accordance with *Definition 1*, we obtain either $d\varphi(\Omega) = du_1 + du_2 + \ldots + du_w$ or $d\varphi(\Omega) = (u_1^+ - u_1^-) + (u_2^+ - u_2^-) + \ldots + (u_w^+ - u_w^-)$, or

$$d\varphi(\Omega) = -u_1^- + (u_1^+ - u_2^-) + \ldots + (u_{w-1}^+ - u_w^-) + u_w^+. \quad (7)$$

The incidence property of path arcs implies $u_i^+ = u_{i+1}^-, i \in [1, w-1]$. Therefore, expression (7) can be written in the form $d\varphi(\Omega) = u_w^+ - u_1^-$ or $d\varphi(\Omega) = x'' - x'$.

Lemma 3. The difference between the images of two complete paths on a network graph is a *p*-contour.

Proof. The difference between the images of the full paths $\varphi(\Omega_i)$ and $\varphi(\Omega_j)$, being the result of the additive function of adding two elements of the group **P**(**U**), also belongs to this group. On the other hand, the differential of this difference is defined as $d[\varphi(\Omega_i) - \varphi(\Omega_j)] = d\varphi(\Omega_i) - d\varphi(\Omega_j)$. However, by the assertion of *Lemma 2* $d\varphi(\Omega_i) = d\varphi(\Omega_j)$. Hence, $d[\varphi(\Omega_i) - \varphi(\Omega_j)] = 0$, which proves the lemma.

Let us define the mapping α as follows:

$$\alpha(p) = \sum_{u_i \in p}^{n} \gamma_i \cdot \varepsilon(u_i^+). \quad (8)$$

Definition 4. The value of the *p*-contour *r* is the value $\alpha(r)$.

Lemma 4. The difference between the lengths of two complete paths on a network graph is equal to the value *p* of the contour formed by the difference of their images.

Proof. Let $\Omega_i = \{u_{i1}, u_{i2}, \ldots, u_{iw}\}$ and $\Omega_j = \{u_{j1}, u_{j2}, \ldots, u_{js}\}$ be any two complete paths of the graph.

Then their lengths, in accordance with (4), can be written as $L(\Omega_i) = \varepsilon(x') + \sum_{t=1}^{w} \varepsilon(u_{it}^+)$, $L(\Omega_j) = \varepsilon(x') + \sum_{t=1}^{s} \varepsilon(u_{jt}^+)$, and the difference in lengths as

$$L(\Omega_i) - L(\Omega_j) = \sum_{t=1}^{w} \varepsilon(u_{it}^+) - \sum_{t=1}^{s} \varepsilon(u_{jt}^+). \quad (9)$$

On the other hand, in accordance with *Lemma 3*, the *p*-contour defined by the images of these full paths can be written, taking into account expression (5), as $r = \varphi(\Omega_i) - \varphi(\Omega_j) = u_{i1} + u_{i2} + \ldots + u_{iw} - u_{j1} - u_{j2} - \ldots - u_{js}$, and the value of this *p*-contour, taking into account (8), as

$$\alpha(r) = \sum_{t=1}^{w} \varepsilon(u_{it}^+) - \sum_{t=1}^{s} \varepsilon(u_{jt}^+). \quad (10)$$

Comparing expressions (9) and (10), we conclude that the lemma is true.

Definition 5. An elementary *p*-contour $g^{(ab)}$ with respect to an arc of a graph $u = (a, b) \in \mathbf{U}$ is an element of the group **P(U)**, defined as the sum of this arc with the difference between the images of the shortest weighted paths connecting the vertices *x'* and *a*, as well as the vertices *x'* and *b* (see Fig. 2). In this case, the elementary *p*-contours are oriented to the shortest paths, since the main goal is to find *k*-shortest paths of the graph.

Definition 6. An arc u_i is said to be incident to the path Ω_0 at the vertex *b* if the following conditions are met:

1) $u_i \notin \Omega_0$;
2) $\exists u_i \in \Omega_0, u_i^+ = u_j^+ = b$.

The vertex *b* is called the vertex of the section of the path Ω_0 (Fig. 2).

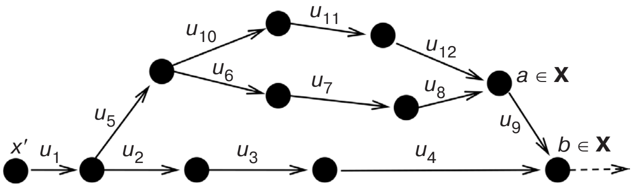


Fig. 2. $\{u_1, u_5, u_6, u_7, u_8\}$ – shortest path between vertices x' and a.
$\{u_1, u_2, u_3, u_8\}$ – shortest path between vertices x' and b.
$\{u_5 + u_6 + u_7 + u_8 + u_9 - u_2 - u_3 - u_4\}$ – elementary p-contour ($g^{(ab)}$), based along the path $\{u_1, u_2, u_3, u_4, \ldots\}$ along the incident arc u_9 at the vertex of the section b: $g^{(ab)} \equiv r^{bi}$

Definition 7. An elementary p-contour with respect to an arc incident to the path Ω at the vertex of the section $b \in \mathbf{X}$ is called based on this path and is denoted as r^b (Fig. 2)[1].

In connection with the consideration of the full paths of the graph in the framework of the network planning problem, the p-contour based on the path Ω_0, which allows us to determine another full path Ω_s, will be called generating. Based on lemmas 3 and 4, we write the relations

$$\varphi(\Omega') = \varphi(\Omega_0) + r^b,\ L(\Omega') = L(\Omega_0) + \alpha(r^b), \quad (11)$$

which are further used in the algorithm for finding k-shortest paths of the graph.

In order to use the models to efficiently solve numerical problems, an ordered structure of relations between the complete paths of the network graph can be constructed on the basis of the introduced mathematical objects.

With regard to the development of the described algorithm, we consider the problem of finding an ordered set of full paths

$$\{\Omega_0, \Omega_1, \ldots, \Omega_k\},\ L(\Omega_i) \le L(\Omega_{i+1}),\ i \le [0, k-1]. \quad (12)$$

The system of relationships between paths is built relative to the full path Ω_0, which has a minimum length $\Omega_0 : L(\Omega_0) = \min_{\Omega_j \in \mathbf{\Omega}} \{L(\Omega_j)\}$, can be easily obtained using one of the well-known algorithms, for example as given in [13, 14]. This system is described by a hierarchical multilevel structure of shortest path Ω_0 generating p-contours in the form of a graph $\mathbf{G} = (\mathbf{R}_0, \mathbf{V})$, $\mathbf{V} = \mathbf{R}_0 \times \mathbf{R}_0$. Here $\mathbf{R}_0$ is the complete set of generating p-contours, while $\mathbf{V}$ is the set of relations between them that connect the generating p-contours of adjacent levels. So, for the upper 0th level, the connection of p-contours[2] $\mathbf{R}_{00} = \{r_0^b\}$ with p-contours of the next 1st level $\mathbf{R}_{01}$ is represented by the relation

$$r_1^{be} = r_0^b + r^e;\ \alpha(r_1^{be}) = \alpha(r_0^b) + \alpha(r^e), \quad (13)$$

where b is the vertex of the section of the path Ω_0; e is the vertex of the section of the path formed by the p-contour r_0^b (11) located on this path to the left of the vertex b (i.e., $e < b$)[3]. The grouping by levels defines the entire set of generating p-contours $\mathbf{R}_0 = \bigcup_Z \mathbf{R}_{0Z}$.

DESCRIPTION OF THE ALGORITHM

The algorithm is presented below in a less compact form, excluding loops, in order to visually show the finiteness of the number of steps performed.

Step 1. Find the full path Ω_0 with the minimum length.

Step 2. Put the set $\mathbf{R}_{00} = \{r_1, r_2, r_3, \ldots\}$ of generating p-contours of the 0th level in ascending order of their values[4] and exclude from further consideration the p-contours that are below the kth place in the resulting sequence; $\bar{\mathbf{R}} = \mathbf{R}_{00}$. Find the full path Ω_1 through relations (11) using the p-contour r_1, which is the first in the sequence.

Step 3. Include in the set $\bar{\mathbf{R}}$ the p-contours of the next level, determined by the p-contour r_1 through relations (13) over all vertices of the section, excluding the p-contour r_1 itself. Order the set $\bar{\mathbf{R}}$ according to the relation $\alpha(r_i) \le \alpha(r_{i+1})$ and exclude from further consideration the p-contours that are below $(k - 1)$ places in the resulting sequence.

Find the full path Ω_2 through relations (11) using the p-contour r_1.

. . . .

Step(s). Include in the set $\bar{\mathbf{R}}$ p-contours of the next level, determined by the p-contour r_1 through relations (13) over all vertices of the section, excluding the p-contour r_1 itself. Order the set $\bar{\mathbf{R}}$ according to the relation $\alpha(r_i) \le \alpha(r_{i+1})$ and exclude from further consideration p-contours that are in the resulting sequence below $[k - (s - 2)]$ place[5].

[1] Several elementary contours can be based on the vertex b in the presence of several incident arcs.

[2] An additional subscript is introduced to denote the level number in the system of generating p-contours.

[3] This condition is related to the orientation to the input vertex x' when constructing elementary p-contours (see *Definition 5*).

[4] A simplified designation of generating contours was introduced in order to better understand the operation of the algorithm. The subscript shows the place of the contour in the ordered set.

[5] The particular case is when the number of elements of the set $\mathbf{R}_0$ is less than $[k - (i - 2)]$. In this case, contours are not excluded from the set, which does not affect the course of all reasoning and the final result.

Find the full path Ω_{s-1} through relations (11) using the p-contour r_1.

. . . .

Step (k + 1). Include in the set $\bar{\mathbf{R}}$ p-contours of the next level, which are determined by the p-contour r_1 through relations (13) over all vertices of the section, excluding the p-contour r_1 itself. Order the set $\bar{\mathbf{R}}$ according to the relation $\alpha(r_i) \le \alpha(r_{i+1})$ and exclude from further consideration the p-contours that are below $[k - (k - 1)] = 1$ place in the resulting sequence.

Find the full path Ω_k through relations (11) using the p-contour r_1.

The implementation of the algorithm is based on the search for generating p-contours and the full path Ω_0, which has a minimum length. This problem is effectively solved by the dynamic programming method. The main Bellman recursive functional relation for the problem of finding the shortest path[6] $\Omega^{1t}_{\min}$ connecting the input vertex (x_1) with any vertex x_t can be written as

$$L(\Omega^{1t}_{\min}) = \min_{i} \{\varepsilon(x_t) + L(\Omega^{1i}_{\min})\},\ i \in \mathbf{I}_t, \quad (14)$$

where $\mathbf{I}_t$ is a subset of graph vertex numbers defined by the condition $\forall x_i$ ($i \in \mathbf{I}_t$) $\exists u_y$ such that $x_i = u_y^-,\ x_t = u_y^+$. An element from $\mathbf{I}_t$, which is the optimal policy defined by (14), will be denoted as i_o.

To search for generating p-contours based on the vertex of the section t, one more functional relation must be defined in the form

$$L(\Omega^{1t}_j) = \varepsilon(x_t) + L(\Omega^{1j}_{\min}),\ j \in \mathbf{J}_t,\ \mathbf{J}_t = \mathbf{I}_t / i_o, \quad (15)$$

where Ω^{1t}_j is the path connecting vertices x_1 and x_t and passing through vertex x_j (Fig. 3).

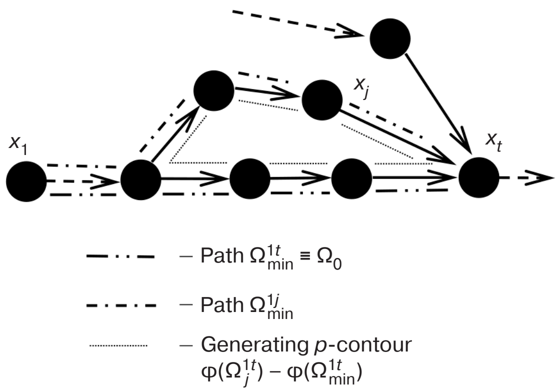


Fig. 3. Graphical explanation for the above calculations

Relation (15) implies the presence of suboptimal policies in the dynamic programming method, whose presence demonstrated by R. Bellman [34]. This relation defines the set of incident arcs $\{(x_j, x_t)\}$, $j \in \mathbf{J}_t$ $\{(x_j, x_t)\}$ at the vertex of the section x_t and the corresponding generating p-contours through the knowledge of paths $\Omega^{1j}_{\min}$ (see *Definition 5*).

A software implementation of the algorithm in the Delphi programming language was performed. The conducted testing of the program confirmed the efficiency and effectiveness of the presented algorithm.

JUSTIFICATION OF THE ALGORITHM

To prove the efficiency of the constructed algorithm, it is necessary to verify the validity of the following provisions:

- firstly, any generating p-contour of the complete path Ω_0 is described by the system $\mathbf{G} = (\mathbf{R}_0, \mathbf{V})$;
- secondly, there is no loss in the final solution during the operation of the algorithm;
- thirdly, the paths found as a result of the algorithm's operation meet the condition (12).

The validity of the first proposition is confirmed by the following theorem.

Theorem 1. Any generating p-contour of the complete path Ω_0 is described by the system $\mathbf{G} = (\mathbf{R}_0, \mathbf{V})$.

Proof. Let r_x be an arbitrarily chosen generating p-contour of the path Ω_0. Let us show that this p-contour is described by one of the levels of the system $\mathbf{G} = (\mathbf{R}_0, \mathbf{V})$.

There necessarily exists an arc u_0 incident to the path Ω_0 with the section vertex b_0 such that $u_0 \in \varphi'(r_x)$. This arc corresponds to an elementary p-contour $g^{u_0} = r^{b_0}$ based on Ω_0.

If $r^{b_0} = r_x$, then the p-contour r_x is described by the system $\mathbf{G} = (\mathbf{R}_0, \mathbf{V})$ at the 0th level. If the equality is not maintained, then $\alpha(r^{b_0}) \le \alpha(r_x)$, which follows from *Definition 5*; therefore, the inequality $L(\Omega_1) \le L(\Omega_x)$ is true, where Ω_1 and Ω_x are full paths defined by p-contours r^{b_0} and r_x through relations (11), i.e.

$$\varphi(\Omega_1) = \varphi(\Omega_0) + r^{b_0},\ L(\Omega_1) = L(\Omega_0) + \alpha(r^{b_0}), \quad (16)$$

$$\varphi(\Omega_x) = \varphi(\Omega_0) + r_x,\ L(\Omega_x) = L(\Omega_0) + \alpha(r_x). \quad (17)$$

Subtracting equalities (16) and (17) term by term, and having transformed the result, we obtain $\varphi(\Omega_x) = \varphi(\Omega_1) + (r_x - r^{b_0})$, where $(r_x - r^{b_0})$ is the generating p-contour of the path Ω_1. Moreover, its value (obtained by the same subtraction of equalities (16) and (17)) is equal to $\alpha(r_x) - \alpha(r^{b_0}) \ge 0$.

Since $u_0 \in \Omega_1$, then the vertex of the section b_1 of the p-contour $(r_x - r^{b_0})$ is to the left of b_0.

Let us go to the next similar step of calculations. There is an arc u_1 incident to the path Ω_1 with the section

[6] The superscripts show the numbers of connected vertices.

vertex b_1 such that $u_1 \in \varphi'(r_x - r^{b_0})$. This arc corresponds to an elementary p-contour $g^{u_1} = r^{b_1}$ based on Ω_0.

If $r^{b_1} = r_x - r^{b_0}$, then p-contour $r_x = r^{b_0} + r^{b_1}$, i.e. is described by the system $\mathbf{G} = (\mathbf{R}_0, \mathbf{V})$ at the 1st level. If the equality is not maintained, then $\alpha(r^{b_1}) \le \alpha(r_x - r^{b_0})$ or $\alpha(r_x) \ge \alpha(r^{b_0}) + \alpha(r^{b_1})$, and, hence, $L(\Omega_2) \le L(\Omega_x)$, where Ω_2 is the full path determined by the p-contour r^{b_1} through relations (11) as $\varphi(\Omega_2) = \varphi(\Omega_1) + r^{b_1}$ or taking into account (16)

$$\varphi(\Omega_2) = \varphi(\Omega_0) + r^{b_0} + r^{b_1},\ L(\Omega_2) =$$
$$= L(\Omega_0) + \alpha(r^{b_0} + r^{b_1}). \tag{18}$$

As a result, by analogy, there is a p-contour of the path Ω_2 generating p in the form

$$(r_x - r^{b_0} - r^{b_1}),\ \alpha(r_x - r^{b_0}) - \alpha(r^{b_1}) \ge 0.$$

Thus, at any Nth step calculations, the non-fulfillment of the equality condition

$$r^{b_N} = r_x - r^{b_0} - r^{b_1} - \ldots - r^{b_{N-1}} \tag{19}$$

leads to the next step. However, since the number of steps is limited, therefore, at a certain step, equality (19) will be maintained (i.e., r^{b_N} will be an elementary p-contour). This means that the p-contour r_x of the path Ω_0 will be described by the system $\mathbf{G} = (\mathbf{R}_0, \mathbf{V})$ at the $(N - 1)$ level.

The limited number of steps is confirmed by the following considerations. Firstly, the incident arc chosen at any step belongs to the path Ω_x; secondly, as noted above, the section vertex is located to the left of that chosen at the previous step. Since the number of arcs of the path Ω_x is limited, the number of steps in the main reasoning is also limited.

Since the p-contour r_x was chosen arbitrarily, the theorem can be considered to have been proved.

To validate the second statement, we introduce into consideration the vector $\mathbf{Q}_j = [\alpha(r_{1j}), \alpha(r_{2j}), \ldots, \alpha(r_{(k+1-j)j})]$, whose elements are the values of p-contours that make up the set $\bar{\mathbf{R}}$ when the algorithm works at the stage of determining the path Ω_j (the number of the path determines the second subscript in the designation of these p-contours). Recall that the elements of the vector $\mathbf{Q}_j$ form a non-decreasing sequence.

We also introduce a vector $\mathbf{W} = [\alpha(r_{k1}), \alpha(r_{(k-1)2}), \ldots, \alpha(r_{1k})]$ (here, the semantic meaning of the indices corresponds to the vector $\mathbf{Q}_j$), whose elements $\alpha(r_{(k+1-j)j})$ are the values of p-contours that are at the last place in the set $\bar{\mathbf{R}}$ at the stage of determining the path Ω_j.

Theorem 2. The elements of the vector $\mathbf{W}$ form a non-increasing sequence.

Proof. Consider an arbitrarily chosen jth element of the vector $\mathbf{W}$, i.e., $\alpha(r_{(k+1-j)j})$, which is also the last element of the vector $\mathbf{Q}_j$. We are interested in the process of transition during the operation of the algorithm from the vector $\mathbf{Q}_j$ to the vector $\mathbf{Q}_{j+1}$ looking to the appearance of a new element $\alpha(r_{(k-j)(j+1)})$ of the vector $\mathbf{W}$. It consists of the following stages:

- the first element of the vector $\mathbf{Q}_j$ is excluded;
- new elements are added to the remaining elements, corresponding to the lower level p-contours included in the consideration;
- ordering of the resulting set;
- exclusion from further consideration of extra p-contours.

Exclusion of the first element of the vector $\mathbf{Q}_j$ cannot affect the choice of the element $\alpha(r_{(k-j)(j+1)})$.

The result of the procedures that should follow will not change if they are performed in a slightly different order. To the remaining elements of the vector $\mathbf{Q}_j$ (their number now corresponds to the required number of elements of the vector $\mathbf{Q}_{j+1}$), we will add one element of the new set of p-contour values. After each such addition, we will arrange the set $\bar{\mathbf{R}}$ and eliminate the extra element.

The elements of the vector $\mathbf{Q}_{j+1}$ before adding new elements are determined as follows: $\alpha(r_{1(j+1)}) = \alpha(r_{2j})$, $\alpha(r_{2(j+1)}) = \alpha(r_{3j})$, ..., $\alpha(r_{(k-j)(j+1)}) = \alpha(r_{(k+1-j)j})$. For the added element $\alpha(r_y)$, there are two possible cases: either $\alpha(r_y) \ge \alpha(r_{(k+1-j)j})$, or $\alpha(r_{(t+1)j}) \ge \alpha(r_y) \ge \alpha(r_{tj})$, $t \in [2, k-j]$. In the first case, it is the element $\alpha(r_y)$ that is excluded from further consideration, which will not change the ordered position of the remaining elements of the vector $\mathbf{Q}_j$. In the second case, following ordering, the element $\alpha(r_y)$ will take the place of the element $\alpha(r_{(t+1)j})$, the element $\alpha(r_{(t+1)j})$ will replace element $\alpha(r_{(t+2)j})$, etc.

Ultimately, the element $\alpha(r_{(k-j)j})$ will take the place of the element $\alpha(r_{(k+1-j)j})$ that will be removed. But, since $\alpha(r_{(k-j)j}) \le \alpha(r_{(k+1-j)j})$ (due to the definition of the vector $\mathbf{Q}_j$), then in this case there will be no increase in the value of the last element either, i.e., the element $\alpha(r_{(k-j)(j+1)})$ cannot be greater than the element $\alpha(r_{(k+1-j)j})$.

Since an arbitrary element of the vector $\mathbf{W}$ was chosen as the jth element, *the theorem can be considered to have been proved.*

Using *Theorem 2*, it is easy to prove that there are no losses in the final solution during the operation of the algorithm. Indeed, losses can occur only in the procedure of excluding p-contours from further consideration after they are ordered at each step[7]. However, at any step of

[7] The exclusion of the contour that is the first does not lead to losses, since on the one hand it is used to obtain a solution, and on the other hand, all lower-level contours generated by it fall into consideration.

the algorithm, the values of the eliminated p-contours are not less than the element $\alpha(r_{(k+1-j)j})$ of the vector $\mathbf{Q}_j$. Based on *Theorem 2*, we can conclude that these values are also certainly not less than the subsequent elements of the vector $\mathbf{W}$: $\alpha(r_{(k-j)(j+1)})$, $\alpha(r_{(k-j-1)(j+2)})$, ..., $\alpha(r_{1k})$, i.e., they should not be considered at the remaining steps of the algorithm and cannot be included in the final decision.

To validate the last statement, we introduce into consideration a vector consisting of the first elements of the vectors $\mathbf{Q}_j$, $j \in [1, k]$: $\bar{\mathbf{W}} = [\alpha(r_{11}), \alpha(r_{12}), ..., \alpha(r_{1k})]$, participating in the formation of the final solution through relations (11). Then the validity of condition (12) is affirmed by the following theorem.

Theorem 3. Vector elements $\bar{\mathbf{W}}$ form a non-decreasing sequence.

Proof. Consider an arbitrarily chosen jth element $\alpha(r_{1j})$ of the vector $\bar{\mathbf{W}}$. This element is also the first element of the vector $\mathbf{Q}_j$. In the process of transition from the vector $\mathbf{Q}_j$ to the vector $\mathbf{Q}_{j+1}$, the added new set of p-contours is characterized by their values certainly being not less than $\alpha(r_{1j})$; this is because they are at the next level after the p-contour and are determined by this p-contour through relations (13).

After ordering the set $\bar{\mathbf{R}}$, the element $\alpha(r_{1(j+1)})$ will be replaced by the element $\alpha(r_{2j})$ or by the smallest from the added set $\alpha_{\min}(r_{\text{add}})$, i.e., $\alpha(r_{1(j+1)})$ = $= \min\{\alpha(r_{2j}), \alpha_{\min}(r_{\text{add}})\}$. But since $\alpha(r_{2j}) \geq \alpha(r_{1j})$, $\alpha_{\min}(r_{\text{add}}) \geq \alpha(r_{1j})$, then $\alpha(r_{1(j+1)}) \geq \alpha(r_{1j})$.

Since an arbitrary element of the vector $\bar{\mathbf{W}}$ was chosen as the jth one, the last inequality *proves the theorem.*

The performance of the algorithm was evaluated in comparison with the "double sweep" algorithm, according to which its author conducted serious computational experiments [35]. Moreover, the analytical performance assessment for algorithms of this category is far from real results, since the computation time strongly depends on the configuration of the networks used; moreover, the generalized operations performed during the execution of the algorithm cannot be unambiguously mapped into elementary operations of addition and comparison. Thus, for the mentioned algorithm, which belongs to the class of the most productive algorithms of this category, the analytical estimated computation time is of the order of $O(kN^3)$ [19, 20], while computational experiments show other results [35]; instead of a linear one, a polynomial dependence is observed of computation time τ on the number of shortest paths k:

$$\tau = 0.8457 + 0.1616\,k + 0.0260\,k^2.$$

At the same time, it should be noted that studies [35] were carried out not on network- but directed graphs, whose vertices formed a lattice structure with contours. The dimension of the graphs was taken into account through the dimensions and configuration of the lattice formed by the vertices. The weights of the arcs were generated by random integers in the range up to 100. In addition, the value of k was significantly limited.

Therefore, in connection with the search for subcritical paths on network graphs, a series of computational experiments was carried out on an Intel microprocessor with a clock frequency of 1.7 GHz in order to compare the algorithm obtained in this work with the Double sweep algorithm. The results of the dependence of the computation time (t) on the number of subcritical paths (k) and the dimensional characteristic of the network graphs (z) are shown in Fig. 4. The dimensional characteristic z is the product of the total number of graph arcs and the number of arcs in the path of maximum length connecting the input and output vertices of the network.

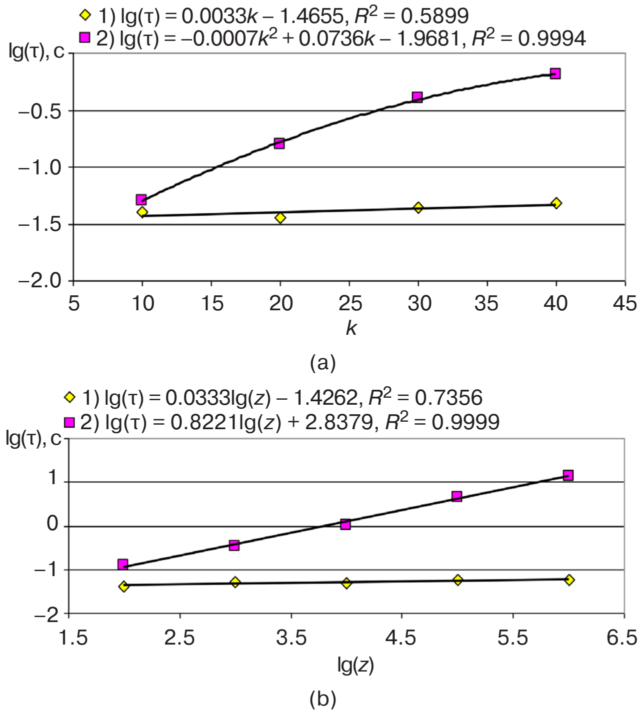


Fig. 4. Comparative analysis of algorithms: (a) $z = 1600$; (b) $k = 40$.
1 – Algorithm used in this work; 2 – Double sweep algorithm

The higher performance of the algorithm presented in the paper is explained by its special orientation to the considered class of network graphs, while other algorithms, including the Double sweep one, are more universal in relation to computed graphs.

CONCLUSIONS

The presented algorithm implements the search for k-shortest paths without contourless directed graphs having a strict order relation used as models (so-called network graphs) in network planning and project management problems. This can be used to find subcritical paths on these models in order to align and optimize the resources used in a project.

The above-described features of the graphs were instrumental in building a multilevel structure of relations between specially introduced abstractions (p-contours) that display the structural elements of graphs. This served as the basis for the development of the algorithm, which was numerically implemented by the dynamic programming method and extended through the use of an additional functional relation.

The narrow focus of the algorithm on graphs used in project management determined its high performance, which was confirmed by a series of comparative computational experiments. The efficiency of the algorithm is stable relative to the size of the computed graphs and the number of subcritical paths. For this reason, it can be recommended for practical use in project management information systems.

REFERENCES

1. Hagstrom J.N. Computing the probability distribution of project duration in a PERT network. *Networks*. 1990;20(2): 231–244. https://doi.org/10.1002/net.3230200208
2. Kamburowski J., Michael D.J., Stallmann M.F.M. Minimizing the complexity of an activity network. *Networks*. 2000;36(1):47–52. https://doi.org/10.1002/1097-0037(200008)36:1%3C47::AID-NET5%3E3.0.CO;2-Q
3. Bianco L., Caramia M. A new formulation of the resource-unconstrained project scheduling problem with generalized precedence relations to minimize the completion time. *Networks*. 2010;56(4):263–271. https://doi.org/10.1002/net.20388
4. Brennan M. *PERT and CPM: a selected bibliography.* Monticello, Ill.: Council of Planning Librarians, 1968. 11 p.
5. Alagheband A., Soukhakian M.A. An efficient algorithm for calculating the exact overall time distribution function of a project with uncertain task durations. *Indian Journal of Science and Technology*. 2012;5(9):3310–3316. https://doi.org/10.17485/ijst/2012/v5i9.20
6. Damci A., Polat G., Akin F.D., et al. Use of float consumption rate in resource leveling of construction projects. *Front. Eng. Manag.* 2022;9:135–147. https://doi.org/10.1007/s42524-020-0118-0
7. Willis R.J. Critical path analysis and resource constrained project scheduling – Theory and practice. *Eur. J. Oper. Res*. 1985;21(2):149–155. https://doi.org/10.1016/0377-2217(85)90026-8
8. Bowers J.A. Criticality in resource constrained networks. *J. Oper. Res. Soc*. 1995;46(1):80–91. https://doi.org/10.1057/jors.1995.9
9. Shtub A. The trade-off between the net present cost of a project and the probability to complete it on schedule. *J. Oper. Manag*. 1986;6(3–4):461–470. https://doi.org/10.1016/0272-6963(86)90017-3
10. Shtub A. The integration of CPM and material management in project management. *Constr. Manag. Econ*. 1988;6(4):261–272. https://doi.org/10.1080/01446198800000023
11. Ananthanarayanan P.S. Project Technology and Management. In: S. Seetharaman (Ed.). *Treatise on Process Metallurgy*. V. 3. *Industrial Processes*. Elsevier; 2014. P. 1145–1191. https://doi.org/10.1016/B978-0-08-096988-6.00038-9
12. García-Heredia D., Molina E., Laguna M., et al. A solution method for the shared resource-constrained multi-shortest path problem. *Expert Systems with Applications*. 2021;182: 115193. https://doi.org/10.1016/j.eswa.2021.115193
13. Fulkerson D.R. Expected critical path lengths in PERT networks: *Oper. Res*. 1962;10(6):745–912. https://doi.org/10.1287/opre.10.6.808
14. Bellman R. On a routing problem. *Quart. Appl. Math.* 1958;16(1):87–90. https://doi.org/10.1090/qam/102435
15. Stern R., Goldenberg M., Saffidine A., et al. Heuristic search for one-to-many shortest path queries. *Ann. Math. Artif. Intell.* 2021;89:1175–1214. https://doi.org/10.1007/s10472-021-09775-x
16. Dreyfus S.E. An appraisal of some shortest-path algorithms. *Operations Research.* 1969;17(3):395–412. https://doi.org/10.1287/opre.17.3.395
17. Clarke S., Krikorian A., Rausan J. Computing the *N* best loopless paths in a network. *J. Soc. Indust. Appl. Math*. 1963;11(4):1096–1102. https://doi.org/10.1137/0111081
18. Pollack M. Solutions of the kth best route through a network – A review. *J. Math. Anal. Appl*. 1961;3(3): 547–559. https://doi.org/10.1016/0022-247X(61)90076-2
19. Shier D.R. Iterative methods for determining the k shortest paths in a network. *Networks*. 1976;6(3):205–229. https://doi.org/10.1002/net.3230060303
20. Shier D.R. On algorithms for finding the *k* shortest paths in a network. *Networks*. 1979;9(3):195–214. https://doi.org/10.1002/net.3230090303
21. Minieka E., Shier D.R. A note on an algebra for the *k* best routes in a network. *IMA J. Appl. Math*. 1973;11(2): 145–149. https://doi.org/10.1093/imamat/11.2.145
22. Eppstein D. Finding the *k* shortest paths. *SIAM J. Comput*. 1999;28(2):652–673. https://doi.org/10.1137/S0097539795290477
23. Yen J.Y. Finding the *K* shortest loopless paths in a network. *Manag. Sci.* 1971;17(11–Theory Series):712–716. Available from URL: http://www.jstor.org/stable/2629312
24. Hershberger J., Maxel M., Suri S. Finding the *k* shortest simple paths: A new algorithm and its implementation. *ACM Trans. Algor.* 2007;3(4): Art. 45(19 p.) https://doi.org/10.1145/1290672.1290682
25. Chen B.Y., Chen X.-W., Chen H.-P., et al. A fast algorithm for finding *K* shortest paths using generalized spur path reuse technique. *Transactions in GIS.* 2021;25(1): 516–533. https://doi.org/10.1111/tgis.12699
26. Hu X.-B., Zhang C., Zhang G.-P., et al. Finding the *k* shortest paths by ripple-spreading algorithms. *Eng. Appl. Artif. Intell*. 2020;87:Art.103229. https://doi.org/10.1016/j.engappai.2019.08.023
27. Hamed A.Y. A genetic algorithm for finding the *k* shortest paths in a network. *Egypt. Inform. J*. 2010;11(2):75–79. https://doi.org/10.1016/j.eij.2010.10.004
28. Hu X.-B., Zhou J., Li H., et al. Finding the *k* shortest paths for co-evolutionary path optimization. In: *IEEE Symposium Series on Computational Intelligence*. November 18–21, 2018; Bangalore, India. P. 1906–1912. https://doi.org/10.1109/SSCI.2018.8628928

29. Liu G., Qiu Z., Chen W. An iterative algorithm for single-pair *K* shortest paths computation. *WSEAS Transactions on Information Science and Applications*. 2015;12(Art. 30):305–314. Available from URL: https://wseas.org/wseas/cms.action?id=10185
30. Guo J., Jia L. A new algorithm for finding the *K* shortest paths in a time-schedule network with constraints on arcs. *J. Algorithms Comput. Technol*. 2017;11(2):170–177. https://doi.org/10.1177/1748301816680470
31. Xu W., He S., Song R., et al. Finding the *K* shortest paths in a schedule-based transit network. *Comput. Oper. Res.* 2012;39(8):1812–1826. https://doi.org/10.1016/j.cor.2010.02.005
32. Jin W., Chen S., Jiang H. Finding the *K* shortest paths in a time-schedule network with constraints on arcs. *Comput. Oper. Res.* 2013;40(12):2975–2982. https://doi.org/10.1016/j.cor.2013.07.005
33. Kaufmann A., Debazeille G. *La méthode du chemin critique*. Paris: Dunod; 1964. 170 p.
34. Bellman R., Kalaba R. On *k*th best policies. *J. Soc. Ind. Appl. Math.* 1960;8(4)582–588. Available from URL: https://www.jstor.org/stable/2099049
35. Shier D.R. Computational Experience with an algorithm for finding the *k* shortest paths in a network. *J. Res. Natl. Bureau Stand*. 1974;78B(3)139–165. https://doi.org/10.6028/JRES.078B.020

About the author

Mikhail A. Anfyorov, Dr. Sci. (Eng.), Professor, Domain-Specific Information Systems Department, Institute of Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: anfyorov@inbox.ru. https://orcid.org/0000-0003-2853-6184

Об авторе

Анфёров Михаил Анисимович, д.т.н., профессор, профессор кафедры «Предметно-ориентированные информационные системы» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: anfyorov@inbox.ru. https://orcid.org/0000-0003-2853-6184

Translated from Russian into English by Evgenii I. Shklovskii
Edited for English language and spelling by Thomas A. Beavitt

UDC 004.685.5
https://doi.org/10.32362/2500-316X-2023-11-1-70-80

RESEARCH ARTICLE

Logical-semantic definition of a production process digital twin

Alexander V. Rechkalov [1], **Alexander V. Artyukhov** [1], **Gennady G. Kulikov** [2, @]

[1] *United Engine Corporation, Moscow, 109147 Russia*
[2] *Ufa State Aviation Technical University, Ufa, 450008 Russia*
[@] *Corresponding author, e-mail: gennadyg_98@yahoo.com*

Abstract
Objectives. A methodology currently being developed for generalizing and presenting knowledge about studied subject-oriented areas is based on a "model hypothesis" for determining the distinguished objects and their relationships. Such system models can be used in any information system to define the knowledge about subject-oriented areas. Systems engineering methods already make it possible to create information models of real objects supplemented by virtual components, and vice versa, i.e., models of virtual objects supplemented by real components. So, for example, the availability of information models of technological and production facilities and their connections with real equipment allow the creation and management of real-virtual production processes (PP) in accordance with Industry 4.0 methodologies. From a theoretical aspect, the development of system models of objects and their connections in subject-oriented areas is based on the problem of a formal consistent description (grammatical calculus) of the functional regularities of a given set of objects and their relationships. The purpose of the study is to develop an approach and principles of methodology for system modeling of production facilities and their connections to provide closed-loop control (forecasting, planning, accounting, regulation, etc.) in the production environments of machinery enterprises taking the form of their digital twins (DTs).
Methods. The basic provisions of the theory of sets and graph theory—in particular, the provisions of the theory of categories of sets—are used according to the formal logic and control theories. System Engineering methods are also applied in the organization and management of machinery production.
Results. The approach to the formation of a metastructure of production process digital twin (PPDT) based on the models of production facilities and their relationships is substantiated. The procedure and rules for constructing a PPDT system model are developed along with an approach to the structural and parametric identification of DT models, taking logical and semantic restrictions into account.
Conclusions. A presented example for identifying the basic set of objects of the organization of the PPDT based on the logical-semantic analysis of production activities and the provisions of the unified standards of the unified system for technological preparation of production of machinery production as a researched subject area confirms the main provisions of the proposed methodology for constructing the PPDT.

Keywords: system models, production processes, information systems, digital twin, decision support system, smart factory, planning system

• **Submitted:** 08.07.2022 • **Revised:** 19.09.2022 • **Accepted:** 14.11.2022

For citation: Rechkalov A.V., Artyukhov A.V., Kulikov G.G. Logical-semantic definition of a production process digital twin. *Russ. Technol. J.* 2023;11(1):70–80. https://doi.org/10.32362/2500-316X-2023-11-1-70-80

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

Логико-семантическое определение цифрового двойника производственного процесса

А.В. Речкалов [1],
А.В. Артюхов [1],
Г.Г. Куликов [2, @]

[1] *Объединенная двигателестроительная корпорация, Москва, 109147 Россия*
[2] *Уфимский государственный авиационный технический университет, Уфа, 450008 Россия*
[@] *Автор для переписки, e-mail: gennadyg_98@yahoo.com*

Резюме

Цели. В настоящее время активно развивается методология обобщения и представления знаний об исследуемых предметно-ориентированных областях (ПОО) на базе «модельной гипотезы» определения выделяемых объектов и их связей. Такие системные модели определяют знания о ПОО в любой информационной системе. Методы системной инженерии уже сегодня позволяют создавать информационные модели реальных объектов, дополненные виртуальными составляющими, и наоборот, модели виртуальных объектов, дополненные реальными составляющими. Так, например, наличие информационных моделей технологических и производственных объектов и их связей с реальным оборудованием позволит создавать и управлять реально-виртуальными производственными процессами (ПП) в соответствии с методологией Industry 4. В теоретическом аспекте в основе разработки системных моделей объектов и их связей в ПОО лежит проблема формального непротиворечивого описания (грамматического исчисления) функциональных закономерностей данного множества объектов и их связей. Цель исследования – разработать подход и принципы методологии системного моделирования производственных объектов и их связей для замкнутого управления (прогнозирования, планирования, учета, регулирования и др.) в производственной среде машиностроительного предприятия в форме их цифровых двойников (ЦД).

Методы. Использованы базовые положения теории множеств, теории графов, и, в частности, положения теории категорий множеств, положения теории формальной логики, теории управления. Применены методы теории системной инженерии, методы организации и управления машиностроительным производством.

Результаты. Обоснован подход к формированию метаструктуры цифрового двойника производственного процесса (ЦД ПП) на основе моделей производственных объектов и их связей, разработаны порядок и правила построения системной модели ЦД ПП, разработан подход к структурно-параметрической идентификации моделей ЦД с учетом логико-семантических ограничений.

Выводы. Представлен пример идентификации базового множества объектов ЦД организации ПП на основе логико-семантического анализа производственной деятельности и положений единых стандартов Единой системы технологической подготовки производства для машиностроительного производства как исследуемой предметной области, подтверждающий основные положения предлагаемой методологии построения ЦД ПП.

Ключевые слова: системные модели, производственные процессы, информационные системы, цифровой двойник, digital twin, система поддержки принятия решения, smart factory, система планирования

• **Поступила:** 08.07.2022 • **Доработана:** 19.09.2022 • **Принята к опубликованию:** 14.11.2022

Для цитирования: Речкалов А.В., Артюхов А.В., Куликов Г.Г. Логико-семантическое определение цифрового двойника производственного процесса. *Russ. Technol. J.* 2023;11(1):70–80. https://doi.org/10.32362/2500-316X-2023-11-1-70-80

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

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Many societies are undergoing processes of fundamental changes determined by the requirements of the fourth industrial revolution (Industry 4.0), which implies an active transformation of existing systems of industrial production, technology management, transportation, supply of resources, etc.[1] As noted in [1], the basis and primary infrastructure of this new way of life is formed by information technology (IT). At the same time, within this framework, it is of fundamental importance to develop a comprehensive, integrated understanding of how technologies will change our lives, implying a transformation of the environment of our understanding[2]. According to research conducted by the analytical company IDC, by 2025 the share of the digital economy will grow to 58.2% in the total volume of the world economy [2].

In the context of ever more complex industrial technologies, increasing competition, and accelerating the development of new technology, the introduction of new methods of coordination and interaction both in the external environment and within individual enterprises is becoming not only an urgent need, but also the key to the vital transformation of industrial models [3]. Digital technologies significantly expand interaction possibilities, increase production and supply efficiency, improve the distribution and use of resources, as well as implementing new methods of inventory management. This implies the concomitant possibility of carrying out a system-wide optimization of the management of production processes (PP) and the use of resources.

In recent years, the development of IT is increasingly focused on the formation of technologies for the development and use of a digital twin (DT). The system of DTs, unlike traditional solutions focused on periodic recalculations of plans, can take into account a variety of emergency situations that can be predicted by analyzing a huge data stream. The ability to process and analyze data on the current state of production in accordance with plans in real time allows one to quickly respond to deviations in production and eliminate emerging problems.

Thus, the DT system will allow modeling not only the state of the PP, but also its dynamics. The main task of the digital system is to provide real-time control of all factors affecting the PP[3].

1. PRODUCTION PROCESS DIGITAL TWIN CONCEPT

The concept of a DT in the science and machinery production is quite new and does not yet have an established terminology [4]. For example, the article entitled “Digital Twin of Organization, DTO”[4] presents several options for formulating the concept of a DT:

- a software analogue of a physical device for simulating the internal processes, technical characteristics and behavior of a real object under the influence of interference and the environment;
- a virtual prototype of real production assets—wells, turbines, wind electrical turbines, etc.;
- a digital representation of an object sufficient to meet the requirements of a set of use cases;
- a digital model of a specific physical element or process with connections to data to support the convergence of physical and virtual states at an appropriate rate;
- a model that describes as accurately as possible the real cause-and-effect relationships between the production, economic, financial and organizational indicators of the company.

A more general formulation can be given: “A digital twin is a digital representation of an object or system of the real world”[5]. If we define the concept of a DT fully, then a DT is a digital system representation of a real object, process or system in a virtual environment.

[1] Artyukhov A.V. *Methods and models of the organization of the production process of a multi-product machinery enterprise as a control object*: Cand. Sci. Thesis (Eng.). Samara; 2017. 20 p. (in Russ.).

[2] Khuzmiev I.K. Information technology is the infrastructure of the fourth industrial revolution. *Russia: trends and development prospects.* 2017;12(3):274–277 (in Russ.).

[3] Within the framework of this article, we mean the production process.

[4] Digital Twin of Organization, DTO. https://www.tadviser.ru/index.php/Статья:Цифровой_двойник_(Digital_Twin_of_Organization,_DTO). Accessed September 23, 2022 (in Russ.).

[5] Digital Twin. https://www.gartner.com/en/information-technology/glossary/digital-twin. Accessed September 23, 2022.

It is believed that initially the concept of a DT in management was reported by Michael Greaves at the Product Lifecycle Management forum at the University of Michigan in early 2002. The concept was based on the mutual correspondence and interaction of the physical system and its virtual display based on virtual interconnection [5].

If initially there appeared DTs of products that associatively determine the means and objects of labor (OLs) in the virtual space of the enterprise, then, when forming an intellectual enterprise (Smart Factory) or DT of a production system (DTPS) it becomes necessary to formalize the production process digital twin (PPDT) [6]. At the same time, the PP can be represented as a process of interaction of means of labor (MLs), OLs, and labor resources (LRs) in space and time in order to obtain the result of the process, i.e., the product (Pr)[6] (Fig. 1).

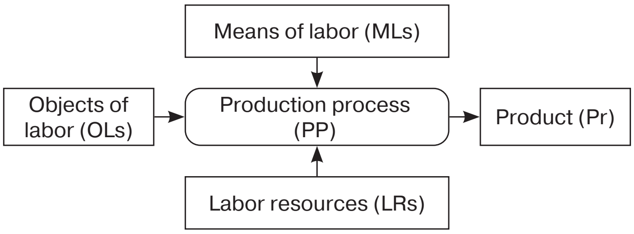


Fig. 1. Relationship of the elements of the production process

The OLs as PP elements comprise materials, components, semi-finished products, parts and assembly units; in essence, and their models are DT-components (Component Twin). Models of equipment, tools and LPs are DTs of a resource (Asset Twin). Models of the result of the PP are the product DT (System Unit Twin). Models of the process of interaction of these elements are the PPDT (Process Twin) [5]. As a whole, the totality of these elements represents DTPS (Fig. 2).

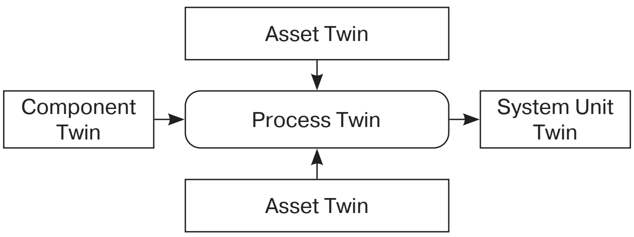


Fig. 2. Relationship of DTPS elements

As a result, we can agree with the following formulation of the DTPS concept: "The digital twin of a production system is a digital model that is constantly updated and changed as its physical counterpart changes in order to synchronously present data on the status, operating conditions, product configuration, and state of resources"[7]. The definition of a DT as a software-hardware complex that implements a complex dynamic model for researching and managing the activities of a sociotechnical system allows us to represent the PP as a control loop, where the PP itself acts as a control object, and its DT acts as a control process with feedback (Fig. 3).

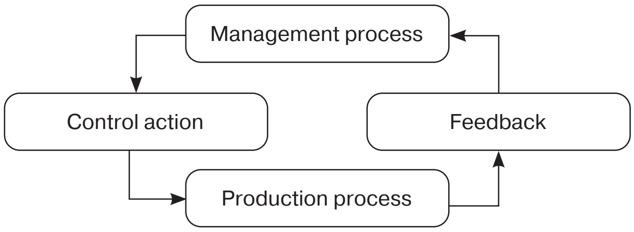


Fig. 3. PP as a control loop

Proceeding from this provision, the principles for the formation of a PPDT based on models of traditional automatic control systems can be formulated as follows: a DT should reflect the current state of the PP and carry out a forecast of this state in the form of phase trajectories (in the space states) of plans based on a detailed analysis of perturbing influences, with variants developed for the behavior of the system along with recommendations for decision-making. In this case, all actions are performed in real time.

At the same time, it is noted that for the practical implementation of the DT concept, the most difficult is the PPDT model, which implements a number of fundamental functions [6, 7]:

- real-time display of PPs occurring in the PS;
- carrying out calculations for making managerial decisions;
- implementation of decision support processes, for example, "what if" based on modeling the behavior of PPs.

2. PPDT STRUCTURE

The PPDT, defined as a reflection of the process of interaction between MLs, OLs, and LRs in order to obtain Pr, basically contains five groups of models:

- a group of product models, including the composition and relationships of OLs within the product, including quantitative, temporal, cost and quality characteristics that determine the use value of the result of the PP;
- a group of models of labor objects, including models of materials used in the PP, semi-finished products,

[6] Artyukhov A.V. *Methods and models of the organization of the production process of a multi-product machinery enterprise as a control object*: Cand. Sci. Thesis (Eng.). Samara; 2017. 20 p. (in Russ.).

[7] PNST 429-2020. *Smart manufacturing. Production digital twins*. Part 1. *General provisions*. Moscow: Standartinform; 2020. 8 p. (in Russ.).

purchased components, as well as parts and assembly units that are in the process of production, fully embodied in the product;

- a group of models of labor means, including models of characteristics and parameters of each separate resource within the framework of the participation of resources in a technological operation;
- a group of models of LRs, including models of characteristics and parameters of each individual resource that determine the participation of resources in a technological operation;
- a group of models of technological processes as a set or sequence of technological operations, including quantitative, temporal, cost and qualitative characteristics of the relationship between OLs and resources.

Taking into account that in this case the MLs, OLs, LRs, and Pr models contain only data necessary for the PP implementation, these models reflect the technological process that determines the PP static (structural) organization. The concept of "static" in this case is relative. Although all the above models support the changes inherent in the corresponding physical objects, these changes are reflected within the framework of the PPDT regardless of its dynamic characteristics. Accordingly, the formulation of the concept of a technological process as a complex of models of a technological process, containing information about technological processes as targeted actions in order to change and (or) determine the state of the OL (which can be attributed to the finished product, its component or to the methods of processing, shaping and assembly[8]), corresponding to the content of the static model of the PP or the DT of the PP organization.

In general, models are defined as static, describing objects in stationary modes of their operation, and dynamic, describing transient processes. The dynamic nature of the PPDT model is reflected in the standard: "Digital twin of production allows real-time production control to dynamically manage production volume and meet the production plan"[9]. In this case, the time requirements of the product are input variables, while the requirements for other resources are output variables; the model describes the transient mode of the PP.

Quantitative and temporal characteristics of the interaction of resources in the PP, which are determined by the demand for finished products, form the initial data for the planning model or DT of planning the PP.

The structure of the planning model primarily depends on the structure of the initial data generated from the results of demand planning. An analysis of planning and management models based on descriptions of manufacturing resource planning (MRP) algorithms and enterprise resource planning (ERP) systems allows two types of demand data to be distinguished—i.e., volume-calendar and order-based—as well as to determine a set of PP management models. Correspondingly, the DT of PP planning can be represented by two linked DTs: the DT of volume-calendar planning of PP and the DT of order-based planning of PP, as well as the complex of feedback and control models by the DT of control of PP [8].

The volume-calendar planning DT implements the following main functions:

- enlarged planning of production capacities;
- enlarged planning of purchased products and materials;
- creation of the main production calendar.

The DT of order-by-order planning implements the following main functions:

- planning production orders;
- power distribution planning;
- planning orders for the repair of products;
- planning orders for storage and distribution;
- planning purchase orders.

The DT of production control implements the following main functions:

- production accounting;
- decision support;
- control of direct production costs;
- quality control management;
- tool distribution control.

The analysis of ERP systems can also be used to differentiate the PP as a control object in the form of a set of internally homogeneous, relatively independent, but interconnected processes:

- PP;
- product repair process;
- process of storing OLs;
- process of formation of costs;
- quality control process;
- power distribution process;
- process of allocating labour;
- tool allocation process.

Correspondingly, the expanded model of the PPDT takes the form of a set of interconnected control loops, which can be defined as functional areas of a DT (Fig. 4).

Thus, the PPDT comprises a set of DTs:

- PP organization;
- volume-calendar planning of production;
- order-by-order production planning;
- production management.

[8] GOST 3.1109-82. *Interstate standard. Unified system of technological documentation. Terms and definitions of basic concepts*. Moscow: Izdatel'stvo standartov; 1981. 15 p. (in Russ.).

[9] PNST 429-2020. *Smart manufacturing. Twins of digital production*. Part 1. *General provisions*. Moscow: Standartinform; 2020. 8 p. (in Russ.).

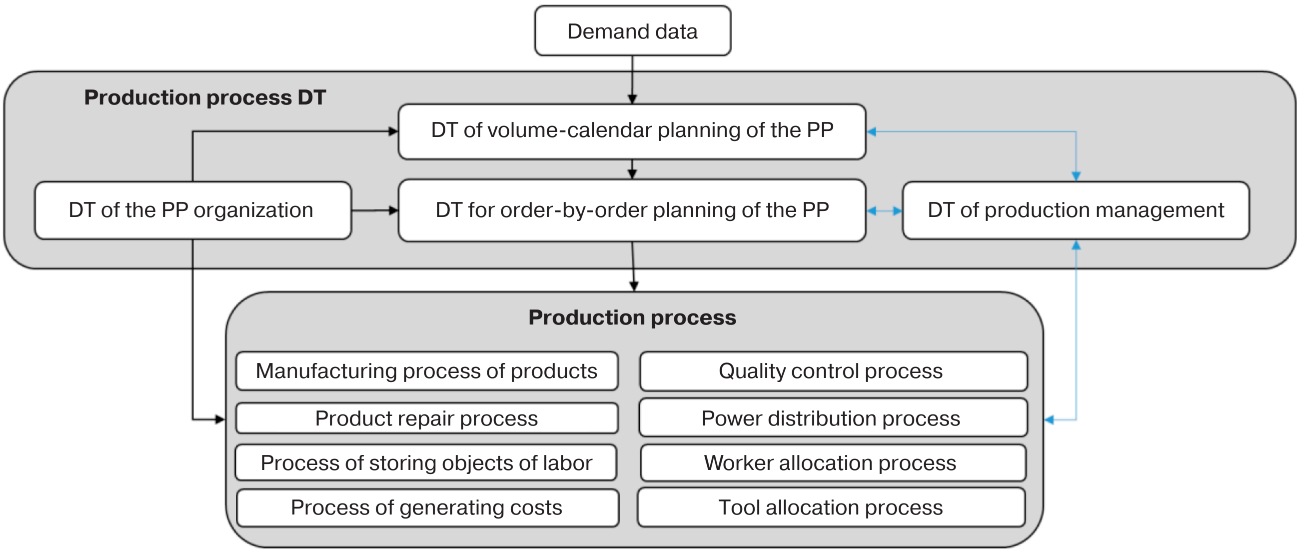


Fig. 4. Expanded top-level structure of the PPDT

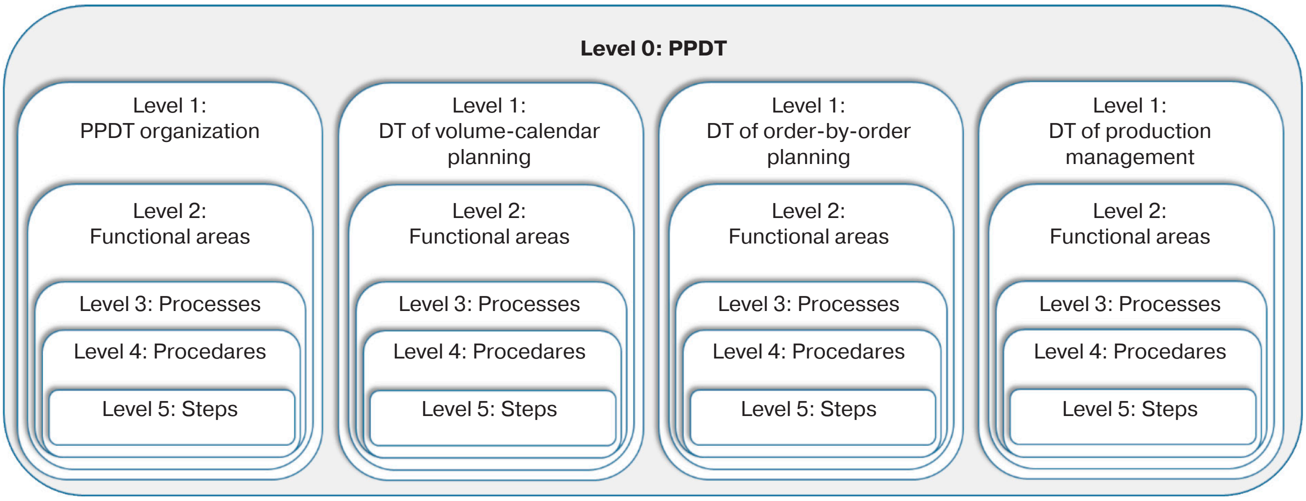


Fig. 5. PPDT generalized structure

Each DT consists of control loops of the corresponding functional areas. Within the framework of a detailed description of the models, each functional area can contain a three-level structure: processes, procedures, and steps. Correspondingly, the PPDT comprises a set of models formed on the basis of the composition of the functions of each DT, as well as the corresponding functional area and hierarchical description of the models (Fig. 5).

3. PPDT IDENTIFICATION

In accordance with Ashby's law of requisite variety [9], optimal control is achieved under the condition that the variety of the control action corresponds to the variety of the controlled one. This property is provided by two factors: full knowledge of the behavior of the controlled system on the part of the control system and the ability to turn this knowledge into an adequate reflection within the DT, i.e., provide the generated models with the requirements of identifiability and traceability. Identifiability is a property that a model must satisfy in order to be able to draw accurate inferences[10]. A model is identifiable if it is theoretically possible to know the true values of the main parameters of this model after receiving an infinite number of observations from it. Typically, a model is identifiable only under certain technical constraints, in which case the set of these requirements is referred to as identification conditions.

The concept of system identification can be defined as a set of methods for constructing mathematical models of a dynamic system based on observational data[11]. Another

[10] *Identifiability*. Wikipedia. https://ru.wikibrief.org/wiki/Identifiability. Accessed September 09, 2022 (in Russ.).

[11] *Systems identification*. Wikipedia. https://en.wikipedia.org/wiki/System_identification. Accessed September 09, 2022 (in Russ.).

definition is as follows: identification is the process of constructing models of objects of various nature. In this case, the identification process consists of two interrelated stages: identification of the structure of models and identification of parameters in models of the selected structure. When constructing the model structure (or a set of competing or complementary structures), a priori information about the object is used. For each class of objects, banks of structures with related information are formed[12].

Coming from the features of the content of the DT, the identification process should be based on the following fundamental provisions:

- *Identification of the DT of the organization of production*, as containing a static model of the PP, is based on the analysis of the structure and parameters of the technological process documented in the standards of the Unified System of Technological Documentation (USTD).
- *Identification of the DT of the volume-calendar process and the DT of the order-by-order planning process* is based on the structure of processes defined in the international standards for production planning and management by the American Production and Inventory Control Society, the MRP/ERP methodology and the parameters determined based on the analysis of the totality of PPs.
- *Identification of the DT of production management processes* in terms of production accounting, analysis and decision-making processes is based on the Ashby principles [9], which involve ensuring maximum detail of the planning and accounting characteristics of the PP that determine the maintenance of homeostasis of the PS.

Thus, in order to ensure the identifiability of PPDT models, it is necessary to determine the structure and set of parameters that provide the identifiability of each type of PP objects (MLs, OLs, LRs, and Pr), interaction processes, and a list of technical restrictions that determine the conditions for identification.

The approach to the identification of the structure of the PPDT based on the provisions of the theory of sets, and, in particular, the provisions of the theory of categories, as well as the provisions of formal logic, will make it possible to form methods for identifying the metastructure of the DT based on the formalization of expert knowledge. At the same time, it is important to note that the conceptual scheme of the top-level representation of the DT structure, as a rule, consists of semantic attributive descriptions in the form of hypertext and production knowledge bases that form the basis of the structure of production management expert systems.

[12] Boyko R.S. On modeling discrete-continuous processes. *Young scientist*. 2010;4(15):93–98 (in Russ.).

4. AN APPROACH TO DT IDENTIFICATION: CASE STUDY OF A DT OF A PP ORGANIZATION

Let us consider an approach to the identification of the DT of the PP organization based on the formalization of knowledge formulated in the standards of the Unified System for Technological Preparation of Production (USTPP)[13, 14] for determining the semantic boundaries of the studied subject area of the PP based on text documents using a graph-analytical representation of PP objects. In this case, MLs (*ML*), OLs (*OL*), LRs (*LR*), and Pr (*Pr*) comprise sets of objects, while the PP represents sets (relations) of morphisms. At the same time, if we exclude from consideration the dynamic characteristics of the PP (amount and time), then we form a graphical representation of the PPDT organization or a graph of the ontological model that defines the boundaries of the subject area and the structure of the PPDT organization. For a more correct representation of the structure of objects in the MLs, we can allocate capacities (equipment, warehouses, and production premises) (*Caps*) and tools and fixtures (*Tools*) (Fig. 6).

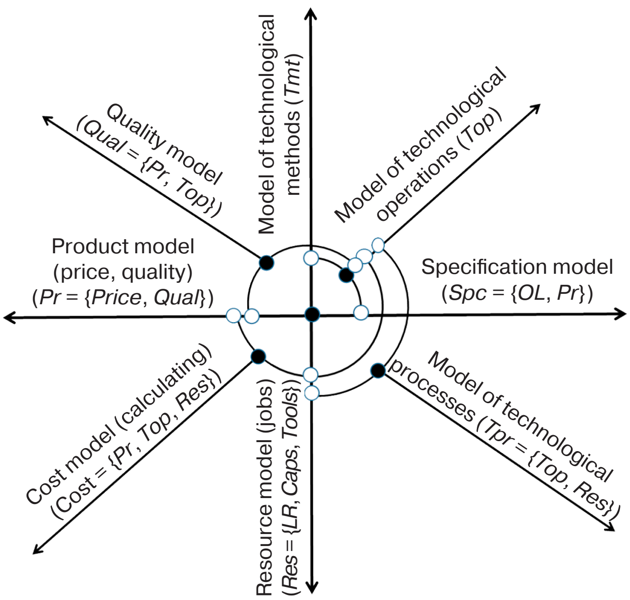


Fig. 6. Graphical representation of the structure of the DT of the PP organization

At the first stage, the textual and tabular representation of the data of the sets of objects reflected in the USTD standards is analyzed. For example, the set of technological methods

[13] GOST 3.1109-82. *Interstate standard. Unified system for technological documentation. Terms and definitions of main concepts*. Moscow: Izdatel'stvo standartov: 1981. 15 p. (in Russ.).

[14] GOST 14.004-83. *Technological preparation of production. Terms and definitions of basic concepts*. Moscow: Izdatel'stvo standartov; 1983. 8 p. (in Russ.).

Table 1. Structure of the system model of the PPDT organization. Part 1

Objects of labor models		Technological methods	Models of technological operations
Objects of Labor (OL)	OL structure		
Item Material Main material Auxiliary material Semi-product Blank Initial workpiece Stamped product Casting Forging Accessory Part Assembly unit Assembly kit Ready product Aggregate Refurbished product Blend Alloy Die block Defective product Technological losses Secondary material resources Used waste Unused waste Irretrievable waste	OL identification: A – Anonymous OL N – Numbered OL B – Numbered batch of OL C – Custom OL Traceable entry type: A – A A – B A – N A – C B – B B – N B – C N – N N – C C – C	Casting Molding Cutting material Forging Volumetric stamping Sheet stamping Surface plastic deformation Machining Thermal processing Electrophysical processing Electrochemical processing Electroplating processing Locksmith processing Assembling Welding Riveting Soldering Gluing Coating Technical control Tests Acquisition Dismantling Transportation Storage Repair Waste recycling	MTO1 MTO2 MTO3 MTO4 MTO5 MTO6 MTO7 MTO8 MTO9

$Tmt = \{tmt_1, tmt_2, \ldots, tmt_n\}$ represents a list of elements of the structure and characteristics of the methods of interaction of OLs, while the set of specifications $Spc = (spc_1, spc_2, ..., spc_n)$ represents an interconnected list of OLs: $OL = (OL_1, OL_2, ..., OL_n)$, i.e., purchased materials, semi-finished products, purchased workpieces, as well as manufactured parts and assembly units that are in operation[15]. At the same time, for each pair of elements of the set Spc and the set Tmt, a set of morphisms is formed: technological operations $Top = \{top_1, top_2, ..., top_n\}$, i.e., $Tmt \subset Top$ and $Spc \subset Top$.

The set of production workers with indication of their professions and qualification characteristics $LR = \{LR_1, LR_2, \ldots, LR_n\}$, the set of capacities $Caps = \{cap_1, cap_2, \ldots, cap_n\}$, the set of tools $Tools = \{tool_1, tool_2, \ldots, tool_n\}$ are included in the set resources (jobs) $LR \subset Res$, $Caps \subset Res$, $Tools \subset Res$. Other sets of the presented structure of the PPDT organization are formed similarly.

At the next stage, the characteristics of each object of the set are generated on the basis of the graph-analytical and syntax analysis of the USTPP documents. The result of identification of models of the PPDT organization is the formal structure of the system model of the PPDT organization (Tables 1 and 2)[16].

The PP organization model comprises a metamodel that combines the elementary models. The metamodel of the organization of the PP as presented in Tables 1 and 2 also reflects the variety of functional interconnections of the elements of the PP: elementary OLs, elementary MLs, and elementary technological processes (operations), as well as elementary interoperational PPs that determine the composition and content of the PPDT organization.

CONCLUSIONS

The main results obtained in the presented work are as follows:

1. Formulation of logical-semantic requirements for the creation of system models of logistics, production, service, and other processes in the form of their DT comprising the necessary conditions for constructing

[15] GOST 3.1109-82. *Interstate standard. Unified system for technological documentation. Terms and definitions of main concepts*. Moscow: Izdatel'stvo standartov; 1981. 15 p. (in Russ.).

[16] Artyukhov A.V. *Methods and models of the organization of the production process of a multi-product machinery enterprise as a control object*: Cand. Sci. Thesis (Eng.). Samara; 2017. 20 p. (in Russ.).

Table 2. Structure of the system model of the PPDT organization. Part 2

Model of organizational structure	Workplace models		Models of organization of technological process operations	Models of interoperational processes
	Workplace structure	Workplace type		
Divisions: - warehouses of purchased OL - warehouses of sold OL - warehouses for obsolete OL - warehouses of OL in production - warehouses of finish-processed and ready-to-assemble OLs - own consignment warehouses - consignment warehouses not own - production shops - production sites	Group of mechanisms Mechanism Group of workers Worker	Simple worksplace Multi-station workplace Integrated workplace	Model of a single-object non-reconfigurable process Model of a single-object reconfigurable process Model of a multi-object reconfigurable process	Models for the formation of production batches of OL Models for the formation of transfer batches of OL Models for the formation of safety stock Models of processing-waiting OL

a cybernetic control model with feedback to meet the required stability margins and quality.

2. Based on the formalization of the results of the logical-semantic analysis of the PP and the management process, an approach has been developed to the formation of a structural-parametric model of the PPDT, taking into account the existing external restrictions (boundaries).
3. Based on the above approach, the metastructure of the PPDT has been developed; it has been shown that the existing structures of production facilities and PP can be coordinated (harmonized) taking into account the requirements of the international standards ISO "Smart Manufacturing."[17]
4. It is shown that the application of the main provisions of the theory of sets—and, in particular, the provisions of the theory of categories of sets and the provisions of formal logic—forms the basis for an approach to identify the model of the PPDT based on the formalization of expert knowledge.
5. A case study of identification of the PPDT organization based on the logical and semantic analysis of the provisions of the USTPP standards is presented.

Authors' contributions

A.V. Rechkalov has proposed an approach to the digital transformation of automated production control systems of operating machinery enterprises (ME) based on the formalization of the logical-semantic model of the production process (PP). He showed that the existing metastructures of production facilities and PP can be harmonized with the requirements of ISO "Smart Manufacturing."

A.V. Artyukhov has developed a methodology and algorithm to design system models of planning, logistics, production, service, and other ME processes in the form of their digital twins (DT) based on structural and parametric identification under the condition of their traceability and controllability. He showed that on the basis of ME PPDT it is expedient to form metastructures of the distributed (parallel) PP for ME united in a machinery holding. He proposed an approach to the formation of the structure of an information management system with feedback.

G.G. Kulikov has proposed to use the basic provisions of the theory of sets, in particular, the provisions of the theory of categories of sets and the provisions of mathematical logic for a formal description of the PPDT.

[17] https://www.iso.org/standard/81277.html. Accessed September 23, 2022.

REFERENCES

1. Rechkalov A.V., Artyukhov A.V., Kulikov G.G., Novikov V.N. The concept of transformation of the model of planning and management processes based on the digital twin of the production system in the industrial model of a machine-building enterprise. *Vestnik UGATU*. 2022;26(1–95):120–135 (in Russ.). https://doi.org/10.54708/19926502_2022_26195120
2. Shvab K. *Chetvertaya promyshlennaya revolyutsiya = The Fourth Industrial Revolution*: transl. from Eng. Moscow: Eksmo; 2018. 285 p. (in Russ.). ISBN 978-5-699-98379-7 [Schwab K. *The Fourth Industrial Revolution*. Geneva, Switzerland: World Economic Forum; 2016. 192 p.]
3. Tolstykh T.O., Gamidullaeva L.A., Shkarupeta E.V. Key factors of development of the industrial enterprises in the conditions of the industry 4.0. *Ekonomika v promyshlennosti = Russian Journal of Industrial Economics*. 2018;11(1):11–19 (in Russ.). https://doi.org/10.17073/2072-1633-2018-1-11-19
4. Dozortsev V.M. Digital twins in industry: genesis, composition, terminology, technologies, platforms, prospects. Part 1. The emergence and development of digital twins. How do the existing definitions reflect the content and functions of digital twins? *Avtomatizatsiya v promyshlennosti = Automation in Industry*. 2020;9:3–11 (in Russ.). https://doi.org/10.25728/avtprom.2020.09.01
5. Prokhorov A., Lysachev M. *Tsifrovoi dvoinik. Analiz, trendy, mirovoi opyt* (*Digital twin. Analysis, trends, world experience*). Prof. Borovkov A. (Sci. Ed.). 1st ed. Moscow: Al'yansPrint; 2020. 401 p. (in Russ.). ISBN 978-5-98094-008-9
6. Frolov E.B., Klimov A.S., Zin Min Khtun. MES – the basis for the creation of a "digital twins" for production systems. *Vestnik MGTU "Stankin*." 2019;2(49):52–56 (in Russ.).
7. Solomentsev Yu.M., Frolov E.B. "Digital twin" of the production system is a promising tool for improving the efficiency of the machine park of a machine-building enterprise. *Stanochnyi Park*. 2018;8:36–39 (in Russ.).
8. Gavrilov D.A. *Upravlenie proizvodstvom na baze standarta MRP II. Seriya: Teoriya i praktika menedzhmenta* (*Production Management Based on the MRPII Standard. Series: Theory and Practice of Management*). St. Petersburg: Peter; 2003. 352 p. (in Russ.).
9. Eshbi U.R. *Vvedenie v kibernetiku* (*Introduction to Cybernetics*): transl. from Eng. Moscow: Lenand; 2021. 432 p. (in Russ.). ISBN 978-5-453-00197-2 [Ashby W.R. *An Introduction to Cybernetics*. London: Champan & Hall; 1956.]

СПИСОК ЛИТЕРАТУРЫ

1. Речкалов А.В., Артюхов А.В., Куликов Г.Г., Новиков В.Н. Концепция системного представления предметной области при формировании цифрового двойника производственного процесса машиностроительного предприятия. *Вестник УГАТУ*. 2022;26(1–95):120–135. https://doi.org/10.54708/19926502_2022_26195120
2. Шваб К. *Четвертая промышленная революция = The Fourth Industrial Revolution*: пер. с англ. М.: Эксмо; 2018. 285 с. ISBN 978-5-699-98379-7
3. Толстых Т.О., Гамидуллаева Л.А., Шкарупета Е.В. Ключевые факторы развития промышленных предприятий в условиях цифрового производства и индустрии 4.0. *Экономика в промышленности*. 2018;11(1):11–19. https://doi.org/10.17073/2072-1633-2018-1-11-19
4. Дозорцев В.М. Цифровые двойники в промышленности: генезис, состав, терминология, технологии, платформы, перспективы. Часть 1. Возникновение и становление цифровых двойников. Как существующие определения отражают содержание и функции цифровых двойников? *Автоматизация в промышленности*. 2020;9:3–11 https://doi.org/10.25728/avtprom.2020.09.01
5. Прохоров А., Лысачев М. *Цифровой двойник. Анализ, тренды, мировой опыт*; под научн. ред. проф. Боровкова А. 1-е изд., исправл. и доп. М.: ООО «АльянсПринт»; 2020. 401 с. ISBN 978-5-98094-008-9
6. Фролов Е.Б., Климов А.С., Зин Мин Хтун. MES – основа для создания «цифрового двойника» производственной системы. *Вестник МГТУ СТАНКИН*. 2019;2(49):52–56.
7. Соломенцев Ю.М., Фролов Е.Б. «Цифровой двойник» производственной системы – перспективный инструмент повышения эффективности станочного парка машиностроительного предприятия. *Станочный парк*. 2018;8:36–39.
8. Гаврилов Д.А. *Управление производством на базе стандарта MRP II. Серия: Теория и практика менеджмента*. СПб.: Питер; 2003. 352 с.
9. Эшби У.Р. *Введение в кибернетику*: пер. с англ. М.: Ленанд; 2021. 432 с. ISBN 978-5-453-00197-2

About the authors

Alexander V. Rechkalov, Dr. Sci. (Eng.), Professor, Expert, United Engine Corporation (11, Mayakovskogo per., Moscow, 109147 Russia). E-mail: av@rechkalov.ru. Scopus Author ID 6507553020.

Alexander V. Artyukhov, Cand. Sci. (Eng.), General Director, United Engine Corporation (11, Mayakovskogo per., Moscow, 109147 Russia). artyukhov@uecrus.com. Scopus Author ID 57457052400.

Gennady G. Kulikov, Dr. Sci. (Eng.), Professor, Dept. of Automated Control Systems, Ufa State Aviation Technical University (12, K. Marksa ul., Ufa, The Republic of Bashkortostan, 450008 Russia). E-mail: gennadyg_98@yahoo.com. Scopus Author ID 7102426624.

Об авторах

Речкалов Александр Васильевич, д.т.н., профессор, эксперт, АО «Объединенная двигателестроительная корпорация» (109147, Россия, Москва, пер. Маяковского, д. 11). E-mail: av@rechkalov.ru. Scopus Author ID 6507553020.

Артюхов Александр Викторович, к.т.н., генеральный директор, АО «Объединенная двигателестроительная корпорация» (109147, Россия, Москва, пер. Маяковского, д. 11). E-mail: artyukhov@uecrus.com. Scopus Author ID 57457052400.

Куликов Геннадий Григорьевич, д.т.н., профессор кафедры автоматизированных систем управления, ФГБОУ ВО «Уфимский государственный авиационный технический университет» (450008, Россия, Республика Башкортостан, Уфа, ул. К. Маркса, д. 12). E-mail: gennadyg_98@yahoo.com. Scopus Author ID 7102426624.

Translated from Russian into English by Evgenii I. Shklovskii
Edited for English language and spelling by Thomas A. Beavitt

MIREA – Russian Technological University.
78, Vernadskogo pr., Moscow, 119454 Russian Federation.
Publication date January 31, 2023.
Not for sale.

МИРЭА – Российский технологический университет.
119454, РФ, г. Москва, пр-т Вернадского, д. 78.
Дата опубликования 31.01.2023 г.
Не для продажи.