

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ»
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МІЖНАРОДНИЙ НАУКОВО–ТЕХНІЧНИЙ ЖУРНАЛ

№ 3

2020

ЗАСНОВАНО У ЛИПНІ 2001 р.

РЕДАКЦІЙНА КОЛЕГІЯ:

Головний редактор

М.З. ЗГУРОВСЬКИЙ, акад. НАН України

Заступник головного редактора

Н.Д. ПАНКРАТОВА, чл.-кор. НАН України

Члени редколегії:

П.І. АНДОН, акад. НАН України

А.В. АНІСІМОВ, чл.-кор. НАН України

Г.-В. ВЕБЕР, проф., Турція

О.Д. ГВІШАНИ, акад. РАН, Росія

П.О. КАСЬЯНОВ, проф., д.ф.-м.н.,
Україна

В.М. КУНЦЕВИЧ, акад. НАН України

О.А. ПАВЛОВ, проф., д.т.н., Україна

Л. САКАЛАУСКАС, проф., Литва

А.М. САЛЕМ, проф., Єгипет

І.В. СЕРГІЄНКО, акад. НАН України

Х.-М. ТЕОДОРЕСКУ, акад. Румунської
Академії

Е.О. ФАЙНБЕРГ, проф., США

Я.С. ЯЦКІВ, акад. НАН України

АДРЕСА РЕДАКЦІЇ:

03056, м. Київ,

просп. Перемоги, 37, корп. 35,

ННК «ІПСА» КПІ ім. Ігоря Сікорського

Тел.: 204-81-44; факс: 204-81-44

E-mail: journal.iasa@gmail.com

http://journal.iasa.kpi.ua

У номері:

• **Прогресивні інформаційні технології, високопродуктивні комп'ютерні системи**

• **Проблеми прийняття рішень та управління в економічних, технічних, екологічних і соціальних системах**

• **Теоретичні та прикладні проблеми інтелектуальних систем підтримання прийняття рішень**

• **Методи оптимізації, оптимальне управління і теорія ігор**

• **Математичні методи, моделі, проблеми і технології дослідження складних систем**

NATIONAL ACADEMY OF SCIENCES OF UKRAINE
EDUCATIONAL AND SCIENTIFIC COMPLEX
«INSTITUTE FOR APPLIED SYSTEM ANALYSIS»
OF THE NATIONAL TECHNICAL UNIVERSITY OF UKRAINE
«IGOR SIKORSKY KYIV POLYTECHNIC INSTITUTE»

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

INTERNATIONAL SCIENTIFIC AND TECHNICAL JOURNAL

№ 3

2020

IT IS FOUNDED IN JULY 2001

EDITORIAL BOARD:

The editor – in – chief

M.Z. ZGUROVSKY, Academician of
NASU

Deputy editor – in – chief

N.D. PANKRATOVA, Correspondent
member of NASU

Associate editors:

F.I. ANDON, Academician of
NASU

A.V. ANISIMOV, Correspondent
member of NASU

E.A. FEINBERG, Prof., USA

A.D. GVISHIANI, Academician of RAS,

P.O. KASYANOV, Prof., Ukraine

V.M. KUNTHEVICH, Academician of NASU

A.A. PAVLOV, Prof., Ukraine

L. SAKALAUŠKAS, Prof., Lithuania

A.M. SALEM, Prof., Egypt

I.V. SERGIENKO, Academician of NASU

H.-N. TEODORESCU, Academician of
Romanian Academy

G.-W. WEBER, Prof., Turkey

Ya.S. YATSKIV, Academician of NASU

THE EDITION ADDRESS:

03056, Kyiv,
av. Peremogy, 37, building 35,
Institute for Applied System Analysis
at the Igor Sikorsky Kyiv Polytechnic Institute
Phone: **204-81-44**; Fax: **204-81-44**
E-mail: journal.iasa@gmail.com
<http://journal.iasa.kpi.ua>

In the issue:

• **Progressive information technologies, high-efficiency computer systems**

• **Decision making and control in economic, technical, ecological and social systems**

• **Theoretical and applied problems of intellectual systems for decision making support**

• **Methods of optimization, optimum control and theory of games**

• **Mathematical methods, models, problems and technologies for complex systems research**

Шановні читачі!

Навчально-науковий комплекс «Інститут прикладного системного аналізу» Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» видає міжнародний науково-технічний журнал

«СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ».

Журнал публікує праці теоретичного та прикладного характеру в широкому спектрі проблем, що стосуються системних досліджень та інформаційних технологій.

Провідні тематичні розділи журналу:

Теоретичні та прикладні проблеми і методи системного аналізу; теоретичні та прикладні проблеми інформатики; автоматизовані системи управління; прогресивні інформаційні технології, високопродуктивні комп'ютерні системи; проблеми прийняття рішень і управління в економічних, технічних, екологічних і соціальних системах; теоретичні та прикладні проблеми інтелектуальних систем підтримання прийняття рішень; проблемно і функціонально орієнтовані комп'ютерні системи та мережі; методи оптимізації, оптимальне управління і теорія ігор; математичні методи, моделі, проблеми і технології дослідження складних систем; методи аналізу та управління системами в умовах ризику і невизначеності; евристичні методи та алгоритми в системному аналізі та управлінні; нові методи в системному аналізі, інформатиці та теорії прийняття рішень; науково-методичні проблеми в освіті.

Головний редактор журналу — ректор Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», академік НАН України Михайло Захарович Згуровський.

Журнал «Системні дослідження та інформаційні технології» включено до переліку фахових видань ВАК України.

Журнал «Системні дослідження та інформаційні технології» входить до таких наукометричних баз даних: Scopus, EBSCO, Google Scholar, DOAJ, Index Copernicus, реферативна база даних «Україніка наукова», український реферативний журнал «Джерело», наукова періодика України.

Статті публікуються українською, російською та англійською мовами.

Журнал можна передплатити. **Наш індекс 23918.** Якщо ви не встигли передплатити журнал, його можна придбати безпосередньо в редакції. Наша адреса: 03056, м. Київ, просп. Перемоги, 37, корп. 35.

Завідувачка редакції **С.М. Шевченко**

Редакторка **Р.М. Шульженко**

Молодша редакторка **Л.О. Тарин**

Комп'ютерна верстка, дизайн **А.А. Патюхи**

Свідоцтво про реєстрацію КВ № 23234–13074 ПР від 22.03.2018 р.

Підписано до друку 07.12.2020. Формат 70x108 1/16. Папір офс. Гарнітура Times.

Спосіб друку – цифровий. Ум. друк. арк. 14,411. Обл.-вид. арк. 28,56. Наклад 113 пр. Зам. № 11/04

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Свідоцтво про державну реєстрацію: ДК № 5354 від 25.05.2017 р.

просп. Перемоги, 37, м. Київ, 03056.

ФОП Пилипенко Н.М., вул. Мічуріна, б. 2/7, м. Київ, 01014.

Виписка з Єдиного державного реєстру № 2 070 000 0000 0214697 від 17.05.2019 р.,

тел.: (044) 361 78 68.

Dear Readers!

Educational and Scientific Complex «Institute for Applied System Analysis» of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» is published of the international scientific and technical journal

«SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES».

The Journal is printing works of a theoretical and applied character on a wide spectrum of problems, connected with system researches and information technologies.

The main thematic sections of the Journal are the following:

Theoretical and applied problems and methods of system analysis; theoretical and applied problems of computer science; automated control systems; progressive information technologies, high-efficiency computer systems; decision making and control in economic, technical, ecological and social systems; theoretical and applied problems of intellectual systems for decision making support; problem- and function-oriented computer systems and networks; methods of optimization, optimum control and theory of games; mathematical methods, models, problems and technologies for complex systems research; methods of system analysis and control in conditions of risk and uncertainty; heuristic methods and algorithms in system analysis and control; new methods in system analysis, computer science and theory of decision making; scientific and methodical problems in education.

The editor-in-chief of the Journal is rector of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», academician of the NASU Michael Zaharovich Zgurovsky.

The articles to be published in the Journal in Ukrainian, Russian and English languages are accepted. Information printed in the Journal is included in the Catalogue of periodicals of Ukraine.

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

3 • 2020

ЗМІСТ

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, ВИСОКОПРОДУКТИВНІ КОМП'ЮТЕРНІ СИСТЕМИ	
<i>Крувуй S.L., Pogorilyu S.D., Slynko M.S., Kramov A.A.</i> Method of semantic application verification in GPGPU technology	7
<i>Мельник І.В., Починок А.В.</i> Дослідження класу алгебричних функцій для інтерполяції межових траєкторій короткофокусних електронних пучків	23
<i>Мухомід А.В., Курило Е.А.</i> Расширенная интеграция системы электронного документооборота с пакетом офисных программ (работа с персоналом)	40
ПРОБЛЕМИ ПРИЙНЯТТЯ РІШЕНЬ ТА УПРАВЛІННЯ В ЕКОНОМІЧНИХ, ТЕХНІЧНИХ, ЕКОЛОГІЧНИХ І СОЦІАЛЬНИХ СИСТЕМАХ	
<i>Drogobyskij I.N.</i> Systematicity of organizational management	51
ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМАННЯ ПРИЙНЯТТЯ РІШЕНЬ	
<i>Bodyanskiy Eu., Zaychenko Yu., Hamidov G., Kuleshova N.</i> Multilayer GMDH-neuro-fuzzy network based on extended neo-fuzzy neurons and its application in online facial expression recognition	66
<i>Рудзевич А.-М.П.</i> Методи машинного навчання в сентимент-аналізі текстової інформації на прикладі настроїв користувачів стосовно кандидатів у президенти України 2019	78
МЕТОДИ ОПТИМІЗАЦІЇ, ОПТИМАЛЬНЕ УПРАВЛІННЯ І ТЕОРІЯ ІГОР	
<i>Sineglazov V., Riazanovskiy K., Chumachenko O.</i> Multicriteria conditional optimization based on genetic algorithms	89
<i>Статкевич В.М.</i> Регулярные выражения для некоторых языков сетей Петри в задаче о производителе и потребителе	105
МАТЕМАТИЧНІ МЕТОДИ, МОДЕЛІ, ПРОБЛЕМИ І ТЕХНОЛОГІЇ ДОСЛІДЖЕННЯ СКЛАДНИХ СИСТЕМ	
<i>Matsuki Y., Bidyuk P.I.</i> Simulating the rotation of a black hole and antigravity	124
<i>Kirilyuk V.S., Levchuk O.I., Gavrilenko V.V., Viter M.B.</i> Modeling of contact interaction of a heated plane rigid elliptical punch with a transversally isotropic elastic half-space	138
Відомості про авторів	149

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

3 • 2020

CONTENT

PROGRESSIVE INFORMATION TECHNOLOGIES, HIGH-EFFICIENCY COMPUTER SYSTEMS	
<i>Kryvyi S.L., Pogorilyy S.D., Slynko M.S., Kramov A.A.</i> Method of semantic application verification in GPGPU technology	7
<i>Melnyk I.V., Pochynok A.V.</i> Investigation of the class of algebraical functions for interpolation of boundary trajectories of short-focus electron beams	23
<i>Mukhoid O.V., Kurylo O.O.</i> Extensible integration of the electronic document management system with productivity software (hr management)	40
DECISION MAKING AND CONTROL IN ECONOMIC, TECHNICAL, ECOLOGICAL AND SOCIAL SYSTEMS	
<i>Drogobyskij I.N.</i> Systematicity of organizational management	51
THEORETICAL AND APPLIED PROBLEMS OF INTELLECTUAL SYSTEMS FOR DECISION MAKING SUPPORT	
<i>Bodyanskiy Eu., Zaychenko Yu., Hamidov G., Kuleshova N.</i> Multilayer GMDH-neuro-fuzzy network based on extended neo-fuzzy neurons and its application in online facial expression recognition	66
<i>Rudzevych A.-M.P.</i> Machine learning in sentiment analysis of text information on the example of user attitudes regarding candidates for Ukrainian presidential elections 2019	78
METHODS OF OPTIMIZATION, OPTIMUM CONTROL AND THEORY OF GAMES	
<i>Sineglazov V., Riazanovskiy K., Chumachenko O.</i> Multicriteria conditional optimization based on genetic algorithms	89
<i>Statkevych V.M.</i> Regular expressions for some Petri net languages for the producer/consumer problem	105
MATHEMATICAL METHODS, MODELS, PROBLEMS AND TECHNOLOGIES FOR COMPLEX SYSTEMS RESEARCH	
<i>Matsuki Y., Bidyuk P.I.</i> Simulating the rotation of a black hole and antigravity	124
<i>Kirilyuk V.S., Levchuk O.I., Gavrilenko V.V., Viter M.B.</i> Modeling of contact interaction of a heated plane rigid elliptical punch with a transversally isotropic elastic half-space	138
Information about the authors	149

METHOD OF SEMANTIC APPLICATION VERIFICATION IN GPGPU TECHNOLOGY

S.L. KRYVYI, S.D. POGORILYY, M.S. SLYNKO, A.A. KRAMOV

Abstract. An application development and verification method for massively parallel systems using NVIDIA GPUs is proposed. The method allows creating models at different levels of abstraction using the apparatus of marked transition systems. The compositions (product) of such systems are transformed into a Petri net, which are then analyzed by appropriate means. The proposed method allows specifying model properties by temporal logic formulas. This allows studying the properties of massively parallel systems which is almost impossible to analyze manually, since the number of execution threads in the latest NVIDIA video adapter architectures (Pascal, Volta, Turing, Ampere) is measured in hundreds of thousands or millions.

Keywords: CUDA, graphical processing units (GPU), General Purpose Graphics Computing (GPGPU), transition system, Petri net, model design.

INTRODUCTION

The main feature of the latest graphical processing units (GPU) is the availability of a set of streaming multiprocessors (SM) that were used previously in image processing algorithms and tasks only. General Purpose Graphics Computing (GPGPU) technology is based on the use of a combination of GPUs working in parallel to process data using general-purpose algorithms (scientific or other, but not necessarily related to image processing). The latest graphics architectures from NVIDIA include Pascal, Volta, Amper, Thuring [1]. For this day, Volta is one of the most powerful GPU architectures, which is an indicator of achievements in the field of high-performance artificial intelligence calculations (the GTX TITAN Z graphics adapter, built on the top of two powerful GK110 cores, can provide peak performance up to 8 teraflops, and each core can implement 2880 stream processors, which in total gives 5760 stream processors).

NVIDIA, in addition, develops a series of video adapters focused on scientific applications and use in high-performance (cluster) computing. These GPUs lack some graphic-specific features and are widely used in the scientific field. This led to a significant increase in the number of supercomputers included in the TOP500 world most efficient computers [2] that utilize NVIDIA video adapters.

Nowadays high-performance computing (HPC) trends get shifted from using clusters consisting of general-purpose modules to more specialized accelerated

components (in other words, from universal CPUs to other units — GPU, FPGA, etc.), that is, to less functional and less power consuming modules. Accelerators, unlike universal CPUs, can not run an operating system, and rely on external systems for I/O operations or task scheduling. Their advantage in productivity lies solely in the fact that these elements are used in large groups simultaneously. The article proposes methods of modelling the properties of GPU accelerator architectures at different abstraction levels.

In terms of programming, success of the most common GPGPU technologies (CUDA in particular) lies in the fact that they encapsulate the SIMD nature of GPU hardware. In most cases, the developer deals with individual streams that work with scalar data instead of warps [1] working with vectors.

Development of the complicated, massively parallel systems that utilize video adapters requires new scientific methods to justify both the architecture of the system and applications for them. Since modern multi-threaded applications have hundreds of thousands and millions of threads, solving parallelization tasks for such systems makes impossible to use traditional engineering design approaches and requires the use of a mathematical apparatus and formalization methods to substantiate the decisions made. Model justification is an effective way of algorithmic study of the parallel algorithms properties. One of the options for implementing such a justification is the use of the apparatus of algorithmic algebras [3], which allow formulating schemes of algorithms in the form of algebraic expressions that depend on various parameters, including software and hardware platforms, paradigms of parallel programming, etc. This paper focuses on using the transition systems (TS) [4] and their compositions as the main mathematical model, which allow creating models at different abstraction levels. Their properties can be investigated by translating in Petri networks (PN) [5] and can be specified by temporal logic formulas [6].

MODERN NVIDIA GPU ARCHITECTURES

CUDA (Computational Unified Device Architecture) is a parallel computing architecture developed by NVIDIA to facilitate the GPGPU programming by using high-level APIs. Since the pilot CUDA platform had been introduced more than 10 years ago, each new generation of NVIDIA GPUs provided better application performance (for example, in floating point operations), increased energy efficiency, added important new computing capabilities, and streamlined graphics processor programming. Today, NVIDIA GPUs are leading computing devices that have, in some way, defined the artificial intelligence revolution (AI). Nowadays GPGPU technology accelerates deep learning; high-precision text, voice, and other media data recognition systems; are used in the areas of molecular modeling, modelling of medical products, medical diagnosis, financial modelling, and others.

Application instructions that are executed in a GPU-based heterogeneous environment can be logically decomposed into the following parts:

- “host” instructions — blocks which are executed on the CPU;
- “kernel” functions — instruction blocks which are executed on the GPUs.

Host blocks define the context of the kernel functions execution, transferring data between the computer’s RAM and GPU memory.

All NVIDIA GPU architectures execute instructions in groups of 32 thread (known as warps) using the SIMT model (Single Instruction, Multiple Thread), that is, one instruction is executed by many threads simultaneously, although behaviour of each individual thread is not limited by anything. However, architectures prior to Pascal include, among others, a software counter and a mask common to all threads of the warp that determines which threads are active at any given moment. This means that in the case of a execution ow branching, each execution path uses only a subset of all threads, while the rest are deactivated. Once execution paths are converged, threads of a warp start being executed simultaneously again.

Such an implementation model gets rid of the necessity to track each individual thread state separately. However, tracing only a warp in general means reducing the level of parallelism if branching is present, as described above. In turn, this prevents the data exchange between the threads within a single warp, if those threads are at different execution stages, or if they execute the instructions in different branches. That means that threads of different warps may execute instructions in parallel, but threads of a single warp sometimes have to execute instructions sequentially. Thus, algorithms that require data sharing at a high level of detail or that utilize synchronization tools (such as mutexes) can suffer from deadlocks. Therefore, developers have to rely on algorithms with minimal blocking support while using the NVIDIA GPUs of Pascal architecture or earlier (that is, more than 60% of devices, because Volta architecture was released only in 2017 and Turing — in 2018).

More modern architectures, starting from Volta, include independent flow planning, which stores instruction counters and call stacks for each thread separately, which can be utilized for optimal resource usage or to allow one thread to wait for data from another. To increase the level of parallelism, Volta includes an optimizer that defines how to group active threads within the warp within the SIMT module. This maintains the high performance of the SIMT approach, as in previous NVIDIA GPUs, but with much greater flexibility: the threads can now perform various branching paths within a single warp. The verification methods proposed are illustrated on a simple system model analysis example.

MODEL JUSTIFICATION AND VERIFICATION

Testing process has always been the main method of increasing the reliability of programs, developed using traditional methods. Edsger Dijkstra once said: “Program testing can be used to show the presence of bugs, but never to show their absence!”. In addition, testing can not detect typical synchronization errors of parallel programs. Parallel programs may for years retain errors that manifest themselves after a long usage period as a reaction to a specific combination of numerous factors that have arisen (for example, due to the unpredictable rates of individual threads/processes execution in parallel programs). However, if any of the system properties can be expressed formally, for example, in the form of a mathematical logic formula, then analysis of this property can be performed by verification methods. Normally process of the system verification consists of the following parts:

1. Construction of a mathematical model of the system under analysis.

2. Definition of the properties to be checked in the form of a formal text (also known as specification).

3. Building a formal proof of the presence or absence of the property being verified.

Usually, mathematical model of a system is a graph whose vertices are called states and represent situations (or situation classes), in which the system may be present at different times; whose edges can have labels depicting the actions system can perform. The functioning of the system in this model is represented by transitions along the edges of the graph from one state to another. If the passable edge has a label, then this label represents the action of the system, executed when passing from the state at the beginning of the edge to the state at its end.

The choice of the abstraction level for the system modelling depends on many factors (algorithmic solvability, astronomical dimensions of the model, the absence of effective methods of formal analysis of properties etc.). In this regard, the informal rules of this choice are reduced to the following: system model should not be over-specified, because the excessive model complexity may cause significant computational problems during its formal analysis. On the other hand, system model should not be oversimplified: it should reflect those aspects of the system that are relevant to the properties being verified, and preserve all the properties of the simulated system which are of interest for analysis.

Model checking (MC, [7]) approach is used to find a formal proof that the model does not meet its specification. In this paper we propose a new method of justifying that the model satisfies the specification. This paper focuses on the method of checking the conformity of a model an its specification, which uses the apparatus of TS and PN. Definition of a simple and labeled TS is given below, while definition of PN is well known and can be found in [7] if necessary.

Definition 1. A simple transition system is $A = (S, R, \alpha, \beta)$, where

S — finite or infinite set of states;

R — finite or infinite set of transitions;

α, β — two mappings from S to R , which make a correspondence between a transition $t \in R$ and two states $\alpha(t)$ and $\beta(t)$, which are called respectively the beginning and end of the transition t .

The transition t with beginning s and end s' is written as follows: $s \rightarrow s'$. Sometimes transitions may have a common beginning or end or both. This means that the pair $\alpha, \beta : R \rightarrow S$ is not necessarily an injective function. TS A is called finite if sets S, R are finite. If the set of states defines an initial state, such TS is labeled as $A = (S, R, \alpha, \beta, s_0)$ and is called initial TS. Such model as a simple TS may be enough to study properties of a model at a certain level of abstraction, but when it is necessary to carry out a more detailed analysis, labeled TS is more appropriate.

Definition 2. Let X be an alphabet. The labeled transition system (LTS) is an ordered six $A = (S, T, \alpha, \beta, s_0, h)$ where $(S, T, \alpha, \beta, s_0)$ represent a TS, and h is a mapping from T to X , which makes a correspondence between each transition t and its label $h(t) \in X$. LTS is finite if sets S, T, X are finite. Transition label $h(t)$ may also be called an action, and the transition itself is written as follows:

$(s, h(t), s')$ or $s \xrightarrow{h(t)} s'$. Transition label set is often accompanied by a special label τ , which represents an internal action of the system that is not visible at a given level of modelling.

Using LTS as a model of a real system allows analysing properties of actions associated with transitions, which is impossible by using a simple TS. Analysis and verification of the applications for graphic video adapters (NVIDIA in particular) was chosen as a subject of the study, because this area perfectly illustrates the impossibility of manual verification, as the number of threads allocated for solving the problem is measured by hundreds of thousands (in Pascal/Volta architectures). The use of LTS to construct a high-level model for substantiating the properties of a CUDA application was described in [8].

CUDA APPLICATION EXECUTION MODEL

A generalized execution model in the NVIDIA CUDA architecture, based on LTS and Petri nets was presented in [9]. The main emphasis was put on obtaining a high-level model without a detailed examination of the labels semantics of each transition system. The following briefly recalls the main details: three LTS were emerged as a result of CUDA application decomposition:

LTS $A = (S_1 = \{a_0, a_1, a_2, a_3\}, R_1, \alpha_1, \beta_1, a_0, h_1)$ — represents the warp that contains a set of instructions and sequentially executes them, where $X_1 = \{r_1, r_2, r_3, r_4\}$;

LTS $B = (S_2 = \{b_0, b_1, b_2, b_3, b_4, b_5\}, R_2, \alpha_2, \beta_2, b_0, h_2)$ — represents a generalized information instruction, where $X_2 = \{p_1, p_2, p_3, p_4, p_5, p_6\}$;

LTS $C = (S_3 = \{c_0, c_1, c_2, c_3\}, R_3, \alpha_3, \beta_3, c_0, h_3)$ — represents a thread block execution process (warp scheduling) on the SM, where $X_3 = \{q_1, q_2, q_3, q_4\}$.

Transition label functions h_1, h_2, h_3 are shown below in the charts (Fig. 1, 2, 3). The main activities modelled at this level of abstractions include planning, choosing warp and instruction for it and providing exclusive access to computational resources for the execution time of each warp-instruction tuple. This paper shows creation of a model of the lower level of abstraction, which will allow us to analyze the properties of actions associated with transitions, and the properties of the model in particular states.

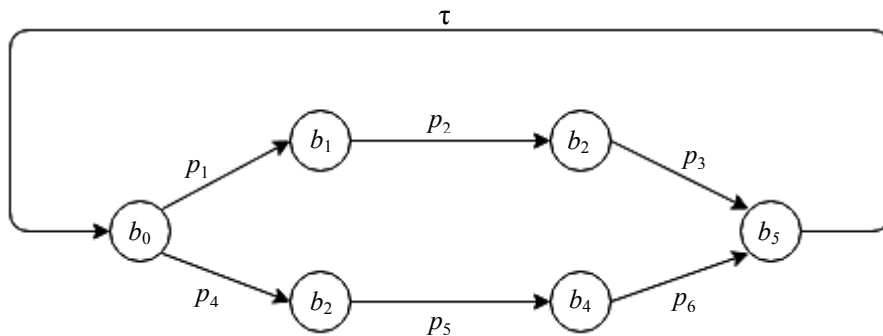


Fig. 1. LTS B of the information instruction

Definition 1. General TS (GTS) is a TS $A = (S, X, R, s_0, AP, L)$, where

S — set of states;

X — set of actions associated with transitions;

$R \subseteq S \times X \times S$ — transition relation;

s_0 — initial state from S ;

AP — set of propositional formulas associated with states;

$L: S \rightarrow B(AP)$ — state label function, where $B(AP)$ is the power set of AP .

We introduce a set of Boolean variables:

$$V_1 = \{instrAr, instrMem, instrArExec, instrMemExec, instrArFin, instrMemFin\},$$

where

$$h_1(\tau) = \{instrAr = 0, instrArExec = 0, instrArFin = 0, instrMem = 0, instrMemExec = 0, instrMemFin = 0\};$$

$$h_1(p_1) = \{instrAr = 1\} \text{ — arithmetical instruction selection;}$$

$$h_1(p_4) = \{instrMem = 1\} \text{ — global memory access selection;}$$

$$h_1(p_2) = \{instrArExec = 1\} \text{ — arithmetical instruction execution;}$$

$$h_1(p_5) = \{instrMemExec = 1\} \text{ — memory access execution;}$$

$$h_1(p_3) = \{instrArFin = 1\} \text{ — retrieving results of an arithmetical operation;}$$

$$h_1(p_6) = \{instrMemFin = 1\} \text{ — retrieving results of memory access operation.}$$

We define the state label function L_1 :

$$L_1(b_0) = \{instrAr \vee instrMem = 0\};$$

$$L_1(b_1) = \{instrAr = 1\};$$

$$L_1(b_2) = \{instrArExec = 1\};$$

$$L_1(b_3) = \{instrMem = 1\};$$

$$L_1(b_4) = \{instrMemExec = 1\};$$

$$L_1(b_5) = \{instrMemFin \vee instrArFin = 1\}.$$

We introduce a set of Boolean variables $V_2 = \{warpActive, warpBusy, warpFin\}$ that are set by the following transition labels (Fig. 2):

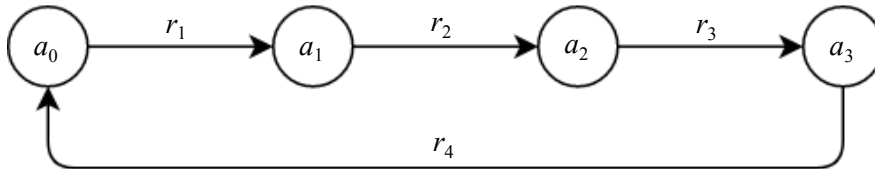


Fig. 2. LTS A of the instruction execution within a warp

$h_2(r_1) = \{warpActive = 1\}$ — warp activation — given warp gains access to SM resources to execute a single instruction;

$h_2(r_2) = \{warpBusy = 1\}$ — instruction is being executed by a warp;

$h_2(r_3) = \{warpFin = 1\}$ — instruction execution is finished;

$h_2(r_4) = \{warpActive = 0, warpBusy = 0, warpFin = 0\}$ — warp deactivation.

State label function L_2 is defined as following:

$L_2(a_0) = \{warpActive \vee warpBusy \vee warpFin = 0\};$

$L_2(a_1) = \{warpActive = 1\};$

$L_2(a_2) = \{warpBusy = 1\};$

$L_2(a_3) = \{warpFin = 1\}.$

We introduce a set of Boolean variables $V_3 = \{warpInstrSel, warpInstrExec, warpInstrFin\}$ that are set by the following transition labels (Fig. 3):

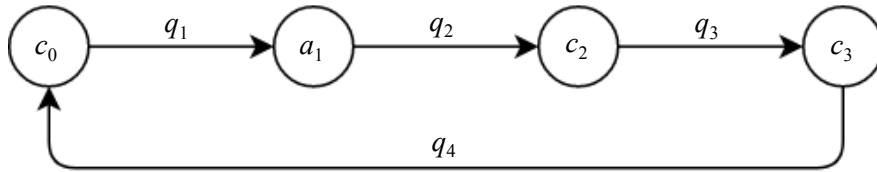


Fig. 3. LTS C of the warp scheduler

$h_3(q_1) = \{warpInstrSel = 1\}$ — warp and instruction selection;

$h_3(q_2) = \{warpInstrExec = 1\}$ — execution of the selected instruction by a selected warp;

$h_3(q_3) = \{warpInstrFin = 1\}$ — confirmation of instruction execution finish;

$h_3(q_4) = \{warpInstrSel = 0, warpInstrExec = 1, warpInstrFin = 1\}$ — transition to the next iteration.

State label function L_3 is defined as following:

$L_3(c_0) = \{warpInstrSel \vee warpInstrExec \vee warpInstrFin = 0\};$

$L_3(c_1) = \{warpInstrSel = 1\};$

$L_3(c_2) = \{warpInstrExec = 1\};$

$L_3(c_3) = \{warpInstrFin = 1\}.$

Integration of multiple TS into a holistic system that orchestrates the joint work of all subsystems is performed depending on the requirements of the component interaction model (synchronous, asynchronous, parallel, sequential). These interaction methods are introduced using different TS composition types and the general notion of TS. To analyze the model properties, we consider the concept of parallel composition of the TS.

There are several options of TS composition that model parallel functioning of multiple TS. The simplest one is the composition in which TSs work in parallel, but do not interact with each other. Such a composition is based on the concept of alternating actions (interleaving), which are performed by different subsystems of the composition. In this case, the order of actions performed by each TS is preserved. This is a common way of modelling parallel interactions, which is based on the assumption that the result of a parallel execution of operations coincides with the result of their sequential execution. The formal definition of this composition is as follows.

Let $A_i = (S_i, X_i, R_i, s^i_0, AP_i, L_i)$ — transition systems, where $i = 1, 2, \dots, n$. Parallel composition of TS A_1, A_2, \dots, A_n with interleaving is TS

$A = A_1 \parallel A_2 \parallel \dots \parallel A_n = (S, X, R, s_0, AP, L)$, where $S = S_1 \times S_2 \times \dots \times S_n$, $s_0 = (s^1_0, s^2_0, \dots, s^n_0)$, $X = X_1 \cup X_2 \cup \dots \cup X_n$, and transitions from R are defined as follows: transition $((s_1, \dots, s_n), x, (s'_1, \dots, s'_n)) \in R$ if and only if state (s'_1, \dots, s'_n) differs from (s_1, \dots, s_n) by the value of not more than one component, that is there exists $i \in \{1, 2, \dots, n\}$ such that $(s_i, x, s'_i) \in R_i$, where $x \in X_i \cup \{\tau\}$.

Note that in this composition TS can be repeated, that is, a given A_i may be included in TS A more than once. In addition, some of the TS A_i may be compositions of other TS: $A_i = A_{i1} \parallel A_{i2} \parallel \dots \parallel A_{in}$. Therefore one can describe the whole system in a structured form by using a parallel composition. However, there may be achievable states in the resulting TS which are not desirable. Therefore a parallel composition with alternating actions does not always reflect the real situation when the TSs should interact with each other. Handshaking composition concept is a more adequate approach to describe parallel interaction. Such option describes a situation when different TS are synchronized by the actions in which multiple TS participate simultaneously (data exchange etc.). Those actions are indicated by the same symbol in alphabets of all TS which take part in the interaction.

Definition 2. Let $A_i = (S_i, X_i, R_i, s^i_0, AP_i, L_i)$ — transition systems, where $i=1,2$. Parallel composition of TS A_1, A_2 with handshaking is a TS $A = A_1 \parallel A_2 = (S, X, R, s_0, AP, L)$, where $S = S_1 \times S_2$, $s_0 = (s^1_0, s^2_0)$, $X = X_1 \cup X_2$. Transitions that belong to T are defined as following:

if $x \in X_1 \cap X_2$ and $(s_1, x, s'_1) \in R_1$, $(s_2, x, s'_2) \in R_2$, then $((s_1, s_2), x, (s'_1, s'_2)) \in R$;

if $x \in X_1 \setminus X_2$ and $(s_1, x, s'_1) \in R_1$, then $((s_1, s_2), x, (s'_1, s_2)) \in R$;

if $x \in X_2 \setminus X_1$ and $(s_2, x, s'_2) \in R_2$, then $((s_1, s_2), x, (s_1, s'_2)) \in R$.

The following composition summarizes both previous concepts and is called a synchronized parallel composition or a synchronous product of TS (Fig. 4).

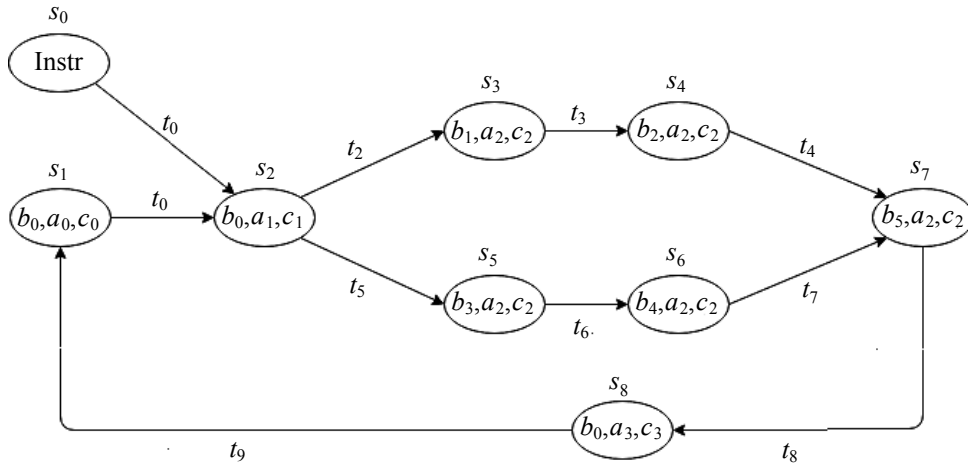


Fig. 4. GTS A of synchronous product of $LTS_A \times LTS_B \times LTS_C$

Definition 3. $A = (S_s, X_s, R_s, s_0, AP_s, L_s) = A_1 \times \dots \times A_n$, where $S_s = S_1 \times \dots \times S_n$, $s_0 = (s_0^1, \dots, s_0^n)$ is called a synchronous product of TS A_1, A_2, \dots, A_n where $A_i = (S_i, A_i, R_i, s_0^i, AP_i, L_i)$. Set of transitions is divided into two classes: asynchronous and synchronous. When synchronous transition from state s_1, \dots, s_n occurs, some of its components change simultaneously while the rest remains unchanged. To describe this the symbol ε is added to labels of each transition set R_i (i.e. $X_i \cup \{\varepsilon\}$) along with a corresponding transition (s_i, ε, s_i) . Such a symbol shows that state does not change during a particular transition. Subset R of $R_1 \times \dots \times R_n$ is called the synchronization constraints set. If TS is labeled, then the set of global transitions R corresponds to the set of transition labels X . In other words, $X \subseteq X_1 \times \dots \times X_n$, where X_i is an alphabet of LTS A_i , $i = 1, \dots, n$. An arbitrary element $A = (A_1, \dots, A_n, R)$ of the set X is called the LTS synchronization vector.

We build a model of application execution in the CUDA architecture in the form of synchronous product with the following global transitions (initial and final states of the transitions are omitted, since they are present in the model):

$$R = \{t_1 = (\varepsilon, r_1, q_1), t_2 = (p_1, r_2, q_2), t_3 = (p_2, \varepsilon, \varepsilon), t_4 = (p_3, \varepsilon, \varepsilon); \\ t_5 = (p_4, r_2, q_2), t_6 = (p_5, \varepsilon, \varepsilon), t_7 = (p_6, \varepsilon, \varepsilon), t_8 = (\tau, r_3, q_3), t_9 = (\varepsilon, r_4, q_4)\}.$$

DETERMINATION OF SEMANTIC CORRECTNESS OF THE MODEL ACTIONS

Definition function of L_i allow to find the context of correct system functioning at each transition in each state. For example, if LTS_B is in state b_0 , values of both $instrAr$ and $instrMem$ should be 0. If system was moved to a state b_0 after a number of allowed transitions and $instrMem$ flag is set, this will mean that system is not semantically correct, even if the corresponding PN meets all reliability criteria.

To analyze the semantic correctness of a model, it is necessary to determine the context of each global state and consider its conformity to the system under analysis. As shown above, transition system state context is specified by the state label function L_i , and the global state context can be retrieved by combining local state contexts of the components of the synchronous products. Each of the global states can be described as follows:

$$L(s_1) = L_1(b_0) \wedge L_2(a_0) \wedge L_3(c_0) = \\ = \{instrAr \vee instrMem = 0\} \wedge \{warpActive = 0\} \wedge \{warpInstrSel = 0\} \text{ —}$$

instruction is absent, warp is not active, scheduler is ready for planning;

$$L(s_2) = L_1(b_0) \wedge L_2(a_1) \wedge L_3(c_1) = \\ = \{instrAr \vee instrMem = 0\} \wedge \{warpActive = 1\} \wedge \{warpInstrSel = 1\} \text{ —}$$

warp and instruction are selected, but instruction type is not yet determined;

$$L(s_3) = L_1(b_1) \wedge L_2(a_2) \wedge L_3(c_2) = \\ = \{instrAr = 1\} \wedge \{warpBusy = 1\} \wedge \{warpInstrExec = 1\} \text{ —}$$

arithmetic instruction is selected for a given warp;

$$\begin{aligned} L(s_4) &= L_1(b_2) \wedge L_2(a_2) \wedge L_3(c_2) = \\ &= \{instrArExec = 1\} \wedge \{warpBusy = 1\} \wedge \{warpInstrExec = 1\} \text{ ---} \end{aligned}$$

given warp executes arithmetic operation;

$$\begin{aligned} L(s_5) &= L_1(b_3) \wedge L_2(a_2) \wedge L_3(c_2) = \\ &= \{instrMem = 1\} \wedge \{warpBusy = 1\} \wedge \{warpInstrExec = 1\} \text{ ---} \end{aligned}$$

memory access instruction is selected for a given warp;

$$\begin{aligned} L(s_6) &= L_1(b_4) \wedge L_2(a_2) \wedge L_3(c_2) = \\ &= \{instrMemExec = 1\} \wedge \{warpBusy = 1\} \wedge \{warpInstrExec = 1\} \text{ ---} \end{aligned}$$

given warp executes memory access instruction;

$$\begin{aligned} L(s_7) &= L_1(b_5) \wedge L_2(a_2) \wedge L_3(c_2) = \\ &= \{instrArFin \vee instrMemFin = 1\} \wedge \{warpBusy = 1\} \wedge \{warpInstrFin = 1\} \text{ ---} \end{aligned}$$

current instruction is executed regardless of its type;

$$\begin{aligned} L(s_8) &= L_1(b_0) \wedge L_2(a_3) \wedge L_3(c_3) = \\ &= \{instrAr \vee instrMem = 0\} \wedge \{warpFin = 1\} \wedge \{warpInstrFin = 1\} \text{ ---} \end{aligned}$$

instruction is absent, warp and scheduler finished their work.

All received contexts are semantically correct in accordance with the materials [1], thus, the set of global transitions of synchronous products is defined correctly.

In addition to the above, the semantic correctness of the model is provided by the following properties:

- *mutual exclusion*: a single instruction can be of one type only — either arithmetic or memory access;
- *fairness*: if the warp is active, scheduler must provide an instruction to be executed;
- *liveliness*: if the warp is active, one of the available types of instruction should be executed;
- *deactivation*: warp scheduler does not activate the warp if there are no instructions for execution.

Sequence $\sigma = s_0\alpha_1s_1\alpha_2\dots\alpha_ns_n$ is called a finite execution in GTS, where $(s_i, \alpha_i, s_{i+1}) \in T$ for $i = 0, 1, \dots, n-1$.

The following finite execution paths exist in the received LTS:

$$Path_1 = (s_1t_1s_2t_2s_3t_3s_4t_4s_7t_8s_8t_9s_1)^*;$$

$$Path_2 = (s_1t_1s_2t_5s_5t_6s_6t_7s_7t_8s_8t_9s_1)^*,$$

where $*$ denotes the iterative operation of regular language. The language that corresponds to the paths above is:

$$L(Path) = (L(Path_1) \vee L(Path_2))^* = (t_1(t_2t_3t_4 \vee t_5t_6t_7)t_8t_9)^*.$$

And this language must correspond to the formula:

$$L(s_1)L(s_2)L(s_3)L(s_4) \vee L(s_5)L(s_6)L(s_7)L(s_8).$$

Let's check the semantic correctness properties of the model:

- *mutual exclusion*: the following expression is always true:
 $L(s_3) \vee L(s_5) = L_2(a_2)L_3(c_2)(L_1(b_1) \vee L_1(b_3))$;
- *fairness*: always $L_2(a_1) \rightarrow L_3(c_3)$;
- *liveliness*: always $L_2(a_1) \rightarrow L_1(b_1)L_2(a_2)L_3(c_2) \vee L_1(b_3)L_2(a_2)L_3(c_2)$;
- *deactivation*: $L_3(c_0) \rightarrow L_1(b_0)L_2(a_0)$.

FORMALIZATION OF THE GTS ANALYSIS PROCESS

Analysis process presented in the previous section can be generalized and reused for any GTS.

Definition 1. The subset of a set $B(AP)^*$, where $B(AP)$ is the power set of the set AP , is called the linear-temporal property P over the set of atomic propositional formulas AP . Consequently, $P \subseteq (B(AP))^*$.

In our case set $(B(AP))^*$ is a set of words of finite length constructed from concatenated formulas of $B(AP)$. Assume there is $A = (S, X, R, I, AP, L)$ and $\pi = s_0s_1\dots s_n$ is a sequence of states. Such sequence is called a path fragment if $s_{i+1} \in Post(s_i)$ where $Post(s_i) = \cup Post(s_i, x)$, and $Post(s_i, x) = \{s' \in S : (s_i, x, s') \in R\}$, $x \in X$. Sequence $\pi = s_0s_1\dots s_n$ for which $Post(s_n) = \emptyset$ is called the maximal path fragment.

Definition 2. Word $L(s_0)L(s_1)\dots L(s_n)$ is called a trace ($trace(\pi)$) of a finite sequence π . Consequently, the set of traces is a set of finite words over the alphabet of the propositional formulas $B(AP)$ which are executed in states of this sequence. Denote $trace(\Pi) = \{trace(\pi) | \pi \in \Pi\}$, $trace(s) = trace(Path(s))$ and $Traces(A) = \cup_{s \in I} trace(s)$ where $Path(s)$ is maximal fragment of path π that begins in state s .

In our case, the set of AP propositional formulas includes the following items:
 $AP = \{instrAr, instrMem, instrArExec, instrMemExec, instrArFin, instrMemFin, warpActive, warpInstrSel, warpBusy, warpInstrExec, warpFin, warpInstrFin\}$.

The following words are traces of GTS A :

$$p_1 = L(s_2)L(s_3)L(s_4)L(s_7)L(s_8)L(s_1);$$

$$p_2 = L(s_2)L(s_5)L(s_6)L(s_7)L(s_8)L(s_1).$$

Let's define bad prefixes in these words as prefixes that violate the truth of $L(s_i)$, which mean the following words:

$$p'_1 = \overline{L(s_2)L(s_3)} \dots L(s_1);$$

$$p'_2 = L(s_2)\overline{L(s_3)} \dots L(s_1);$$

$$p'_3 = L(s_2)\overline{L(s_5)} \dots L(s_1);$$

.....

$$p'_k = L(s_2)\overline{L(s_1)}.$$

Therefore language $BadPref(A) = (p_1 \vee p_2)^* (p'_1 \vee \dots \vee p'_k)$ is regular and is accepted by a finite automaton $B = (Q, AP, f, Q_0, F)$ where $Q_0 \cap F = \emptyset$ that is shown at Fig. 5.

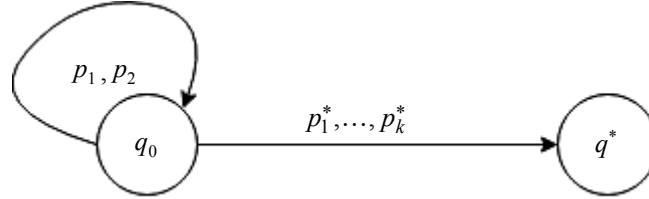


Fig. 5. Automaton B that accepts bad prefixes, $q_0 \in Q_0$, $q_0 \notin F$, $q^* \in F$

We will connect the GTS from Fig. 4 and automaton B with such a product $A \times B$ that produces GTS \bar{A} as result so that:

$$S' = S \times Q,$$

R' — the smallest relation defined by the rule

$$\frac{s \xrightarrow{x} s' \wedge q \xrightarrow{L(s')} q'}{(s, q) \xrightarrow{x} (s', q')},$$

$$I' = \{(s_0, q_0) \mid s_0 \in I \wedge \exists q_0 \in Q_0 : q_0 \xrightarrow{L(s_0)} q\},$$

$$AP' = Q,$$

$$LS \times Q \rightarrow B(Q) \text{ where } L'(s, q) = \{q\}.$$

Then the correctness of GTS A functioning is expressed as a condition

$$Traces(A) \cap BadPref(A) = Traces(A) \cap L(B) = \emptyset,$$

where $L(B)$ is the language accepted by the automaton B . Consequently, if P is a property whose execution guarantees the correct functioning of GTS A , then $Traces(A) \cap BadPref(P) = \emptyset$. Thus, GTS \bar{A} will look as shown on Fig. 6. There is a transition from each vertex, except for (s_0, q_0) , to the state (s_2, q^*) . Such transitions represent one of the bad prefix traces. For the sake of clarity only 4 examples of bad prefixes are left in Fig. 6.

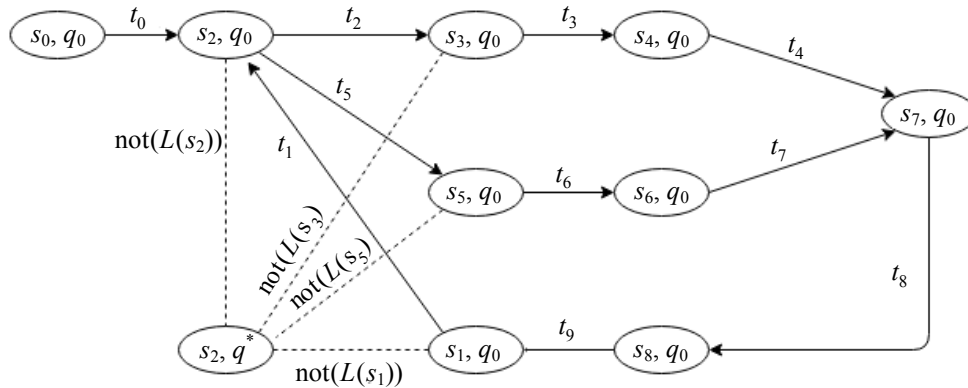


Fig. 6. GTS of $A \times B$ product, where $q^* \in F$

DETERMINATION OF MODEL CORRECTNESS ON THE HIGH LEVEL OF ABSTRACTION

In addition to semantic verification, it is important to check the redundancy of the system, deadlock/trap balance etc. To do this, we use verification at the highest level of abstraction, without need of transition label semantics analysis.

The translation of the received synchronous product into the PN gives the network shown in Fig. 7. It is described in [10], [11] that the TS product semantics and the semantics of the corresponding PN are consistent in the sense that a sequence of global transitions represents the global history of the TS product A if and only if it is an admissible sequence of transitions in the PN. Accordingly, elements of the set R become transitions of the PN, and the global states of the TS product (the set of states of each TS involved in the synchronous product before or after the global transition) become the places of the received network. We build a PN by using the synchronization constraints set, and such a network simulates the interoperability of all subsystems. Recall that NVIDIA video adapters operate with multiple warps at the same time, so there is a situation of synchronous and asynchronous execution, since different instructions may have different execution times and will not reach synchronization location at the same time. Therefore using the Petri net apparatus is expedient.

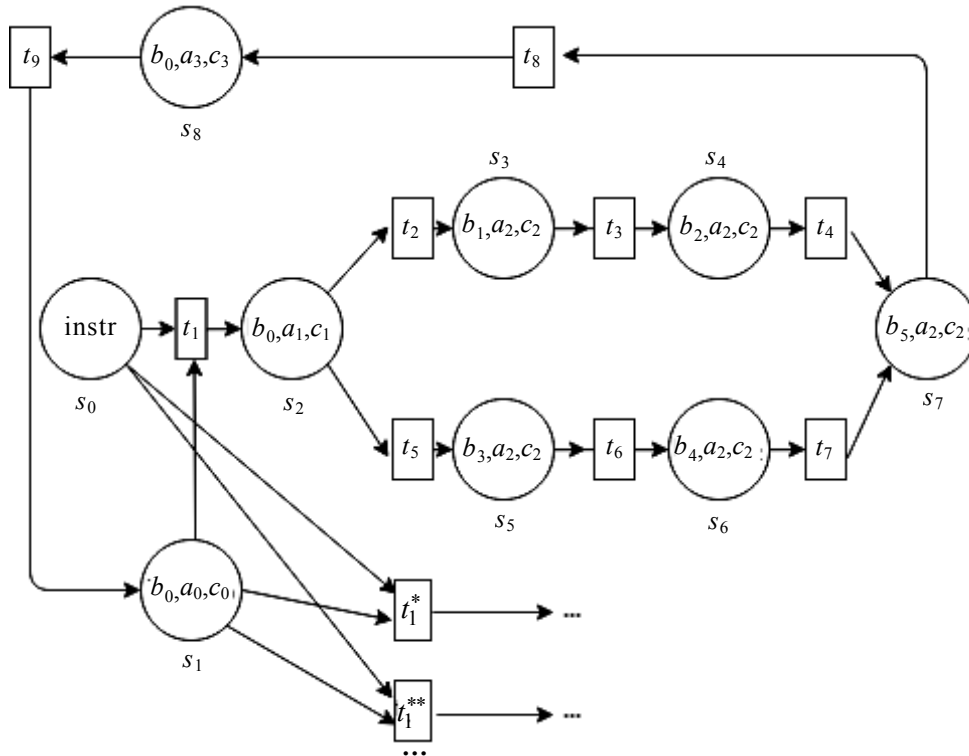


Fig. 7. PN that represents synchronous product $LTS_A \times LTS_B \times LTS_C$

Let's analyze the presence and number of dead transitions in PN for one warp case. To do this, we form a state equation $A \cdot x + M_0 - M_k =$

$= A \cdot x + d = 0$ where $M_0 = (1, 1, 0, 0, 0, 0, 0, 0, 0, 0)$, $M_k = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0)$, $d = -(M_k - M_0)$ (see Table 1).

Table 1. Petri net state equation matrix

s	t									$-(M_k - M_0)$
	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	
s_0	-1	0	0	0	0	0	0	0	0	1
s_1	-1	0	0	0	0	0	0	0	1	0
s_2	1	-1	0	0	-1	0	0	0	0	0
s_3	0	1	-1	0	0	0	0	0	0	0
s_4	0	0	1	-1	0	0	0	0	0	0
s_5	0	0	0	0	1	-1	0	0	0	0
s_6	0	0	0	0	0	1	-1	0	0	0
s_7	0	0	0	1	0	0	1	-1	0	0
s_8	0	0	0	0	0	0	0	1	-1	0

Applying the TSS-algorithm [9] to solve the state equation with the above matrix, we obtain the following solutions (Table 2).

Table 2. Petri net state equation solutions

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
1	1	1	1	0	0	0	1	1
1	0	0	0	1	1	1	1	1

As can be seen from the set of solutions, all transitions in the PN with the initial and final markings given above are alive (the value corresponding to each transition is positive at least in one of the solutions). In addition, the property of *mutual exclusion* has been verified: transitions that correspond only to one of the possible types of instructions are performed at each point of time.

The analysis of the properties of the received PN also showed the absence of deadlocks, verified limitation and controllability [10]. As described above, this analysis did not take into account the properties of actions in transitions and properties associated with the states of the model.

CONCLUSION

This paper proposes the use of the TS mathematical apparatus to obtain a formalized system specification in the GPGPU technology. The advantages of the existing model include the ability to reduce synchronous product to a PN, which allows for further verification by automated means. The ability to study the characteristics of the model created by the combination of LTS and PN apparatuses is shown. The model was analyzed to verify there are no dead transitions and places (without taking into account the semantics of the transition labels), and a separate analysis was performed to verify the semantic correctness of the model actions.

As a result of these actions, the proof of the correct construction of the model was obtained. The developed approach allows to simplify and reduce the processes of verification and testing of multi-threaded applications in computer systems that utilize video adapters.

REFERENCES

1. Nvidia Data Center – Nvidia, 2018. [Online]. Available: <https://www.nvidia.com/en-us/data-center/>. Accessed on: 2019, March 14.
2. TOP500 Lists - TOP500 Supercomputer Sites, 2018. [Online]. Available: <https://www.top500.org/lists>. Accessed on: 2019, March 14.
3. A.V. Anisimov, S.D. Pogorilyy, and D.Yu. Vitel, “About the Issue of Algorithms formalized Design for Parallel Computer Architectures”, *Applied and Computational Mathematics*, vol. 12, no. 2, pp.140–151, 2013.
4. A. Arnold, *Finite Transition Systems: Semantics of Communicating Systems*. Paris, France: Prentice Hall, 1994, 177 p.
5. T. Murata, “Petri nets: properties, analysis and applications”, in *Proc. of the IEEE*, 77:541.80, 1989.
6. M. Ben-Ari, *Mathematical Logic for Computer Science*. UK: Prentice Hall International Ltd, 1993, 305 p.
7. E.M. Clarke, Jr., O. Grumberg, and D.A. Peled, *Model Checking*. USA: MIT Press, 1999.
8. S.D. Pogorilyy, S.L. Kryvyi, and M.S. Slynko, “Model justification of GPU-based applications”, *Control Systems and Computers*, vol. 4, pp. 46–56, 2018.
9. S.L. Kryvyi, *Linear Diophantine constraints and their applications*. Chernivtsi: Bukrek Publishing House, 2015.
10. S.L. Kryvyi, S.D. Pogorilyy, and M.S. Slynko, “Transition systems as method of designing applications in GPGPU technology”, in *Proc. 11-th international scientific and practical conference on programming UkrPROG'2018*.
11. S.L. Kryvyi et al., “Design of Grid Structures on the Basis of Transition Systems with the Substantiation of the Correctness of Their Operation”, *Cybernetics and Systems Analysis*, vol. 53, no. 1, pp.105–114, New York, USA: Springer Science + Business Media, January 2017.

Received 15.10.2020

INFORMATION ON THE ARTICLE

Serhii L. Kryvyi, ORCID: 0000-0003-4231-0691, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: sl.krivoi@gmail.com

Sergiy D. Pogorilyy, ORCID: 0000-0002-6497-5056, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: sdp77@i.ua, sdp@univ.net.ua

Maksym S. Slynko, ORCID: 0000-0001-9667-8729, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: maxim.slinko@gmail.com

Artem A. Kramov, ORCID: 0000-0003-3631-1268, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: artemkramov@gmail.com

МЕТОД СЕМАНТИЧНОЇ ВЕРИФІКАЦІЇ ЗАСТОСУВАНЬ У ТЕХНОЛОГІЇ GPGPU /
С.Л. Кривий, С.Д. Погорілий, М.С. Слинко, А.А. Крамов

Анотація. Запропоновано метод розроблення та верифікації застосувань для систем з масовим паралелізмом на основі відеоадаптерів від компанії NVIDIA,

який дозволяє створювати абстракції різних рівнів за допомогою апарата розмічених транзійних систем. Композиції таких систем трансформуються в мережі Петрі, які далі аналізуються відповідними засобами. Метод також дає змогу створювати моделі на різних рівнях абстракції, а їх властивості можуть специфікуватися формулами темпоральної логіки. Це дозволяє досліджувати властивості систем з масовим паралелізмом, які майже неможливо аналізувати вручну, оскільки кількість потоків у новітніх архітектурах відеоадаптерів (Pascal, Volta, Ampere, Тюрінг), виділених для виконання коду, вимірюється сотнями тисяч або мільйонами.

Ключові слова: CUDA, графічні процесори (GPU), графічні обчислення загального призначення (GPGPU), транзійна система, мережа Петрі, побутова модель.

МЕТОД СЕМАНТИЧЕСКОЙ ВЕРИФИКАЦИИ ПРИЛОЖЕНИЙ В ТЕХНОЛОГИИ GPGPU / С.Л. Крывый, С.Д. Погорелый, М.С. Слинко, А.А. Крамов

Аннотация. Предложен метод разработки и верификации приложений для систем с массовым параллелизмом на основе видеоадаптеров от компании NVIDIA, который позволяет создавать абстракции различных уровней с помощью аппарата размеченных транзисционных систем. Композиции таких систем трансформируются в сети Петри, которые далее анализируются соответствующими средствами. Метод позволяет создавать модели на различных уровнях абстракции, а их свойства могут специфицироваться формулами темпоральной логики. Это позволяет исследовать свойства систем с массовым параллелизмом, которые практически невозможно анализировать вручную, так как количество потоков в новейших архитектурах видеоадаптеров (Pascal, Volta, Ampere, Тьюринг), выделенных для выполнения кода, измеряется сотнями тысяч или миллионами.

Ключевые слова: CUDA, графические процессоры (GPU), графические вычисления общего назначения (GPGPU), транзисционная система, сеть Петри, построение модели.

ДОСЛІДЖЕННЯ КЛАСУ АЛГЕБРИЧНИХ ФУНКЦІЙ ДЛЯ ІНТЕРПОЛЯЦІЇ МЕЖОВИХ ТРАЄКТОРІЙ КОРОТКОФОКУСНИХ ЕЛЕКТРОННИХ ПУЧКІВ

І.В. МЕЛЬНИК, А.В. ПОЧИНОК

Анотація. Досліджено метод інтерполяції залежності радіуса електронного пучка від положення площини зрізу по поздовжній координаті, оснований на використанні спеціального класу інтерполяційних функцій. Як клас інтерполяційних функцій, які відповідають яружним залежностям радіуса короткофокусного електронного пучка від відстані, обрано корені степеня n з поліномів того ж самого степеня. Результати тестування показали, що, залежно від обраного набору відлікових точок, їх положення відносно області мінімуму та кількості, відносна похибка інтерполяції лежить у межах 0,7 – 8%. Отримані результати математичного моделювання фокальних параметрів електронних пучків є вкрай важливими для фахівців, які займаються розробленням сучасного високоефективного електронно-променевого обладнання та його впровадженням у промислове виробництво.

Ключові слова: електронний пучок, транспортування електронного пучка, фокальна область, інтерполяція, клас алгебричних функцій.

ВСТУП

Сьогодні електронно-променевої технології широко впроваджуються у різних галузях промисловості, зокрема в машинобудуванні, авіаційній та космічній промисловості, електронній промисловості та приладобудуванні [1–6]. Головними перевагами електронно-променевих технологій над іншими способами високотемпературного оброблення речовини, зокрема лазерними та плазмовими технологіями, є такі [1–3]:

- 1) висока загальна потужність електронного пучка, значення якої може досягати сотень кіловатів;
- 2) висока питома потужність електронного пучка, значення якої може становити 10^6 Вт/м²;
- 3) проведення технологічних операцій, пов'язаних з нагріванням речовини, у вакуумі, що забезпечує високу хімічну чистоту оброблюваної речовини;
- 4) простота керування просторовими та енергетичними параметрами електронних пучків з використанням електричних та магнітних полів.

Крім наведених переваг електронно-променевих технологій, важливою є також можливість забезпечення роботи джерел електронів в імпульсному режимі, яка дозволяє використовувати різноманітні фізичні ефекти, пов'язані з обробленням речовини тепловими імпульсами заданої потужності, тривалості та частоти [1–3].

Джерела електронів технологічного використання розробляються за двома головними напрямками: 1) удосконалення та впровадження у промис-

ловість традиційних джерел електронів з розжарюваними катодами [1–3]; 2) розроблення джерел електронів, які працюють за новими фізичними принципами. Серед таких джерел електронів окреме місце займають джерела, у яких емісія заряджених частинок здійснюється з іонізованого газу, зокрема джерела електронів високовольтного тліючого розряду (ВТР) [4–7].

Зрозуміло, що розроблення та впровадження у промисловість сучасних електронно-променевих технологій неможливі без ефективного використання методів математичного моделювання. Зокрема, у працях [1, 8–11] відмічалась важливість проведення інтерполяції та апроксимації енергетичних і просторових параметрів електронних пучків. Оцінки параметрів електронних пучків з використанням методів інтерполяції та апроксимації є особливо важливими для таких технологічних процесів, як зварювання, розмірне оброблення виробів та сучасні адитивні технології з наплавленням металів на плоску поверхню. Для таких процесів передбачення закономірності розподілу густини струму електронного пучка у фокальній площині з метою оптимізації теплової дії електронного пучка є вкрай важливою науково-технічною проблемою, особливо для джерел електронів, які працюють у складних фізичних умовах низького вакууму, зокрема для джерел електронів на основі ВТР [9–11].

У загальному вигляді завдання інтерполяції межових траєкторій короткофокусних електронних пучків, які формуються електронними гарматами ВТР, та відповідні методи інтерполяції розглядалися у працях [9, 10]. У цій роботі розглядатиметься окремий важливий клас алгебричних функцій, властивості яких дозволяють за одну ітерацію проводити інтерполяцію межових траєкторій електронних пучків, що формуються джерелами електронів ВТР, без розбиття цих траєкторій на окремі відрізки та з невеликими похибками. Для розв'язання поставленого завдання інтерполяції як базові точки та як достовірні числові дані бралися результати числових розрахунків просторових параметрів електронних пучків, що формуються джерелами електронів ВТР, за умови їх транспортування у низькому вакуумі з компенсацією просторового заряду електронів іонами залишкового газу [12].

ПОСТАНОВКА ЗАВДАННЯ

У загальному вигляді завдання інтерполяції межової траєкторії короткофокусного електронного пучка за умови його транспортування в низькому вакуумі було поставлено у праці [12]. Така постановка завдання ґрунтується на загальнотеоретичних положеннях, сформульованих наприкінці ХХ ст. у працях [8, 13, 14], які залишаються актуальними дотепер. За будь-яких фізичних умов завдання інтерполяції значення радіуса короткофокусного електронного пучка r відносно поздовжньої координати h можна формулювати таким чином. Необхідно знайти таку функцію $r(h)$, яка має один глобальний мінімум $r_{\min} = r(h_{\phi})$ і є симетричною відносно фокальної площини $h = h_{\phi}$, тобто задовольняє умову [8, 13]

$$\frac{dr(h_{\phi} - \delta h)}{dh} = -\frac{dr(h_{\phi} + \delta h)}{dh}, \quad (1)$$

де δh — нескінченно мала зміна величини h .

Важливим завданням є також визначення оптимальної кількості базових точок для розв'язування завдання інтерполяції з наперед заданою точністю [15, 16]. Наприклад, у працях [9, 10] розглянуто лише окремий випадок $n_{\text{int}} = 3$, а трьох базових точок зазвичай може виявитись недостатньо для досягнення потрібної точності.

Завдання інтерполяції з використанням формули (1) зазвичай розв'язують для точних розрахункових даних, які не мають додаткової експериментальної похибки [15, 16]. У роботі використано фізико-математичну модель межової траєкторії короткофокусного електронного пучка, який транспортується в низькому вакуумі за умови компенсації просторового заряду електронів іонами залишкового газу, описану у праці [12].

Важливою є також оцінка точності інтерполяції щодо даних числових розрахунків, згідно з якими взято базові точки [15, 16].

ЧИСЛОВІ СПІВВІДНОШЕННЯ ДЛЯ РОЗРАХУНКУ МЕЖОВОЇ ТРАЄКТОРІЇ ЕЛЕКТРОННОГО ПУЧКА ТА РЕЗУЛЬТАТИ ЧИСЛОВОГО МОДЕЛЮВАННЯ

Як показано й обґрунтовано у праці [12], для числового моделювання процесу транспортування короткофокусного електронного пучка в низькому вакуумі необхідно враховувати такі фізичні ефекти та процеси [8, 13, 14]:

- іонізацію залишкового газу електронами пучка;
- розсіювання електронів на атомах залишкового газу за моделлю Резерфорда;
- дефокусування пучка під дією власного просторового заряду електронів;
- фокусування пучка під дією власного просторового заряду іонів;
- магнітне самофокусування електронного пучка в іонізованому газі або пінч-ефект.

У праці [12] на основі проведених числових експериментів обґрунтовано, що магнітне самофокусування істотно не впливає на межову траєкторію короткофокусного електронного пучка. Теоретичні оцінки та результати комп'ютерних розрахунків показали, що для пучків з кутом збіжності, більшим за 15° , енергією електронів 10–30 кеВ та струмом 1,5–2 А вплив магнітного самофокусування на межову траєкторію пучка становить не більше ніж 0,1%.

Розглянемо відповідну систему рівнянь, числове розв'язування якої дозволяє знайти межову траєкторію короткофокусного електронного пучка за умови його транспортування в низькому вакуумі та компенсації просторового заряду електронів іонами залишкового газу.

1. Система алгебро-диференціальних рівнянь для моделювання процесу вільного дрейфу електронів пучка з урахуванням їх власного просторового заряду та компенсації іонами залишкового газу [8, 12–14]:

$$f = \frac{n_e}{n_{i0} - n_e}; C = \frac{I_{\Pi}(1 - f - \beta^2)}{4\pi\epsilon_0\sqrt{\frac{2e}{m_e}U_{\text{пр}}^{3/2}}}; \frac{d^2r_{\Pi}}{dz^2} = \frac{C}{r_{\Pi}};$$

$$\theta = \frac{dr_n}{dz} + \theta_s; n_e = \frac{I_{\Pi}}{\pi r_{\Pi}^2}; v_e = \sqrt{\frac{2eU_{\text{пр}}}{m_e}} \quad (2)$$

де I_{Π} — струм електронного пучка; r_{Π} — радіус електронного пучка; $U_{\text{пр}}$ — прискорювальна напруга; f — рівень іонізації залишкового газу; n_e — концентрація електронів пучка; n_{i0} — концентрація іонів залишкового газу на осі симетрії пучка; m_e — маса електрона; ϵ_0 — діелектрична стала; $\beta = v_e / c$ — відношення усередненої швидкості електронів пучка v_e до швидкості світла c .

2. Алгебричне рівняння для визначення концентрації іонів залишкового газу на осі симетрії електронного пучка n_{i0} [8, 12–14]:

$$n_{i0} = \sqrt{\pi} r_{\Pi}^2 B_i p n_e \sqrt{\frac{M\epsilon_0 n_e}{m_e U_{\text{пр}}}} \exp\left(-\frac{U_{\text{пр}}}{\epsilon_0 n_e r_{\Pi}^2}\right), \quad (3)$$

де p — тиск залишкового газу; B_i — рівень іонізації газу.

3. Система алгебричних рівнянь для розрахунку середнього кута розсіювання електронів пучка на атомах залишкового газу згідно з моделлю Резерфорда має вигляд [8, 12–14]:

$$\text{tg}\left(\frac{\theta_{\min}}{2}\right) = \frac{10^{-4} Z_a^{4/3}}{2\gamma\beta^2}; \text{tg}\left(\frac{\theta_{\max}}{2}\right) = \frac{Z_a^{3/2}}{2\gamma\beta^2};$$

$$\bar{\theta}^2 = \frac{8\pi r_{\Pi}^2 n Z_a^2 dz}{\beta^4 \gamma^2} \ln\left(\frac{\theta_{\max}}{\theta_{\min}}\right); \gamma = \sqrt{1 - \frac{v_e^2}{c^2}}, \quad (4)$$

де θ_{\min} , θ_{\max} — мінімальний та максимальний кут розсіювання відповідно; Z_a — заряд ядра атомів залишкового газу; n — концентрація атомів залишкового газу; dz — довжина пробігу електронів у поздовжньому напрямку на поточній ітерації; $\bar{\theta}$ — середній кут розсіювання електронів пучка; γ — релятивістський фактор.

Результати розрахунку межевої траєкторії електронного пучка для різних значень прискорювальної напруги за умови $I_{\Pi} = 0,5 \text{ А}$ і $p = 0,1 \text{ Па}$ показано на рис. 1 [10].

Як видно з отриманих результатів комп'ютерних розрахунків, залежності для межевої траєкторії короткофокусного електронного пучка від відстані $r(z)$ мають один глобальний мінімум у фокальній площині, а за межами області максимуму ці залежності близькі до лінійних. Згідно з такою поведінкою функції, заданої даними числового моделювання, будемо шукати відповідну функцію інтерполяції.

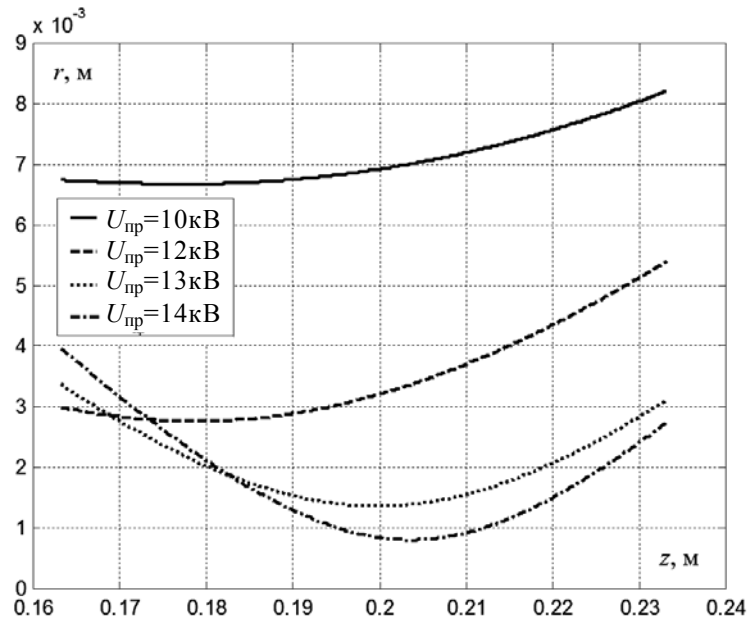


Рис. 1. Результати розрахунку межевої траєкторії короткофокусного електронного пучка, сформованого джерелом електронів ВТР, за умови використання азоту як робочого газу

УЗАГАЛЬНЕНА ПОСТАНОВКА ЗАВДАННЯ ІНТЕРПОЛЯЦІЇ ТА ВИЗНАЧЕННЯ КЛАСУ ІНТЕРПОЛЯЦІЙНИХ ФУНКЦІЙ

Згідно з графічними залежностями (рис. 1) межові траєкторії електронного пучка за умови його транспортування в низькому вакуумі з компенсацією просторового заряду електронів іонами залишкового газу можна інтерполювати алгебричними функціями, які мають один глобальний мінімум, а за межами області мінімуму залежність $r(z)$ має лінійний характер. Із теорії алгебричних функцій відомо, що саме таким умовам відповідає клас функцій, які формуються як корені степеня n з поліномів степеня n , тобто [15, 16]

$$r(h) = \sqrt[n]{C_n h_i^n + C_{n-1} h_i^{n-1} + \dots + C_2 h_i^2 + C_1 h_i + C_0}, \quad (5)$$

де n — порядок функції інтерполяції.

Для використання функції інтерполяції (5) необхідно задати набір базових відлікових точок r_i таким чином, щоб однозначно визначити значення коефіцієнтів C_0, \dots, C_n . Зрозуміло, що для функції апроксимації порядку n кількість відлікових точок має бути $n+1$. Тоді для набору базових відлікових точок (r_i, h_i) , $i \in [0, \dots, n]$ можна на основі співвідношення (5) записати несуперечливу та замкнену систему з $n+1$ рівнянь, у якій кількість невідомих також становить $n+1$:

$$r_i = \sqrt[n]{C_n h_i^n + C_{n-1} h_i^{n-1} + \dots + C_2 h_i^2 + C_1 h_i + C_0}, \quad i \in [0, \dots, n]. \quad (6)$$

Після піднесення всіх рівнянь системи (6) до степеня n отримуємо відповідну систему лінійних рівнянь для коефіцієнтів C_0, \dots, C_n :

$$C_n h_i^n + C_{n-1} h_i^{n-1} + \dots + C_2 h_i^2 + C_1 h_i + C_0 = r_i^n, \quad i \in [0, \dots, n]. \quad (7)$$

За умови відомих координат базових точок $P_1(h_1, r_1), P_2(h_2, r_2), \dots, P_n(h_n, r_n)$ та $P_{n+1}(h_{n+1}, r_{n+1})$ можна розрахувати відповідні коефіцієнти C_0, \dots, C_n функції інтерполяції (5) як розв’язок системи лінійних рівнянь (7), і така задача, за умови додатних значень координат r_1, \dots, r_n , завжди має однозначний розв’язок. Для розв’язування системи рівнянь (7) можна використати будь-який з відомих методів розв’язання систем лінійних рівнянь, наприклад метод виключення змінних Гауса–Зейделя. Тоді розв’язок системи рівнянь (7) можна шукати у вигляді аналітичних співвідношень для коефіцієнтів C_0, \dots, C_n , що дозволяє використовувати функцію інтерполяції (5) для будь-якого визначеного набору базових точок.

Розглянемо аналітичні співвідношення для визначення коефіцієнтів C_0, \dots, C_n для функцій, заданих співвідношенням (5) за умов $n = 2$, $n = 3$, $n = 4$ і $n = 5$.

АНАЛІТИЧНІ СПІВВІДНОШЕННЯ ДЛЯ РОЗРАХУНКУ КОЕФІЦІЄНТІВ ФУНКЦІЙ ІНТЕРПОЛЯЦІЇ МЕЖОВОЇ ТРАЄКТОРІЇ ЕЛЕКТРОННОГО ПУЧКА ВІД ДРУГОГО ДО П’ЯТОГО ПОРЯДКУ

Функція інтерполяції другого порядку. У разі $n = 2$ функція інтерполяції (5) має такий вигляд:

$$r(h) = \sqrt{C_2 h^2 + C_1 h + C_0}. \quad (8)$$

Узагальнену систему рівнянь (7) для аналітичного співвідношення (8) можна переписати у вигляді системи з трьох лінійних рівнянь з трьома невідомими, відповідно кількість відлікових точок дорівнює 3. Якщо ці точки мають координати $P_1(h_1, r_1), P_2(h_2, r_2)$ і $P_3(h_3, r_3)$, систему лінійних рівнянь (7) для визначення коефіцієнтів C_0, C_1 і C_2 можна згідно зі співвідношенням (8) переписати у вигляді

$$\begin{cases} C_2 h_1^2 + C_1 h_1 + C_0 = r_1^2; \\ C_2 h_2^2 + C_1 h_2 + C_0 = r_2^2; \\ C_2 h_3^2 + C_1 h_3 + C_0 = r_3^2. \end{cases} \quad (9)$$

Аналітичний розв’язок системи рівнянь (9) матиме вигляд:

$$\begin{aligned} k_1 &= \frac{r_2^2 - r_1^2}{h_2 - h_1}; \quad k_2 = \frac{r_3^2 - r_1^2}{h_3 - h_1}; \quad k_3 = \frac{h_2 + h_1}{h_3 - h_2}; \\ C_2 &= \frac{k_2 - k_1}{h_3 - h_2}; \quad C_1 = k_1 - (k_2 - k_1)k_3; \\ C_0 &= r_1^2 - \frac{k_2 - k_1}{h_3 - h_2} h_1^2 - (k_1 - (k_2 - k_1)k_3)h_1. \end{aligned} \quad (10)$$

За умови відомих поліноміальних коефіцієнтів C_0, C_1 і C_2 , які визначаються співвідношенням (10), положення фокуса електронного пучка h_{f2} можна визначити як точку мінімуму функції (5), а фокальний радіус електронного пучка — як значення цієї функції (r_{f2}, h_{f2}) . Оскільки функція квадратного кореня для додатних значень h завжди існує і є монотонно зростаючою, мінімум функції $r(h)$, заданої співвідношенням (8), відповідає мінімуму параболічної функції $C_2 h^2 + C_1 h + C_0$. Тоді фокальні параметри електронного пучка за умови використання інтерполяційної функції (8) та відомих значень поліноміальних коефіцієнтів, визначених співвідношенням (10), можна розрахувати так:

$$h_{f2} = -\frac{C_1}{2C_2}; \quad r_{f2} = \sqrt{C_0 - \frac{C_1^2}{4C_2}}. \quad (11)$$

Функція інтерполяції третього порядку. У разі $n = 3$ функція інтерполяції (5) має вигляд

$$r(h) = \sqrt[3]{C_3 h^3 + C_2 h^2 + C_1 h + C_0}. \quad (12)$$

Узагальнену систему рівнянь (7) для аналітичного співвідношення (12) можна переписати у вигляді системи з чотирьох лінійних рівнянь з чотирма невідомими відповідно, кількість відлікових точок тоді дорівнюватиме 4. Якщо ці точки мають координати $P_1(h_1, r_1), P_2(h_2, r_2), P_3(h_3, r_3)$ і $P_4(h_4, r_4)$, систему лінійних рівнянь (7) для визначення коефіцієнтів C_0, C_1, C_2 та C_3 можна, згідно зі співвідношенням (12), переписати таким чином:

$$\begin{cases} C_3 h_1^3 + C_2 h_1^2 + C_1 h_1 + C_0 = r_1^3; \\ C_3 h_2^3 + C_2 h_2^2 + C_1 h_2 + C_0 = r_2^3; \\ C_3 h_3^3 + C_2 h_3^2 + C_1 h_3 + C_0 = r_3^3; \\ C_3 h_4^3 + C_2 h_4^2 + C_1 h_4 + C_0 = r_4^3. \end{cases} \quad (13)$$

Отримана система лінійних рівнянь (13) має такий аналітичний розв'язок:

$$\begin{aligned} k_1 &= \frac{r_3^3 - r_1^3}{h_3 - h_1} \frac{r_2^3 - r_1^3}{h_2 - h_1}; \quad k_2 = \frac{r_4^3 - r_1^3}{h_4 - h_1} \frac{r_2^3 - r_1^3}{h_2 - h_1}; \quad k_3 = \frac{h_2^2 - h_3^2 - h_3 h_1 + h_2 h_1}{h_2 - h_3}; \\ k_4 &= \frac{h_2^2 - h_4^2 - h_4 h_1 + h_2 h_1}{h_2 - h_4}; \quad k_5 = \frac{r_2^3 - r_1^3}{h_2 - h_1}; \quad k_6 = \frac{k_1 - k_2}{k_3 - k_4}; \\ C_3 &= \frac{k_1 - k_2}{k_3 - k_4}; \quad C_2 = k_1 - \frac{k_1 - k_2}{k_3 - k_4} k_3; \end{aligned} \quad (14)$$

$$C_1 = k_5 - (h_2^2 + h_2 h_1 + h_1^2) k_6 - (h_2 + h_1)(k_1 - k_6 k_3); \quad C_0 = r_1^3 - C_3 h_1^3 - C_2 h_1^2 - C_1 h_1.$$

У разі використання інтерполяційної функції (12) за умови відомих поліноміальних коефіцієнтів, визначених співвідношеннями (14), положення

фокуса електронного пучка h_{f3} можна обчислити як корінь квадратного рівняння із $3C_3h^2 + 2C_2h + C_1 = 0$. Тоді фокальний радіус пучка визначається як (r_{f3}, h_{f3}) за співвідношенням (12). Відповідно маємо:

$$h_{f3} = -\frac{C_2}{3C_3}, \quad r_{f3} = \sqrt[3]{C_0 + \frac{2C_2^3}{27C_3^2} - \frac{C_1C_2}{3C_3}}. \quad (15)$$

Функція інтерполяції четвертого порядку. У разі $n = 4$ функція інтерполяції (5) має вигляд

$$r(h) = \sqrt[4]{C_4h^4 + C_3h^3 + C_2h^2 + C_1h + C_0}. \quad (16)$$

У цьому випадку система лінійних рівнянь (7) для визначення коефіцієнтів C_0, C_1, C_2, C_3 і C_4 записується для п'яти базових точок $P_1(h_1, r_1), P_2(h_2, r_2), P_3(h_3, r_3), P_4(h_4, r_4)$ і $P_5(h_5, r_5)$ у такій формі:

$$\begin{cases} C_4h_1^4 + C_3h_1^3 + C_2h_1^2 + C_1h_1 + C_0 = r_1^4; \\ C_4h_2^4 + C_3h_2^3 + C_2h_2^2 + C_1h_2 + C_0 = r_2^4; \\ C_4h_3^4 + C_3h_3^3 + C_2h_3^2 + C_1h_3 + C_0 = r_3^4; \\ C_4h_4^4 + C_3h_4^3 + C_2h_4^2 + C_1h_4 + C_0 = r_4^4; \\ C_4h_5^4 + C_3h_5^3 + C_2h_5^2 + C_1h_5 + C_0 = r_5^4. \end{cases} \quad (17)$$

Для спрощення розв'язування системи лінійних рівнянь (17) уведемо коефіцієнти $a_{k,l}$, де k — номер ітерації у разі розв'язування системи рівнянь (17) методом Гауса–Зейделя; l — номер рівняння в системі (17), який відповідає індексам змінних h і r . Тоді можна записати відповідні співвідношення для індексів $a_{k,l}$:

$$\begin{aligned} a_{1,2} &= \frac{r_2^4 - r_1^4}{h_2 - h_1}; \quad a_{1,3} = \frac{r_3^4 - r_1^4}{h_3 - h_1}; \quad a_{1,4} = \frac{r_4^4 - r_1^4}{h_4 - h_1}; \quad a_{1,5} = \frac{r_5^4 - r_1^4}{h_5 - h_1}; \\ a_{2,3} &= \frac{a_{1,3} - a_{1,2}}{h_3 - h_2}; \quad a_{2,4} = \frac{a_{1,4} - a_{1,2}}{h_4 - h_2}; \quad a_{2,5} = \frac{a_{1,5} - a_{1,2}}{h_5 - h_2}. \end{aligned} \quad (18)$$

Розглянемо також коефіцієнти $b_{k,m,l}$, де дискретні змінні k і l мають ті самі значення, що і для коефіцієнтів $a_{k,l}$ у співвідношеннях (18), а дискретна змінна m відповідає степеню змінної h у співвідношеннях системи рівнянь (17). Відповідні аналітичні співвідношення для коефіцієнтів $b_{k,m,l}$, а також для тих коефіцієнтів $a_{k,l}$, які безпосередньо залежать від $b_{k,m,l}$, можна записати таким чином:

$$\begin{aligned} b_{2,3,3} &= \frac{h_3^3 - h_2^3 + h_3^2h_1 - h_2^2h_1 + h_1^2h_3 - h_1^2h_2}{h_3 - h_2}; \quad b_{2,2,3} = \frac{h_3^2 - h_2^2 + h_3h_1 - h_2h_1}{h_3 - h_2}; \\ b_{2,3,4} &= \frac{h_4^3 - h_2^3 + h_4^2h_1 - h_2^2h_1 + h_1^2h_4 - h_1^2h_2}{h_4 - h_2}; \quad b_{2,2,4} = \frac{h_4^2 - h_2^2 + h_4h_1 - h_2h_1}{h_4 - h_2}; \end{aligned}$$

$$b_{2,3,5} = \frac{h_5^3 - h_2^3 + h_5^2 h_1 - h_2^2 h_1 + h_1^2 h_5 - h_1^2 h_2}{h_5 - h_2}; \quad b_{2,2,5} = \frac{h_5^2 - h_2^2 + h_5 h_1 - h_2 h_1}{h_5 - h_2}; \quad (19)$$

$$b_{3,3,4} = \frac{b_{2,3,4} - b_{2,3,3}}{b_{2,2,4} - b_{2,2,3}}; \quad b_{3,3,5} = \frac{b_{2,3,5} - b_{2,3,3}}{b_{2,2,5} - b_{2,2,3}};$$

$$a_{3,4} = \frac{a_{2,4} - a_{2,3}}{b_{2,2,4} - b_{2,2,3}}; \quad a_{3,5} = \frac{a_{2,5} - a_{2,3}}{b_{2,2,5} - b_{2,2,3}}.$$

З урахуванням співвідношень (18) і (19) аналітичні вирази для коефіцієнтів C_0, C_1, C_2, C_3 і C_4 інтерполяційної функції (16) визначаються розв'язуванням системи лінійних рівнянь (17):

$$C_4 = \frac{a_{3,5} - a_{3,4}}{b_{3,5} - b_{3,4}}; \quad C_3 = \frac{a_{2,4} - a_{2,3}}{b_{2,2,4} - b_{2,2,3}} - \left(\frac{b_{2,3,4} - b_{2,3,3}}{b_{2,2,4} - b_{2,2,3}} \right) \left(\frac{a_{3,5} - a_{3,4}}{b_{3,5} - b_{3,4}} \right);$$

$$C_2 = a_{2,3} - b_{2,3,3} C_4 - b_{2,2,3} C_3; \quad (20)$$

$$C_1 = a_{1,2} - C_4(h_2^3 + h_2^2 h_1 + h_1^2 h_2 + h_1^3) - C_3(h_2^2 + h_2 h_1 + h_1^2) - C_2(h_2 + h_1);$$

$$C_0 = r_1^4 - C_4 h_1^4 - C_3 h_1^3 - C_2 h_1^2 - C_1 h_1.$$

Функція інтерполяції п'ятого порядку. У разі $n = 5$ функція інтерполяції (5) має вигляд:

$$r(h) = \sqrt[5]{C_5 h^5 + C_4 h^4 + C_3 h^3 + C_2 h^2 + C_1 h + C_0}. \quad (21)$$

Для функції інтерполяції, заданої співвідношенням (21) узагальнена система лінійних рівнянь (7) для визначення коефіцієнтів C_0, C_1, C_2, C_3, C_4 і C_5 , записується для шести базових точок $P_1(h_1, r_1), P_2(h_2, r_2), P_3(h_3, r_3), P_4(h_4, r_4), P_5(h_5, r_5)$ і $P_6(h_6, r_6)$:

$$\begin{cases} C_5 h_1^5 + C_4 h_1^4 + C_3 h_1^3 + C_2 h_1^2 + C_1 h_1 + C_0 = r_1^5; \\ C_5 h_2^5 + C_4 h_2^4 + C_3 h_2^3 + C_2 h_2^2 + C_1 h_2 + C_0 = r_2^5; \\ C_5 h_3^5 + C_4 h_3^4 + C_3 h_3^3 + C_2 h_3^2 + C_1 h_3 + C_0 = r_3^5; \\ C_5 h_4^5 + C_4 h_4^4 + C_3 h_4^3 + C_2 h_4^2 + C_1 h_4 + C_0 = r_4^5; \\ C_5 h_5^5 + C_4 h_5^4 + C_3 h_5^3 + C_2 h_5^2 + C_1 h_5 + C_0 = r_5^5; \\ C_5 h_6^5 + C_4 h_6^4 + C_3 h_6^3 + C_2 h_6^2 + C_1 h_6 + C_0 = r_6^5. \end{cases} \quad (22)$$

Перевага використання коефіцієнтів a і b для пошуку коефіцієнтів функції інтерполяції полягає не лише у тому, що спрощуються відповідні аналітичні вирази. Оскільки системи рівнянь (17) і (22) є ідентичними і всі їх рівняння відрізняються лише першим доданком $C_5 h_i^5$, ітераційний процес методу Гауса–Зейделя в обох випадках однаковий. Тому більшість обчисле-

них коефіцієнтів функції інтерполяції четвертого порядку також є правильними і для функції п'ятого порядку. Насамперед відрізнятися будуть коефіцієнти a для першої ітерації $a_{1,l}$, оскільки у правій частині системи рівнянь (22) замість значень r_l^4 стоять значення r_l^5 . Відповідно перші чотири рівняння системи (18) для формування розв'язку системи рівнянь (22) необхідно переписати у такий спосіб:

$$a_{1,2} = \frac{r_2^5 - r_1^5}{h_2 - h_1}; \quad a_{1,3} = \frac{r_3^5 - r_1^5}{h_3 - h_1}; \quad a_{1,4} = \frac{r_4^5 - r_1^5}{h_4 - h_1}; \quad a_{1,5} = \frac{r_5^5 - r_1^5}{h_5 - h_1}. \quad (23)$$

Аналіз закономірності формування отриманих аналітичних співвідношень (18) і (23) свідчить про те, що у загальному випадку для функції інтерполяції порядку n коефіцієнти $a_{1,l}$ можна обчислити з використанням такого універсального аналітичного співвідношення:

$$a_{1,l} = \frac{r_l^n - r_1^n}{h_l - h_1}. \quad (24)$$

Проте інші аналітичні вирази систем рівнянь (18) і (19), отримані для значень $k=2$ і $k=3$, $l=1, \dots, 5$ та $m=1, \dots, 4$, є правильними і для функції апроксимації п'ятого порядку, оскільки ітераційний процес Гауса–Зейделя для розв'язування систем рівнянь (18) і (22) на другій і третій ітераціях має однакові кроки. Інші аналітичні вирази для коефіцієнтів $a_{k,l}$ і $b_{k,m,l}$, які надалі використовуються для формування розв'язку системи рівнянь (22), мають такий вигляд:

$$\begin{aligned} a_{1,6} &= \frac{r_6^5 - r_1^5}{h_6 - h_1}; \quad a_{2,6} = \frac{a_{1,6} - a_{1,2}}{h_6 - h_2}; \quad a_{3,6} = \frac{a_{2,6} - a_{2,3}}{b_{2,2,6} - b_{2,2,3}}; \\ b_{2,2,6} &= \frac{h_6^2 - h_2^2 + h_6 h_1 - h_2 h_1}{h_6 - h_2}; \\ b_{2,4,3} &= \frac{h_3^4 - h_2^4 + h_3^3 h_1 - h_2^3 h_1 + h_3^2 h_1^2 - h_2^2 h_1^2 + h_1^3 h_3 - h_1^3 h_2}{h_3 - h_2}; \\ b_{2,4,4} &= \frac{h_4^4 - h_2^4 + h_4^3 h_1 - h_2^3 h_1 + h_4^2 h_1^2 - h_2^2 h_1^2 + h_1^3 h_4 - h_1^3 h_2}{h_4 - h_2}; \\ b_{2,4,5} &= \frac{h_5^4 - h_2^4 + h_5^3 h_1 - h_2^3 h_1 + h_5^2 h_1^2 - h_2^2 h_1^2 + h_1^3 h_5 - h_1^3 h_2}{h_5 - h_2}; \\ b_{2,4,6} &= \frac{h_6^4 - h_2^4 + h_6^3 h_1 - h_2^3 h_1 + h_6^2 h_1^2 - h_2^2 h_1^2 + h_1^3 h_6 - h_1^3 h_2}{h_6 - h_2}; \\ b_{2,3,6} &= \frac{h_6^3 - h_2^3 + h_6^2 h_1 - h_2^2 h_1 + h_1^2 h_6 - h_1^2 h_2}{h_6 - h_2}; \\ b_{3,4,4} &= \frac{b_{2,4,4} - b_{2,4,3}}{b_{2,2,4} - b_{2,2,3}}; \quad b_{3,4,6} = \frac{b_{2,4,6} - b_{2,4,3}}{b_{2,2,6} - b_{2,2,3}}; \quad b_{3,3,6} = \frac{b_{2,3,6} - b_{2,3,3}}{b_{2,2,6} - b_{2,2,3}}. \end{aligned} \quad (25)$$

За умови відомих коефіцієнтів $a_{k,l}$ і $b_{k,m,l}$, які задаються аналітичними співвідношеннями (18) і (25), розв'язок системи рівнянь (22) для значень коефіцієнтів C_5, C_4, C_3, C_2, C_1 і C_0 має такий вигляд:

$$\begin{aligned}
 C_5 &= \frac{a_{4,6} - a_{4,5}}{b_{4,4,6} - b_{4,4,5}}; \quad C_4 = b_{4,4,5} \frac{a_{4,6} - a_{4,5}}{b_{4,4,6} - b_{4,4,5}} - a_{4,5}; \\
 C_3 &= a_{3,4} - b_{3,4,4}C_5 - b_{3,3,4}C_4; \quad C_2 = a_{2,3} - b_{2,4,3}C_5 - b_{2,3,3}C_4 - b_{2,2,3}C_3; \\
 C_1 &= a_{1,2} - C_5(h_2^4 + h_2^3h_1 + h_2^2h_1^2 + h_1^3h_2 + h_1^4) - \\
 &\quad - C_4(h_2^3 + h_2^2h_1 + h_1^2h_2 + h_1^3) - C_3(h_2^2 + h_2h_1 + h_1^2) - C_2(h_2 + h_1); \\
 C_0 &= r_1^5 - C_5h_1^5 - C_4h_1^4 - C_3h_1^3 - C_2h_1^2 - C_1h_1.
 \end{aligned} \tag{26}$$

Результати інтерполяції, з використанням отриманих співвідношень (8)–(26), межових траєкторій короткофокусних електронних пучків, за умови їх транспортування в низькому вакуумі з компенсацією власного просторового заряду електронів іонами залишкового газу, числові дані для яких задаються числовим розв'язком системи рівнянь (2)–(4) та графічними залежностями (рис. 1), розглянуто далі. Там же буде оцінена похибка інтерполяції як відносна різниця між результатами числового моделювання, які в цьому випадку вважаються еталонними, та значеннями функцій інтерполяції (8), (12), (16) і (21), коефіцієнти яких розраховуються за співвідношеннями (10), (14), (18–20) і (18, 19, 23, 25, 26) відповідно.

РЕЗУЛЬТАТИ ІНТЕРПОЛЯЦІЇ МЕЖОВОЇ ТРАЄКТОРІЇ ЕЛЕКТРОННОГО ПУЧКА ТА АНАЛІЗ ПОХИБКИ ІНТЕРПОЛЯЦІЇ

Результати інтерполяції межової траєкторії короткофокусного електронного пучка для даних числових розрахунків, поданих у вигляді графічних залежностей на рис. 1, показано на рис. 2. Залежності для відносної похибки інтерполяції зображено на рис. 3. Відносну похибку інтерполяції розраховано таким чином:

$$\varepsilon(h) = \frac{|r_{\text{чис}}(h) - r_{\text{інт}}(h)|}{r_{\text{чис}}(h)} \cdot 100, \tag{27}$$

де $r_{\text{чис}}$ — результати числових розрахунків, які для задання інтерполяції вважаються достовірними; $r_{\text{інт}}$ — результати інтерполяції. Як видно з рис. 2 і 3, інтерполяція даних числових розрахунків виконувалась для значень прискорювальної напруги 10, 12 і 13 кВ поліномами другого, третього, четвертого та п'ятого порядків з використанням співвідношень (8), (12), (16) і (21).

Із графічних залежностей (рис. 3) чітко видно, що точність інтерполяції, у разі використання класу запропонованих функцій, є надто високою, оскільки максимальна похибка інтерполяції, яка визначається співвідношенням (27), не перевищує 8%. Найменша похибка інтерполяції, менша за 0,6%, отримана для прискорювальної напруги 10 кВ, що можна пояснити

лінійним характером межевої траєкторії електронного пучка за межами області мінімуму та вдалим вибором набору відлікових точок, відмічених на рис. 2, а.

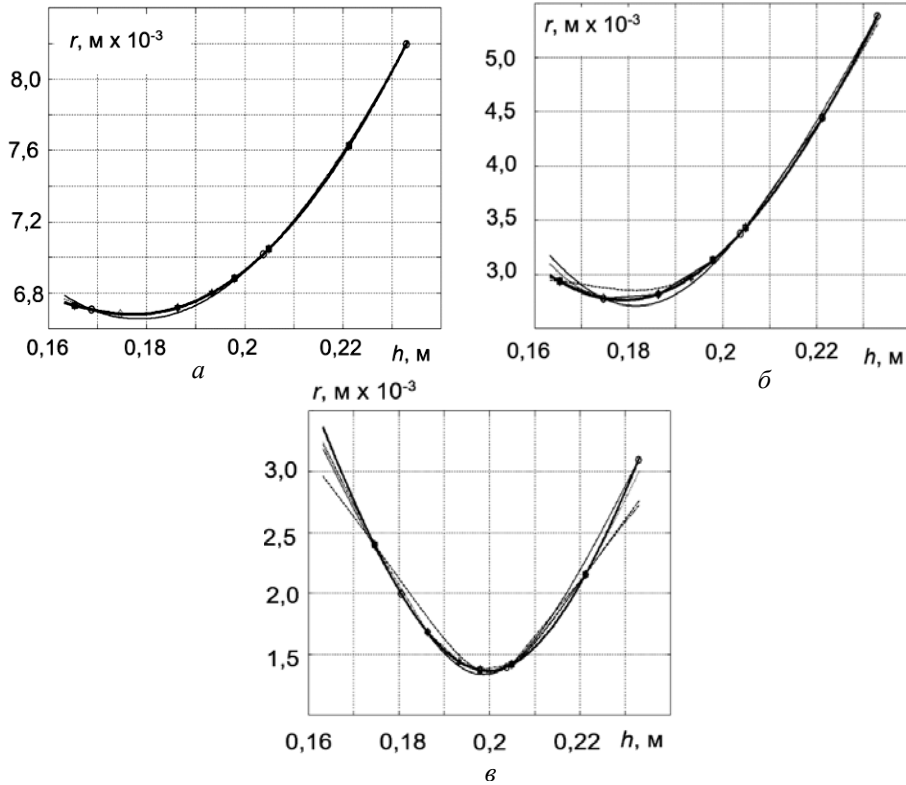


Рис. 2. Результати інтерполяції межевої траєкторії електронного пучка для розрахункових даних, показаних на рис. 1. Жирна суцільна лінія — результати числового моделювання. Результати інтерполяції: тонка суцільна лінія — функція другого порядку, штрихова лінія — функція третього порядку, пунктирна лінія — функція четвертого порядку, штрихпунктирна лінія — функція п'ятого порядку. а — $U_{\text{пр}} = 10$ кВ, б — $U_{\text{пр}} = 12$ кВ, в — $U_{\text{пр}} = 13$ кВ. Базові точки функцій інтерполяції: коло — для функції другого порядку, квадрат — для функції третього порядку, ромб — для функції четвертого порядку, шестикутник — для функції п'ятого порядку. Струм електронного пучка — 0,5 А, робочий тиск — 0,1 Па

Висока точність інтерполяції для будь-яких значень прискорювальної напруги пояснюється передусім правильно обраним класом функцій, заданих співвідношенням (5), оскільки ці функціональні залежності мають ті самі характерні особливості, що й результати числового моделювання, подані на рис. 1. Однаковий характер залежностей, які підлягають інтерполяції, та функцій інтерполяції полягає у тому, що вони мають один глобальний мінімум, а за межами області мінімуму ці залежності дуже близькі до лінійних.

Із використанням співвідношень (11), (15) оцінювалось положення фокуса електронного пучка та його фокального радіуса. Для проведення таких оцінок через обраний набір числових даних, отриманих у результаті виконаних числових розрахунків за співвідношеннями (2)–(4), з використанням алгоритмів розв'язання екстремальних задач [16] визначався мінімум функції $r(h_f) = r_f$. З урахуванням формалізму дискретної математики відповід-

ний алгоритм пошуку положення екстремуму функції дискретних відліків та її мінімального значення можна записати в такій формі [16]:

$$r_f = \min(r_i), h_f = \operatorname{argmin}(r_i), i = 1, \dots, N, \quad (28)$$

де N — кількість дискретних точок у наборі числових даних, який аналізується.

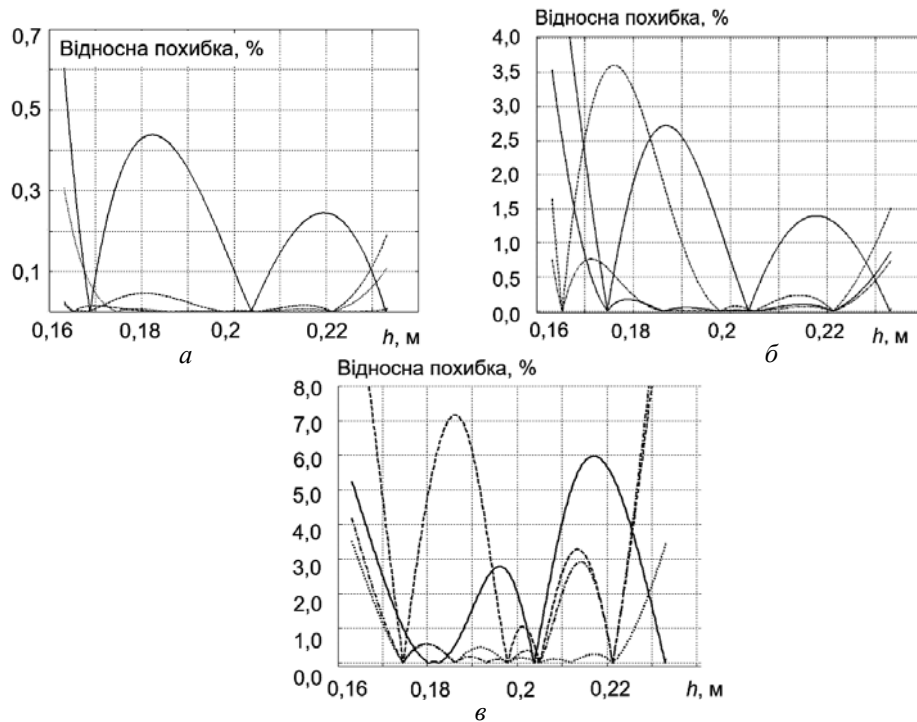


Рис. 3. Відносна похибка інтерполяції межової траєкторії електронного пучка для розрахункових даних, показаних на рис. 1. Суцільна лінія — функція другого порядку, штрихова лінія — функція третього порядку, пунктирна лінія — функція четвертого порядку, штрихпунктирна лінія — функція п'ятого порядку; a — $U_{\text{пр}} = 10$ кВ, b — $U_{\text{пр}} = 12$ кВ, v — $U_{\text{пр}} = 13$ кВ. Струм електронного пучка — 0,5 А, робочий тиск — 0,1 Па

Числові результати розрахунків, виконаних за співвідношеннями (11), (15), (28), наведено у таблиці.

Результати розрахунку фокальних параметрів електронного пучка через аналіз даних числового моделювання та за функціями інтерполяції його межової траєкторії другого і третього порядків

Прискорювальна напруга, кВ	Положення фокуса пучка h_f			Фокальний радіус r_f		
	Еталонне значення, м	Функція другого порядку, значення, м/похибка, %	Функція третього порядку, значення, м/похибка, %	Еталонне значення, мм	Функція другого порядку, значення, мм / похибка, %	Функція третього порядку, значення, мм / похибка, %
10	0,185	0,186 / 0,5	0,184 / 0,5	6,71	6,73 / 0,2	6,72 / 0,15
12	0,181	0,185 / 2,2	0,183 / 1,1	2,83	2,75 / 2,8	2,8 / 1,1
13	0,21	0,2 / 4,8	0,206 / 1,9	1,41	1,4 / 0,71	1,405 / 0,35

ОБГОВОРЕННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

Із загальних основ теорії інтерполяції зрозуміло, що мале значення отриманої похибки для даних числового моделювання можна отримати лише за умови правильного вибору набору базових точок в області фокальної площини електронного пучка, де залежність $r(h)$ досягає мінімальної величини [15, 16]. Тестові експерименти показали, що, оскільки згідно з узагальненим співвідношенням (1) залежність $r(h)$ є дзеркально-симетричною відносно лінії $h = h_f$, дзеркальне розташування базових точок відносно цієї лінії з додатковою точкою в області розташування фокуса пучка зазвичай є найкращим рішенням щодо отримання мінімальної похибки інтерполяції. З іншого боку, якщо фокус електронного пучка розташований на початку або в кінці числового інтервалу за змінною h $[h_{\text{поч}}, h_{\text{кін}}]$, який розглядається, достатньо взяти лише одну базову точку на інтервалі $[h_{\text{поч}}, h_f]$ або $[h_f, h_{\text{кін}}]$, оскільки за межами фокальної області залежність $r(h)$ завжди близька до лінійної. Тому у випадку, коли фокальна площина електронного пучка для обраного набору числових даних розташована в середині інтервалу $[h_{\text{поч}}, h_{\text{кін}}]$, краще обирати функцію інтерполяції парного порядку, наприклад другого або четвертого. Навпаки, функції інтерполяції непарного порядку, наприклад третього або п'ятого, краще обирати, якщо фокальна площина електронного пучка h_f розташована на початку або в кінці інтервалу $[h_{\text{поч}}, h_{\text{кін}}]$, який розглядається. Сформовані в результаті проведення тестових експериментів рекомендації щодо обрання порядку функції інтерполяції та розташування базових точок дозволили створити простий алгоритм для розв'язування завдання інтерполяції в автоматичному режимі. Базові точки, показані на рис. 2, розташовувались з використанням цього алгоритму. Якщо необхідно взяти лише одну базову точку, вона розташовувалась всередині інтервалу, а якщо потрібна кількість точок N , відповідний інтервал ділиться на $N + 1$ частину.

У будь-якому разі тестові експерименти показали, що з використанням функціональних залежностей (5) від другого до п'ятого порядку можна досягти досить високої точності інтерполяції межевої траєкторії короткофокусних електронних пучків, а відносна похибка інтерполяції становить від 0,7% до 8%.

Із розглянутого у роботі теоретичного матеріалу зрозуміло, що, оскільки ітераційний процес Гауса–Зейделя повністю повторюється для пошуку коефіцієнтів функцій (5) різних порядків, існують певні закономірності щодо формування аналітичних залежностей для обчислення поліноміальних коефіцієнтів C_0, \dots, C_n . Наприклад, співвідношення (18) і (19), отримані для функції четвертого порядку, були використані для формування аналітичних виразів (25), призначених для визначення поліноміальних коефіцієнтів функції п'ятого порядку через співвідношення (26). Тобто, з теоретичної точки зору, можна отримати відповідні аналітичні співвідношення і для функцій (5) більш високих порядків. Проте цілком зрозуміло, що в разі збільшення порядку n функції інтерполяції до значень, більших за 20, обчислювальна складність завдання інтерполяції непомірно зростає, а тоді вона вже не оправдується можливим зменшенням похибки, яке стає неістотним [15, 16].

Із практичного погляду запропонований метод інтерполяції межових траєкторій короткофокусних електронних пучків є дуже ефективним аналі-

тичним інструментом, за допомогою якого можна оцінювати можливості використання електронних гармат ВТР для виконання відповідних технологічних операцій, пов'язаних з електронно-променевим нагріванням речовини та відповідних виробів.

ВИСНОВКИ

Описано метод інтерполяції межової траєкторії короткофокусних електронних пучків за умови їх транспортування в низькому вакуумі з компенсацією просторового заряду електронів іонами залишкового газу. Виконаний аналіз результатів числового моделювання межової траєкторії електронних пучків для різних значень прискорювальної напруги, тиску залишкового газу та струму пучка показав, що ці залежності мають яружний характер з одним глобальним мінімумом у площині фокуса пучка, а за межами фокальної області вони близькі до лінійних залежностей. Подальший аналіз показав, що для інтерполяційного описання таких функціональних залежностей з невисокою похибкою можуть бути використані аналітичні функції, які у загальному вигляді описуються співвідношенням (5). З використанням методу Гауса–Зейделя отримано аналітичні співвідношення для пошуку поліноміальних коефіцієнтів функцій (5) від другого до п'ятого порядку. Тестові експерименти показали, що похибка інтерполяції для даних числового моделювання залежить від вибору базових точок і становить 0,7–8%. Для функцій інтерполяції другого і третього порядків отримано аналітичні співвідношення (11), (15), які дозволяють безпосередньо оцінювати фокальні параметри електронного пучка. Проведено оцінки точності визначення фокальних параметрів електронного пучка за результатами інтерполяції, відносна похибка розрахунків за співвідношеннями (11), (15) не перевищувала 5%. Отримані результати інтерполяції дозволяють на початковому етапі проектування досить точно оцінювати технологічні можливості електронних гармат ВТР та узгоджувати їх з вимогами відповідного технологічного процесу. Такі результати є вкрай цікавими для фахівців, які займаються розробленням новітнього електронно-променевого технологічного обладнання та його впровадженням у промислове виробництво.

ЛІТЕРАТУРА

1. Н.Н. Рыкалин, И.В. Зуев, и А.А. Углов, *Основы электронно-лучевой обработки материалов*. Москва: Машиностроение, 1978, 239 с.
2. З. Шиллер, У. Гайзиг, и З. Панцер, *Электронно-лучевая технология*. Москва: Энергия, 1980, 528 с.
3. *Электронно-лучевая сварка*, под общ. ред. Б.Е. Патона. Киев: Наукова думка, 1987, 256 с.
4. М.А. Завьялов, Ю.Е. Крейндель, А.А. Новиков, и Л.П. Шантурин, *Плазменные процессы в технологических электронных пушках*. Москва: Атомиздат, 1989, 256 с.
5. Y.E. Krasik et al., “High-current electron sources based on gaseous discharges”, *Vacuum*, vol. 77, no. 4, pp. 391–398, 2003.
6. V.A. Gruzdev, V.G. Zalesski, D.A. Antonovich, and V.P. Golubev, “Universal plasma electron source”, *Vacuum*, vol. 77, no. 4, pp. 399–406, 2003.
7. S. Denbnovetskiy et al., “Principles of operation of high voltage glow discharge electron guns and particularities of its technological application”, in *Proceedings of SPIE The International Society of Optical Engineering, 2017*, pp. 10445–10455.
8. J.D. Lawson, *The Physics of Charged-Particle Beams*. Clarendon Press, Oxford, 1977.
9. И.В. Мельник и А.В. Починок, “Интерполяция граничной траектории электронного пучка в прифокальной области линейными и квадратичными функциями

- с использованием арифметико-логических выражений”, *Вісник Херсон. нац. техн. ун-ту*, вип. 2 (69), ч. 2, с. 23–30, 2019.
10. I. Melnyk, S. Tugay, and A. Pochynok, “Interpolation Functions for Describing the Boundary Trajectories of Electron Beams Propagated in Ionised Gas”, in *15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET - 2020) Conference Proceedings*. Available: https://www.researchgate.net/publication/341248002_Interpolation_Functions_for_Describing_the_Boundary_Trajectories_of_Electron_Beams_Propagated_in_Ionised_Gas
 11. I. Melnyk, S. Tuhai, and A. Pochynok, “Interpolation of the Boundary Trajectories of Electron Beams by the Roots from Polynomial Functions of Corresponded Order”, *2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO). Conference Proceedings*, pp. 28–33.
 12. С.В. Денбовецкий, В.И. Мельник, И.В. Мельник, и Б.А. Тугай, “Моделирование транспортировки короткофокусных электронных пучков из низкого в высокий вакуум с учетом разброса тепловых скоростей электронов”, *Прикладная физика*, № 3, с. 84–90, 2010.
 13. M. Szilagyı, “Electron and Ion Optics”, *Springer Science & Business Media*, 2012.
 14. С.И. Молоковский и Д.И. Сушков, *Интенсивные электронные и ионные пучки*. Москва: Энергоатомиздат, 1991, 304 с.
 15. А.А. Самарский и А.В. Гулин, *Численные методы: учеб. пособие для вузов*. Москва: Наука, 1989, 432 с.
 16. Ф.П. Васильев, *Численные методы решения экстремальных задач: учеб. пособие для вузов*. Москва: Наука, 1988, 552 с.

Надійшла 31.05.2020

INFORMATION ON THE ARTICLE

Igor V. Melnyk, ORCID: 0000-0003-0220-0615, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: imelnik@phbme.kpi.ua

Alina V. Pochynok, ORCID: 0000-0001-9531-7593, Scientific and Education Institute of Information Technology of University of State Fiscal Service of Ukraine, e-mail: alina_pochynok@yahoo.com

INVESTIGATION OF THE CLASS OF ALGEBRAICAL FUNCTIONS FOR INTERPOLATION OF BOUNDARY TRAJECTORIES OF SHORT-FOCUS ELECTRON BEAMS / I.V. Melnyk, A.V. Pochynok

Abstract. The method for interpolating the dependence of the electron beam radius on the cutting plane position by the longitudinal coordinate, based on using the special class of interpolation functions, was investigated in this article. The roots of n -th order from the polynomials of the same order are used as the class of interpolation functions, which correspond to the ravine dependence of the radius of a short-focus electron beam on the distance of propagation. The results of testing have shown that, depending on choosing the set of basic points, including its position relatively to the minimum and their quantity, the relative interpolation error is between 0,7% and 8%. Obtained results of mathematical simulation of focal parameters of electron beams have the significant practical importance for the specialists, involved in the development and introduction into industrial production of modern highly effective electron-beam technological equipment.

Keywords: electron beam, electron beam guiding, focal region, interpolation, class of algebraic functions.

ИССЛЕДОВАНИЕ КЛАССА АЛГЕБРАИЧЕСКИХ ФУНКЦИЙ ДЛЯ ИНТЕРПОЛЯЦИИ ГРАНИЧНЫХ ТРАЕКТОРИЙ КОРОТКОФОКУСНЫХ ЭЛЕКТРОННЫХ ПУЧКОВ / И.В. Мельник, А.В. Починюк

Аннотация. Исследован метод интерполяции зависимости радиуса электронного пучка от положения плоскости среза по продольной координате, основанный на использовании специального класса интерполяционных функций.

Как класс интерполяционных функций, которые соответствуют овражным зависимостям радиуса короткофокусного электронного пучка от расстояния, выбраны корни степени n из полиномов той же самой степени. Результаты тестирования показали, что, в зависимости от выбранного набора отсчётных точек, их положения относительно области минимума и количества, относительная ошибка интерполяции лежит в пределах 0,7–8%. Полученные результаты математического моделирования фокальных параметров электронных пучков имеют большое практическое значение для специалистов, занимающихся разработкой и внедрением в промышленное производство современного высокоэффективного электронно-лучевого технологического оборудования.

Ключевые слова: электронный пучок, транспортирование электронного пучка, прифокальная область, интерполяция, класс алгебраических функций.

REFERENCES

1. N.N. Rykalin, I.V. Zuev, and A.A. Uglov, *Fundamentals of Electron-Beam Materials Treatment*. Moscow: Mashinostroenie, 1978.
2. S. Shiller, U. Geisig, and S. Panzer, *Electron-beam technology*. Moscow: Energy, 1980.
3. *Electron Beam Welding*, edited by B.E. Pathon. Kyiv: Naukova Dumka, 1987.
4. M.A. Zavalov, Yu.E. Kreindel, A.A. Novikov, and L.P. Shanturin, *Plasma Processes in the Technological Electron Guns*. Moscow: Atomizdat, 1989.
5. Y.E. Krasik et al., “High-current electron sources based on gaseous discharges”, *Vacuum*, vol. 77, no. 4, pp. 391–398, 2003.
6. V.A. Gruzdev, V.G. Zalesski, D.A. Antonovich, and V.P. Golubev, “Universal plasma electron source”, *Vacuum*, vol. 77, no. 4, pp. 399–406, 2003.
7. S. Denbnovetskiy et al., “Principles of operation of high voltage glow discharge electron guns and particularities of its technological application”, *Proceedings of SPIE The International Society of Optical Engineering*, pp. 10445–10455, 2017.
8. J.D. Lawson, *The Physics of Charged-Particle Beams*. Oxford: Clarendon Press, 1977.
9. I.V. Melnyk and A.V. Pochynok, “Interpolation of Boundary Trajectory of Electron Beam in the Near-Focus Region by the Linear and Square Functions with Using of Arithmetic-Logical Equations”, *Bulletin of Kherson National Technical University*, issue 2 (69), part 2, pp. 23–30.
10. I. Melnik, S. Tugay, and A. Pochynok, “Interpolation Functions for Describing the Boundary Trajectories of Electron Beams Propagated in Ionised Gas”, *15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET–2020) Conference Proceedings*. [Online]. Available: https://www.researchgate.net/publication/341248002_Interpolation_Functions_for_Describing_the_Boundary_Trajectories_of_Electron_Beams_Propagated_in_Ionised_Gas
11. I. Melnyk, S. Tuhai, and A. Pochynok, “Interpolation of the Boundary Trajectories of Electron Beams by the Roots from Polynomial Functions of Corresponded Order”, *2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO) Conference Proceedings*, pp. 28–33.
12. S.V. Denbnovetskiy, V.I. Melnik, I.V. Melnik, and B.A. Tugai, “Simulation of Guiding Short-Focus Electron Beams from the Low to High Vacuum with Taking into Account the Dissipation of Electrons Velocity”, *Applied Physics*, no. 3, pp. 84–90, 2010.
13. M. Szilagy, *Electron and Ion Optics*. Springer Science & Business Media, 2012.
14. S.I. Molokovskiy and D.I. Sushkov, *Intensive Electron and Ion Beams*. Moscow: Energoatomizdat, 1991.
15. A.A. Samarskiy and A.V. Gulin, *Numerical Methods: Tutorial Book for Higher Education Institutions*. Moscow: Nauka, 1989.
16. F.P. Vasiliev, *Numerical Methods for Solving the Extremal Problems: Tutorial Book for Higher Education Institutions*. Moscow: Nauka, 1988.

РАСШИРЕННАЯ ИНТЕГРАЦИЯ СИСТЕМЫ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА С ПАКЕТОМ ОФИСНЫХ ПРОГРАММ (РАБОТА С ПЕРСОНАЛОМ)

А.В. МУХОИД, Е.А. КУРИЛО

Аннотация. Посвящено проблематике ведения отпусков на предприятии и автоматизации этого процесса. Описан типовой жизненный цикл документа, приведены основные недостатки жизненного цикла бумажного документооборота. Поставлено задание подробно рассмотреть существующий алгоритм ведения отпусков. Проведена оценка трудоемкости процесса и приведены недостатки данного алгоритма, в частности неоптимальное использование рабочего времени и возможность появления ошибок. Детально описаны этапы проведения автоматизации, способы решения поставленной задачи. Приведены механизмы получения необходимых данных и проведен анализ полученных результатов. Представлен перспективный план дальнейшей автоматизации процесса ведения отпусков.

Ключевые слова: система электронного документооборота, прикладная разработка, автоматизация процесса, ведение отпусков.

ВВЕДЕНИЕ

Большинство организаций применяют в своей работе канцелярский документооборот. Поиск и заполнение необходимой информации для регулярно составляемых документов (еженедельно, ежемесячно) отнимают много времени из-за однотипной рутинной работы, при этом не исключается вероятность возникновения ошибок в связи с человеческим фактором.

Типовой жизненный цикл документа содержит такие этапы: создание документа, согласование с соисполнителями, утверждение руководителями, ознакомление с документом ответственных исполнителей, передача документа в отделы для проработки.

В рамках указанного жизненного цикла бумажного документооборота возникают следующие недостатки:

- большие временные затраты для составления документов вручную;
- возможность появления ошибок в документе в связи с человеческим фактором;
- продолжительное время согласования документов;
- потеря документов в процессе согласования или рассылки в подразделения [1].

В связи с приведенными сложностями множество организаций переходит на электронный документооборот. Внедрение программного комплекса системы электронного документооборота позволяет уменьшить время на создание и согласование документов, исключить или минимизировать веро-

ятность утери документов, снизить затраты предприятия на печать, копирование и хранение документации [2]. Такой способ позволяет экономить время и ресурсы. В настоящее время используется множество систем электронного документооборота, которые автоматизируют процесс создания и согласования документов.

Рассмотрим систему электронного документооборота (СЭД) Directum, используемую предприятием АО «ВНИИАЭН». Инструментарий системы позволяет автоматизировать простые задания, используя прикладные формы и шаблоны, а также согласовывать документы с помощью заранее запрограммированных маршрутов.

И все же стандартный функционал Directum не позволяет автоматизировать все потребности предприятия по ведению делового документооборота. Поэтому со временем с помощью инструментов системы создавались новые типовые маршруты согласования документов, карточки различных видов документов, справочники, а также другие вспомогательные функции и сценарии, настроенные под делопроизводство предприятия.

С помощью СЭД автоматизируются ведение документации по договорам, деловая переписка между организациями и подразделениями, кадровая работа — учет рабочего времени, прием и увольнение сотрудников, ведение отпусков [3].

Ведение отпусков — трудоемкий процесс, в котором задействованы все подразделения предприятия. Целью является автоматизация процесса ведения отпусков в целом путем создания единого общего пространства данных — справочника «График отпусков» как базы данных. Эта база будет использоваться совместно для формирования документов, создаваемых различными подразделениями — бухгалтерско-финансовым отделом, отделом планово-экономического анализа, отделом по работе с персоналом. Таким образом, будут устранены разночтения и возможные ошибки, возникающие в результате создания и согласования отдельных документов, содержащих, по сути, аналогичные данные. Алгоритм, который необходимо реализовать, приведен на плановой схеме ведения отпусков (рис. 1).

Рассмотрим состояние ведения отпусков на предприятии до проведения автоматизации.

В конце года подразделения направляют на согласование служебные записки об ежегодном плановом отпуске на следующий год. Далее формируется и согласовывается общий приказ на ежегодный отпуск. Основываясь на данных об отпусках, указанных в общем приказе на год, подразделения формируют и согласовывают служебные записки на плановый отпуск ежемесячно. В случае изменения даты отпусков подразделения направляют служебные записки на внеплановый отпуск, на основании которых отдел по работе с персоналом формирует приказ на внеплановый отпуск. Последним этапом является создание отделом по работе с персоналом общего приказа на отпуск ежемесячно, учитывая данные из служебных записок от подразделений.

Алгоритм представлен в виде фактической схемы ведения отпусков на предприятии (рис. 2).

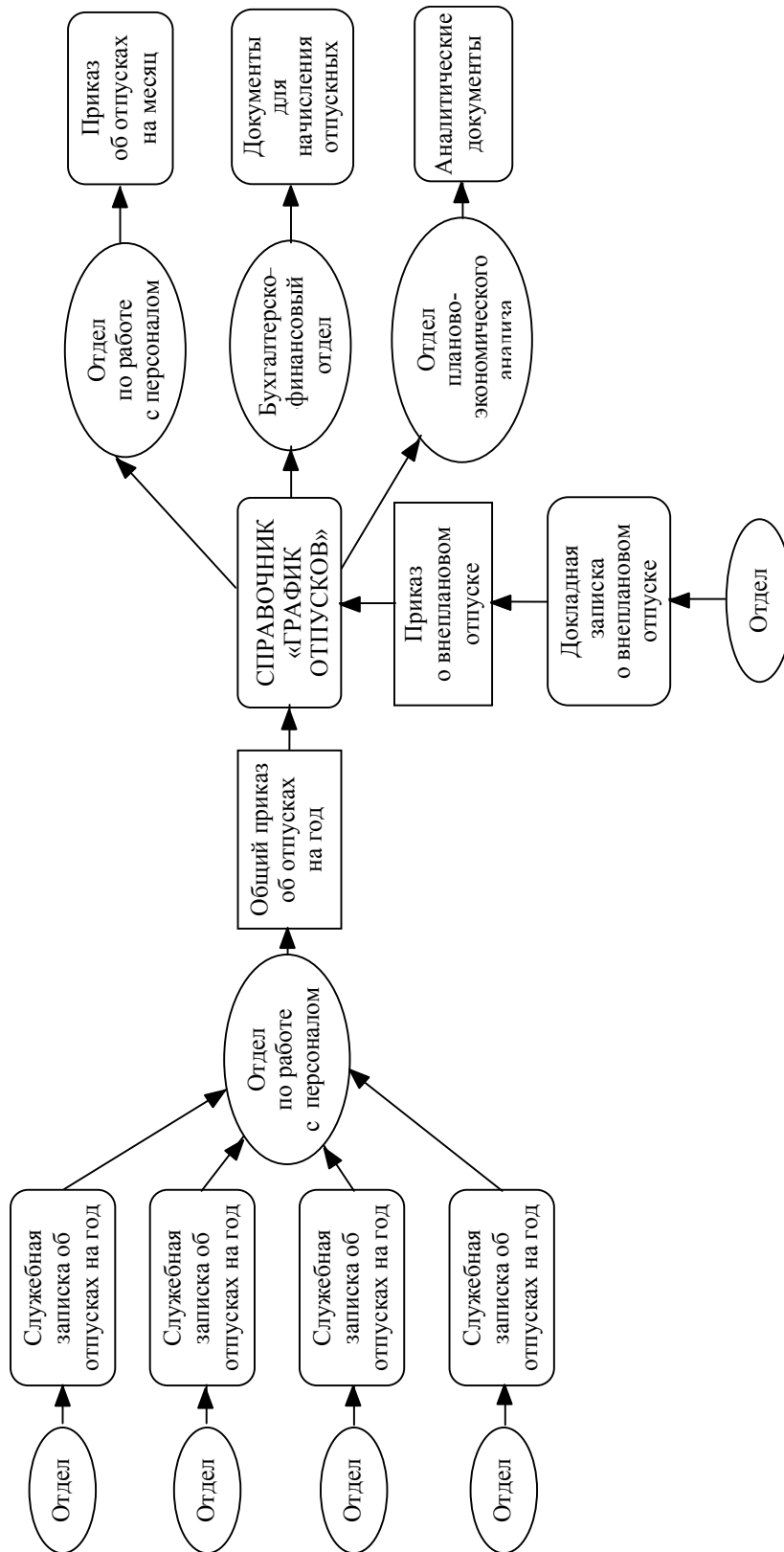


Рис. 1. Плановая схема ведения отпусков

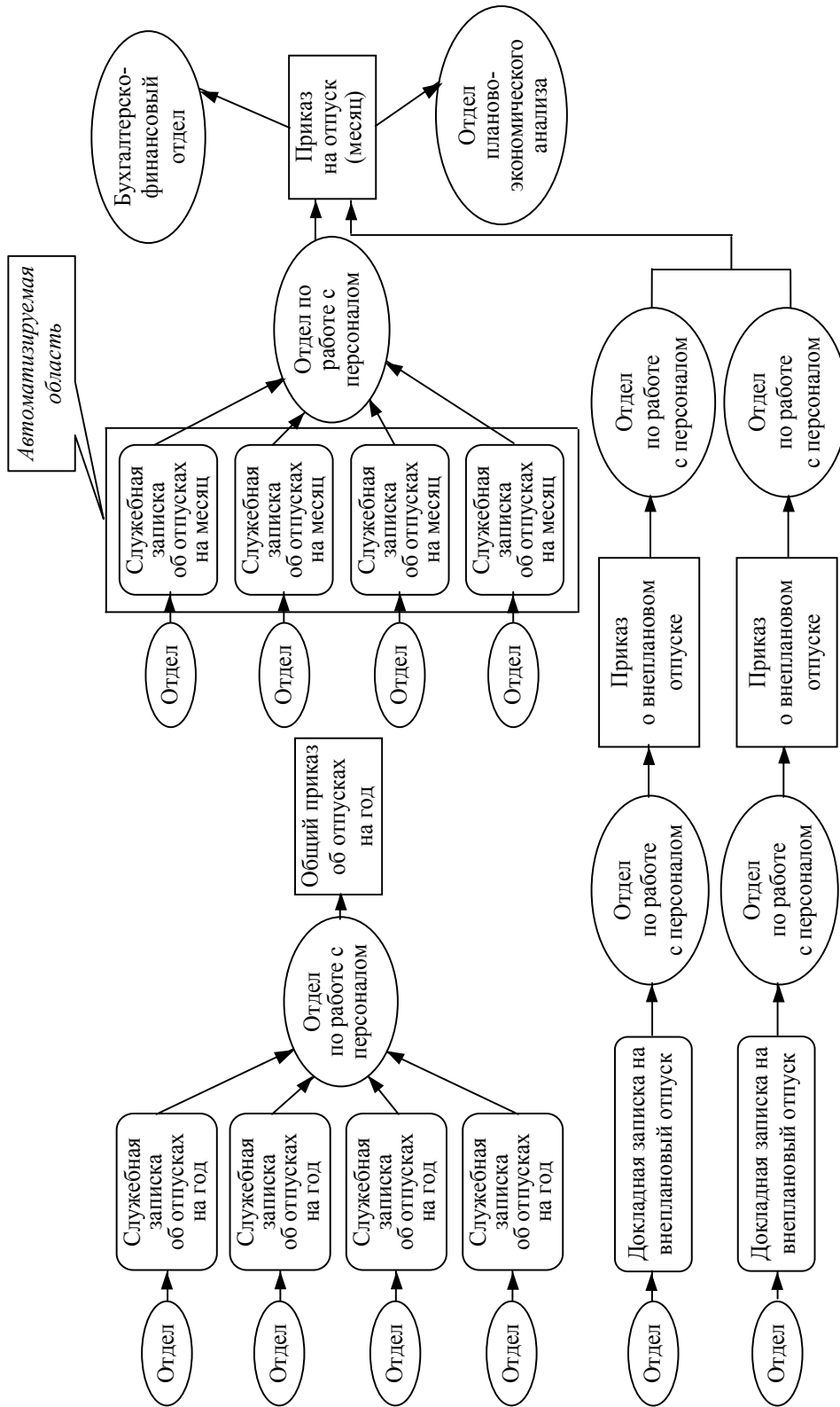


Рис. 2. Фактическая схема ведения отпусков на предприятии

ПОСТАНОВКА ЗАДАЧИ

Учитывая массовое использование информации об отпусках подразделениями, было принято решение автоматизировать данную задачу. Поскольку модуль ведения отпусков с необходимыми характеристиками и функциями в используемой системе электронного документооборота не существовало, он был разработан сотрудниками отдела информационных технологий с помощью прикладного инструментария системы. Первым шагом для перехода на плановую схему ведения отпусков была выбрана автоматизация создания служебной записки.

Этот документ подается на согласование ежемесячно руководителями подразделений. Составление документа до автоматизации составляло от 30 до 60 минут в зависимости от количества сотрудников конкретного подразделения. Что касается необходимой для составления служебной записки об отпусках информации, то она содержится в разных источниках. Даты отпуска каждого сотрудника зафиксированы в общем приказе на год; должности, полные имена и табельные номера работников — в различных справочниках системы. После получения необходимой информации составляется служебная записка на отпуск, при этом текст служебной записки заполнялся вручную. Недостатки ручного заполнения основного текста спровоцировали идею автоматизировать подобную рутинную работу.

В результате проведения работ по автоматизации отпусков была создана программа, которая автоматически отслеживает наступление дат отпуска, собирает необходимую информацию по каждому сотруднику и ежемесячно по запросу формирует сводный документ, в который автоматически переносит найденную информацию в шаблонном виде. Так как на предприятии есть несколько подразделений, необходимо реализовать выбор сотрудников для конкретного подразделения, учитывая автора документа, и формировать служебную записку для работников выбранного подразделения.

Для составления служебной записки на отпуск необходимо создать документ в системе Directum, затем заполнить карточку документа, в которой указываются адресат и тема документа (рис. 3). Дата, номер служебной записки, автор письма, подразделение автора и месяц заполняются автоматически.

В карточке по умолчанию указывается следующий месяц, на который будет составлена служебная записка на отпуск. При необходимости можно изменить месяц, на который формируется документ.

После сохранения карточки создается служебная записка на отпуск.

Пример сформированного документа с автоматически заполненными полями и основным текстом показан на рис. 4.

Поля адресат, номер служебной записки, дата, тема и автор документа заполняются автоматически, перенося информацию из карточки документа.

Основной текст служебной записки представляет собой комбинацию данных из различных источников. Автоматизация заполнения основной части и является задачей, которую необходимо решить.

Для начала в соответствии с приказом об отпусках на год необходимо заполнить справочник «График отпусков», в котором будут записаны даты отпусков всех сотрудников (рис. 5).

Для каждого сотрудника создается карточка, куда заносятся даты отпусков.

Служ. и докладн. записки

*Вид эл. документа: Службная записка

*№ документа: 054/11 *Дата документа: 10.04.2020

*Инициатор: Кобзарь М.В.

*Тема (о чём?): Щодо відпусток

Подразделение: Відділ №11

Примечание:

*Наименование: С/з №054/11 Щодо відпусток

*Приложение: Microsoft Word

*Автор: Кобзарь Максим Володимирович

*Дата создания: 10.04.2020 13:12:24

Дата посл. изм.: 15.04.2020 7:47:48

Признак:

*Адресат: Лисак Валентина Вікторівна

Выбрать несколько адресатов Заполнить отпуски

*№ *Адресаты

Пуск

ИД: 222101 Просмотр

Рис. 3. Карточка электронного документа

АТ «ВНДІАЕН»

Начальнику ВУП та СП
Лисак В.В.

СЛУЖБОВА ЗАПИСКА

10.04.2020
Щодо відпусток

Повідомляю Вам про надання наступної відпустки на Травень 2020 року:

- Завідувач сектору Бойко Андрій Олександрович табельний №3039 з 04.05.2020 по 08.05.2020 включно;
- Інженер-програміст 1 кат. Дегтярьова Олена Миколаївна табельний №1621 з 12.05.2020 по 15.05.2020 включно;
- Інженер-конструктор 1 кат. Кінашевська Аяна Володимирівна табельний №1607 з 04.05.2020 по 08.05.2020 включно;
- Інженер-програміст 3 кат. Курило Олена Олександрівна табельний №1632 з 12.05.2020 по 15.05.2020 включно;
- Завідувач сектору Мухойд Олександр Вікторович табельний №1625 з 26.05.2020 по 29.05.2020 включно;
- Провідний інженер-електронік Сирота Олександр Анатолійович табельний №3030 з 25.05.2020 по 29.05.2020 включно;

Завідувач відділу №11

М.В.Кобзарь

Рис. 4. Сформированный документ

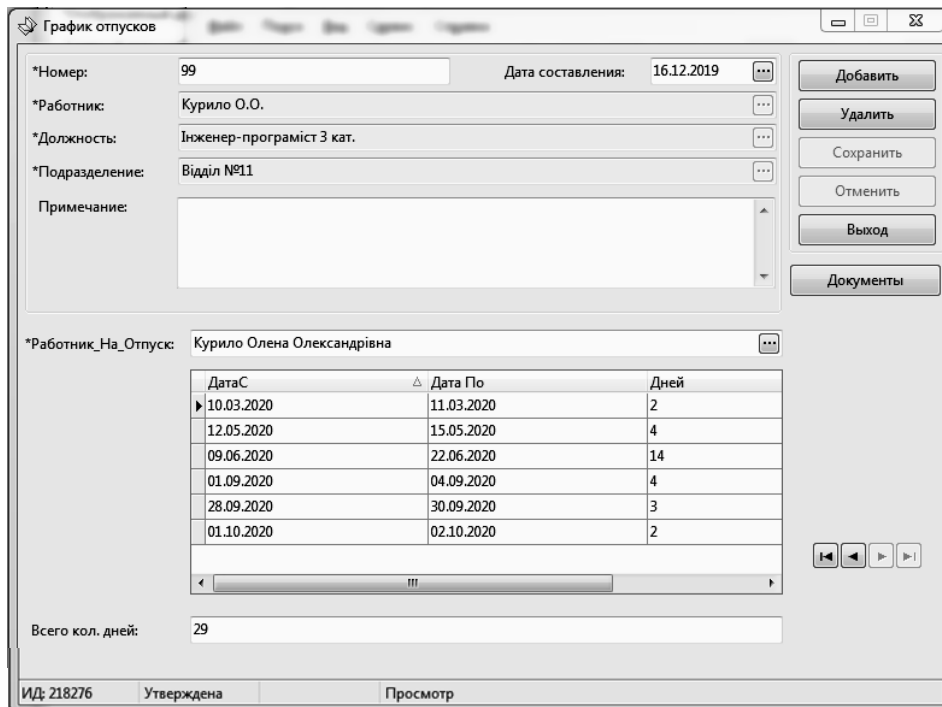


Рис. 5. Карточка «Графика отпусков»

Затем необходимо получить данные о работнике — фамилия, имя, отчество, должность и табельный номер.

РЕШЕНИЕ ЗАДАЧИ С ПОМОЩЬЮ СЦЕНАРИЯ

На первом этапе автоматизации информация размещалась непосредственно в сценарии карточки документа, при выполнении которого в момент создания служебной записки формировались необходимые данные для основной части. Информация была изложена в следующем виде и должна указываться в коде сценария для каждого сотрудника организации:

```

if User.Name == 'kurilo_e' FIO = 'Курило О.О.' Dolgnost =
'Инженер-програмист 3 кат. відділу №11' Tabel = '1632' endif
if User.Name == 'babak' FIO = 'Бабак В.О.' Dolgnost =
'Инженер-програмист відділу№11' Tabel = '1633' endif
if User.Name == 'anischenko' FIO = 'Анищенко В.М.' Dolgnost =
'Инженер-програмист 3 кат. відділу №11' Tabel = '3004' endif
if User.Name == 'shifrin' FIO = 'Шифрин Д.М.' Dolgnost =
'Головний конструктор з інформаційних технологій' Tabel = '1600'
endif

```

Такой подход имел ряд недостатков. Во-первых, программисту нужно было постоянно контролировать актуальность данных в коде сценария, отслеживать прием на работу новых сотрудников и заполнять соответствующие данные. Во-вторых, из-за большого объема данных и множества условных операторов увеличивался объем кода, что существенно ухудшало его читабельность и восприятие.

Поскольку необходимые данные содержатся в справочниках СЭД, было принято решение модифицировать проблемный участок кода.

РЕШЕНИЕ ЗАДАЧИ С ПОМОЩЬЮ API

На втором этапе автоматизации получение информации реализовывалось с помощью API — функций системы. Программный интерфейс приложения (API) DIRECTUM предоставляет несколько методов для работы с источниками данных системы [4]. Один из них — метод для работы со справочниками. За основу взята штатная функция CreateReference(), с помощью которой было реализовано обращение к справочнику «Особові картки працівників» с целью получения значений таких реквизитов, как должность, полное имя, табельный номер и название подразделения.

В коде сценария обращение к справочнику для получения должности сотрудника было реализовано следующим образом:

```
RefBook = CreateReference('ОсобоваКартка')
text = ""
foreach Line in RefBook
    tmp = Line.Requisites('Должность').Value
    if tmp <<>> null
        text = text & tmp & CR
    endif
endforeach
EditText( text )
```

При этом способе из справочника формировался массив данных, из которого затем информация вносилась в шаблон документа, выбор необходимых данных происходил уже при занесении непосредственно в документ.

Это решение помогло избавиться от необходимости контроля актуальности данных в коде сценария, а также улучшить читабельность кода, но возник более существенный недостаток — время работы скрипта прямо пропорционально увеличивалось в зависимости от количества пользователей в списке отпускников для периода, и было равно 20 секундам для одного отпускника. Так как в среднем количество сотрудников подразделения составляет 10 человек, для получения данных для документа потребовалось более трех минут ожидания. А если учитывать, что в перспективе планируется создавать один общий документ для всех подразделений, то в период, когда количество отпускников будет составлять до 50 сотрудников, время на программный сбор информации может составить более 15 минут.

РЕШЕНИЕ ЗАДАЧИ С ПОМОЩЬЮ SQL-ЗАПРОСОВ

Поскольку такое быстрое действие программы является неприемлемым, а метод использования API функций оказался неэффективным в данной ситуации, было принято решение отказаться от программного интерфейса доступа к справочникам системы, а взаимодействовать непосредственно с базой данных. Для извлечения данных из базы используется SQL — язык структурированных запросов [5].

Далее приведен SQL-запрос для получения необходимого массива информации из базы данных:

```
SQLQuery = "
SELECT DISTINCT
```

```

        ok.Dop + ' ' + ok.Dop2 + ' ' + ok.Dop3 AS FullNameUK,
        d.DataS AS DateStart,
        d.DataPo AS DateEnd,
        ok.Tabelnyui AS PersonNumb,
        pos.NameAn AS Position
    FROM MBAnalit gr
        JOIN MBAnValR5 d ON (d.Analit = gr.Analit)
        JOIN MBAnalit ok ON (ok.Vid = 3443 AND ok.Polzovatel =
gr.RabotnikLOtp)
        LEFT JOIN MBAnalit pos ON (pos.Vid = 3300 AND
pos.Analit = ok.Dolzhnost)
        JOIN MBAnalit dep ON (dep.Vid = 446 AND dep.Analit =
gr.Podr)
    WHERE gr.Vid = 3505
        AND ( (d.DataS BETWEEN '" & dateBegin & "' AND '" &
dateEnd & "')
        OR (d.DataPo BETWEEN '" & dateBegin & "' AND '" &
dateEnd & "') )
        AND dep.Kod LIKE '%" & subDivision & "'
"
    Query = CreateQuery()
    Query.CommandText = SQLQuery
    Query.Open

    foreach Record in Query
        position = Record.Fields('Position').Value
        fullNameUK = Record.Fields('FullNameUK').Value
        personNumb = Record.Fields('PersonNumb').Value
        dateStart = Record.Fields('DateStart').Value
        dateEnd = Record.Fields('DateEnd').Value
        vacationer = "- " & position & " " & fullNameUK & " та-
бельный №" & personNumb & CR
        dateRange = "з " & dateStart & " по " & dateEnd & " включно;"
        if in(buffer; vacationer; false)
            buffer[0; i] = ""
            buffer[1; i] = dateRange
        else
            buffer[0; i] = vacationer
            buffer[1; i] = dateRange
        endif
        i = i + 1
    endforeach

```

Данный способ представляет собой получение информации с помощью SQL-запроса, необходимая информация выбирается, минуя API интерфейс, взаимодействие происходит непосредственно с таблицами базы данных СЭД. В коде запроса был указан поиск из нескольких таблиц данных — «График отпусков», «Работники», «Должности», «Подразделения». Для ускорения работы запроса указывались точные параметры поиска, конкретный временной промежуток и подразделение.

Вследствие высокой скорости работы SQL-запроса это решение полностью избавило модуль от описанных выше недостатков и повысило его быстродействие. На получение информации из базы вне зависимости от количества сотрудников уходит всего несколько секунд.

ВЫВОДЫ

Таким образом, главной целью автоматизации являлось получение данных для служебной записки на отпуск и формирование полностью оформленного документа. Создание и заполнение данного документа до автоматизации составляло от 30 до 60 минут в зависимости от количества сотрудников конкретного подразделения.

В ходе автоматизации создания служебной записки об отпусках и организации работы модуля с помощью SQL-запросов существенно повысилось быстродействие программы. После проведения автоматизации документ формируется за несколько секунд.

Задачу автоматизации ведения отпусков можно считать частично выполненной. Получение информации для служебной записки об отпусках происходит в автоматическом режиме, что позволило существенно уменьшить время на поиск информации и заполнение документа, а также избежать ошибок при ручном вводе данных.

В дальнейшем планируется усовершенствование модуля ведения отпусков — использование данных из единого источника, которым будет справочник «График отпусков», всеми участниками процесса, автоматизированное формирование данных для бухгалтерского учета и отдела планово-экономического анализа, формирование общего приказа на отпуск.

ЛИТЕРАТУРА

1. И.К. Корнеев, *Информационные технологии в работе с документами*, изд-во “Проспект”, 2018, 304 с.
2. Н.Ф. Алтухова, А.Л. Дзюбенко, В.В. Лосева, и Ю.Б. Чечиков, *Системы электронного документооборота*, изд-во “КноРус”, 2019, 208 с.
3. С.Л. Кузнецов, *Современные технологии документационного обеспечения управления*, изд-во “Термика”, 2017, 470 с.
4. *Онлайн руководство разработчика Directum*. Доступно: https://club.directum.ru/webhelp/directum/5.8/index.html?collection_dirdev.htm
5. М. Грабер, *SQL*, изд-во “Лори”, 2017, 400 с.

Поступила 01.09.2020

INFORMATION ON THE ARTICLE

Oleksandr V. Mukhoid, JSC “VNIIAEN”, Ukraine, e-mail: muhoid@vniiaen.sumy.ua

Olena O. Kurylo, JSC “VNIIAEN”, Ukraine, e-mail: sapr@vniiaen.sumy.ua

EXTENSIBLE INTEGRATION OF THE ELECTRONIC DOCUMENT MANAGEMENT SYSTEM WITH PRODUCTIVITY SOFTWARE (HR MANAGEMENT) / O.V. Mukhoid, O.O. Kurylo

Abstract. This article deals with the vacation tracking in the company and the automation of this process. A typical life cycle of a document and the main disadvantages of the life cycle of paperwork are given in the article. The authors' goal was to consider in detail the currently existing vacation tracking algorithm. The complexity of the process was assessed. The article also summarizes the disadvantages of this algorithm, such as the sub-optimal use of working time and the possibility of a human error. Further, the stages of automation, methods of solving the

problem are described in detail. The mechanisms for obtaining the necessary data are presented, and obtained results are analyzed. It also provides a long-term plan for further automation of the vacation tracking process.

Keywords: electronic document management system, application development, process automation, vacation tracking.

РОЗШИРЕНА ІНТЕГРАЦІЯ СИСТЕМИ ЕЛЕКТРОННОГО ДОКУМЕНТООБІГУ З ПАКЕТОМ ОФІСНИХ ПРОГРАМ (РОБОТА З ПЕРСОНАЛОМ) / О.В. Мухойд, О.О. Курило

Анотація. Присвячено проблематиці ведення відпусток на підприємстві та автоматизації цього процесу. Описано типовий життєвий цикл документа, наведено основні недоліки життєвого циклу паперового документообігу. Поставлено завдання детально розглянути існуючий алгоритм ведення відпусток. Виконано оцінку трудомісткості процесу і наведено недоліки даного алгоритму, зокрема неоптимальне використання робочого часу і можливість появи помилок. Детально описано етапи проведення автоматизації, способи вирішення поставленого завдання. Наведено механізми отримання необхідних даних, проаналізовано отримані результати. Подано перспективний план подальшої автоматизації процесу ведення відпусток.

Ключові слова: система електронного документообігу, прикладне розроблення, автоматизація процесу, ведення відпусток.

REFERENCES

1. I.K. Korneev, *Information Technologies in Document Management*, 2018.
2. Yu.B. Chechikov, A.L. Dziubenko, V.V. Loseva, and Yu.B. Chechikov, *Electronic Document Management Systems*, 2019.
3. S.L. Kuznetsov, *Modern Technologies of Management Document Support*, 2017.
4. *Online developer guide Directum*. Available: https://club.directum.ru/webhelp/directum/5.8/index.html?collection_dirdev.htm
5. M. Graber, *SQL*, Lory, 2017, 400 p.

SYSTEMATICITY OF ORGANIZATIONAL MANAGEMENT

I.N. DROGOBYTSKIJ

Abstract. Organizational management is the weakest element in the evolution of the modern civilization. It creates a lot of problems at all levels of its hierarchy that have not been solved for a long time. Overcoming these problems requires streamlining, improving and enhancing the instrumental methodological arsenal of organizational management, creating a system and thus guaranteeing results and efficiency of its practical use. In this article, we make an attempt to formulate the ideas and content for further studies. Integrating them in the systemizing stream, i.e. building/restoring the systematic character of organizational management, this article addresses its basics, such as definitions, fundamental principles, the style difference and system typology that should create main landmarks in the field of a new scientific approach to an interested reader.

Keywords: system, management, administration, science, art, skills, vocation, style of management, systematic management typology.

INTRODUCTION

Imperfection of modern management is registered by many scientific, commercial and political authorities. It penetrates the whole modern social structure and is visible at all its levels. So, at a mega level, the level of human civilization, this imperfection leads to a multitude of unsolved civilizational problems — poverty, starvation, inequality, narcotic addiction, environmental problems, that we egoistically delegate to the next generations to solve.

At a macro level, the level of individual countries, imperfection of management has a national touch that in every single case leads to a series of specific (national) problems. In Russia, for instance, it is the resource-based national economy, challenging transition to the innovative trajectory, imbalance in the industry structure, low labor efficiency, horrendous social segregation, extreme bureaucratization, all those are on the agenda for a long time and no solution is found.

The registered problems are transitioned to a meso level, the level of individual industries and economic regions, get multiplied and intensified there and lead to uneven and unbalanced growth of many industry-based and regional systems, their structural incompleteness and instability, limited internal options for enhanced reproductive performance and many other problems calling for immediate solutions.

At a micro level, the level of business enterprises, such imperfection of organizational management is most notably shown as an imbalance of their internal operational mechanisms, uncoordinated interaction with other entities that leads to violation of their integrity, suppression of reproductive processes and growth of social tension between staff, managers and owners. Besides, the dominance of autocratic leadership style is a reason for highly subjective managerial decisions, it leads to deconsolidation of workforce and doesn't help employees to seamlessly combine their mental abilities and their roles in the reproduction process [13].

Still, the most notable impact that the imperfection of organizational management has is registered with the last participant of economic activities — individual — that is positioned at a nano level and completes the economic hierarchy. Individual is under a constant pressure of both their own problems (discontent with the workplace, job conditions, financial remuneration, growing qualification requirements) and the problems of all the higher economic levels. Targeted actions towards individuals should take into consideration their current mental state [6] and their established behavioral patterns. Besides, the following pattern should be regarded: the lower the individual is at the socioeconomic ranking, the more difficult it is to make him or her change their behavior. So, the uncertainty of organizational management at each level of economic hierarchy is the essential trait of the current situation.

As it is known, any appearing/existing problem is a direct consequence of non-consistency (inadequate consistency) in the researched field [8, p. 12]. Hence, to alter the negative tendencies in the organizational management dynamics, we need to add (restore) its consistency, i.e. *systematize* it. We need to move to a new, higher level of consistency, give life to a fading system of organizational management and provide its further self-development, self-sustainment and self-education.

SYSTEM DEFINITION OF ORGANIZATIONAL MANAGEMENT

Before moving directly to consistency of organizational management, let's start with its definition — what does “organizational management” consist of? The semantic analysis of this phrase instantly leads us to a following conclusion. As “management” means “administration”, the word “organizational” should obviously mean the area of applying these administrative activities, or creating a special environment where purposeful managerial decisions will be made, taken and implemented. As for the first option, the adjective is associated with organization as an economic system classified as an object, for the second — with organization as an economic system classified as an environment. So, the phrase “organizational management” refers to a special type of professional activity aimed at managing economic systems where the key role belongs to people. In that case the concepts of organizational management and organizational administration can be considered synonyms and used simultaneously [14].

Still, to systemize the management science and for practical purposes it is necessary to have a clear definition of their subject. As it was already mentioned, the concepts of “management” and “administration” used to define human influence on development of enterprises, production processes, innovative projects, organized markets, creative communities, professional organizations — everything that covers the variety of organizational systems — are quite obvious and

don't require further explanations. It is a reality for practicing managers, so they don't bother about definitions while creating new regulations and processes or giving verbal directions to staff members. But for scientists that study the theory and/or methodology of organizational management, the absence of a clear definition for the researched subject causes major problems. When the definition is absent, all regulations, requirements and recommendations don't always have a solid theoretical basis and more often remind theoretical fantasies than scientific results. In order to find a generally acceptable definition for organizational management we will try to define and analyze its main features that reflect the subject of this purposeful type of human activity.

Firstly, organizational management is an information process aimed to reduce entropy of the managed system. Entropy is the feature of a system that describes the degree of its order and inner organization: the higher is the level of organization the lower is the entropy and vice versa. With no management applied, the entropy is increasing, the system loses its structure and order, turns into chaos. Organizational management prevents the system from degrading and shifting to ambiguity. Fig. 1 shows the classic contour of organizational management. All arrows represent information flows. I.e. procurement, transformation, analysis and usage of information are inherently connected to management through all its life cycle. Information is both the "daily bread" for organizational management (flows $S(t)$, $S_{pl}(t)$ and $\Delta S(t)$) and its final accord (flow $u(t+1)$). Therefore, organizational management doesn't just consume information but also produces it.

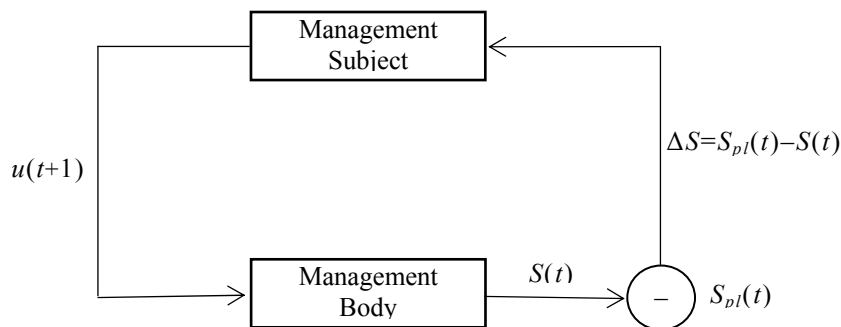


Fig. 1. Classic organizational management contour: $S_{pl}(t)$ — planned system status at the time point t ; $S(t)$ — actual system status at the time point t ; ΔS — difference between the planned and the actual system status; $u(t+1)$ — administrative impact on the system at the next $(t+1)$ administrative step

Secondly, administrative activity is a manifestation of the manager's/team's conscious desire to make a certain impact on the managed system, object, process, project or background. Herewith they choose the most efficient (in their perception) instrument from a wide range of available administrative methods — issue an order or an instruction, give verbal

directions, negotiate a contract for certain works or initiate a whole prearranged action plan. Hence, management is a fruit of human brain. Emotions and instincts that definitely accompany the processes of making, taking and implementing administrative decisions are of a marginal significance.

Thirdly, the nature of organizational management is energetic. The ability of a manager to influence the managed system can't be explained other than by its energy influence. Thus, if a person moves an item from one place to another, they

perform a work; if they give a direction to another person to move the item — they already perform a managerial activity. Administrative impact in this case is similar to an actuated electric relay switch. Its momentum is energetically low but leads to much more significant energetic results.

Despite not knowing yet the nature of managerial energy and the ways to measure it, we still use it since long, in a rather efficient way. The energetic imperatives discussed on the pages of scientific publications — will, authority, enforcement and influence — are nothing else but certain types of managerial energy [1, 7, 9]. Successful implementation of any managerial decision requires a certain amount of each energy type and a proper combination of them. The ability to concentrate a required amount of a case-specific managerial energy in a proper place and at a proper time is the sign of high-level managerial skills that deserve encouragement.

Fourthly, organizational management is a goal-oriented process. The goal of the managed system is some kind of a beacon that a manager uses to organize and drive other people's efforts in order to continuously move towards that goal and finally reach it. Management with no specific goal results in fruitless wandering of the managed system in space and time, capable solely to fulfill the wannabe manager's maniacal ambitions. Lucius Annaeus Seneca has said once a great phrase about this: "...When a man does not know what harbor he is making for, no wind is the right wind".

Those features of management — information, conscious, energetic and targeting — are nothing else but signs of consistency. Looking at the seemingly continuous managerial process U from the point of the systems theory, we can see its elements (managerial impacts) u_i that are applied not randomly but in a very specific order

$$U = \langle u_1, u_1, \dots, u_i \rangle .$$

More so, each managerial impact is a symbiosis of three acts

$$u_i = \langle u_{i_1}, u_{i_2}, u_{i_3} \rangle , \quad i = 1, 2, \dots,$$

where

u_{i_1} is an act of making a managerial decision;

u_{i_2} is an act of taking a managerial decision;

u_{i_3} is an act of implementing a managerial decision.

Such conscious, sequence-aligned and goal-oriented synergy of elements is the most valid evidence of consistency [8]. Hence, managerial activity is systematic, and this fact allows us to use the whole lot of the powerful system methodology for its analysis. First, it can be used to restore the consistency of organizational management at a certain enterprise in case it was lost.

The above-mentioned concepts related to organizational management are quite enough to synthesize its new, system definition. It can be formulated as follows: organizational management is a cognitive, informational and energy impact applied to the system and its immediate surroundings that pursues the goal to resolve the problems preventing its harmonious functioning and dynamic development.

This definition explains that the goal of organizational management is to discover, configure, find solutions and directly eliminate problems arising on the way of a functioning and dynamically developing economic system. Furthermore,

a functioning system is associated with expanded production of goods and services in accordance with its mission (intended purpose) while development relates to modifying the inner structure of the system to make it better fulfill its mission and reach assigned strategic goals.

No management is needed where no problems are registered, and no problems are registered when there is no dynamic development (life). So, to manage is to set and clear a mission-life path in space and time for the managed system to follow. Metaphorically speaking, managers in economic systems perform the same functions as sweepers in curling, ice-breaking ships in winter navigation or evangelists in religion. They should neutralize all inner troubles and outer turmoil and provide comfortable conditions for the managed system, allowing it to function and develop. Such interpretation of organizational management explains why management goes far beyond the physical boundaries of managed systems.

Organizational management is currently very popular. Each year new forces are getting involved, managerial scope and complexity are growing, number of educational institutions that produce managers for the national economy is increasing. In order to provide the required methodological support for the latter, many manuals, study guides, audio and video tutorials are produced. Being issued by various authors that belong to different scientific traditions, those educational materials sometimes contradict each other and blur the current situation instead of clarifying it.

The point is, scientific world hasn't reached the consensus yet about the nature of organizational management. Some regard management as an art of finding the right way to develop the managed system at the competitive market, others interpret it as an act of force that makes the system strictly follow (function and develop) a fixed path in space and time, while some connect it to the self-regulation and self-development of the system, mediated by its managers. Depending on the author's preferences about the nature of management, we get a varying content that cannot be brought to a common standard. This is the reason why knowledge in management is not codified, certified or measured, as it is in engineering, medicine or military science. Teaching in management still uses more qualitative references — authoritative, intuitive, balanced — than quantitative. So, the first step in systematizing organizational management is to structure and zone its domain.

FUNDAMENTALS OF ZONING IN ORGANIZATIONAL MANAGEMENT

Systematizing of organization management is a multifaceted process of adding (restoring) consistency to the methods and instruments used in making, taking and implementing managerial decisions. While doing so, we should take into consideration the natural identity of managerial activities. Following the Henry Mintzberg theory [15, p. 143), I dare assume that this identity is based on four fundamental cornerstones — craft, science, art and vocation. The weight of each cornerstone in different management entities, whether it is a team or a standalone manager, is very individual, and it gives a distinctive charm to their managerial activity. Such differences can be taken as a basis for structuring the management domain.

Craft is present in those areas of human activity where production or management knowledge is not subject to formalization and thus are unalienable

from its bearer. Transition of knowledge, skills and experience in such areas is possible only when master-teacher and apprentice-student work in tandem. It appears that all kinds of professional activities in various historical periods have witnessed the crafting stage in passing knowledge, skills and experience. The crafting period ended only when the required knowledge got described by algorithms, standards and procedures in instructions, regulations, manuals and guides, when professional trainings got concentrated in special educational institutions, from trade schools to universities.

In management this transition from craftsmanship to professionalism isn't complete yet. While in manufacturing, military art and construction the humankind has already switched to professional management trainings, the same kind of trainings in culture, sports, medicine, governmental and municipal administration has started in the current millennium, while the politics, science and art still train their managers and administrators through craftsmanship methods. Even the domains of human activity where management trainings have been set to professional footing, "craftsmen" are still present, which is not legislatively permitted in the other areas. For example, it is hard to imagine a person with no professional background that would legitimately practice as a doctor or a lawyer.

Science in any domain is associated with systematization, development and practical application of knowledge obtained while researching its subject. Humankind has been trying to understand the scientific nature of management during the whole period of its history [14]. During the last 100 years these attempts have acquired a specific purposefulness and materialized in a whole range of theoretical concepts, including the Frederick Taylor theory of scientific management, the Chester Barnard theory of organization, the George Mayo and Fritz Roethlisberger theory of human relationships, the Herbert Simon informational paradigm of management, the Michael Porter concept of strategic management, The Thomas Peters, Robert Waterman, Edgar Schein and Jeffrey Pfeffer theory of organizational culture, the integrational concept of management by Peter Draker, Peter Senge, Coimbatore Prahalad, Gary Hamel, George Stalk, James Evans and Zinovy Shulman, the Ichak Adizes style theory of management, the concept of behavioral economy by Daniel Kahneman, Amos Tverski, Richard Tyler and others [1, 5, 10, 16]. Such an impressive list of theoretical concepts in management is another proof that this domain lacks of a generally acknowledged scientific theory for its main subject — organizational management.

The absence of the general theory of management partially explains poor condition of its practical part. Practicing managers that make, take and implement specific managerial decisions can rely on anything: theoretical concepts of management that are traditional for their company or call to them, main commandments of their religion, utilitarian philosophy, their own intuition, their current mood etc. As a result, organizational management is often chaotic, with incoherent actions and absence of inner logic in administrative decisions.

Art plays a huge role in modern management; it is associated with inspiration caused by the manager's rising spiritual and physical energy. In turn, the art of management is built on intuition that is perceived as a solution of a certain administrative task through subconscious analysis of the current situation, with no understanding of its inner logics and/or insufficient information and/or absence of a standard search-for-solution pattern. Strong concentration, aggregation of all available information, high level of knowledge and expertise in the activity field of the managed entity form the triad of necessary conditions required for the

search of an intuitive (creative) solution for any administrative task. Sufficient conditions arise from the manager's higher nervous activity as a result of a heightened attention level, activated memory, broadened perception and new level of consciousness, all referring to the field of psychology.

Intuitive management requires a high amount of knowledge in behavioural science, human relationships, management ethics as well as ability to adapt this knowledge to the context. Professional expertise in the field of the managed entity is a must for a person in the managerial position, and the higher the level of expertise is, the better. Surely, intuitive management is referred as one of the new types of human activity that lack of standard, formal algorithms of making and taking administrative decisions, or one of the areas where such formalization is basically impossible. Still, that doesn't mean that traditional areas where such algorithms exist and constantly develop, give no place for intuitive management. In this case we should talk only about the balance of formal and intuitive, traditional and creative, old and new, craft and art [7]. Wishing to make new steps in order to improve the practical aspects of management, we have to not only address familiar patterns and standard procedures but also perceive the hidden reality, making scientific hypotheses and showing creative initiative.

Vocation in management is often considered as a desire to perform as a manager. In my earnest belief, inner motivation to manage is just a necessary condition of vocation. True vocation, along with the desire, should also include sufficient conditions, namely specific abilities and required (at least basic) level of knowledge.

Vocation-driven managers are always suitable for their position. Trying to fulfill their ambitions, on the one hand, and justify the credit of confidence from staff members/entity/executive who has appointed them to the position, on the other hand, they have a creative view on their responsibilities, they experiment and, at the end, achieve visible success. Such managers are usually not afraid to take responsibility, are able to set balanced relationships between staff members, create an enabling environment and, of course, are well regarded by both their colleagues and external stakeholders.

It is not uncommon that in situations of extreme gravity the managing position goes to a professional with huge expertise but no desire to manage. Such members of management team are called "necessity managers" [19]. They usually combine the managing responsibilities with the work they have done before being promoted (for example, university president keeps teaching students and being the head of department, director of a medical center keeps treating patients, film director keeps acting and performs one of the leading parts along with directing a show or a movie). Considering the fact that management is not a job that can be done without sincere desire, such "necessity managers" do not stay long in their positions. They "bear the weight" of managing till they find someone suitable to replace them and convince staff members/entity/executive who has appointed them, of their choice, or eventually they stop being "necessity managers" and fully dedicate themselves to management.

The worst case is when the managerial position goes to a person with "reduced" vocation — they have motivation to manage but lack of abilities and knowledge — not unusual in the Russian practice. This self-assured know-nothing is not just unable to provide further development of the managed system but even can't support its existing level of operations. The arrival of such manager is usually the sign of a system starting to fade.

MANAGEMENT STYLISTICS

It is reasonable to suggest that the four basic elements of management, mentioned above, somehow define the stylistic focus of organizational management in each single case. They, so to speak, create a certain stylistic circuit that encompasses every existing and described, as well as unknown and undescribed styles (manners) of practical management. Below you can find one of the options to organize such circuit.

Consecutive analysis of different style dominants, set in pairings, makes us assume that they all create certain continuous multitudes where one style feature slips into another: “science – art”, “science – craft”, “science – vocation”, “art – craft”, “art – vocation”, “vocation – craft”. Combining these continuous sequences, we get a certain closed area (circle), or better say, sphere that cuts the space of organizational management from unorganized surroundings (Fig. 2). Each corner — “craft”, “science”, “art”, “vocation” — is associated with a certain gravity force that tends to give its style tinge to the style of management, typical for the manager caught in the organizational management force field. So, the corners in the mentioned diagram refer to “pure” organizational management styles that are almost never present in real life, while the whole closed space is distributed between existing practical styles of management in accordance with their preference for one or another corner. Therefore, we can define 13 style variations of management.

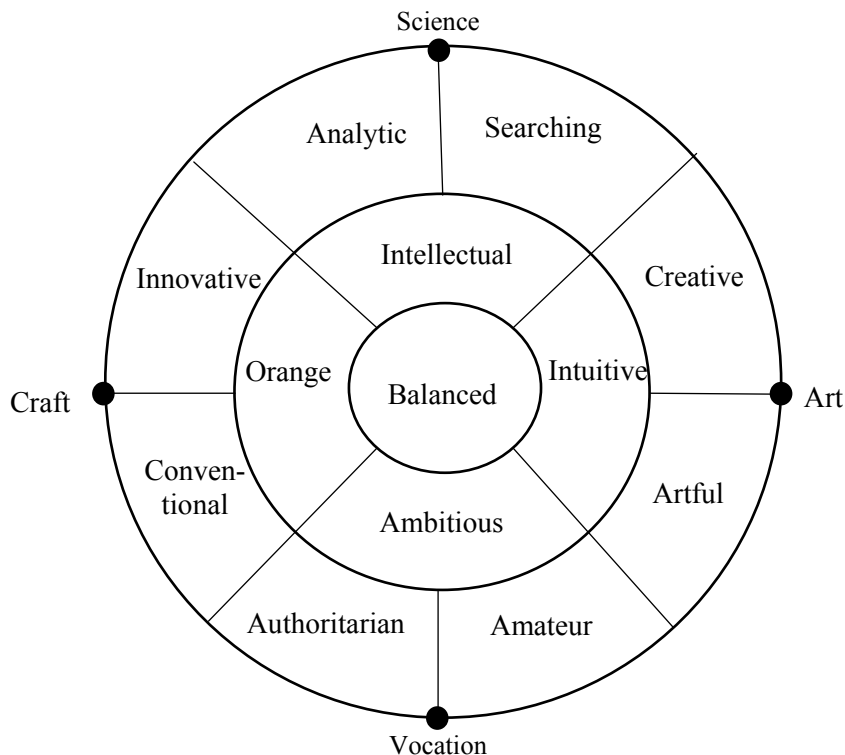


Fig. 2. Style space of organizational management

This diagram has a clear sectoral structure. The base of each sector is formed by a curve with the center in the corner that refers to one of four fundamental

elements of management — craft, science, art or vocation. All four sectors end in the same point (or, better say, same area) that is associated with the balanced or systematic style of organizational management. Each basic corner at the border of the organized management space is adjoined by two style variations of management, while another one, together with the cross-functional balanced variation of management, marks the smooth transition of its style feature to the opposite sector.

So, the “science” corner is closely bordered on the *analytical* style where the science methods are primarily used to define current achievements, development trends in the managed entity that generally don’t cross the border of existing traditions and experience, and on the *searching* style where, along with the science methods, there is a clear inclination to intuitive elements that work as indicators of creative approach that gravitates towards arts. Transition to the opposite sector is made through the *intellectual* style of management based on the mental model of managerial activities [6], the model that reflects individual features of making, taking and implementing managerial decisions.

The “art” corner is closely bordered on the creative and artful styles of management. In *creative* management art works as an instrument for searching unconventional, unorthodox approaches to management that still have a scientific base and a practical application in other domains of human activity. The *artful* style of management implies synthesis and implementation of managerial decisions “from scratch”. It doesn’t suppose any inherent scientific ground for these decisions. Such ground may appear later as a result of a theoretical conceptualization of the managerial decision made by a talented creative person with the vocation. The *intuitive* style of management is a derivative of creative and artful and serves as a first step of transitioning to the opposite sector, it tends to fit the decisions already made, to the Procrustean bed of the traditional (craft) approach.

The “vocation” corner is bordered on authoritarian and amateur styles of management. The *authoritarian* style is defined by a highly centralized executive team, disregard of staff opinions, strict discipline, rejection of subordinates’ initiatives. The *amateur* style of management is bordered with the artful style but it doesn’t require from the manager to have a mandatory methodological background and a solid knowledge base. This style is included in the general style diagram to reflect the existence of people who serve as managers without any professional managerial education. The *ambitious* style of management takes an intermediate position at the “vocation – science” axis. It is typical for high-tech enterprises. Ambitious leaders that rarely disclose their thoughts, so their subordinates frequently have to guess what exactly they have to do usually lead them.

The “craft” corner is bordered with the conventional and the innovative styles of management. The *conventional* style of organizational management is typical for many traditional areas of business. The etymology of the word speaks for itself: making, taking and implementing managerial decisions is performed in accordance with existing conventions and doesn’t suppose creative breakthroughs. It is not by chance that this style of management stands close to the authoritarian style that is considered the standard of management. The *innovative* style of management is effected within the borders of the traditional pattern of making, taking and implementing managerial decisions but allows (and stimulates) improvisation targeted to increase the efficiency and effectiveness of management. As every managerial decision is based on three acts (making, taking

and implementing the decision), the abilities of the innovative style of management are almost unlimited. The Orange style of management takes the intermediate position at the “craft – art” axis and defines the highest level of conventional and innovative styles. It unconventionally interprets the conventional pattern, using both scientific approaches and elements of creativity (art).

The central position of the diagram is taken by the *balanced*, or *systematic* style of management that seamlessly combines all four fundamental basics — craft, science, art and vocation. In managerial activities it should be associated with the ideal reference that all managers must strive to achieve. It requires a radical revision of study plans to provide such managers for the national economy. A decent place should be given to new courses — management psychology, management culture, theory of organizational systems, theory of practice etc. — many of those are to be created yet [7, p. 134].

Unfortunately, in real life we have to deal with a highly unbalanced management. Some cases show total predomination of a certain element, and the management gravitates towards degenerative styles that can affect the managed entity in a negative way: a degenerative science style can lead to excess mercantilism; a degenerative art style — to narcissism (self-admiration); pure vocation can degenerate into “absolute monarchy” while pure craft — to conservatism, when the manager refuses to go beyond their personal managerial experience. Even a combination of just two elements leads to a troubled management. Science and art without craft and vocation lead to *distracted* style of management, out of synch from reality. Science and craft without art and vocation mean *withered* (spiritless) style of management. Craft and vocation without art and science can lead to the *fading* style of management, while a combination of art and vocation without science and craft creates the *adventurous* style of management that can bring disasters and terminate the business at any moment.

Success is available only in case of a more or less balanced management, based on all four management dominants. In this regard, it is strongly recommended to evaluate the balance of organizational management, performed by each manager and the managerial team as a whole [12].

SYSTEM CLASSIFICATION OF ORGANIZATIONAL MANAGEMENT

As it is known, axiomatics and content of organizational management are defined by the management body and the management subject. Consistency levels of organizational management are discerned depending on their structuredness, organization and systematization.

Management body is an agglomeration of all components of the managed system that require managerial actions: elements, their interconnections, outbound connections, factors that influence the system and its surroundings, technological processes, financial flows etc. All mentioned components must be bound in a general operated agglomeration or connected to the managed domain. Certain moments from the list can move the boundaries of the domain far beyond the borders of the managed system. If the elements are not connected, at least partially, or their connection is not holistic, we are dealing with a non-systematic domain of various levels of non-consistency.

Management subject is usually formed by a group of people who have been delegated the authority to make, take and implement managerial decisions for the managed system. There are various levels of the management entity consistency, depending on how their executive authorities for certain elements of the managed domain are distributed, the way they interact with each other while making, taking and implementing managerial decisions, and the frequency of applying for external services. The highest level of consistency is attributed to the complementary management team able to be flexible and adaptable to the changes in activities and growth of the managed system. Not hard to guess that the main responsibility for the consistency of the management subject goes to the organizational structure [5, 15]. This is where each member of the management team gets their specific area of management, their authorities, instruments of coercion and influence, as well as determines relationships with other team members.

Certain system types of management are defined, depending on the level of consistency in the management subject and the management entity. The Table presents a basic typology of organizational management consistency that can be significantly extended in both directions if needed.

Basic Typology of Organizational Management Consistency

Management Subject	Management Entity	
	Non-systematic	Systematic
Non-systematic	Manual management	Institutional management
Systematic	Strategic management	System management

If a non-systematic Subject is managed by a non-systematic subject, we can see *manual management*. For every specific problem arising throughout the functioning and development of the system, an individual decision is made by one of the members of the disconnected management team who can yet be affiliated to a certain group of people. Should the same problem arising again, the team member and the decision can be completely different.

If a well-structured and systematically organized subject takes the decision for a problem arising in a non-systematic management entity, we are dealing with *institutional management*. That means that decision-making is delegated to a certain member of the management team, and they take it in accordance with existing rules, regulations, traditions and other institutions shared with other participants of the management process. There is a strict navigation of problems in the management subject (team), defined by the manager authorities appointed through the distribution of rights and duties within the management team.

If a problem arises in a well-structured management subject but the managerial decision leading to its elimination or reduction of its negative influence on the managed system and/or its surroundings is made by a not completely structured management team (management entity), based on the principles, rules and patterns traditional for their strategy of growth, we witness *strategic management*. It is not hard to guess that the detailed strategy for the foreseeable future is a result of the properly structured and systematized management entity.

Finally, if a problem arises in a systematized subject and an integral management team “packed” in a flexible organizational structure takes the decision, we can see *systematic (balanced) management*. In that case the decision-making person (DMP) has credible resources to consider all nuances of the current condition of the system, predict desired values of managed parameters for the foreseeable future and find the solution that will allow it to function and develop harmoniously at the current stage.

In order to implement system management it is essential to ensure that reviewing and systematization of the management subject, on the one hand, and creation and system structuring of the management team, on the other hand, are effected concurrently and stay harmoniously united throughout the whole life cycle of the managed system. The practical implementation of this rule requires its further elaboration to a certain set of rules (principles) that should be taken in consideration while adding consistency to the management subject and management entity. We will briefly mention its key moments.

Organizational management subject should be planned very thoroughly, consequently drawing its parts with consideration to every element, connection, factor, technological process, financial flow, social aspect that will be subjected to management afterwards. At this stage it becomes obvious that some parts need to be expanded (broadened), others, to the contrary, are to be reduced and made more compact, while some will need to be built from the scratch. As a result of connecting all parts into an integral entity, we get the expected systematic management subject that will become the scene for the management entity’s activities in the foreseeable future.

Depending of the entity that requires management, we need to create a team of managers and build an adequate organizational management structure. We should take into consideration not only professional requirements for the managers that will be responsible for this or that element (subsystem) of the management subject but also the desired style of management that depends on the subsystem type and on the management subject. Such two-dimensional specifics of the management team members (their profiles and management styles) allows us to create a holistic, functional team where the weaknesses of some members are compensated by the strengths of others, and their joint efforts are goal-oriented and systematic [7, 9].

With the task of matching the management subject area and the management team’s powers of authority, timelines of their life cycles are to be taken into consideration. If the management team’s (management entity’s) life cycle is significantly shorter than the management subject (managed system) life cycle, there is a risk of time-servers who will care only about their own interests at the expense of the managed system and its surroundings. In that case management designers should implement strict formalized procedures for recruiting, arranging and rotating executives. If the management team’s life cycle is significantly longer than the life cycle of the managed system — a typical situation for project and process systems [11, 12, 13], then we should think about properly using the highly qualified executives after the life cycle of the system is over.

Systematization of the management subject and the management entity culminates into the system synthesis, or coherent integration of those essential components of organizational management into a holistic management system. De-

pending on its system type, on the one hand, and predominant trends in its super-system, on the other hand, a general prognosis for the management system is made for the foreseeable future, principles (rules) of interaction with external environment are formulated, strategic vision perspectives are defined.

The objects with the system management have a real opportunity to delegate management activities to external, highly professional companies that focus on remote services. Remote system management supports long-term consistency, balance and integrity of the managed system, allowing it to switch from total monitoring and control of the area to discrete management that implies sampling analysis of key parameters and managerial actions in certain control points. In the interim the distantly controlled system is left to its own devices that allows it to fully deploy its inner reserves of self-education, self-improvement and self-development.

CONCLUSION

The main reason of numerous unsolvable problems that accompany the development of the modern civilization is the non-consistency of organizational management. To overcome it, we need to systematize methodology, instruments, technologies and processes that are used in making, taking and implementing managerial decisions. Our humble attempts to draw a picture of the future systematization, made in this article, convince us that it is a very hard and bold scientific task that requires many intellectual, material and financial resources. Still, the studies on the systematization of organizational management are to be started immediately, so that we could find faster the ways to solve unsolvable tasks.

REFERENCES

1. I. Adizes, *The Ideal Executive: Why You Cannot Be One and What to Do About It*. Moscow, 2007, 263 p. (in Russian).
2. A.V. Buhvalov and V.S. Kat'kalo, "From the Origins of Researches in Management to New Paradigms: Analysis of Innovative Companies", *Russian Journal of Management*, no. 6 (4), pp. 49–60, 2012 (in Russian).
3. V.A. Dresvyanikov, "The Concept of Intellectual Management", *Management in Russia and Abroad*, no. 1, pp. 100–105, 2014 (in Russian).
4. K.S. Drogobyckaya and I.N. Drogobyckij, *Architectural Models of Economic Systems*. Moscow, 2014, 301 p. (in Russian).
5. A.I. Drogobyckij, *Corporate Management in the Knowledge Economy*. Moscow, 2008, 150 p. (in Russian).
6. I.N. Drogobyckij, "Mental Models in Economy", *Economics Science in Modern Russia*, no. 1 (84), pp. 7–14, 2019.
7. I.N. Drogobyckij, *System Cybernation in Organizational Management*. Moscow, 2016, 333 p. (in Russian).
8. I.N. Drogobyckij, *System Analysis in Economics*. Moscow, 2017, 423 p. (in Russian).
9. V.M. Zotov, *Intellectual Property Managing System in Hi-Tech Holding Groups*. Moscow, 2018, 159 p. (in Russian).
10. D. Kahneman, *Thinking, Fast and Slow*. Moscow, 2016, 653 p. (in Russian).
11. G.B. Kleiner, "New Theory of Economic Systems and Its Practical Application", *Journal of Russian Academy of Science*, no. 9, pp. 89–100, 2011 (in Russian).

12. G.B. Kleiner and M.A. Rybachuk, *Systematically Balanced Economy*. Moscow, 2017, 320 p. (in Russian).
13. G.B. Kleiner, "System Paradigm and System Management", *Russian Journal of Management*, no. 6 (3), pp. 27–50, 2008 (in Russian).
14. I.I. Mazur, V.D. Shapiro, and N.G. Olderogge, *General History of Management*. Moscow, 2018, 888 p. (in Russian).
15. H. Mintzberg, *Managing*. Moscow, 2011, 288 p. (in Russian).
16. R. Thaler, *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Moscow, 2018, 240 p. (in Russian).
17. M. Buckingham, "What Great Managers Do", *Harvard business review*, no. 83(3), pp. 70–79, 2005 (in English).
18. P. Goleman, "Leadership that Gets Results", *Harvard business review*, March–April, pp. 78–90, 2000 (in English).
19. T. Watson, "How Do Managers Think", *Management learning*, no. 27 (3), pp.78–90, 1996 (in English).

Received 19.10.2020

INFORMATION ON THE ARTICLE

Ivan N. Drogobyskij, ORCID: 0000-0002-9982-0361, Financial University, Moscow, e-mail: dinind@mail.ru

ОСНОВЫ СИСТЕМИЗАЦІЇ ОРГАНІЗАЦІЙНОГО МЕНЕДЖМЕНТУ / І.М. Дрогобицький

Анотація. Організаційний менеджмент є найслабшою ланкою в розвитку сучасної цивілізації, що породжує низку проблем на всіх рівнях її ієрархії. Для їх вирішення необхідно впорядкувати, удосконалити і розширити наявний інструментально-методичний арсенал організаційного менеджменту, надати йому системного характеру, забезпечивши результативність та ефективність його практичного використання. Зроблено спробу окреслити перелік і зміст очікуваних робіт. Інтегруючи їх з метою системізації, тобто процесу надання/відновлення системності організаційного менеджменту, розглянуто вихідні положення цього процесу – визначення, фундаментальні основи, різноманітність стилів та системна типологія – які покликані зорієнтувати зацікавленого читача предметною галуззю нового наукового напрямку.

Ключові слова: система, менеджмент, управління, наука, мистецтво, ремесло, покликання, стиль управління.

ОСНОВЫ СИСТЕМИЗАЦИИ ОРГАНИЗАЦИОННОГО МЕНЕДЖМЕНТА / И.Н. Дрогобыцкий

Аннотация. Организационный менеджмент является самым слабым звеном в развитии современной цивилизации, что порождает множество проблем на всех уровнях её иерархии. Для их решения необходимо упорядочить, совершенствовать и расширить имеющийся инструментально-методологический арсенал организационного менеджмента, придать ему системный характер, обеспечив тем самым результативность и эффективность его практического применения. Предпринята попытка очертить состав и содержание предстоящих работ. Интегрируя их с целью системизации т.е. процесса придания/восстановления системности организационному менеджменту, рассмотрены его исходные предпосылки – определение, фундаментальные основы, стилевое разнообразие и системная типология – которые призваны сориентировать заинтересованного читателя в предметной области нового направления.

Ключевые слова: система, менеджмент, управление, наука, искусство, ремесло, призвание, стиль управления, системная типология управления.

REFERENCES

1. I. Adizes, *The Ideal Executive: Why You Cannot Be One and What to Do About It*. Moscow, 2007, 263 p. (in Russian).
2. A.V. Buhvalov and V.S. Kat'kalo, "From the Origins of Researches in Management to New Paradigms: Analysis of Innovative Companies", *Russian Journal of Management*, no. 6 (4), pp. 49–60, 2012 (in Russian).
3. V.A. Dresvyanikov, "The Concept of Intellectual Management", *Management in Russia and Abroad*, no. 1, pp. 100–105, 2014 (in Russian).
4. K.S. Drogobyckaya and I.N. Drogobyckij, *Architectural Models of Economic Systems*. Moscow, 2014, 301 p. (in Russian).
5. A.I. Drogobyckij, *Corporate Management in the Knowledge Economy*. Moscow, 2008, 150 p. (in Russian).
6. I.N. Drogobyckij, "Menthall Models in Economy", *Economics Science in Modern Russia*, no. 1 (84), pp. 7–14, 2019.
7. I.N. Drogobyckij, *System Cybernation in Organizational Management*. Moscow, 2016, 333 p. (in Russian).
8. I.N. Drogobyckij, *System Analysis in Economics*. Moscow, 2017, 423 p. (in Russian).
9. V.M. Zotov, *Intellectual Property Managing System in Hi-Tech Holding Groups*. Moscow, 2018, 159 p. (in Russian).
10. D. Kahneman, *Thinking, Fast and Slow*. Moscow, 2016, 653 p. (in Russian).
11. G.B. Kleiner, "New Theory of Economic Systems and Its Practical Application", *Journal of Russian Academy of Science*, no. 9, pp. 89–100, 2011 (in Russian).
12. G.B. Kleiner and M.A. Rybachuk, *Systematically Balanced Economy*. Moscow, 2017, 320 p. (in Russian).
13. G.B. Kleiner, "System Paradigm and System Management", *Russian Journal of Management*, no. 6 (3), pp. 27–50, 2008 (in Russian).
14. I.I. Mazur, V.D. Shapiro, and N.G. Olderogge, *General History of Management*. Moscow, 2018, 888 p. (in Russian).
15. H. Mintzberg, *Managing*. Moscow, 2011, 288 p. (in Russian).
16. R. Thaler, *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Moscow, 2018, 240 p. (in Russian).
17. M. Buckingham, "What Great Managers Do", *Harward business review*, no. 83(3), pp. 70–79, 2005 (in English).
18. P. Goleman, "Leadership that Gets Results", *Harward business review*, March–April, pp. 78–90, 2000 (in English).
19. T. Watson, "How Do Managers Think", *Management learning*, no. 27 (3), pp.78–90, 1996 (in English).

**MULTILAYER GMDH-NEURO-FUZZY NETWORK BASED ON
EXTENDED NEO-FUZZY NEURONS AND ITS APPLICATION
IN ONLINE FACIAL EXPRESSION RECOGNITION**

Ye. BODYANSKIY, Yu. ZAYCHENKO, G. HAMIDOV, N. KULISHOVA

Abstract. Real-time image recognition is required in many important practical problems. Interaction with users in online mode requires flexibility and adaptability from applications. The Group Method of Data Handling (GMDH) allows changing the model structure and adjusting the system architecture to the characteristics of each task under consideration. Moreover, the approximating properties of neo-fuzzy neurons used as elements of the system provide the high recognition accuracy under conditions of short data samples. This paper proposes a multilayer GMDH-neuro-fuzzy network based on extended neo-fuzzy neurons. The learning algorithm has filtering and tracking properties, guarantees the required speed important for real-time applications. The effectiveness of the proposed system is confirmed for the human emotions recognition.

Keywords: Group Method of Data Handling, extended neo-fuzzy neuron, online image recognition, facial expression recognition.

INTRODUCTION

Information technologies are actively being introduced into education, business, healthcare, entertainment and other spheres of human life. This requires technology interactivity, to conduct continuous two-way cooperation between person and computer or mobile device. One of the promising areas for such intellectual interfaces' development is the approach that uses the recognition of people, their age, sex, state of health, emotional status on the real time video. This complex technical problem already finds its own solutions [1–9]. Frequently, these decisions use the machine learning and neuro-fuzzy approach.

As a technical problem, the task of a user emotional status recognition by video is reduced to characteristic features detecting, and to the collected data classification. This problem is related to the fact that machine learning algorithms in this task require the training data sets in which the samples number can be tens or even hundreds of thousands. The forming of such sets is a serious, time-consuming task, significantly increasing the projects developing cost and implementation duration.

The prospective methods of recognition by short datasets are fuzzy systems and GMDH. Earlier it was proved that neural networks are universal approxima-

tors and have some remarkable properties, such as parallel information processing, ability to work with incomplete noisy input data and learning possibilities to achieve the desired output.

The GMDH, from the other side, uses the principle of self-organization that allows constructing an optimal structure of the mathematical model during the algorithm operation. It's very promising to combine advantages of these both approaches for the solution of the problem — development an efficient model structure. GMDH-neural networks whose nodes are active neurons [10–12], N-adalines [13], R-neurons [14–16], Q-neurons [17] are known. At the junction of the fuzzy GMDH [18] and neural networks, the GMDH-neuro-fuzzy system [15, 19] and the GMDH-neo-fuzzy system [20] were created. These systems have proven their effectiveness in solving a wide range of problems, but have lost the main advantages of the original GMDH: a small number of evaluated parameters in each node. In this regard, it seems promising to develop a GMDH-neo-fuzzy system that combines the advantages of traditional GMDH and hybrid computational intelligence systems, and is trained using simple procedures to ensure high speed of online image recognition.

The goal of the present paper is a synthesis of the GMDH neo-fuzzy system for the online image recognition.

THE EXTENDED NEO-FUZZY NEURON AS A NODE OF GMDH-NEURO-FUZZY SYSTEM

Takeshi Yamakawa and co-authors in [21–23] proposed the architecture of neo-fuzzy neuron (NFN). The authors of the NFN admit among its most important advantages, the high learning rate, computational simplicity, the possibility to find the learning criterion global minima in real-time processing. Besides, NFN is characterized by fuzzy linguistic “if-then” rules. The neo-fuzzy neuron is a nonlinear multi-input single-output system shown in Fig. 1.

It realizes the following mapping

$$f_i(x_i) = \sum_{j=1}^h w_{ji} \mu_{ji}(x_i)$$

and implements fuzzy inference

IF x_i IS x_{ji} THEN THE OUTPUT IS w_{ji} ,

where x_{ji} is a fuzzy set with membership function $\mu_{ji}(x_i)$, w_{ji} is a singleton synaptic weight in consequent [2]. As it can be seen nonlinear synapse in fact actualize Takagi–Sugeno fuzzy inference of zero order. The membership functions $\mu_{ji}(x_i)$ in the antecedent could be B-splines or triangular functions, for example, like this

$$\mu_{ji} = \begin{cases} \frac{x_i - c_{j-1,i}}{c_{j,i} - c_{j-1,i}}, & \text{if } x_i \in [c_{j-1,i}, c_{j,i}]; \\ \frac{c_{j+1,i} - x_i}{c_{j+1,i} - c_{j,i}}, & \text{if } x_i \in [c_{j,i}, c_{j+1,i}]; \\ 0, & \text{otherwise,} \end{cases}$$

where c_{ji} — the centers of membership functions, usually distributed on interval $[0, 1]$. This contributes to simplify the fuzzy inference. An input signal x_i activates only two neighboring membership functions simultaneously and the sum of the grades equals to unity, providing Ruspini partition:

$$\mu_{j-1,i}(x_i) + \mu_{ji}(x_i) = \mu_{ji}(x_i) + \mu_{j+1,i}(x_i) = 1.$$

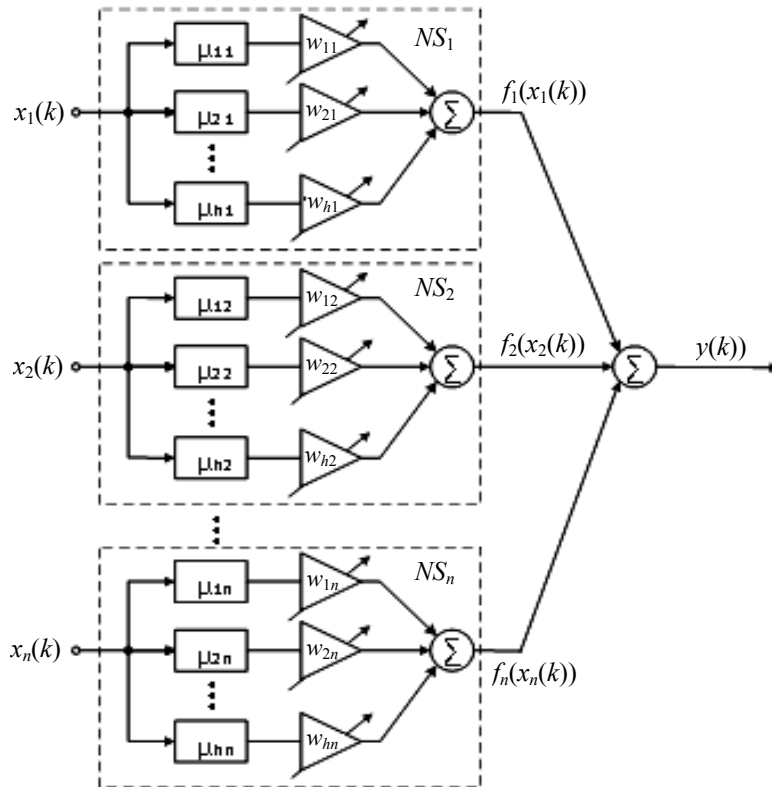


Fig. 1. Neo-fuzzy neuron

The inference result can be produced by arbitrary defuzzification method. Center-of-Gravity method gives output in the simple form:

$$f_i(x_i) = w_{ji}\mu_{ji}(x_i) + w_{j+1,i}\mu_{j+1,i}(x_i).$$

It is possible to improve approximating properties of such a system by using a structural unit, called by authors as “extended nonlinear synapse” (ENS_i) (Fig. 2) and synthesized on its basis the “extended neo-fuzzy neuron” [24–27] (ENFN). ENFN contains ENS_i as elements instead of usual nonlinear synapses NS_i.

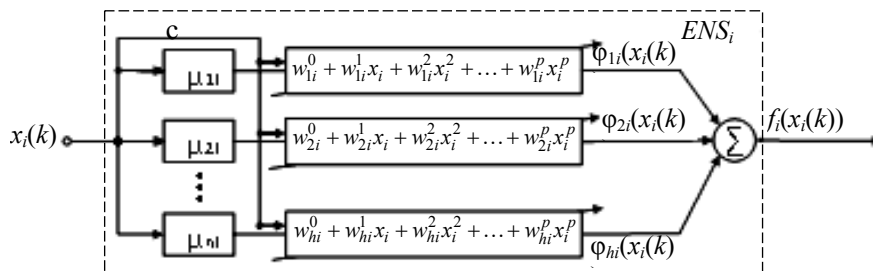


Fig. 2. Extended non-linear synapse

By introducing the additional variables:

$$\begin{aligned}
 y_{li}(x_i) &= \mu_{li}(x_i)(w_{li}^0 + w_{li}^1 x_i + w_{li}^2 x_i^2 + \dots + w_{li}^p x_i^p); \\
 f_i(x_i) &= \sum_{l=1}^h \mu_{li}(x_i)(w_{li}^0 + w_{li}^1 x_i + w_{li}^2 x_i^2 + \dots + w_{li}^p x_i^p) = \\
 &= w_{1i}^0 \mu_{1i}(x_i) + w_{1i}^1 x_i \mu_{1i}(x_i) + \dots + w_{1i}^p x_i^p \mu_{1i}(x_i) + \\
 &+ w_{2i}^0 \mu_{2i}(x_i) + \dots + w_{2i}^p x_i^p \mu_{2i}(x_i) + \dots + w_{hi}^p x_i^p \mu_{hi}(x_i); \\
 w_i &= (w_{1i}^0, w_{1i}^1, \dots, w_{1i}^p, w_{2i}^0, \dots, w_{2i}^p, \dots, w_{hi}^p)^T;
 \end{aligned}$$

$$\tilde{\mu}_i(x_i) = (\mu_{1i}(x_i), x_i \mu_{1i}(x_i), \dots, x_i^p \mu_{1i}(x_i), \mu_{2i}(x_i), \dots, x_i^p \mu_{2i}(x_i), \dots, x_i^p \mu_{hi}(x_i))^T,$$

we can write

$$\begin{aligned}
 f_i(x_i) &= w_i^T \tilde{\mu}_i(x_i), \\
 \hat{y} &= \sum_{i=1}^n f_i(x_i) = \sum_{i=1}^n w_i^T \tilde{\mu}_i(x_i) = \tilde{w}^T \tilde{\mu}(x),
 \end{aligned}$$

where $\tilde{w}^T = (w_1^T, \dots, w_n^T)^T$;

$$\tilde{\mu}(x) = (\tilde{\mu}_1^T(x_1), \dots, \tilde{\mu}_i^T(x_i), \dots, \tilde{\mu}_n^T(x_n))^T.$$

It's easy to see that ENFN contains $(p+1)hn$ adjusting synaptic weights and fuzzy output, implemented by each ENS_i, has the form:

IF x_i IS x_{li} THEN THE OUTPUT IS $w_{li}^0 + w_{li}^1 x_i + \dots + w_{li}^p x_i^p, l = 1, 2, \dots, h$,

i.e. essentially coincides with p -order Takagi–Sugeno inference.

Fig. 3 shows the architecture of an extended neo-fuzzy neuron.

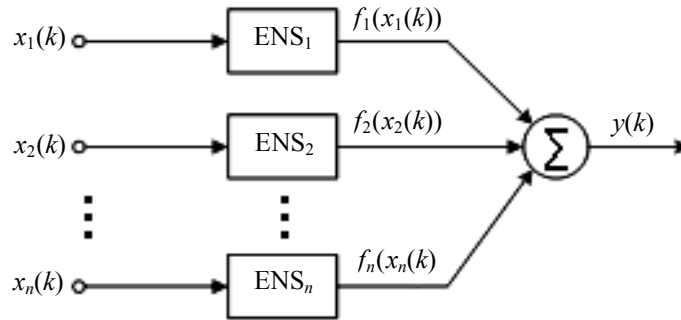


Fig. 3. Extended neo-fuzzy neuron

THE NEO-FUZZY NEURON LEARNING ALGORITHM

As a goal function for NFN learning the local quadratic error function is used:

$$E(k) = \frac{1}{2} (y(k) - \hat{y}(k))^2 = \frac{1}{2} e^2(k) = \frac{1}{2} \left(y(k) - \sum_{i=1}^n \sum_{j=1}^h w_{ji} \mu_{ji}(x_i(k)) \right)^2,$$

where $y(k)$ — external reference signal.

It is minimized in the gradient stepwise algorithm in the form:

$$w_{ji}(k) = w_{ji}(k-1) + \eta e(k) \mu_{ji}(x_i(k)) = \\ = w_{ji}(k-1) + \eta \left(y(k) - \sum_{i=1}^n \sum_{j=1}^h w_{ji}(k-1) \mu_{ji}(x_i(k)) \right) \mu_{ji}(x_i(k)),$$

$e(k)$ — learning error; η — the scalar learning rate parameter.

To increase the speed of training process Kaczmarz–Widrow–Hoff one-step learning algorithm [28–32] can be used:

$$w(k) = w(k-1) + \frac{y(k) - w^T(k-1)\mu(x(k))}{\|\mu(x(k))\|^2} \mu(x(k)),$$

where $\mu(x(k)) = (\mu_{11}(x_1(k)), \dots, \mu_{h1}(x_1(k)), \dots, \mu_{h2}(x_2(k)), \dots,$

$$\dots, \mu_{ji}(x_i(k)), \dots, \mu_{hn}(x_n(k)))^T,$$

$$w(k-1) = (w_{11}(k-1), \dots, w_{h1}(k-1), \dots,$$

$$\dots, w_{h2}(k-1), \dots, w_{ji}(k-1), \dots, w_{hn}(k-1))^T \text{ —}$$

$(nh \times 1)$ — vectors generated by input variables.

The learning algorithm exponentially weighted form:

$$\begin{cases} w(k) = w(k-1) + r^{-1}(k)(y(k) - w^T(k-1)\mu(x(k)))\mu(x(k)); \\ r(k) = \alpha r(k-1) + \|\mu(x(k))\|^2, \quad 0 \leq \alpha \leq 1, \end{cases}$$

which has filtering and tracking properties can be effectively used in stochastic and nonstationary situation.

THE NEURO-FUZZY SYSTEM AND ITS ARCHITECTURE OPTIMIZATION USING THE GROUP METHOD OF DATA HANDLING

The neuro-fuzzy system under consideration is a multilayer feedforward architecture that consists of extended neo-fuzzy neurons and shown on Fig. 4.

A $(n \times 1)$ -dimensional input signals vector $x = (x_1, x_2, \dots, x_n)^T$ arrives at system zero receptive layer, and is transmitted then to first hidden layer containing $n_1 = C_n^2$ neuron nodes with only two inputs. At the outputs of first hidden layer nodes $N^{[1]}$, output signals $\hat{y}_l^{[1]}, l = 1, 2, \dots, \frac{n}{2}(n-1) = C_n^2$ are formed. Further, these signals are sent to selection block $SB^{[1]}$ of first hidden layer, which selects from output signals set $\hat{y}_l^{[1]}$ the most accurate n_1^* ($n_1^* < n_1$) of them in the accepted criterion sense, most often the mean square error $\sigma_{y_l^{[1]}}^2$. From these n_1^* best outputs of first hidden layer $\hat{y}_l^{[1]*}, n_2$ ($n \leq n_2 \leq 2n$ usually) pairwise combi-

nations are formed, which are fed to second hidden layer formed by neurons $N^{[2]}$ similar to neurons $N^{[1]}$. From output signals of this layer $\hat{y}_l^{[2]}$, the selection block of the second hidden layer $SB^{[2]}$ selects only those that exceed the best signal of the first hidden layer $\hat{y}_l^{[1]*}$, for example, in terms of accuracy $\sigma_{y_l^{[2]}}^2$.

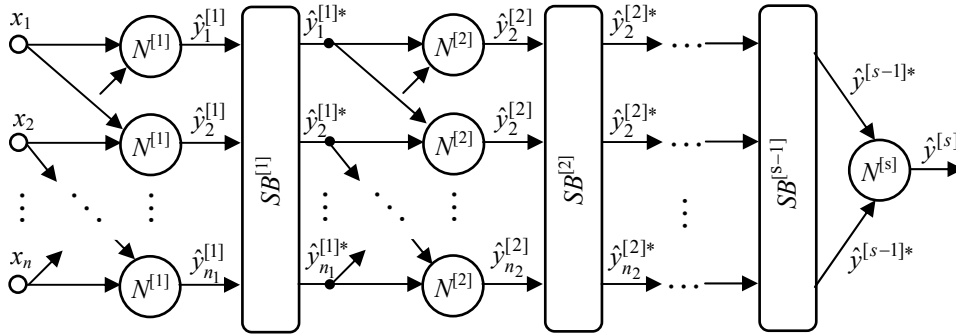


Fig. 4. GMDH-neuro-fuzzy system based on extended neo-fuzzy neurons

The third hidden layer generates signals that are superior in accuracy to the best signal $\hat{y}_l^{[2]*}$ etc. The process of system forming occurs until the selection block $SB^{[s-1]}$ generates at its output only two signals $\hat{y}_1^{[s-1]*}$ and $\hat{y}_2^{[s-1]*}$. These signals that are applied to a single output neuron $N^{[s]}$ are used to calculate the single system output signal $\hat{y}^{[s]}$.

As the nodes of the GMDH system, one can use various neurons types with necessary approximating capabilities. However, at the same time, the main advantage of the original GMDH method — the ability to real-time work in the presence of short training sets — may be lost.

This paper proposes to use the extended multidimensional neo-fuzzy neurons [24–27] as GMDH system nodes. Its architecture is a Takagi–Sugeno–Kang neuro-fuzzy system [33–35] with two inputs x_1 and x_2 , five sequentially connected information processing layers and one output \hat{y}_l . A two-dimensional vector of input signals $x(k) = (x_1(k), x_2(k))^T$ to be processed is fed to the input of the node ($k=1, 2, \dots, N$ — the observation number in the training set or the current discrete time). The node first layer contains $2h$ membership functions $\mu_{p_i}(x_i), \mu_{p_j}(x_j)$, $p=1, 2, \dots, h$ and implements fuzzification of input variables. The second layer provides aggregation of membership levels calculated in the first layer, contains h multiplication blocks and forms two-dimensional radial basis activation functions $\mu_{p_i}(x_i), \mu_{p_j}(x_j)$. The third layer is a layer of synaptic weights to be adjusted during the training process, while layer outputs are values $w_{ip}^{ij} \mu_{p_i}(x_i), \mu_{p_j}(x_j)$, and weights number is determined by number of membership functions at each input h . The fourth layer is formed by two adders and calculates the sum of second and third layers output signals. Finally, in the fifth neuron output layer normalization is performed, as a result of which the node output signal \hat{y}_l is calculated.

Thus, when a signal $x(k)$ is fed to the neuron node input, elements of the first layer calculate membership levels $0 < \mu_{pi}(x_i) \leq 1, 0 < \mu_{pj}(x_j) \leq 1$. As membership functions, bell-shaped constructions with a not strictly local receptor field are usually used. They avoid the appearance of “holes” in the fuzzified space when using the scattered partition spaces of input variables [35]. The membership functions of the first layer are Gaussian

$$\mu_{pi}(x_i(k)) = \exp\left(-\frac{(x_i(k) - c_{pi})^2}{2\sigma_i^2}\right), \mu_{pj}(x_j) = \exp\left(-\frac{(x_j(k) - c_{pj})^2}{2\sigma_j^2}\right),$$

where c_{pi}, c_{pj} — parameters defining the centres of membership functions; σ_i, σ_j — width parameters of these functions.

The h aggregated signals appear on second layer outputs

$$\tilde{x}_p(k) = \mu_{pi}(x_i(k))\mu_{pj}(x_j(k)),$$

where

$$\begin{aligned} \tilde{x}_p(k) &= \exp\left(-\frac{(x_i(k) - c_{pi})^2}{2\sigma_i^2}\right) \exp\left(-\frac{(x_j(k) - c_{pj})^2}{2\sigma_j^2}\right) = \\ &= \exp\left(-\frac{\|x(k) - c_p\|^2}{2\sigma^2}\right), c_p = (c_{pi}, c_{pj})^T. \end{aligned}$$

The third layer outputs are the values

$$w_{ip}^{ij} \mu_{pi}(x_i(k))\mu_{pj}(x_j(k)) = w_{ip}^{ij} \tilde{x}_p(k),$$

the fourth layer calculates the following two sums:

$$\sum_{p=1}^h w_{ip}^{ij} \mu_{pi}(x_i(k))\mu_{pj}(x_j(k)) = \sum_{p=1}^h w_{ip}^{ij} \tilde{x}_p(k);$$

$$\sum_{p=1}^h \mu_{pi}(x_i(k))\mu_{pj}(x_j(k)) = \sum_{p=1}^h \tilde{x}_p(k)$$

and finally, at node output (fifth layer), a signal is formed

$$\begin{aligned} \hat{y}_l(k) &= \frac{\sum_{p=1}^h w_{ip}^{ij} \mu_{pi}(x_i(k))\mu_{pj}(x_j(k))}{\sum_{p=1}^h \mu_{pi}(x_i(k))\mu_{pj}(x_j(k))} = \frac{\sum_{p=1}^h w_{ip}^{ij} \tilde{x}_p(k)}{\sum_{p=1}^h \tilde{x}_p(k)} = \\ &= \sum_{p=1}^h w_{ip}^{ij} \frac{\tilde{x}_p(k)}{\sum_{p=1}^h \tilde{x}_p(k)} = \sum_{p=1}^h w_{ip}^{ij} \varphi_p^{ij}(x(k)) = (w_l^{ij})^T \varphi^{ij}(x(k)), \end{aligned}$$

where

$$\varphi_p^{ij}(x(k)) = \mu_{pi}(x_i(k))\mu_{pj}(x_j(k)) \left(\sum_{p=1}^h \mu_{pi}(x_i(k))\mu_{pj}(x_j(k)) \right)^{-1};$$

$$w_l^{ij} = (w_{l1}^{ij}, w_{l2}^{ij}, \dots, w_{lh}^{ij})^T;$$

$$\varphi^j(x(k)) = (\varphi_1^j(x(k)), \varphi_2^j(x(k)), \dots, \varphi_p^j(x(k)))^T.$$

It can be noticed that the node implements a non-linear mapping of input signals to output.

THE EXPERIMENTS

The task of a person's facial expression recognition is complex and multi-stage. It includes pre-processing of the image and searching for the face area within the image. After the face area is distinguished, it is possible to recognize the emotion by the face features set. In the practice of faces expressions recognition, several descriptor principles are used. The most common are adaptive appearance models, which use the descriptions based on face image feature points and contours. It is established that such descriptions convey complete information about the person emotional state, even if it is expressed weakly.

Under the emotions influence, the facial muscles reduction leads to the displacement of feature points and this movement can serve as an indicator of basic facial actions. The most commonly used facial expressions are some basic emotions (fear, sadness, happiness, anger, disgust, surprise) and neutral state.

As a base for the feature vector, it is proposed to use a set of 35 characteristic points that can be localized in the facial area using contour detectors (Fig. 5). The neo-fuzzy neurons has one output as the dimensionality of the output data vector. Seven basic emotions were selected for recognition: anger, disgust, fear, surprise, happiness, sadness, neutral expression. Therefore, the output values are $\{1; 2; 3; 4; 5; 6; 7\}$. The character features vector contains the two-dimensional coordinates of feature points position (fig. 5). So, the system input is vector $\{x_i\}_{1 \times 70}$. The order of the polynomial in nonlinear synapse membership function was chosen equal to 4, the number of synapses in the neo-fuzzy neuron was equal to 5.

The proposed architecture ability to recognize individual emotions was investigated using photographs from two open bases — Psychological Image Collection at Stirling (PICS) [36], partly from the Extended Cohn-Kanade (CK+) database [37]. Some images are in public use as objects for recognition.

In this set of photographs, those were selected differ in the person emotional state expression degree — from weakly noticeable to very noticeable.



Fig. 5. Examples of training images and position of characteristic points

In this task, special attention was paid to a learning data set small size. To examine how the proposed architecture and learning algorithm will recognize facial expressions, small photo sets are used. Their dimensions are given in Table 1.

Table 1. Dimensions of training sets of photos for individual emotions

Emotion	Anger	Disgust	Fear	Happiness	Sorrow	Surprise	Neutral
Data set size	49	66	35	45	19	50	80

Then the architecture ability to learn from a mixed set was examined, and sets total size was 344 photos. The number of unrecognized emotions is given in Table 2.

Table 2. The number of unrecognized emotions as a result of GMDH-neuro-fuzzy system based on extended neo-fuzzy neurons learning from a mixed set

Recognition accuracy	Primary emotions						
	Anger	Disgust	Fear	Happiness	Sorrow	Surprise	Neutral
The percentage of unrecognized images, %	2,04	0	5,71	4,44	0	0	2,5

A GMDH-neuro-fuzzy system based on extended neo-fuzzy neurons configured a task model from two rows. Learning error change is shown in Fig. 6.

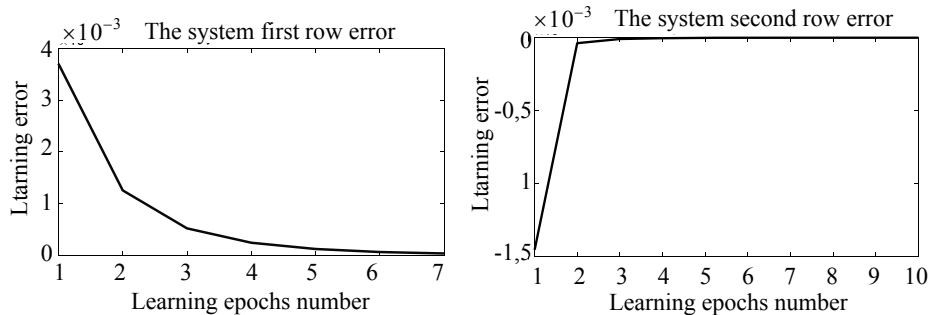


Fig. 6. A GMDH-neuro-fuzzy system based on extended neo-fuzzy neurons learning error

CONCLUSIONS

The article proposes the GMDH system with extended neo-fuzzy neurons as nodes. The system architecture allows modifying the process model structure in real time due to proposed neo-fuzzy nodes-neurons synaptic weights adjusting. A feature of the proposed architecture and its learning algorithm is the ability to work with small training samples.

REFERENCES

1. A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M.R. Wrobel, "Human-Computer Systems Interaction: Backgrounds and Applications", ch. 3, *Emotion Recognition and Its Applications*. Cham: Springer International Publishing, 2014, pp. 1–62.

2. Kaggle. *Challenges in representation learning: Facial recognition challenge*, 2013.
3. G.U. Kharat and S.V. Dudul, "Emotion Recognition from Facial Expression Using Neural Networks", in *Human-Computer Systems Interaction. Advances in Intelligent and Soft Computing*, vol. 60, Z.S. Hippe, J.L. Kulikowski, Eds. Berlin, Heidelberg: Springer, 2009.
4. C. Shan, S. Gong, and P.W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.
5. Ch.-Yi. Lee and Li-Ch. Liao, "Recognition of Facial Expression by Using Neural-Network System with Fuzzified Characteristic Distances Weights", *IEEE Int. Conf. Fuzzy Systems FUZZ-IEEE 2008* [IEEE World Congress on Computational Intelligence], pp. 1694–1699, 2008.
6. N. Kulishova, "Emotion Recognition Using Sigma-Pi Neural Network", *Proc. of 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP)*, Lviv, 2016, pp. 327–331.
7. A. Graves, J. Schmidhuber, C. Mayer, M. Wimmer, and B. Radig, "Facial Expression Recognition with Recurrent Neural Networks", *International Workshop on Cognition for Technical Systems*, Munich, Germany, October 2008.
8. S. Ouelett, "Real-time emotion recognition for gaming using deep convolutional network features", *CoRR*, vol. abs./1408.3750, 2014.
9. B. Kim, J. Roh, S. Dong, and S. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition", *Journal on Multimodal User Interfaces*, pp. 1–17, 2016.
10. A.G. Ivakhnenko, G.A. Ivakhnenko, and J.A. Mueller, "Self-organization of the neural networks with active neurons", *Pattern Recognition and Image Analysis*, vol. 4, no. 2, pp. 177–18, 1994.
11. G.A. Ivakhnenko, "Self-organization of neuronet with active neurons for effect of nuclear test explosion forecasting", *System Analysis Modeling Simulation*, vol. 20, pp. 107–116, 1995.
12. A.G. Ivakhnenko, D. Wuensch, and G.A. Ivakhnenko, "Inductive sorting-out GMDH algorithms with polynomial complexity for active neurons of neural networks", *Neural Networks*, vol. 2, pp. 1169–1173, 1999.
13. D.U. Pham and X. Liu, *Neural networks for Identification, Prediction and Control*. London: Springer-Verlag, 1995, 238 p.
14. E. Lughofer, *Evolving Fuzzy Systems – Methodologies, Advanced Concepts and Applications*, Berlin-Heidelberg, Springer-Verlag, 2011.
15. T. Ohtani, "Automatic variable selection in RBF network and its application to neurofuzzy GMDH", *Proc. Fourth Int. Conf. on Knowledge-Based Intelligent Engineering Systems and Allied Technologies*, vol. 2, pp. 840–843, 2000.
16. Ye. Bodyanskiy, N. Teslenko, and P. Grimm, "Hybrid evolving neural network using kernel activation functions", *Proc. 17th Zittau East West Fuzzy Colloquium*, Zittau/Goerlitz: HS, 2010, pp. 39–46.
17. Ye. Bodyanskiy, O. Vynokurova, and I. Pliss, "Hybrid GMDH-neural network of computational intelligence", *Proc. 3rd Int. Workshop on Inductive Modeling*, Krynica, Poland, 2009, pp. 100–107.
18. Yu. Zaychenko, "The fuzzy Group Method of Data Handling and its application for economical processes forecasting", *Scientific Inquiry*, vol. 7, no.1, pp. 83–96, 2006.
19. T. Ohtani, H. Ichihashi, T. Miyoshi, K. Nagasaka, and Y. Kanaumi, "Structural learning of neurofuzzy GMDH with Minkowski norm", *Proc. 1998 Second Int. Conf. on Knowledge-Based Intelligent Electronic Systems*, vol. 2, pp. 100–107, 1998.
20. Ye. Bodyanskiy, Yu. Zaychenko, E. Pavlikovskaya, M. Samarina, and Ye. Viktorov, "The neo-fuzzy neural network structure optimization using GMDH fog the solving

- forecasting and classification problems”, *Proc. Int. Workshop on Inductive Modeling*, Krynica, Poland, 2009, pp. 77–89.
21. J. Miki and T. Yamakawa, “Analog implementation of neo-fuzzy neuron and its on-board learning”, in *Computational Intelligence and Applications*, Ed. N.E. Mastorakis, Piraeus: WSES Press, 1999, pp. 144–149.
 22. T. Yamakawa, E. Uchino, J. Miki and H. Kusanagi, “A neo-fuzzy neuron and its application to system identification and prediction of the system behavior”, *Proc. 2-nd Int. Conf. on Fuzzy Logic and Neural Networks “IIZUKA-92”*, Iizuka, Japan, 1992, pp. 477–483.
 23. E. Uchino and T. Yamakawa, “Soft computing based signal prediction, restoration and filtering”, in *Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks and Genetic Algorithms*, Ed. Da Ruan, Boston: Kluwer Academic Publishers, 1997, pp. 331–349.
 24. Ye.V. Bodyanskiy and N.Ye. Kulishova, “Extended neo-fuzzy neuron in the task of images filtering”, *Radioelectronics. Computer Science. Control*, no. 1(32), pp. 112–119, 2014.
 25. Z. Hu, Ye. Bodyanskiy, N. Kulishova, O.A. Tyshchenko, “Multidimensional Extended Neo-Fuzzy Neuron for Facial Expression Recognition”, *International Journal of Intelligent Systems and Applications (IJISA)*, vol. 9, no. 9, pp. 29–36, 2017.
 26. Ye. Bodyanskiy, N. Kulishova and D. Malysheva, “The Extended Neo-Fuzzy System of Computational Intelligence and its Fast Learning for Emotions Online Recognition”, *Proc. of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, Lviv, Ukraine, August 21–25, 2018, pp.473–478.
 27. Ye. Bodyanskiy, N. Kulishova and O. Chala, “The Extended Multidimensional Neo-Fuzzy System and its Fast Learning in Pattern Recognition Tasks”, *Data*, no. 3, pp. 63–73, 2018.
 28. S. Kaczmarz, “Angenaeherte Ausloesung von Systemen Linearer Gleichungen”, *Bull. Int. Acad. Polon. Sci, Let. A*, pp. 355–357, 1937.
 29. S. Kaczmarz, “Approximate solution of systems of linear equations, *Int. J. Control*, vol. 53, pp. 1269–1271, 1993.
 30. Ye. Bodyanskiy, I. Kokshenev and V. Kolodyazhniy, “An adaptive learning algorithm for a neo-fuzzy neuron”, *Proc. of the 3rd Conference of the European Society for Fuzzy Logic and Technology*, pp. 375–379, 2005.
 31. V. Kolodyazhniy, Ye. Bodyanskiy and P. Otto, “Universal approximator employing neo-fuzzy neurons”, *Computational Intelligence: Theory and Applications*, Ed. By B. Reusch, Berlin-Heidelberg: Springer, 2005, pp. 631–640.
 32. B. Widrow and Jr.M.E. Hoff, “Adaptive switching circuits”, *1960 URE WESCON Convention Record*, part 4. N.-Y.: IRE, 1960, pp. 96–104.
 33. L.-X. Wang and J.M. Mendel, “Fuzzy basis functions, universal approximation and orthogonal least-squares learning”, *IEEE Trans. on Neural Networks*, vol. 3, no. 5, pp. 807–814, 1992.
 34. L.-X. Wang, *Adaptive Fuzzy Systems and Control. Design and Statistical Analysis*. Upper Saddle River: Prentice Hall, 1994, 256 p.
 35. R. J.-S. Jang, C.-T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Upper Saddle River: Prentice Hall, 1997, 640 p.
 36. *2D face sets*. Available: http://pics.psych.stir.ac.uk/2D_face_sets.htm
 37. P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression”, *Proceedings of IEEE workshop on CVPR for Human Communicative Behavior Analysis*, San Francisco, USA, 2010.

Received 19.10.2020

INFORMATION ON THE ARTICLE

Yevgeniy V. Bodyanskiy, ORCID: 0000-0001-5418-2143, Kharkiv National University of Radioelectronics, Ukraine, e-mail: yevgeniy.bodyanskiy@nure.ua

Yuriy P. Zaychenko, ORCID: 0000-0001-9662-3269, Educational and Scientific Complex “Institute for Applied System Analysis” of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: zaychenkoyuri@ukr.net

Galib Hamidov, head of information technologies department of “Azershig”, Azerbaijan, e-mail: galib.hamidov@gmail.com

Nonna Ye. Kulishova, ORCID: 0000-0001-7921-3110, Kharkiv National University of Radioelectronics, Ukraine, e-mail: nokuliaux@gmail.com

БАГАТОШАРОВА МГУА-НЕЙРО-ФАЗЗИ МЕРЕЖА НА ОСНОВІ РОЗШИРЕНИХ НЕЧІТКИХ НЕЙРОНІВ ТА ЇЇ ВИКОРИСТАННЯ ДЛЯ ОНЛАЙН РОЗПІЗНАВАННЯ ВИРАЗІВ ОБЛИЧЧЯ / С.В. Бодяньський, Ю.П. Зайченко, Г. Гамідов, Н. С. Кулішова

Анотація. Розпізнавання зображень в реальному часі потрібно в багатьох практичних задачах. Взаємодія з користувачами в режимі онлайн потребує гнучкості та адаптації від прикладних програм. Метод групового урахування аргументів (МГУА) дозволяє змінювати структуру моделі та налаштовує її архітектуру відповідно до характеристик кожної задачі. Більш того, апроксимаційні властивості нео-фазі нейронів як структурних елементів системи забезпечують високу точність розпізнавання в умовах коротких вибірок даних. Запропоновано багат шарову МГУА-нейро-фаззі мережу на основі розширених нео-фазі нейронів. Алгоритм навчання має фільтрувальні та відслідковувальні властивості та гарантує необхідну швидкість для застосувань реального часу. Ефективність запропонованої системи підтверджено в задачі розпізнавання людських емоцій.

Ключові слова: метод групового урахування аргументів, розширений нео-фазі нейрон, онлайн розпізнавання зображень, розпізнавання виразів обличчя.

МНОГОСЛОЙНАЯ МГУА-НЕЙРО-ФАЗЗИ СЕТЬ НА ОСНОВЕ РАСШИРЕННЫХ НЕЧЕТКИХ НЕЙРОНОВ И ЕЕ ПРИМЕНЕНИЕ ДЛЯ ОНЛАЙН РАСПОЗНАВАНИЯ ВЫРАЖЕНИЙ ЛИЦА / Е.В. Бодянский, Ю.П. Зайченко, Г. Гамидов, Н.Е. Кулишова

Аннотация. Распознавание изображений в реальном времени требуется в многих практических задачах. Взаимодействие с пользователями в режиме онлайн требует гибкости и адаптации от прикладных программ. Метод группового учета аргументов (МГУА) позволяет изменять структуру модели и настраивает ее архитектуру в соответствии с характеристиками каждой задачи. Более того, аппроксимационные свойства нео-фазис нейронов как структурных элементов системы обеспечивает высокую точность распознавания в условиях коротких выборок данных. Предложена многослойная МГУА-нейро-фазис сеть на основе расширенных нео-фазис нейронов. Алгоритм обучения имеет фильтрующие и отслеживающие свойства и гарантирует необходимую скорость для приложений реального времени. Эффективность предлагаемой системы подтверждена в задаче распознавания человеческих эмоций.

Ключевые слова: метод группового учета аргументов, расширенный нео-фазис нейрон, онлайн распознавание изображений, распознавание выражений лица.

МЕТОДИ МАШИННОГО НАВЧАННЯ В СЕНТИМЕНТ-АНАЛІЗІ ТЕКСТОВОЇ ІНФОРМАЦІЇ НА ПРИКЛАДІ НАСТРОЇВ КОРИСТУВАЧІВ СТОСОВНО КАНДИДАТІВ У ПРЕЗИДЕНТИ УКРАЇНИ 2019

А.-М.П. РУДЗЕВИЧ

Анотація. Описано основні методи машинного навчання для аналізу тональності тексту і виконано порівняльний аналіз їх ефективності. Розглянуто етапи попереднього оброблення тексту, такі як стемінг, видалення стоп-слів, алгоритми переведення тексту у векторну форму: мішок слів (Bag-of-Words), TF-IDF векторайзер та Word2Vec. Дослідження полягало у визначенні тональності тексту коментарів під публікаціями кандидатів у Президенти України (В. Зеленського та П. Порошенка) у період передвиборчих перегонів 2019 р. Для визначення тональності тексту використано три алгоритми: наївний байєсівський класифікатор, метод опорних векторів та згорткову нейронну мережу. Для кожного кандидата побудовано окремі моделі та виконано порівняння якості класифікації (за метрикою F1). Найкращою моделлю для обох вибірок даних виявилась згорткова нейронна мережа.

Ключові слова: машинне навчання, сентимент-аналіз, аналіз тональності тексту, інтелектуальний аналіз тексту.

ВСТУП

Із розвитком інформаційних технологій та стрімкого нагромадження великих масивів даних поширення набула така галузь комп'ютерної лінгвістики, як сентимент-аналіз. Стало можливим автоматично витягати з тексту виражену автором думку, а також оцінювати текст як позитивний, негативний, а за необхідності — виокремлювати конкретні емоції (радість, гнів, сум тощо). Для виокремлення емоційної оцінки автора застосовують підходи з використанням тональних словників і правил або методи машинного навчання.

Сентимент-аналіз (аналіз тональності тексту) — це розділ глибокого аналізу даних (data mining) і галузь комп'ютерної лінгвістики, що займається вилученням думок та емоцій з текстових документів.

Хоча лінгвістика та оброблення природних мов (NLP) мають давню історію, до 2000-х років майже не було досліджень, що стосуються сентимент-аналізу. Але відтоді цю галузь учені почали дуже активно вивчати [11, 12].

Термін «sentiment analysis» уперше був згаданий у праці [1], а вираз «opinion mining» (аналіз думок) — у праці [2]. Вагомий внесок у розвиток сентимент-аналізу зробено у працях [8, 9].

Усі завдання з оброблення природних мов є складними і неоднозначними. Загалом завдання визначення емоційної оцінки тексту є суб'єктивним,

оскільки різні люди по-різному оцінюють одні й ті самі події, а відповідно один і той самий текст. Текст може містити орфографічні помилки, скорочення, аббревіатури, сарказм, емоджі. Однакові слова, вжиті в різному контексті, можуть мати діаметрально протилежне емоційне навантаження. Усе це перешкоджає створенню єдиної моделі, яка правильно класифікуватиме тональність тексту незалежно від тематики.

Сентимент-аналіз набув широкого використання для маркетингових цілей, зокрема для визначення думки клієнта про певний товар або послугу, та кращого орієнтування свого повідомлення на цільову аудиторію. Також набуло популярності аналізування твітів, блогів, текстів новин, оглядів, коментарів для визначення ставлення автора до суб'єкта його висловлення. Для цього застосовують різні методики, включаючи алгоритми оброблення природних мов (NLP), статистику та методи машинного навчання.

У роботі застосуємо сентимент-аналіз для визначення настроїв користувачів стосовно кандидатів у Президенти України 2019. Аналізуватимемо коментарі користувачів у соціальній мережі інстаграм протягом усього часу передвиборчих перегонів на предмет позитивного або негативного ставлення до кандидата і зможемо оцінити як змінювалися настрої в суспільстві. Оскільки українці залишають коментарі як українською, так і російською мовами, будемо аналізувати ці дві мови.

ТЕОРЕТИЧНІ ВІДОМОСТІ

Наївний байєсівський класифікатор є ймовірнісним алгоритмом машинного навчання, заснований на теоремі Байєса, який широко використовується для задач класифікації.

Для задачі визначення тональності прогнозуємо ймовірність того, що документ d належить до класу c . Тут документ є вектором: $a = \{w_1, w_2, \dots, w_n\}$, де w_i — вага i -го терміна; n — розмір словника. Тому згідно з теоремою Байєса маємо формулу

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)}.$$

За цією формулою обчислюється умовна ймовірність для всіх класів. Якщо умовна ймовірність належності документа d до класу c максимальна, то C є найбільш імовірним класом, до якого належить документ d :

$$C = \arg \max_c P(w_1, w_2, \dots, w_n | c)P(c).$$

Знаменник може бути випущений, оскільки для одного і того ж документа d ймовірність $P(d)$ буде однаковою, а отже, її можна не враховувати.

Наївний байєсівський класифікатор спирається на припущення, що всі ознаки x_1, x_2, \dots, x_n документа d не залежать одна від одної. Припускається, що позиція слів у реченні не має значення. Тому умовну ймовірність для ознак x_1, x_2, \dots, x_n , можна подати як

$$P(w_1 | c)(w_2 | c) \times \dots \times (w_n | c) = \prod_i P(w_i | c).$$

Таким чином, для знаходження найбільш імовірного класу для документа $d = \{w_1, w_2, \dots, w_n\}$ за допомогою наївного байєсівського класифікато-

ра необхідно визначити умовні ймовірності належності документа d для кожного з поданих класів окремо і вибрати клас, який має максимальну ймовірність:

$$P(c) \prod [P(w_i | c_j)].$$

Далі оцінимо ймовірність класу $P(c)$. Вона є відношенням кількості документів класу c у навчальній вибірці до загальної кількості документів: $P(c) = \frac{D_c}{D}$, де D_c — кількість документів класу c ; D — загальна кількість документів у вибірці.

Щоб оцінити умовні ймовірності для ознак $\hat{P}(w_i | c_j)$, використовуватимемо формулу

$$\hat{P}(w_i | c) = \frac{\text{count}(w_i, c) + 1}{\sum_{w \in V} (\text{count}(w, c) + 1)},$$

де $\hat{P}(w_i | c_j)$ — відношення кількості слів w_i у класі c_j до загальної кількості слів у цьому класі; V — кількість слів у словнику навчальної вибірки [14].

Метод опорних векторів шукає гіперплощину, яка найкраще розділить вибірку на два класи. Допускається багатокласова класифікація стратегіями one-vs-all і one-vs-one.

Дано вибірку елементів $x_i \in \mathbb{R}_n$ і зіставлені їм класи $y_i \in \{-1, 1\}$. Об'єкти вибірки подано точками. Опорні вектори — це точки даних, розташовані якомога ближче до гіперплощини, у разі їх видалення зміниться положення гіперплощини. Їх вважають критичними елементами набору даних. У простій задачі бінарної класифікації з вибіркою, що лінійно розділяється, гіперплощину можна подати у вигляді лінії, що розділяє вибірку на два класи. Чим далі дані від гіперплощини, тим коректніше вони класифіковані. Кращою гіперплощиною вважається та, відстань $1/\|w\|$ від якої до кожного класу є максимальною, де w — нормальний вектор до роздільної гіперплощини, яку можна записати як множину точок x , що задовольняють рівняння $w \cdot x - b = 0$, де b — допоміжний параметр.

Якщо навчальна вибірка лінійно подільна, можна вибрати дві паралельні гіперплощини так, щоб вони розділили цю множину на два класи. Ділянка між ними називається зазором, маржею. Ці площини описуються рівняннями: $w \cdot x - b = 1$; $w \cdot x - b = -1$.

Мінімізуючи відстань $\|w\|$ і одночасно виключаючи потрапляння даних у зазор, отримуємо задачу мінімізації $\|w\|^2 \rightarrow \min$;

$$y_i(w \cdot x_i - b) \geq 1, \text{ для } 1 \leq i \leq n.$$

Таку задачу вважають еквівалентною пошуку сідлової точки функції Лангранжа; її зводять до задачі квадратичного програмування, де наявні лише двоїсті змінні λ_i .

Розв'язавши цю задачу, можна виразити w і b формулами:

$$w = \sum_{i=1}^n \lambda_i c_i x_i; \quad b = w \cdot x_i - c_i; \quad \lambda_i > 0.$$

Кінцевий класифікатор записується як

$$a(x) \operatorname{sign} \left(\sum_{i=1}^n \lambda_i c_i x_i x - b \right). \quad (1)$$

Якщо вибірка лінійно нероздільна, вектори відображають у простір більшої розмірності. При цьому у формулі (1) скалярний добуток змінюється на одну із функцій нелінійного ядра $K(x_i, x)$ і будується найкраща роздільна гіперплощина.

Згорткова нейронна мережа (ЗНМ). Архітектура згорткової нейронної мережі для класифікації тексту бере за основу звичайну ЗНМ, але дещо спрощену. На вхід подається матриця, кількість її рядків дорівнює кількості слів n у реченні (або документі), кількість стовпців — розмірності k векторного подання слів. Для отримання нової ознаки c_i виконується операція згортки. Згортка полягає в застосуванні фільтра з вагами w на вікні з h слів. Ознака c_i генерується з вікна слів $x_{i:i+h-1}$ за формулою

$$c_i = f(w x_{i:i+h-1} + b),$$

де $x_i \in \mathbb{R}$ — нейрон зміщення; f — нелінійна функція; w — вектор ваг; $x_{i:i+h-1}$ — ковзне вікно.

Фільтр буде застосований до всіх можливих вікон слів у реченні $\{x_i : h, \dots, x_{n-h+1} : n\}$ для отримання карти ознак: $c(w) = [c_1, c_2, \dots, c_{n-h+1}]$.

Потім застосовується фільтр Max Pooling (максимізаційне агрегування), тобто шукається максимум у всій послідовності. Його ідея полягає у виокремленні найважливішої ознаки з найбільшим значенням за кожною картою ознак: $\hat{c} = \max(c(w))$.

Отримані таким чином значення передаються в повнозв'язний шар із функцією активації softmax; на виході маємо розподіл імовірності за класами:

$$P(y = j | x) = \frac{e^{x^T w_j + b_j}}{\sum_{k=1}^K e^{x^T w_k + b_k}}.$$

Для запобігання перенаванчання на цьому шарі використовується метод виключення нейронів (дропаут) з імовірністю p і l_2 -регуляризація.

Зазвичай навчання мережі відбувається з використанням стохастичного градієнтного спуску. Використання дропауту вилучає з нейронної мережі деяку кількість нейронів (на етапі навчання) для запобігання коадаптації нейронів і в результаті отримання кращої узагальнювальної здатності мережі. Дропаут також прискорює процес навчання. Вихід після використання дропауту можна подати у вигляді $y = w(zr) + b$, де $z = [\hat{c}_1, \dots, \hat{c}_m]$, r — вектор, що містить 0 і 1.

Як гіперпараметри мережі виділяють розмір фільтра, імовірність дропауту p , l_2 -регуляризацію і розмір батча. l_2 -регуляризація штрафувати ваги мережі, зменшуючи їх значення, і використовується для запобігання її перенаванчання. Батч використовується для пришвидшення навчання, являючи собою «пакет» випадково обраних ознак у методі стохастичного градієнтного спуску.

МЕТРИКИ ОЦІНЮВАННЯ ЯКОСТІ АЛГОРИТМІВ

Емпіричні дані показують, що показник точності дуже залежить від збалансованості даних. У випадку, коли дані незбалансовані, доцільно перевірити, наскільки ефективно класифікатор класифікує лише частину даних — позитивні або негативні класи даних. Прикладами таких метрик є чутливість (precision) та повнота (recall).

Чутливість доцільно використовувати, коли помилково позитивна класифікація небажана. Вона розраховується за такою формулою: $Precision = TP / (TP + FP)$, де TP — правильно визначений позитивний клас; FP — хибний позитивний клас.

Метрику повноти використовують, коли треба уникнути помилково негативної класифікації. Її обчислюють за формулою $Recall = TP / (TP + FN)$, де TP — правильно визначений позитивний клас; FN — хибно визначений негативний клас.

Також є показник, який є гармонічним середнім двох попередніх оцінок — F_1 -міра: $F_1 = \frac{2 * Precision * Recall}{Precision + Recall}$. Це загальна міра точності моделі, яка поєднує в собі чутливість та повноту. Тобто показник F_1 означає малу кількість хибних позитивних та хибних негативних класифікацій.

ДОСЛІДЖЕННЯ

Дослідження полягає в аналізі емоційного навантаження тексту коментарів із соціальної мережі Інстаграм у період передвиборчих перегонів на пост Президента України в 2019.

Для проведення дослідження зібрано коментарі під публікаціями кандидатів у Президенти України 2019 р. В.О. Зеленського та П.О. Порошенка в період з початку передвиборчої кампанії до другого туру президентських виборів. Загалом зібрано близько 70 тис. записів, а навчальна вибірка містить близько 20 тис. записів (по 10 тис. для кожного кандидата).

Для кожного кандидата навчимо окрему модель, застосовуючи кросвалідацію на п'яти фолдах, а як метрику якості використаємо F_1 -міру. Потім застосуємо найкращу з навчених моделей для класифікації коментарів до публікацій у період з початку передвиборчих перегонів до другого туру виборів (близько 50 тис. записів) і з отриманих результатів дослідимо зміну громадської думки залежно від тодішніх подій.

Для визначення тональності тексту будуть використані три алгоритми: наївний байєсівський класифікатор, метод опорних векторів та згортова нейронна мережа.

Для того щоб дані були придатними для алгоритмів машинного навчання, їх необхідно перетворити у вектори. Для векторизації тексту застосуємо такі алгоритми: Bag-of-Words і TF-IDF векторайзер [13] — для перших двох алгоритмів та Word2Vec — для ЗНМ [7]; порівняємо їх ефективність.

Для практичної реалізації поставленого завдання використовуватимемо мову програмування Python, оскільки ця мова найбільше підходить для машинного навчання. Використаємо бібліотеки sklearn для векторизації тексту

та побудови моделей (НБК і SVM) та keras для побудови згорткової нейронної мережі.

Будемо вирішувати завдання бінарної класифікації, оскільки зроблено припущення, що люди, які залишають коментарі, не є політично нейтральними, тому нейтральних коментарів або зовсім не буде, або їх буде зовсім мало, і ними можна знехтувати.

Класи є незбалансованими. Для В. Зеленського позитивний клас становить 83%, а для П. Порошенка — 38%. Про це варто пам'ятати під час навчання моделі.

Для зібраних даних розставлено мітки класів: 0 — негативний сентимент, 1 — позитивний. Навчальна вибірка містить такі поля: автор, дата, коментар та сентимент.

Перш ніж почати використовувати текст коментарів, з нього потрібно вилучити непотрібну інформацію, а саме [10]:

- видалити згадки, оскільки вони не містять емоційного навантаження;
- видалити знак хештега, але не сам хештег, оскільки він може містити інформацію;
- перевести всі слова до нижнього регістра;
- видалити всі розділові знаки, включаючи знаки запитання та знаки оклику;
- видалити URL-адреси, оскільки вони не містять корисної інформації;
- конвертувати емоджі в одне слово;
- видалити цифри;
- видалити стоп-слова;
- застосувати стемінг, щоб зберегти основу слова без закінчення чи суфіксів.

Оскільки коментарі написані українською та російською мовами, то видалятимемо російські і українські стоп-слова. Для цього застосуємо два стемери: спочатку російський, потім український [4].

Якщо після такого очищення з'являться коментарі без жодного слова, їх буде видалено, оскільки вони не містять інформації про сентимент.

Перейдемо до навчання моделей. Щоб підібрати найкращі параметри, будемо використовувати перехресну перевірку (кросвалідацію) на п'яти фолдах. Шукатимемо такі параметри: кількість n-gram, максимальний поріг відсіву, коефіцієнт регуляризації.

Для ЗНМ задано такі параметри: функцію активації: ReLU, регуляризацію (L2): 3, дропаут: 0,4, розмір батча: 100. Будемо шукати: кількість шарів згортки, розмір ядра згортки та кількість фільтрів.

Зведемо результати навчання алгоритмів до таблиці:

Результати роботи алгоритмів, %

Алгоритми	В. Зеленський	П. Порошенко
Наївний байєсівський класифікатор (Bag-of-Words)	93,9	90
Наївний байєсівський класифікатор (TF-IDF)	93,2	92
SVM (Bag-of-Words)	93,5	90
SVM (TF-IDF)	93	94
Згорткова нейронна мережа (Word2Vec)	95,6	95,5

Отже, всі алгоритми досить точно класифікують дані за правильно підібраних параметрів.

Порівняємо найкращі моделі для кожного кандидата за допомогою коробкового графіка (boxplot). Як видно з рис. 1, 2 найкращою моделлю є ЗНМ.

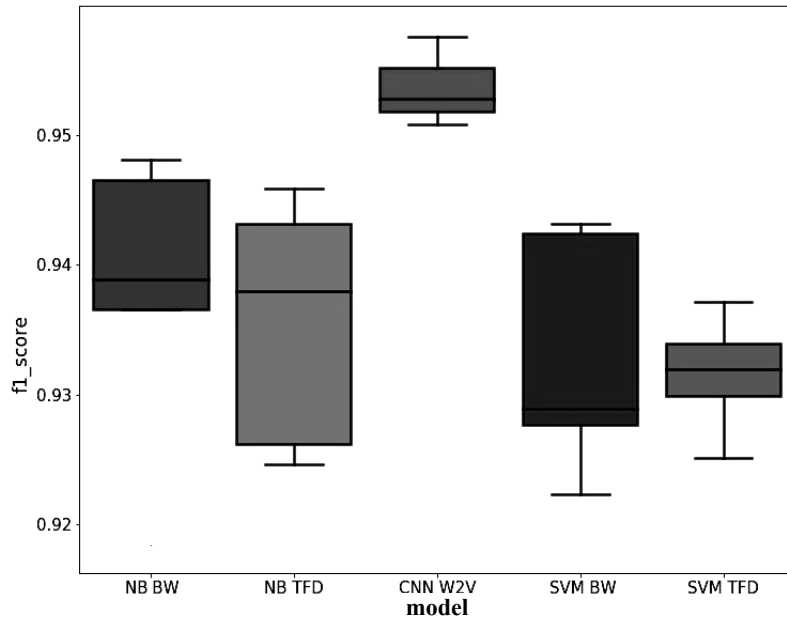


Рис. 1. Вохплот для найкращих моделей (В. Зеленський)

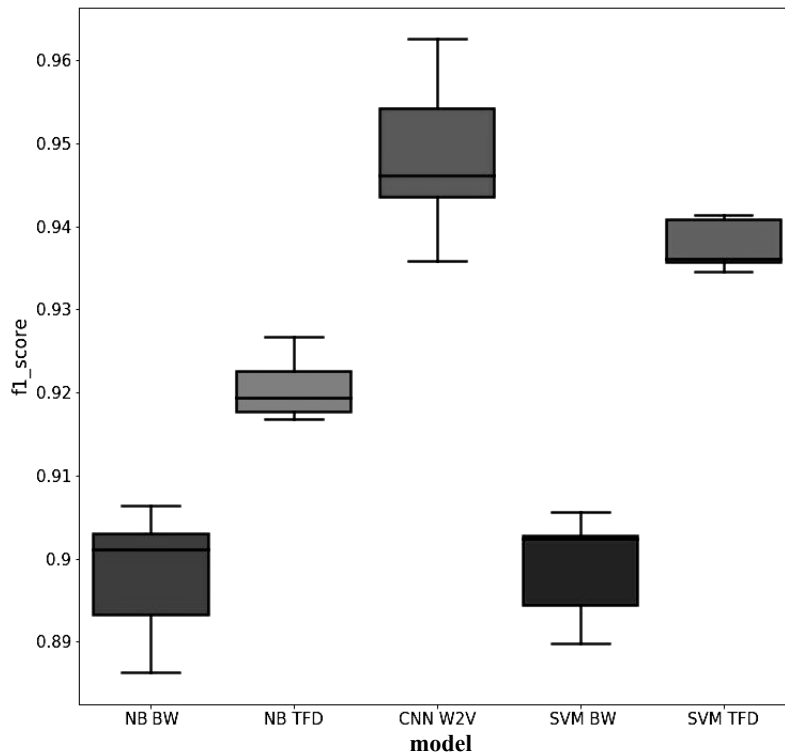


Рис. 2. Вохплот для найкращих моделей (П. Порошенко)

NB BW — наївний байєсівський класифікатор з векторайзером Bag-of-Words.

NB TFD — наївний байєсівський класифікатор з векторайзером TF-IDF.

CNN W2V — згорткова нейронна мережа з векторайзером Word2Vec.

SVM BW — метод опорних векторів з векторайзером Bag-of-Words.

SVM TFD — метод опорних векторів з векторайзером TF-IDF.

Усі моделі мають точність понад 90% (за метрикою F1). Найкращий результат для обох вибірок даних показала ЗНМ з одним шаром згортки — точність 95,5%.

Визначивши найкращу модель, проаналізуємо за її допомогою зміну прихильності громадськості до кандидата за час передвиборчої кампанії. Для цього використаємо нерозмічені коментарі під публікаціями кандидатів у період з початку президентських перегонів до другого туру виборів (03.01.2019–21.04.2019) — близько 50 тис. коментарів. Далі використаємо раніше навчену ЗНМ для класифікації коментарів. Результати класифікації (відсоток позитивного класу) зобразимо на графіку (рис. 3). На графіку на осі X позначено дату публікації поста у соціальній мережі (обиралось таким чином, щоб обидва кандидати мали публікацію в зазначений день), а на осі Y — відсоток позитивних коментарів.

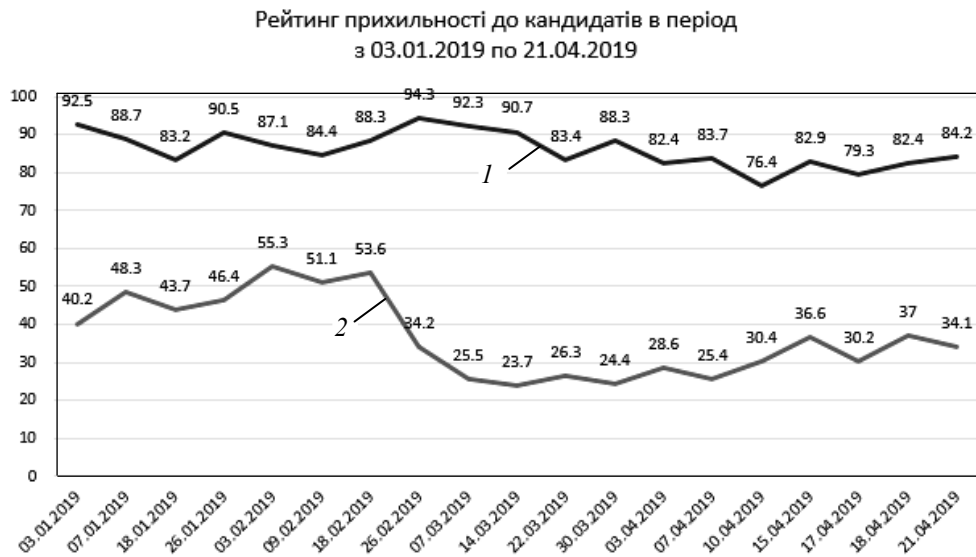


Рис. 3. Графік прихильності до кандидатів у період з 03.01.2019 по 21.04.2019: 1 — В. Зеленський, 2 — П. Порошенко

Графік дозволяє відслідковувати та аналізувати реакцію громадськості на події в автоматичному режимі. Наприклад, на рис. 3 на графіку В. Зеленського у точці за 18.01.2019 спостерігається різкий спад, який імовірно зумовлений оприлюдненням журналістського розслідування, у якому йдеться про те, що В. Зеленський через кіпрську фірму володіє кінокомпаніями в Росії. На тому ж рисунку на графіку П. Порошенка спосте-

рігається різкий спад після 26.02.2019. Ця дата припадає на вихід журналістського розслідування про корупцію в «Укроборонпромі».

ВИСНОВКИ

Досліджено методи машинного навчання для аналізу тональності тексту. Дослідження полягало у визначенні тональності тексту коментарів під публікаціями кандидатів у Президенти України (В. Зеленського та П. Порошенка) у період передвиборчих перегонів 2019 р.

Для визначення тональності тексту використано три алгоритми: наївний байєсівський класифікатор, метод опорних векторів та згорткову нейронну мережу. Оптимальні параметри для моделей підбиралися шляхом кросвалідації.

Для переведення тексту у вектор було використано три векторайзери — Bag-of-Words і TF-IDF — для наївного байєсівського класифікатора та методу опорних векторів; Word2Vec — для ЗНМ. Для кожного кандидата побудовано окремі моделі і порівняно якість класифікації (за метрикою F1).

У результаті дослідження всі моделі показали досить високу точність класифікації. Найбільш точним алгоритмом для даних обох кандидатів виявився ЗНМ з одним згортковим шаром (точність 95,5%).

Проведено дослідження зміни громадської думки в період з 03.01.2019 по 21.04.2019. Для цього зібрано близько 50 тис. коментарів з публікацій у соціальній мережі інстаграм кандидатів у Президенти України та класифіковано їх за допомогою раніше навченої ЗНМ. За результатами аналізу побудовано графік, який дає змогу оцінювати зміну громадської думки у реальному часі, відслідковувати реакцію аудиторії на події і відповідно швидко реагувати на них.

Загалом подано комплексний підхід до розв'язання задачі сентимент-аналізу, включаючи етапи попереднього оброблення тексту, використання різних векторайзерів для надання тексту векторного вигляду, навчання моделей та оцінювання їх якості.

ЛІТЕРАТУРА

1. T. Nasukawa and J.Yi, "Sentiment analysis: Capturing favorability using natural language processing", *Proc. of the 2nd Int. Conf. on Knowledge capture (KCAP)*, pp. 7077, 2003.
2. K. Dave, St. Lawrence, D. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", *Proc. of the Int. Conf. on World Wide Web (WWW)*, pp. 519528, 2003.
3. А. Барсегян, *Технологии анализа данных: Data Mining, Text Mining, Visual Mining, OLAP*, 2 изд., БХВ-Петербург, 2008, 384 p.
4. Vimala Balakrishnan, *Stemming and Lemmatization: A Comparison of Retrieval Performances*, 2014, 204 p.
5. Liu Bing, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, May 2012.
6. Vincent Ng, Claire Cardie, *Weakly Supervised Natural Language Learning Without Redundant Views*, 2003.

7. X. Fulin, D. Yihao, and T. Xiaosheng, “The Architecture of Word2vec and Its Applications”, *Journal of Nanjing*, 2015.
8. Bo Pang and Lillian Lee, *Opinion Mining and Sentiment Analysis*, 2008.
9. Bo Pang and Lillian Lee, *A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts*, 2004.
10. Janyce M. Wiebe, Rebecca F. Bruce, Thomas P. O’Hara, *Development and use of a gold-standard data set for subjectivity classifications*, 1999.
11. Jindal Liu, *Mining comparative sentences and relations*, 2006.
12. Liu Bing, *Sentiment analysis and subjectivity. Handbook of natural language processing*, 2nd ed, Boca Raton: CRC Press, 2010.
13. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, *Efficient estimation of word representations in vector space*, arXiv preprint arXiv:1301.3781. ICLR Workshop, pp. 1–12, 2013.
14. N. Sebe, MS. Lew, I. Cohen, and A. Garg, “Emotion recognition using a cauchy naive bayes classifier”, in *IEEE*, Quebec, 2002.
15. Y. Kim, “Convolutional neural networks for sentence classification”, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, Association for Computational Linguistics, October 2014, pp. 1746–1751.
16. G. Katz, N. Ofek, and B. Shapira, “Context-based sentiment analysis”, *Knowledge-Based Systems. ConSent*, vol. 84, no. 1, pp. 162–178, 2015.

Надійшла 30.07.2020

INFORMATION ON THE ARTICLE

Anna-Mariia P. Rudzevych, Educational and Scientific Complex “Institute for Applied System Analysis” of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: anna.rudzevich@gmail.com

MACHINE LEARNING IN SENTIMENT-ANALYSIS OF TEXT INFORMATION ON THE EXAMPLE OF USER ATTITUDES REGARDING CANDIDATES FOR UKRAINIAN PRESIDENTIAL ELECTIONS 2019 / A.-M. P. Rudzevych

Abstract. The main methods of machine learning for the sentiment analysis of the text are described and a comparative analysis of their effectiveness is performed. The stages of pre-processing of the text, such as stemming, deletion of stop words, algorithms for converting the text to vector form, such as bag-of-words (Bag-of-Words), TF-IDF vectorizer and Word2Vec, are considered. The goal of this study was to determine the sentiment of the comments under the publications of Ukrainian Presidential candidates (V. Zelensky and P. Poroshenko) during the 2019 election campaign. Three algorithms were used to determine the tonality of the text: the naive Bayes classifier, the support vector machine, and the convolutional neural network. Separate models were built for each candidate and a comparison of the classification quality was performed (according to metric F1). The most precise model for both data samples was a convolutional neural network.

Keywords: machine learning, sentiment analysis, text mining.

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В СЕНТИМЕНТ-АНАЛИЗЕ ТЕКСТОВОЙ ИНФОРМАЦИИ НА ПРИМЕРЕ НАСТРОЕНИЙ ПОЛЬЗОВАТЕЛЕЙ ПО ОТНОШЕНИЮ К КАНДИДАТАМ В ПРЕЗИДЕНТЫ УКРАИНЫ 2019 / А.-М. П. Рудзевич

Аннотация. Описаны основные методы машинного обучения для анализа тональности текста и проведен сравнительный анализ их эффективности. Рассмотрены этапы предварительной обработки текста, такие как стемминг, удаление стоп-слов, алгоритмы перевода текста в векторную форму: мешок слов, TF-IDF векторизатор и Word2Vec. Исследование заключалось в определении тональности текста комментариев под публикациями кандидатов в Президенты Украины (В. Зеленского и П. Порошенка) в период предвыборной гонки

2019 г. Для определения тональности текста использованы три алгоритма: наивный байесовский классификатор, метод опорных векторов и сверточная нейронная сеть. Для каждого кандидата построены отдельные модели и проведено сравнение качества классификации (по метрике F1). Лучшей моделью для обеих выборок данных оказалась сверточная нейронная сеть.

Ключевые слова: машинное обучение, sentiment-анализ, анализ тональности текста, интеллектуальный анализ данных.

REFERENCES

1. T. Nasukawa and J.Yi, “Sentiment analysis: Capturing favorability using natural language processing”, *Proc. of the 2nd Int. Conf. on Knowledge capture (KCAP)*, pp. 7077, 2003.
2. K. Dave, St. Lawrence, D. Pennock, “Mining the peanut gallery: Opinion extraction and semantic classification of product reviews”, *Proc. of the Int. Conf. on World Wide Web (WWW)*, pp. 519528, 2003.
3. A. Barsegyan, *Technologies of data analysis: Data Mining, Text Mining, Visual Mining, OLAP*, 2nd ed. BHV-Petersburg, 2008, 384 p.
4. Vimala Balakrishnan, *Stemming and Lemmatization: A Comparison of Retrieval Performances*, 2014, 204 p.
5. Liu Bing, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, May 2012.
6. Vincent Ng, Claire Cardie, *Weakly Supervised Natural Language Learning Without Redundant Views*, 2003.
7. X. Fulin, D. Yihao, and T. Xiaosheng, “The Architecture of Word2vec and Its Applications”, *Journal of Nanjing*, 2015.
8. Bo Pang and Lillian Lee, *Opinion Mining and Sentiment Analysis*, 2008.
9. Bo Pang and Lillian Lee, *A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts*, 2004.
10. Janyce M. Wiebe, Rebecca F. Bruce, Thomas P. O’Hara, *Development and use of a gold-standard data set for subjectivity classifications*, 1999.
11. JindalLiu, *Mining comparative sentences and relations*, 2006.
12. Liu Bing, *Sentiment analysis and subjectivity. Handbook of natural language processing*, 2nd ed., Boca Raton: CRC Press, 2010.
13. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, *Efficient estimation of word representations in vector space*, arXiv preprint arXiv:1301.3781. ICLR Workshop, pp. 1–12, 2013.
14. N. Sebe, MS. Lew, I. Cohen, and A. Garg, “Emotion recognition using a cauchy naive bayes classifier”, in *IEEE*, Quebec, 2002.
15. Y. Kim, “Convolutional neural networks for sentence classification”, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, Association for Computational Linguistics, October 2014, pp. 1746–1751.
16. G. Katz, N. Ofek, and B. Shapira, “Context-based sentiment analysis”, *Knowledge-Based Systems. ConSent*, vol. 84, no. 1, pp. 162–178, 2015.

**MULTICRITERIA CONDITIONAL OPTIMIZATION
BASED ON GENETIC ALGORITHMS**

V. SINEGLAZOV, K. RIAZANOVSKIY, O. CHUMACHENCKO

Abstract. This article takes on solving the problem of multicriteria conditional optimization. This problem is one of the most key tasks of the current time and has its application in many areas. Reuse of various existing algorithms for solving unconstrained optimization is proposed. Different methods of multicriteria unconditional optimization are reviewed. The advantages and disadvantages of each algorithm are analyzed. The algorithms modified to take into account the constraints. Additional algorithms of transition from solving an unconditional optimization problem to a conditional optimization problem are developed. A genetic algorithm SPEA2 was used to test the developed algorithms. Examples of solving the problem at hand using the aforementioned algorithms are presented. A comparative analysis of the final results was conducted.

Keywords: multicriteria optimization, conditional optimization, genetic algorithms, repairing algorithm, SPEA2, Pareto optimization.

INTRODUCTION

There are a lot of applications in existence that require solving a problem of multicriteria optimization (MCO). One of the fields used for this goal is artificial intelligence (AI) studies. An example of this is a process of neural networks “learning” [1].

At this point in time there are a lot of algorithms for solving multicriteria optimization problems. The majority of them are geared towards solving unconditional optimization problems. One of the most effective ways for this is the usage of the genetic algorithms (GA). Those stochastic methods, with a condition of a lot of individuals and learning epochs, allow for reaching good solutions. Because of modern computational methods it’s possible to effectively parallelize the evolutionary algorithms. It allows for easy usage in solving a broad specter of different problems: pattern discovery, signal processing, neural network learning, etc.

The majority of GA are oriented towards solving unconditional problems of MCO, but actual, real problems always have a plethora of constraints that need to be considered, and that results in solving a problem of conditional optimization. This consideration of constraints also requires a modification of the classic GA.

This article is dedicated to development of GA modifications to allow for consideration of constraints and solving problems of conditional MCO.

PROBLEM STATEMENT

A set of allowed vector values of X variables is a limited and an enclosed set (eq. 1).

$$D_X = \{X : G(X) \geq 0\} \subseteq \{X\} = R^n,$$

where $G(X)$ is some limitative vector-function; R^n is an n -dimensional arithmetic space, where n equals the power of a set $X - |X|$.

The target vector-function $F(X) = (f_1(X), f_2(X), \dots, f_{|X|}(X))$ with values in a space of targets $\{F\}$ is described in a range D_X . A problem of minimizing $F(X)$ in this range, which means minimizing each of the individual target functions $f_1(X), f_2(X), \dots, f_{|X|}(X)$ (optimality criteria) (eq. 2).

$$\min_{X \in D_X} F(X) = F(X^*) = F^*,$$

where vector X^* is the solution for a problem at hand.

REVIEW

The solution to the problem of conditional optimization is based on the solution to the problem of unconditional optimization, so it is worth considering methods of unconditional optimization. At this point in time there are two main classes of algorithms: genetic and swarm algorithms. The first class has an advantage in computational difficulty and the amount of learning epochs [11], for this reason genetic algorithms are used in this study.

There are different kinds of GA for solving multicriteria optimization problems. They can be divided in two following groups [5]: lexicographical selection, alternating criterial functions algorithms and algorithms that use Pareto dominance.

The first kind includes a lexicographic tournament selection algorithm and its numerous modifications. This algorithm is also sometimes called “naïve”

It is reasonably fast and simple, but it requires classification of criteria in regards to their importance.

The second kind includes altering objective functions algorithms, which realizes VEGA algorithm (Vector Evaluated Genetic Algorithm) [16]. It is similar to “naïve” algorithms in terms of evaluating of fitness being reliant on corresponding values of various target functions.

This type of GA also includes predator-prey algorithm. It’s more complicated in terms of computing than lexicographic selection, but the results are more accurate. It’s empirically proven that with the number of generations rising, the population of prey is moving towards the Pareto front [14]. A drawback of this algorithm is a chance of losing optimal solutions. There is a number of modifications of this algorithm to remove the drawbacks [9].

The third type includes agent ranging algorithms that is based of Pareto dominance. One of the realizations is a well-known NGS algorithm (Non-dominated Sorting Genetic Algorithm) and its modification, NGS-II [10].

The third kind also includes the Pareto-power algorithms, that utilize agent power and it's wimpiness.

Pareto-power algorithm realizes SPEA (Strength Pareto Evolutionary Algorithm), which evolved into SPEA-2 [4].

The latest agent ranging algorithms based on Pareto-dominance allow for receiving decent approximation of the Pareto front, leaving the level covering and rarefaction untouched, which gives more possible choices of the appropriate solution.

SPEA-2 method is unique and advantageous in the following:

- it includes all of the listed approaches in one algorithm;
- fitness of each individual of the population in this method is decided only in relation to the individuals of the outside set, with no correlation to dominance of individuals in the population;
- despite “the best” individuals of the previous generations being stored in an outside set, all of them are legible for the selection process;
- to prevent premature convergence, SPEA uses a special mechanism for creating niches, where dividing on the basis of fitness is done not because of the distance between individuals, but by Pareto-dominance.

Because of reasons stated above, this study utilizes SPEA2 algorithm.

The provided evolutionary algorithms are good at managing unconditional problems of multicriteria optimization, but during solving problems with constraints, the results are not always decent enough [2, 3]:

- it is possible for them to not include the point of conditional maximum;
- the resulting points may be separated in the search area;
- a part of the solutions may lay beyond the margins of the allowed area.

Based on this, the conclusion would be that the evolutionary algorithms aren't fit enough for solving problems with constraints and require some modifications to be made, those modifications taking into account the specifics of the conditional optimization problem.

Going forward to the problem of conditional optimization, it's possible to single out the following methods of solving it [12].

The first method is based on translating the problem of a conditional optimization to the problem of an unconditional optimization:

- penal functions algorithm [13]. For GA, special algorithms for this method, there is a method of “lethal” penalties, a method of statistical penalties and a method of dynamic penalties;
- algorithm based on sliding admittance [12].

The second method includes algorithms that do not use reduction of the problem down to a problem of the unconditional optimization:

- each restriction is transformed into a separate target function [6–8];
- the procedure of “repairing” using the local search [17, 18, 20];
- Orvosh-Davis reduction method [15];
- modified complex algorithm [19].

There are also hybrid methods. For example, method of the behavioral memory, “repairing” + lethal penalties [12].

The quality of the chosen solutions is dependent on choosing the period of multicriteria optimization algorithm, where the procedure of “repairing” is used.

PROBLEM SOLUTION

To solve the problem of a conditional MCO, the following algorithms were developed and researched:

1. Removal of unfit solutions after the algorithm completed.
2. Removal of unfit solutions after each generation of the population.
3. Taking into account the amount of violations by individuals in Pareto-dominance.
4. Transforming the constraints into additional criteria and solving a multicriteria problem with additional criteria.
5. Solving a multicriteria optimization problem without taking the constraints into account and “repairing” the unfit solutions after the algorithm completed.
6. Transforming the constraints into criteria, solving the multicriteria problem with additional criteria and “repairing” the unfit solutions after the algorithm completed.
7. Solving a multicriteria optimization problem without taking the constraints into account and “repairing” the unfit solutions after each iteration.
8. Transforming the constraints into criteria, solving the multicriteria problem with additional criteria and “repairing” the unfit solutions after each iteration.

For more extensive research and comparison between the aforementioned algorithms, the following is a detailed description of each algorithm.

1. Removal of unfit solutions after the algorithm completed.

This method is the simplest and the easiest. After all the generations of populations, the deletion of unfit individuals follows. But because of the simplicity the amount of fit solutions is not high.

2. Removal of unfit solutions after each generation of the population.

This method is a modification of the former one. The deletion of unfit individuals after every generation is assumed, but, with the usage of the selection operator, a new population is generated with the same size. This way a new population is only generated using the feasible solutions. With large amount of iterations, the amount of fit end-results will be close to the starting population size.

3. Taking into account the amount of constraints violated by individuals in Pareto-dominance.

After every iteration, before computing the fitness-function each individual of the population is checked for fitting in the constraints. If the individual is unfit, it's marked not feasible, and the amount of constraints that it's not fit for is saved.

Then, the modification of Pareto-comparison during the formation of fitness-function follows. The change is, initially the comparison between two individuals their fitness is checked. If one of them violates the constraints, and the other one does not, the former is considered dominated by the latter, without taking into

account its values of criteria vector. If both of them are unfit, the dominated one is chosen on the basis of the amount of constraints it's unfit for. If both are feasible, Pareto-comparison between their criteria vectors follows.

In the process of generation of generations, the amount of individuals that are unfit is reduced, and in large amounts of generations are geared towards 0. But after the learning process, there may be unfit individuals left, so they are marked as unfeasible and simply removed from the end solution.

4. Transforming the constraints into additional criteria and solving a multicriteria problem with additional criteria.

In this method, the transformation of constraints into additional criteria is used. This way, the problem of multicriteria optimization is transformed and takes the following form:

The initial problem: Target functions — $F(X) \rightarrow \text{opt}$, constraints — $G(X) \leq B$.

The transformed problem: Target functions — $F(X) \rightarrow \text{opt}$, $(G(X) - B)^2 \rightarrow \text{opt}$.

Next, the basic SPEA2 algorithm is used.

5. Solving a multicriteria optimization problem without taking into account constraints and “repairing” the unfit solutions after the algorithm completed

This method uses the unfit solution “repairing” algorithm. The “repairing” is conducted using the Pareto local search method [17, 18, 20]. In this method of solving the conditional multicriteria problem utilizes learning all of the populations without taking into account the constraints, and “repairing” all points that are unfit after.

6. Transformation of constraints into criteria, solving the multicriteria problem with multiple additional criteria and “repairing” the unfit solutions after the algorithm completed.

This method also utilizes “repairing” the points at the end, but the learning functions not only for the target criteria, but also with the transforming the constraints into criteria, as it is described in method 4.

7. Solving a multicriteria optimization problem without taking the constraints into account and “repairing” the unfit solutions after each iteration.

This method utilizes learning only on the target functions, but the “repairing” is used after each generation.

8. Transforming the constraints into criteria, solving the multicriteria problem with additional criteria and “repairing” the unfit solutions after each iteration.

The last method uses the “repairing” procedure after every iteration, the constraints are converted into additional criteria.

RESULTS

To analyze the results and to compare the a for ementioned methods, the following conditional multicriteria optimization problem was solved:

$$\text{Minimize} = \begin{cases} f_1(x, y) = 2 + (x - 2)^2 + (y - 1)^2, \\ f_2(x, y) = 9x - (y - 1)^2. \end{cases}$$

$$\text{Constraints} = \begin{cases} g_1(x, y) = x^2 + y^2 \leq 225; \\ g_2(x, y) = x - 3y + 10 \leq 0. \end{cases}$$

$$\text{Variables constraints} = \begin{cases} -20 \leq x; \\ y \leq 20. \end{cases}$$

Real Pareto front looks as shown on Fig. 1.

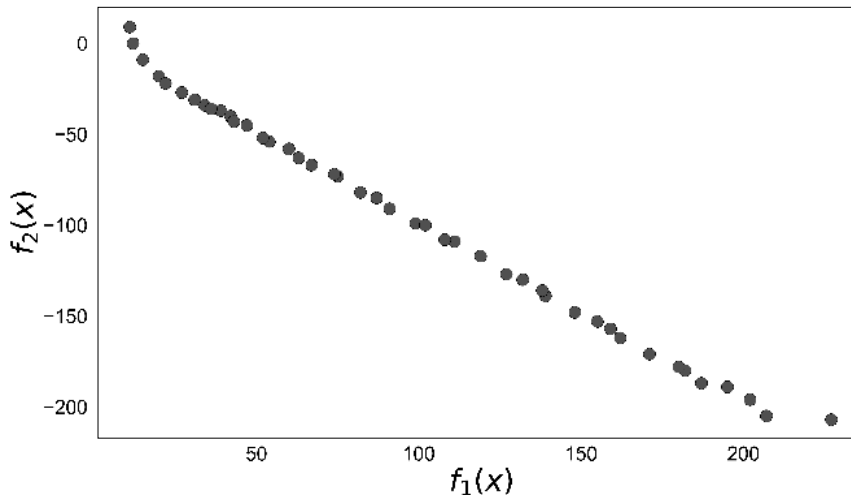


Fig. 1. Real Pareto front of the solved problem

Firstly, consider the detailed results and examples of SPEA2 algorithm working using the 3rd method, *Taking into account the amount of constraints violated by individuals in Pareto-dominance*, to solve the current problem. The amount of individuals in the population is 5, the amount of generations – 20. For each variable memory allocated is 5 bits. Integer numbers are coded into bits, and then into Grey code.

The following is the work of one iteration of the SPEA2 algorithm. For some of the following generations there will be a table of values of some individuals.

Initially, the following random population was generated:

10100	11111	01110	11001
11101	00001	00010	01100
11111	00101		

Firstly, each iteration is allocated a “parent” for new populations, using a selection operator, that is presented in SPEA2 with a binary tournament.

Binary tournament #1.

Two individuals are consecutively, yet randomly chosen:

11111 00101 и 11111 00101. In this case, the individuals chosen are the same, so there’s no better one among them. This individual will be the first “parent”.

Binary tournament #2.

The chosen individuals: 11111 00101 и 00010 01100.

The first individual is dominated by the second, so it is chosen as the second “parent”.

Using the operator “crossover” we get two descendants: 01111 01100 and 10110 00101.

This way, repeating the aforementioned algorithm, 6 more descendants are generated and, with a certain probability, mutated. Then, the values of the target functions and constraints are calculated. The intermediary results are in Table 1.

Table 1. Results of the 1st generation offspring

Variables		Objectives		Constraints	
10001	00111	102	54	125	35
10100	11101	127	-85	160	-22
01000	01110	51	-45	26	2
10101	11101	132	-76	169	-21
11011	00101	43	-43	20	20
00100	00101	252	-142	185	9

Those are added to the current population. The result is a sum of 11 individuals. Next step is to calculate the values of fitness-functions, which are lower than 1. Going forward, if the size of the population is lower than the preassigned value, the dominated elements with the best fitness-function are added, else – the worst individuals are removed. The new population is shown in Table 2.

Table 2. Details of the 2nd population

Gray binary <i>x</i>	<i>x</i>	Gray binary <i>y</i>	<i>y</i>	$f_1(x, y)$	$f_2(x, y)$	$R(i)$	$D(i)$	$F(i)$	$g_1(x, y)$	$g_2(x, y)$	Passes
10100	4	11111	11	106	-64	0	0,013	0,013	137	-19	True
01110	-9	11001	7	159	-117	0	0,013	0,013	130	-10	True
11101	2	00001	-9	102	-82	8	0,015	8,01	85	39	False
00010	-17	01100	-2	372	-162	9	0,003	9,003	293	-1	False
11111	1	00101	-4	28	-16	6	0,009	6,01	17	23	False

After the operations and generation of the new generations, as well as SPEA2 algorithm usage are concluded, the resulting individuals are only non-dominated. The examples of populations on 5th and on the last 20th generations are shown in Table 3 and Table 4 respectively.

Table 3. Details of the 5th population

Gray binary <i>x</i>	<i>x</i>	Gray binary <i>y</i>	<i>y</i>	$f_1(x, y)$	$f_2(x, y)$	$R(i)$	$D(i)$	$F(i)$	$g_1(x, y)$	$g_2(x, y)$	Passes
11110	0	11100	13	150	-144	0	0,026	0,026	169	-29	True
10100	4	11111	11	106	-64	0	0,5	0,5	137	-19	True
10100	4	11110	10	87	-45	0	0,034	0,034	116	-16	True
10100	4	11101	12	127	-86	0	0,031	0,031	160	-22	True
10100	4	11111	11	106	-64	0	0,5	0,5	137	-19	True

Table 4. Details of the 20th population

Gray binary <i>x</i>	<i>x</i>	Gray binary <i>y</i>	<i>y</i>	$f_1(x, y)$	$f_2(x, y)$	$R(i)$	$D(i)$	$F(i)$	$g_1(x, y)$	$g_2(x, y)$	Passes
11010	-1	01011	3	15	-13	0	0,081	0,081	10	0	True
11000	-4	11001	7	74	-72	0	0,046	0,045	65	-15	True
11010	-1	11001	7	47	-45	0	0,038	0,038	50	-12	True
11101	2	11101	12	123	-103	0	0,015	0,014	148	-24	True
11000	-4	11101	12	159	-157	0	0,008	0,008	160	-30	True

It should be noted that, after using this method, every individual fits. A graphical illustration of the real Pareto front and 5 points, the trained individuals, are on Fig. 2.

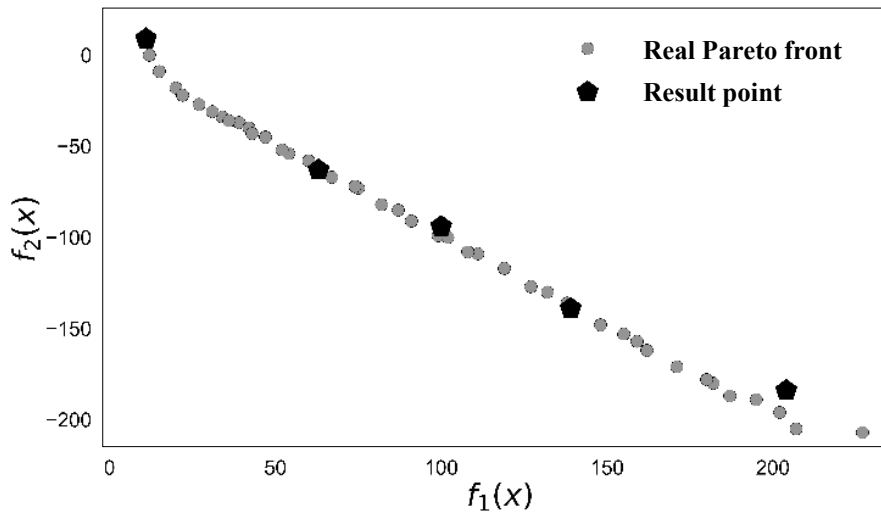


Fig. 2. Real Pareto front and five result individuals

As it is shown, even a small amount of iterations allows for great resulting solutions to a multicriteria optimization problem.

The following is an example of work and the results of each described method on the problem. Each algorithm initially used 100 individuals.

1. Removal of unfit solutions after the algorithm completed.

After learning a population of 100 individuals, only 27 fitted. Fig. 3 shows all of the individuals, and Fig. 4 shows only the fit ones.

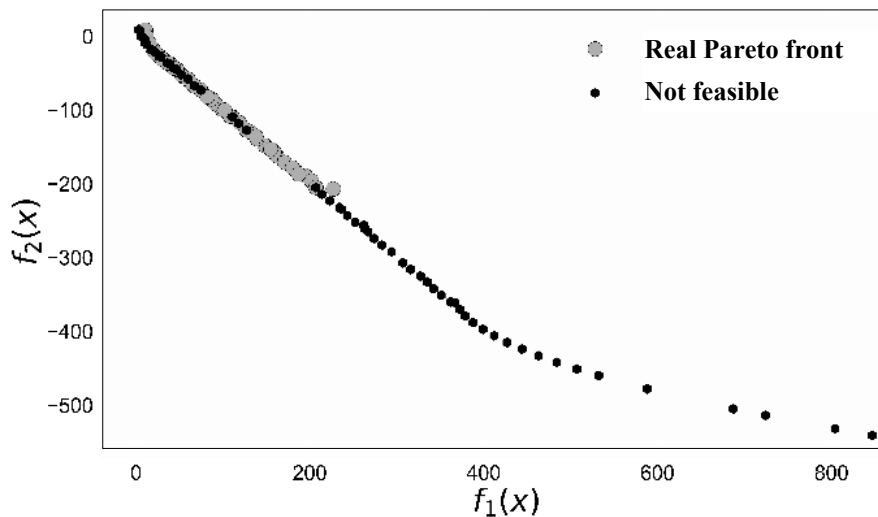


Fig. 3. Unsatisfactory individuals for the 1st method

2. Removal of unfit solutions after every generation of the population. In the end, 43 unique solutions are correct. Fig. 5 is visualization.

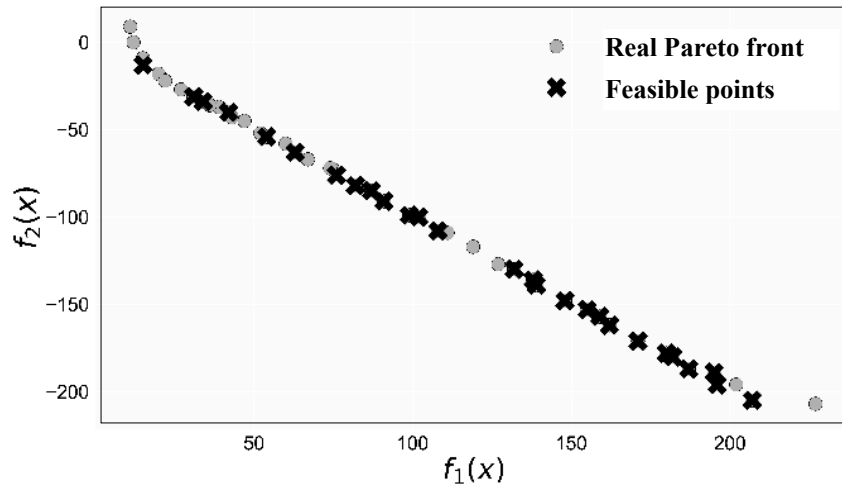


Fig. 4. Resulting individuals satisfying constraints for the 1st method

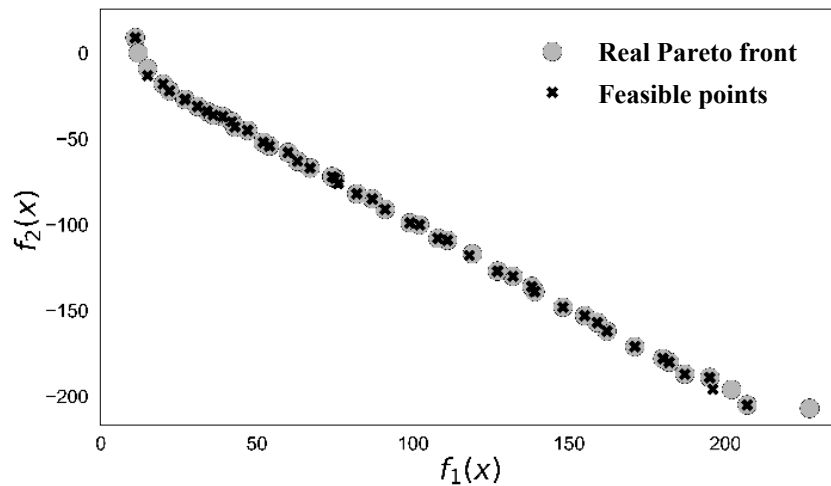


Fig. 5. Resulting individuals for the 2nd method

3. Taking into account the amount of violations by individuals in Pareto-dominance.

This method allows us to get a population with all points being fit, but only 42 of them are unique. The data is shown on Fig. 6.

4. Transforming the constraints into additional criteria and solving a multicriteria problem with additional criteria.

A simple transformation of constraints into criteria shows bad results. Only 19 non-dominated points under 4 criteria (2 real + 2 constraints) have passed. Of them, only 7 are non-dominated by two target criteria. In the end, 93 results were removed. Fig. 7 visualizes the full population, and Fig. 8 shows the end results.

5. Solving a multicriteria optimization problem without taking the constraints into account and “repairing” the unfit solutions after the algorithm completed.

In this case the “repairing” procedure allows us to get the fit points, but all of them are not better than the original set of fit points. 9 unique results were acquired. Fig. 9 and Fig. 10 contains the visualization.

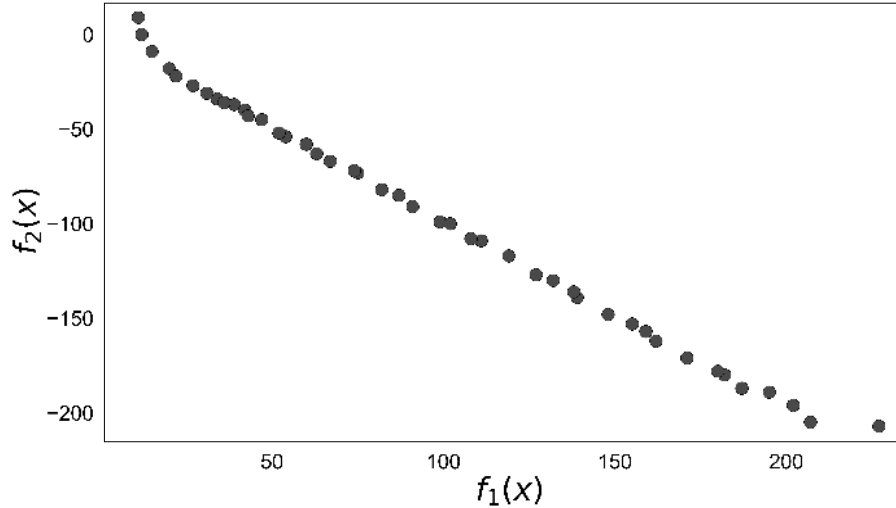


Fig. 6. Resulting population for the 3rd method

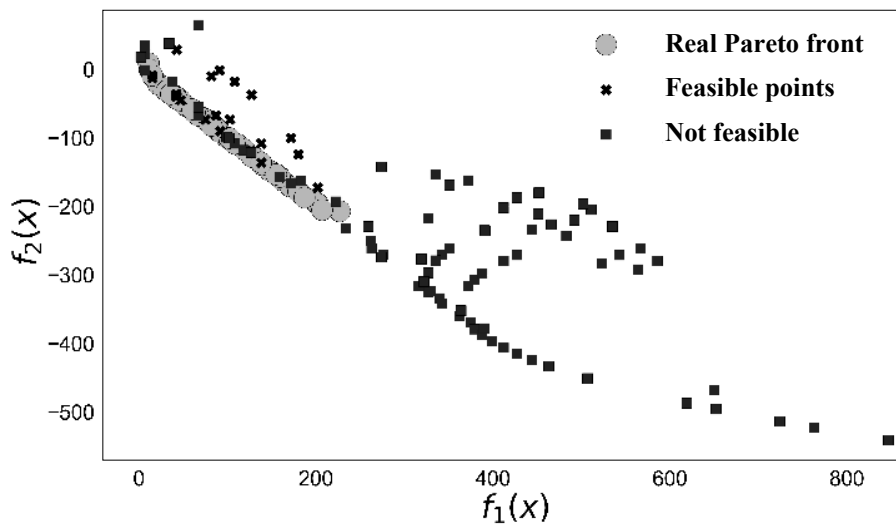


Fig. 7. Resulting population for the 4th method

6. Solving a multicriteria optimization problem without taking the constraints into account and “repairing” the unfit solutions after the algorithm completed.

In this case the “repairing” procedure allows us to get the fit points, but all of them are not better than the original set of fit points. 9 unique results were acquired. Fig. 9 and Fig. 10 contains the visualization.

7. Transforming the constraints into criteria, solving the multicriteria problem with additional criteria and “repairing” the unfit solutions after the algorithm completed.

In this case of an additional transformation into criteria the procedure of “repairing” allows for better results: 34 cured solutions, 11 of them are unique,

19 satisfactory solutions and overall result is 12 solutions. Fig. 11 shows the unfit individuals, and Fig. 12 shows the end results.

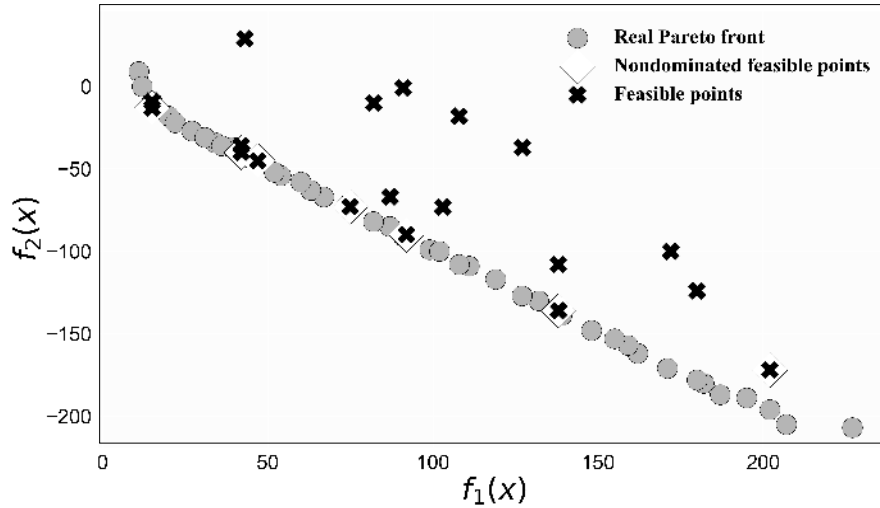


Fig. 8. Resulting non-dominated individuals for the 4th method

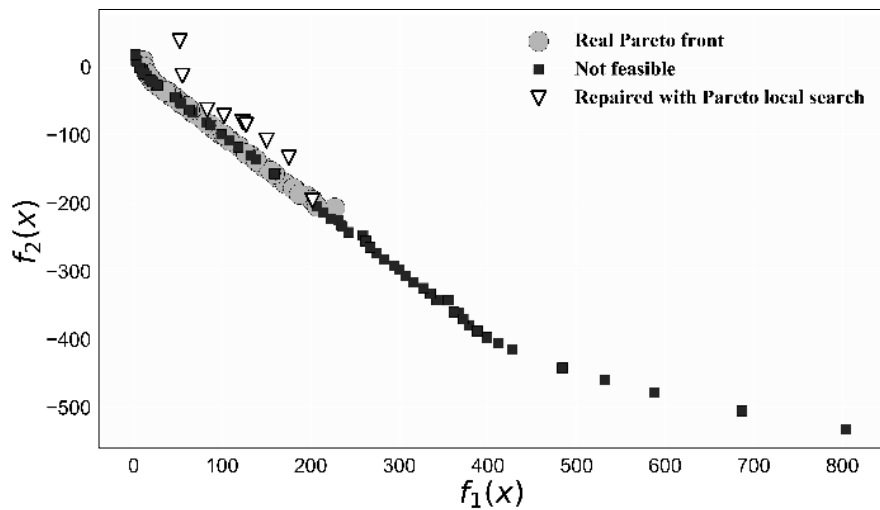


Fig. 9. Not feasible and repaired individuals for the 5th method

In this case of an additional transformation into criteria the procedure of “repairing” allows for better results: 34 cured solutions, 11 of them are unique, 19 satisfactory solutions and overall result is 12 solutions. Fig. 11 shows the unfit individuals, and Fig. 12 shows the end results.

8. Solving a multicriteria optimization problem without taking the constraints into account and “repairing” the unfit solutions after each iteration.

Learning without taking constraints into account and “repairing” after each iteration allows for decent results: 42 unique end solution. Fig. 13 is a visualization.

9. Solving a multicriteria optimization problem without taking the constraints into account and “repairing” the unfit solutions after each iteration.

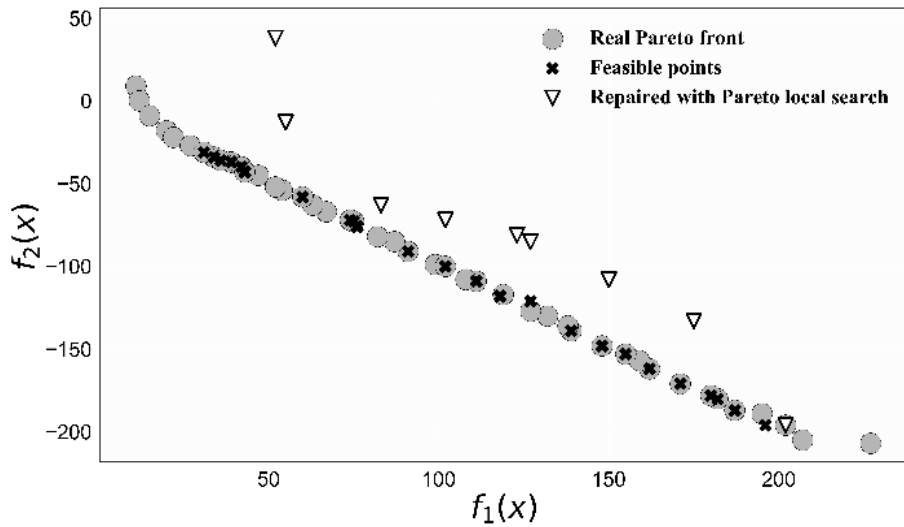


Fig. 10. Resulting individuals for the 5th method

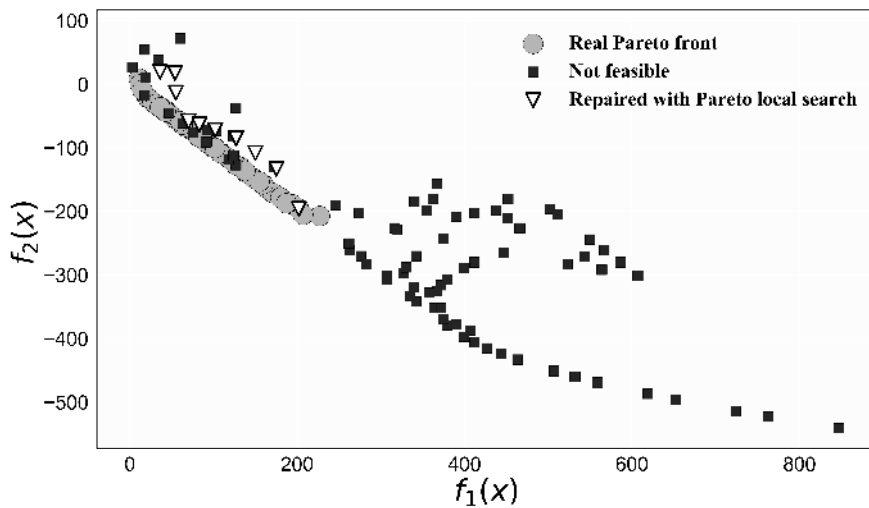


Fig. 11. Not feasible and repaired individuals for the 6th method

Learning without taking constraints into account and “repairing” after each iteration allows for decent results: 42 unique end solution. Fig. 13 is a visualization.

10. Solving a multicriteria optimization problem without taking the constraints into account and “repairing” the unfit solutions after each iteration.

Learning without taking constraints into account and “repairing” after each iteration allows for decent results: 42 unique end solution. Fig. 13 is a visualization.

11. Transforming the constraints into criteria, solving the multicriteria problem with additional criteria and “repairing” the unfit solutions after each iteration.

This method allows to get the whole population non-dominated by 4 criteria, but by the target criteria only 24 individuals are non-dominated. (Fig. 14)

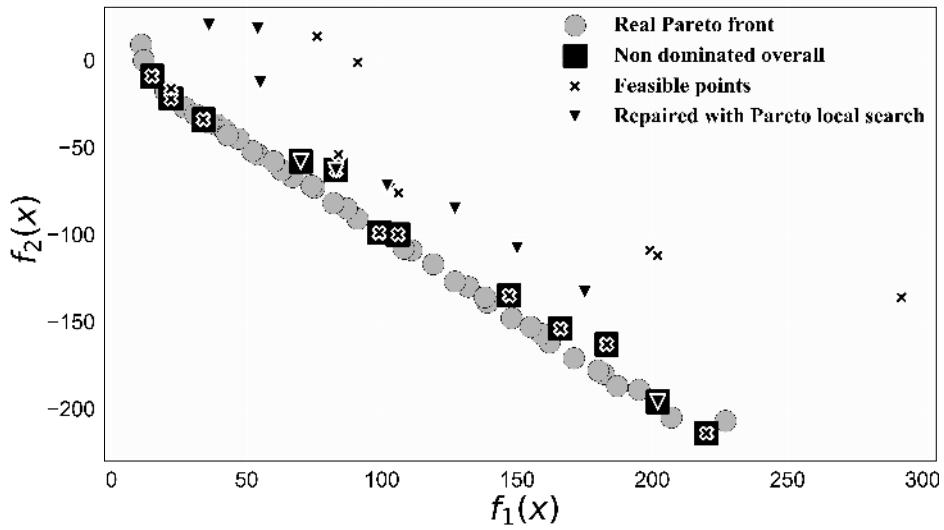


Fig. 12. Non-dominated resulting individuals for the 6th method

These results suggest that of the methods used, the best one is also one of the simplest: removal unfit individuals after each iteration. This method also does not require major SPEA2 modifications.

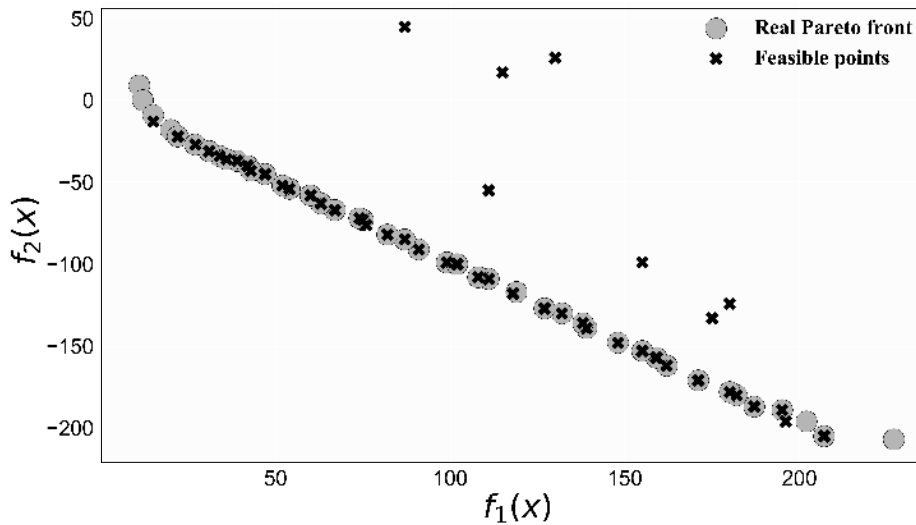


Fig. 13. Resulting non-dominated individuals for the 7th method

The same results were achieved using a method of saving the amount of violated constraints for further dominance checking during the computing of the fitness-function. This method requires some modifications to SPEA2, but those are minor and do not complicate the algorithm too much.

Another method got results, similar to the aforementioned ones. It's a method of "repairing" the individuals after each iteration. It requires much more intermediary computations that complicate the algorithm quite a bit, but the results are better.

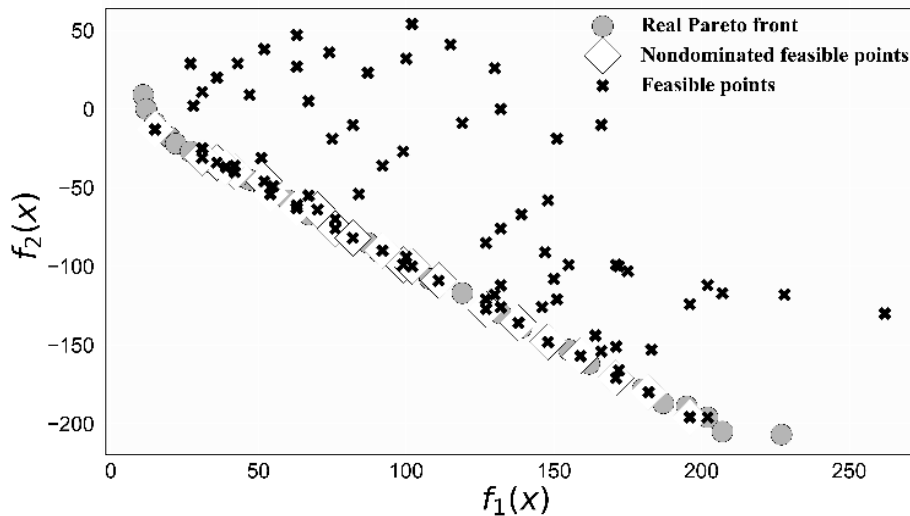


Fig. 14. Resulting individuals for the 8th method

By the means of all gathered data, we are able to come to a conclusion that transforming the constraints into criteria does not give an advantage, and the “repairing” method requires more computational time that does not give high-quality results.

CONCLUSION

The results showed that using the SPEA2 algorithm allows to accurately solve the problem of multicriteria optimization. The methods considered above showed that by modifying SPEA2 it is possible to easily switch from solving the unconditional optimization problem to solving the conditional optimization problem.

The result analysis allows for a following conclusion: the best method for the problem at hand is a simple method of removing the unfit individuals after each iteration as well as the algorithm with repairing after each iteration. Of 100 initial individuals, 43 of the resulting ones were unique and formed a resulting Pareto front of solving the conditional multicriteria optimization problem.

REFERENCES

1. A.P. Braga, R.H. Takahashi, M.A. Costa, and R. Teixeira, “Multi-Objective Algorithms for Neural Networks Learning”, *Multi-Objective Machine Learning. Studies in Computational Intelligence*, vol. 16. Berlin, Heidelberg: Springer, 2016. Available: https://doi.org/10.1007/3-540-33019-4_7
2. Carlos Artemio Coello Coello, “An Empirical Study of Evolutionary Techniques for Multiobjective Optimization in Engineering Design”, *PhD thesis*. Department of Computer Science, Tulane University, New Orleans, LA, April 1996.
3. C. A. Coello Coello, *A comprehensive survey of evolutionary-based multiobjective optimization techniques*. Laboratorio Nacional de Informatica Avanzada, Veracruz, Mexico, 1998. Available: <https://doi.org/10.1007/BF03325101>
4. E. Zitzler, M. Laumanns, and L. Thiele, “Spea2: Improving the strength pareto evolutionary algorithm for multiobjective optimization”, *Evolutionary Methods for De-*

- sign Optimization and Control with Applications to Industrial Problems. International Center for Numerical Methods in Engineering, Athens, Greece, 2001, pp. 95–100.
5. S.V. Groshev, A.P. Karpenko, and V.A. Martynyuk, “The effectiveness of population-based Pareto-approximation algorithms. Experimental comparison”, *on-line journal “Naukovedenie”*, 8(4), 2016. doi: 10.15862/67EVN416.
 6. A.V. Gumennikova, “Hybrid adaptive search algorithm for solving problems of conditional multi-criteria optimization”, *Siberian Journal of Science and Technology*, iss. 5, pp. 70–76, 2004.
 7. A.V. Gumennikova, “On the evolutionary approach to solving multicriteria problems of conditional optimization”, in *VIII international scientific-practical conference “System analysis in the project and management”*, St. Petersburg, 2004, pp. 72–76.
 8. A.V. Gumennikova, “Solving multicriteria problems of conditional and unconditional optimization using genetic algorithms MultiobjectiveGA v.1.0”, *Computer curriculum and innovation*, no. 8, p. 16, 2005.
 9. K. Deb and B.R.N. Uday, “Investigating Predator–Prey algorithms for multi-objective optimization”, *KanGAL*, Kanpur, Indian, Rep. 2005010, Dec. 2005.
 10. K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II”, *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002. doi: 10.1109/4235.996017
 11. Karl O. Jones, “Comparison of genetic algorithm and particle swarm optimization”, in *International Conference on Computer Systems and Technologies*, 2005.
 12. A.P. Karpenko, *Modern search engine optimization algorithms. Algorithms inspired by nature*, Moscow: Publishing House MSTU, 2014.
 13. A.F. Kuri-Morales and J. Gutiérrez-García, “Penalty Function Methods for Constrained Optimization with Genetic Algorithms: A Statistical Analysis”, *Lecture Notes in Computer Science*, vol. 2313. Berlin, Heidelberg: Springer, 2002. Available: https://doi.org/10.1007/3-540-46016-0_12
 14. M. Laumanns, G. Rudolph, and H.P. Schwefel, “A spatial predator-prey approach to multi-objective optimization: A preliminary study”, in *Proceedings of the Parallel Problem Solving from Nature*, V, pp. 241–249, 1998. Available: <https://doi.org/10.1007/BFb0056867>
 15. David Orvosh and Lawrence Davis, “Using a Genetic Algorithm to Optimize Problems with Feasibility Constraints”, *IEEE Conference on Evolutionary Computation – Proceedings*, vol. 2, pp. 548–553, 1994. Available: <https://doi.org/10.1109/ICEC.1994.350001>
 16. J. Schaffer, “Multiple Objective Optimization with Vector Evaluated Genetic Algorithms”, *Proceedings of the First Int. Conference on Genetic Algorithms*, pp. 93–100, 1985.
 17. E.S. Semenkina, O.E. Semenkina, and S.P. Korobeinikov, *Optimization of technical systems. Tutorial*. Krasnoyarsk: SIBMP, 1996.
 18. O.E. Semenkina and V.V. Zhidkov, *Optimization of management of complex systems by the method of generalized local search*. MAKS Press, 2002.
 19. Tomio Umeda and Atsunobu Ichikawa, “A Modified Complex Method for Optimization”, *Industrial & Engineering Chemistry Process Design and Development*, 10 (2), pp. 229–236, 1971. doi: 10.1021/i260038a016
 20. Yu.I. Zhuravlev and Yu.Yu. Finkelstein, “Local Algorithms for Linear Integer Programming Problems”, *Cybernetics problems*, iss. 14, pp. 289–295, 1965.

Received 06.10.2020

INFORMATION ON THE ARTICLE

Victor M. Sineglazov, ORCID: 0000-0002-3297-9060, National Aviation University, Ukraine, e-mail: svm@nau.edu.ua.

Kirill D. Riazanovskiy, ORCID: 0000-0002-8771-8060, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: kir_ryaz@tk.kpi.ua.

Olena I. Chumachenko, ORCID: 0000-0003-3006-7460, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: chumachenko@tk.kpi.ua.

БАГАТОКРИТЕРІАЛЬНА УМОВНА ОПТИМІЗАЦІЯ НА ОСНОВІ ГЕНЕТИЧНИХ АЛГОРИТМІВ / В.М. Синеглазов, К.Д. Рязановський, О.І. Чумаченко

Анотація. Розглянуто проблему багатокритеріальної умовної оптимізації, розв’язання якої натеper є найважливішим завданням для багатьох галузей. Запропоновано повторне використання існуючих алгоритмів розв’язання безумовної оптимізації. Розглянуто різні алгоритми багатокритеріальної безумовної оптимізації. Проаналізовано переваги та недоліки кожного алгоритму. Алгоритми модифіковано для врахування обмежень. Розроблено додаткові алгоритми переходу від розв’язання задачі безумовної оптимізації до задачі умовної оптимізації, для тестування яких використано генетичний алгоритм SPEA2. Наведено приклади вирішення поставленого завдання з використанням згаданих алгоритмів. Виконано порівняльний аналіз остаточних результатів.

Ключові слова: багатокритеріальна оптимізація, умовна оптимізація, генетичний алгоритм, алгоритм лікування, SPEA2, Парето оптимізація.

МНОГОКРИТЕРИАЛЬНАЯ УСЛОВНАЯ ОПТИМИЗАЦИЯ НА ОСНОВЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ / В.М. Синеглазов, К.Д. Рязановский, Е.И. Чумаченко

Аннотация. Рассмотрена проблема многокритериальной условной оптимизации, решение которой является одной из важнейших задач настоящего времени для многих областей. Предложено повторное использование существующих алгоритмов решения безусловной оптимизации. Рассмотрены различные алгоритмы многокритериальной безусловной оптимизации. Проанализированы достоинства и недостатки каждого алгоритма. Алгоритмы модифицированы для учета ограничений. Разработаны дополнительные алгоритмы перехода от решения задачи безусловной оптимизации к задаче условной оптимизации, для тестирования которых использован генетический алгоритм SPEA2. Приведены примеры решения поставленной задачи с использованием упомянутых алгоритмов. Проведен сравнительный анализ окончательных результатов.

Ключевые слова: многокритериальная оптимизация, условная оптимизация, генетический алгоритм, алгоритм лечения, SPEA2, Парето оптимизация.

РЕГУЛЯРНЫЕ ВЫРАЖЕНИЯ ДЛЯ НЕКОТОРЫХ ЯЗЫКОВ СЕТЕЙ ПЕТРИ В ЗАДАЧЕ О ПРОИЗВОДИТЕЛЕ И ПОТРЕБИТЕЛЕ

В.М. СТАТКЕВИЧ

Аннотация. Рассмотрены сеть Петри в задаче о производителе и потребителе (одной из классических задач синхронизации) с ограниченным буфером размера n и регулярные формальные языки L_n , которые она порождает. Для этих языков найдены регулярные выражения в рекурсивном виде, а в случаях ограниченного буфера размера от 1 до 3 — в виде явных формул. По графу достижимости построен конечный автомат, применен метод последовательного удаления вершин. Для высоты итерации (звездной высоты) указанных языков дана оценка сверху, а в случаях ограниченного буфера размера 1 и 2 найдены точные значения. Для указанных языков рассмотрены операции объединения, пересечения, замыкания Клини, конкатенации и разности. Для разности языков $L_n \setminus L_1$ построен конечный автомат и найдены регулярные выражения в рекурсивном виде, а для разности $L_2 \setminus L_1$ — в виде явной формулы.

Ключевые слова: сеть Петри, задача о производителе и потребителе, язык сети Петри, формальный язык, регулярный язык, конечный автомат, регулярное выражение, высота итерации (звездная высота).

ВВЕДЕНИЕ

Сети Петри являются удобным средством для моделирования различных процессов, систем и сетей [1–3]. Одной из классических задач синхронизации является задача о производителе и потребителе, предложенная Э. Дейкстрой [1]; существует несколько различных вариантов этой задачи.

Каждая сеть Петри может порождать язык. В работе [1] определены 12 классов языков и исследована их связь с формальными языками, определяемыми иерархией Хомского (теорию формальных языков см., например, в работах [4–7]).

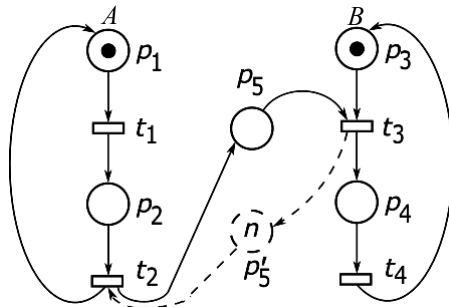
В данной работе рассмотрены некоторые регулярные языки L -типа сетей Петри в задаче о производителе и потребителе с ограниченным буфером, соответствующие регулярные выражения найдены в рекурсивном виде, а в простейших случаях — в виде явных формул. В простейших случаях найдены высоты итерации рассмотренных языков, а в более сложных дана оценка сверху. Некоторые результаты были анонсированы в тезисах [8–9]. Для варианта задачи с неограниченным буфером и соответственно контекстно-свободного языка L -типа автомат с магазинной памятью приведен в тезисах [8], а более полные результаты подробно изложены в тезисах [10].

В работе [11] рассмотрены задачи управления и проектирования сложных систем и предложена теория дискретно-непрерывных сетей, примени-

мая в случаях, когда объект описывается дифференциальными уравнениями в непрерывном времени, а управление дискретно.

ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

В задаче о производителе и потребителе процесс-производитель A производит объекты (фишки) и помещает их в буфер (позицию p_5), а процесс-потребитель B удаляет объект из буфера и использует его (рис. 1). Задача имеет следующие варианты [1, 3]: 1) с неограниченным буфером;



2) с ограниченным буфером размера n , тогда вводится дополнительная позиция p'_5 , имеющая n фишек; 3) с несколькими производителями и несколькими потребителями, тогда в позициях p_1 и p_3 в начальной маркировке μ_0 имеются s и t фишек соответственно; 4) в системе с двумя производителями и двумя потребителями один из потребителей имеет приоритет перед другим; такая система реализуется при помощи ингибиторной дуги.

Рис. 1. Задача о производителе и потребителе

Пусть A — алфавит, A^* — множество слов, т.е. конечных цепочек символов из A . Формальным языком L над алфавитом A называют некоторое подмножество $L \subset A^*$ [4–7]. Каждому переходу из множества переходов T сопоставим символ алфавита A , $\sigma: T \rightarrow A$ — функция помечения. Подобным образом конечной последовательности переходов $w \in T^*$ сопоставляется слово $\sigma(w) \in A^*$. Пусть задано множество заключительных маркировок F . Языком сети Петри L -типа называют множество таких слов $\sigma(w)$, что маркировка сети Петри после запуска последовательности переходов w является заключительной [1]. В рассматриваемой задаче полагаем $T = \{t_1, t_2, t_3, t_4\}$, $A = \{a_1, a_2, a_3, a_4\}$, $\sigma(t_i) = a_i$, $i = 1, \dots, 4$ (таким образом, сеть является свободно помеченной [1]), $F = \{\mu_0\}$.

Множество регулярных выражений над алфавитом A определяется рекурсивно [4, 5, 7]: 1) 0, 1 и любой символ $a \in A$ являются регулярными выражениями; 2) если r и s являются регулярными выражениями, то $(r + s)$, $(r \cdot s)$ и r^* также являются регулярными выражениями; 3) иных регулярных выражений нет. Каждое регулярное выражение r задает язык $L[r] \subset A^*$, определяемый рекурсивно: 1) $L[0] = \emptyset$, $L[1] = \{\epsilon\}$, $L[a] = \{a\}$ для любого символа $a \in A$; 2) $L[r + s] = L[r] \cup L[s]$, $L[r \cdot s] = L[r] \cdot L[s]$, $L[r^*] = (L[r])^*$.

Высота итерации или звездная высота регулярного выражения $sh r$ определяется рекурсивно [5, 12]: 1) $sh 0 = sh a = 0$ для любого символа $a \in A$, $sh 1 = 1$; 2) $sh(r + s) = sh(r \cdot s) = \max(sh r, sh s)$; $sh(r^*) = 1 + sh r$. Высотой итерации регулярного языка L называется минимальная из высот итерации регулярных выражений, задающих язык L .

ЯЗЫКИ, ПОРОЖДАЕМЫЕ СЕТЬЮ С ОГРАНИЧЕННЫМ БУФЕРОМ

РАЗМЕРА n

Рассмотрим сеть Петри с ограниченным буфером размера n , изображенную на рис. 1. Граф достижимости сети, изображенный на рис. 2, имеет блочную структуру.

Порождаемый сетью язык L -типа обозначим L_n . По графу достижимости строим конечный автомат, допускающий язык L_n : вершинам графа (т.е. маркировка сети) сопоставляем состояния автомата, ребрам графа (т.е. переходам сети) — переходы автомата, начальной и заключительной маркировке $\mu_0 = (1,0,1,0,0,n)$ — начальное и допускающее состояние q_0 ($q_0 \in I$, $q_0 \in F$). Отметим, что подобный принцип построения конечного автомата по графу достижимости использовался в работе [13]. Построенный автомат в данном случае является детерминированным:

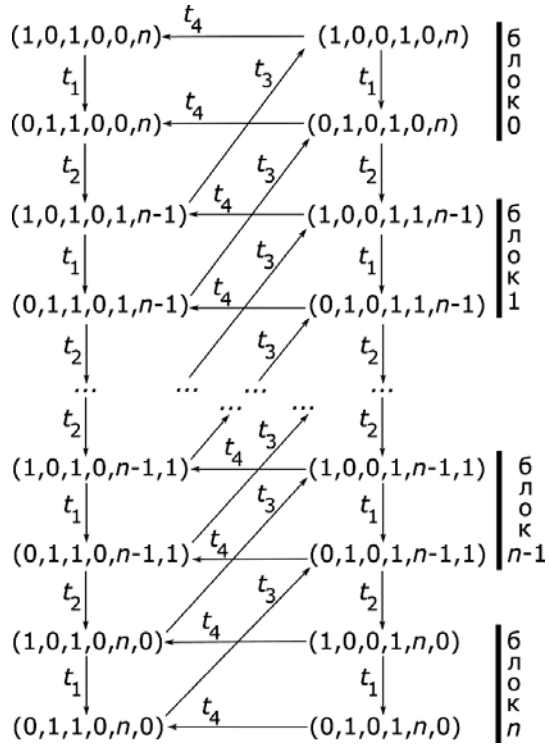


Рис. 2. Граф достижимости сети Петри

$$\langle \{q_0, \dots, q_{4n+3}\}, \{a_1, a_2, a_3, a_4\}, \Delta, \{q_0\}, \{q_0\} \rangle, \quad (1)$$

$$\begin{aligned} \Delta = & \{(q_i, a_1, q_{i+1}) : i \equiv 2, 0 \leq i \leq 4n+2\} \cup \{(q_i, a_2, q_{i+3}) : i \equiv 2, 1 \leq i \leq 4n-1\} \cup \\ & \cup \{(q_i, a_3, q_{i-2}) : i \bmod 4 \in \{0,1\}, 4 \leq i \leq 4n+1\} \cup \\ & \cup \{(q_i, a_4, q_{i-2}) : i \bmod 4 \in \{2,3\}, 2 \leq i \leq 4n+3\} \end{aligned} \quad (2)$$

(здесь и далее $\bmod 4$ означает остаток от деления на 4). Отметим, что в работе [14] представлен метод нахождения регулярного выражения для языка безопасной сети без полного построения графа достижимости, а в качестве примера рассмотрена задача о пяти обедающих мудрецах. Представленный метод основан на редукции сети Петри.

Пример 1. Пусть $n = 1$. Тогда конечный автомат (1), (2) принимает вид

$$\begin{aligned} M_1 = & \langle \{q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7\}, \{a_1, a_2, a_3, a_4\}, \Delta_1, \{q_0\}, \{q_0\} \rangle, \\ \Delta_1 = & \{(q_0, a_1, q_1), (q_2, a_1, q_3), (q_4, a_1, q_5), (q_6, a_1, q_7), (q_1, a_2, q_4), (q_3, a_2, q_6), \\ & (q_4, a_3, q_2), (q_5, a_3, q_3), (q_2, a_4, q_0), (q_3, a_4, q_1), (q_6, a_4, q_4), (q_7, a_4, q_5)\}. \end{aligned}$$

Регулярное выражение находим, удаляя по очереди состояния автомата M_1 ; подробно метод изложен, например, в работах [4, 5, 7]. Удаляя в каждом из двух блоков (рис. 2) левое нижнее и правое верхнее состояния, т.е. состояния q_1, q_2, q_5 и q_6 , получаем обобщенный автомат

$$M_2 = \langle \{q_0, q_3, q_4, q_7\}, \{a_1, a_2, a_3, a_4\}, \{(q_0, a_1 a_2, q_4), (q_4, a_3 a_4, q_0), (q_4, a_1 a_3 + a_3 a_1, q_3), (q_3, a_2 a_4 + a_4 a_2, q_4), (q_3, a_2 a_1, q_7), (q_7, a_4 a_3, q_3)\}, \{q_0\}, \{q_0\} \rangle.$$

Запишем $a_1 \parallel a_3 = a_1 a_3 + a_3 a_1$, $a_2 \parallel a_4 = a_2 a_4 + a_4 a_2$, используя оператор параллельной композиции « \parallel » [1, с. 165], а также последовательно удалим состояния q_7 и q_3 :

$$M_3 = \langle \{q_0, q_4\}, \{a_1, a_2, a_3, a_4\}, \{(q_0, a_1 a_2, q_4), (q_4, a_3 a_4, q_0), (q_4, (a_1 \parallel a_3)(a_2 a_1 a_4 a_3)^* (a_2 \parallel a_4), q_4)\}, \{q_0\}, \{q_0\} \rangle.$$

Следовательно, искомое регулярное выражение имеет вид

$$r_1 = [a_1 a_2 [(a_1 \parallel a_3)(a_2 a_1 a_4 a_3)^* (a_2 \parallel a_4)]^* a_3 a_4]^*,$$

его высота итерации $\text{sh } r_1 = 3$. Для нахождения высоты итерации языка L_1 в автомате M_2 последовательно удаляем состояния q_7 и q_4 :

$$M'_3 = \langle \{q_0, q_3\}, \{a_1, a_2, a_3, a_4\}, \{(q_0, a_1 a_2 a_3 a_4, q_0), (q_0, a_1 a_2 (a_1 \parallel a_3), q_3), (q_3, (a_2 \parallel a_4) a_3 a_4, q_0), (q_3, (a_2 \parallel a_4)(a_1 \parallel a_3) + a_2 a_1 a_4 a_3, q_3)\}, \{q_0\}, \{q_0\} \rangle.$$

Соответствующее регулярное выражение имеет вид

$$[a_1 a_2 a_3 a_4 + a_1 a_2 (a_1 \parallel a_3) [(a_2 \parallel a_4)(a_1 \parallel a_3) + a_2 a_1 a_4 a_3]^* (a_2 \parallel a_4) a_3 a_4]^*,$$

потому высота итерации языка $\text{sh } L_1 = 2$.

Пример 2. Пусть $n = 2$. Тогда конечный автомат (1), (2) принимает вид

$$M_1 = \langle \{q_0, \dots, q_{11}\}, \{a_1, a_2, a_3, a_4\}, \Delta_1, \{q_0\}, \{q_0\} \rangle,$$

$$\Delta_1 = \{(q_i, a_1, q_{i+1}) : i = 0, 2, 4, 6, 8, 10\} \cup \{(q_i, a_2, q_{i+3}) : i = 1, 3, 5, 7\} \cup$$

$$\cup \{(q_i, a_3, q_{i-2}) : i = 4, 5, 8, 9\} \cup \{(q_i, a_4, q_{i-2}) : i = 2, 3, 6, 7, 10, 11\}.$$

Как и в примере 1, удалением в каждом из трех блоков левого нижнего и правого верхнего состояний, т.е. состояний q_1, q_2, q_5, q_6, q_9 и q_{10} , получаем обобщенный автомат

$$M_2 = \langle \{q_0, q_3, q_4, q_7, q_8, q_{11}\}, \{a_1, a_2, a_3, a_4\}, \Delta_2, \{q_0\}, \{q_0\} \rangle,$$

$$\Delta_2 = \{(q_0, a_1 a_2, q_4), (q_4, a_1 a_2, q_8), (q_4, a_3 a_4, q_0), (q_8, a_3 a_4, q_4),$$

$$(q_3, a_2 a_1, q_7), (q_7, a_2 a_1, q_{11}), (q_7, a_4 a_3, q_3), (q_{11}, a_4 a_3, q_7)\},$$

$$(q_4, a_1 \parallel a_3, q_3), (q_8, a_1 \parallel a_3, q_7), (q_3, a_2 \parallel a_4, q_4), (q_7, a_2 \parallel a_4, q_8)\}.$$

Последовательно удаляем состояния q_{11} , q_8 и q_3 :

$$M_3 = \langle \{q_0, q_4, q_7\}, \{a_1, a_2, a_3, a_4\}, \Delta_3, \{q_0\}, \{q_0\} \rangle,$$

$$\Delta_3 = \{(q_0, a_1 a_2, q_4), (q_4, a_3 a_4, q_0), (q_4, a_1 a_2 a_3 a_4 + (a_1 \parallel a_3)(a_2 \parallel a_4), q_4),$$

$$(q_4, a_1 a_2 (a_1 \parallel a_3) + (a_1 \parallel a_3) a_2 a_1, q_7), (q_7, a_4 a_3 (a_2 \parallel a_4) + (a_2 \parallel a_4) a_3 a_4, q_4),$$

$$(q_7, (a_2 \parallel a_4)(a_1 \parallel a_3) + a_2 a_1 a_4 a_3 + a_4 a_3 a_2 a_1, q_7)\}.$$

Удаляя состояние q_7 , а затем q_4 , получаем искомое регулярное выражение

$$r_2 = [a_1 a_2 [a_1 a_2 a_3 a_4 + (a_1 \parallel a_3)(a_2 \parallel a_4) + (a_1 a_2 (a_1 \parallel a_3) + (a_1 \parallel a_3) a_2 a_1) \cdot \\ \cdot [(a_2 \parallel a_4)(a_1 \parallel a_3) + a_2 a_1 a_4 a_3 + a_4 a_3 a_2 a_1]^* (a_4 a_3 (a_2 \parallel a_4) + \\ + (a_2 \parallel a_4) a_3 a_4)]^* a_3 a_4]^*,$$

высота итерации которого $\text{sh } r_2 = 3$. Однако высота итерации языка $\text{sh } L_2 = 2$; в этом убеждаемся, заменяя автомат M_3 на эквивалентный с эпсилон-переходом

$$M'_3 = \langle \{q_0, q'_0, q_4, q_7\}, \{a_1, a_2, a_3, a_4\}, \Delta'_3, \{q_0\}, \{q'_0\} \rangle,$$

$$\Delta'_3 = (\Delta_3 \setminus \{(q_4, a_3 a_4, q_0)\}) \cup \{(q_4, a_3 a_4 a_1 a_2, q_4), (q_0, 1, q'_0), (q_4, a_3 a_4, q'_0)\}$$

и последовательным удалением состояний q_7 и q_4 , получая регулярное выражение

$$r'_2 = 1 + a_1 a_2 [a_1 a_2 a_3 a_4 + a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4) + (a_1 a_2 (a_1 \parallel a_3) + \\ + (a_1 \parallel a_3) a_2 a_1) [(a_2 \parallel a_4)(a_1 \parallel a_3) + a_2 a_1 a_4 a_3 + \\ + a_4 a_3 a_2 a_1]^* (a_4 a_3 (a_2 \parallel a_4) + (a_2 \parallel a_4) a_3 a_4)]^* a_3 a_4$$

высоты итерации $\text{sh } r'_2 = 2$. Подобное преобразование автомата для уменьшения высоты итерации регулярного выражения приведено, например, в работах [5, упр. 5.4.10] и [12, с. 36].

Пример 3. Пусть $n = 3$. Удаляя в автомате (1), (2) в каждом из четырех блоков (рис. 2) левое нижнее и правое верхнее состояния, получаем обобщенный автомат

$$M_1 = \langle \{q_0, q_3, q_4, q_7, q_8, q_{11}, q_{12}, q_{15}\}, \{a_1, a_2, a_3, a_4\}, \Delta_1, \{q_0\}, \{q_0\} \rangle,$$

$$\Delta_1 = \{(q_i, a_1 a_2, q_{i+4}) : i = 0, 4, 8\} \cup \{(q_{i+4}, a_3 a_4, q_i) : i = 0, 4, 8\} \cup$$

$$\cup \{(q_i, a_2 a_1, q_{i+4}) : i = 3, 7, 11\} \cup \{(q_{i+4}, a_4 a_3, q_i) : i = 3, 7, 11\} \cup$$

$$\cup \{(q_i, a_1 \parallel a_3, q_{i-1}) : i = 4, 8, 12\} \cup \{(q_{i-1}, a_2 \parallel a_4, q_i) : i = 4, 8, 12\}.$$

Обозначаем $e_1 = a_1 a_2$, $e_2 = a_3 a_4$, $e_3 = a_1 \parallel a_3$, $e_4 = a_2 \parallel a_4$ и последовательно удаляем состояния q_{15} , q_{12} и q_{11} :

$$M_2 = \langle \{q_0, q_3, q_4, q_7, q_8\}, \{a_1, a_2, a_3, a_4\}, \Delta_2, \{q_0\}, \{q_0\} \rangle,$$

$$\Delta_2 = \{(q_0, e_1, q_4), (q_4, e_1, q_8), (q_4, e_2, q_0), (q_8, e_2, q_4),$$

$$(q_3, e_1^R, q_7), (q_7, e_2^R, q_3), (q_4, e_3, q_3), (q_3, e_4, q_4),$$

$$(q_8, e_1 e_2 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_4 e_2, q_8), (q_7, e_1^R (e_1^R e_2^R + e_4 e_3)^* e_2^R, q_7),$$

$$(q_8, e_3 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_2^R, q_7), (q_7, e_4 + e_1^R (e_1^R e_2^R + e_4 e_3)^* e_4 e_2, q_8)\}$$

(здесь и далее w^R — обращение или зеркальный образ слова w [4, 5, 7]). Отметим, что удаление состояния q_{15} добавляет лишь петлю $(q_{11}, e_1^R e_2^R, q_{11})$, не изменяя другие переходы автомата; удаление состояний q_{12} и q_{11} добавляет лишь петли у состояний q_7 и q_8 и усложняет переходы из q_7 в q_8 и обратно, также не изменяя другие переходы автомата. Удаляем последовательно состояния q_8 и q_3 :

$$M_3 = \langle \{q_0, q_4, q_7\}, \{a_1, a_2, a_3, a_4\}, \Delta_3, \{q_0\}, \{q_0\} \rangle,$$

$$\Delta_3 = \{(q_0, e_1, q_4), (q_4, e_2, q_0), (q_4, e_3 e_4 + e_1 (e_1 e_2 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* e_2, q_4),$$

$$(q_4, e_3 e_1^R + e_1 (e_1 e_2 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* (e_3 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_2^R), q_7),$$

$$(q_7, e_2^R e_4 + (e_4 + e_1^R (e_1^R e_2^R + e_4 e_3)^* e_4 e_2) (e_1 e_2 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* e_2, q_4),$$

$$(q_7, (e_4 + e_1^R (e_1^R e_2^R + e_4 e_3)^* e_4 e_2) (e_1 e_2 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* \cdot$$

$$\cdot (e_3 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_2^R) + e_1^R (e_1^R e_2^R + e_4 e_3)^* e_2^R + e_2^R e_1^R, q_7)\}.$$

Удаляя состояние q_7 , а затем q_4 , получаем искомое регулярное выражение

$$\begin{aligned} r_3 = & [e_1 [e_3 e_4 + e_1 (e_1 e_2 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* e_2 + \\ & + [e_3 e_1^R + e_1 (e_1 e_2 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* (e_3 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_2^R)]] \cdot \\ & \cdot [(e_4 + e_1^R (e_1^R e_2^R + e_4 e_3)^* e_4 e_2) (e_1 e_2 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* \cdot \\ & \cdot (e_3 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_2^R) + e_1^R (e_1^R e_2^R + e_4 e_3)^* e_2^R + e_2^R e_1^R]^* \cdot \\ & \cdot [e_2^R e_4 + (e_4 + e_1^R (e_1^R e_2^R + e_4 e_3)^* e_4 e_2) (e_1 e_2 + e_1 e_3 (e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* e_2]^* e_2]^* \end{aligned}$$

высотой итерации $\text{sh } r_3 = 5$. Заменой же автомата M_3 на эквивалентный с эпсилон-переходом (см. пример 2)

$$M'_3 = \langle \{q_0, q'_0, q_4, q_7\}, \{a_1, a_2, a_3, a_4\}, \Delta'_3, \{q_0\}, \{q'_0\} \rangle,$$

$$\Delta'_3 = (\Delta_3 \setminus \{(q_4, e_2, q_0)\}) \cup \{(q_4, e_2 e_1, q_4), (q_0, 1, q'_0), (q_4, e_2, q'_0)\}$$

и последовательным удалением состояний q_7 и q_4 получаем регулярное выражение

$$\begin{aligned} r'_3 = & 1 + e_1[e_2 e_1 + e_3 e_4 + e_1(e_1 e_2 + e_1 e_3(e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* e_2 + \\ & + [e_3 e_1^R + e_1(e_1 e_2 + e_1 e_3(e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* (e_3 + e_1 e_3(e_1^R e_2^R + e_4 e_3)^* e_2^R)] \cdot \\ & \cdot [(e_4 + e_1^R(e_1^R e_2^R + e_4 e_3)^* e_4 e_2)(e_1 e_2 + e_1 e_3(e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* \cdot \\ & \cdot (e_3 + e_1 e_3(e_1^R e_2^R + e_4 e_3)^* e_2^R) + e_1^R(e_1^R e_2^R + e_4 e_3)^* e_2^R + e_2^R e_1^R]^* \cdot \\ & \cdot [e_2^R e_4 + (e_4 + e_1^R(e_1^R e_2^R + e_4 e_3)^* e_4 e_2)(e_1 e_2 + e_1 e_3(e_1^R e_2^R + e_4 e_3)^* e_4 e_2)^* e_2]^* e_2 \end{aligned}$$

высотой итерации $\text{sh } r'_3 = 4$. Следовательно, $\text{sh } L_3 \leq 4$, однако существует (см. далее замечание 1 и оценку (17)) более точная оценка $\text{sh } L_3 \leq 3$, на основании которой с учетом очевидного вложения $L_2 \subset L_3$ и равенства $\text{sh } L_2 = 2$ (см. пример 2) автор выдвигает следующую гипотезу.

Гипотеза. $\text{sh } L_3 = 3$.

В общем случае регулярное выражение для L_n имеет следующий рекурсивный вид.

Теорема 1. Язык L_n имеет вид

$$L_n = L[(a_1 a_2 (s_{n-1} + r_{n-1} t_{n-1}^* l_{n-1})^* a_3 a_4)^*] = \quad (3)$$

$$= L[1 + a_1 a_2 (a_3 a_4 a_1 a_2 + s_{n-1} + r_{n-1} t_{n-1}^* l_{n-1})^* a_3 a_4], \quad (4)$$

где регулярные выражения s_{n-1} , l_{n-1} , t_{n-1} и r_{n-1} определены рекурсивно:

$$s_0 = 0, \quad l_0 = a_2 \parallel a_4, \quad t_0 = a_2 a_1 a_4 a_3, \quad r_0 = a_1 \parallel a_3, \quad (5)$$

$$s_{k+1} = a_1 a_2 s_k^* a_3 a_4 + a_1 a_2 s_k^* r_k (t_k + l_k s_k^* r_k)^* l_k s_k^* a_3 a_4, \quad (6)$$

$$t_{k+1} = a_2 a_1 (t_k + l_k s_k^* r_k)^* a_4 a_3, \quad (7)$$

$$l_{k+1} = (a_2 \parallel a_4) + a_2 a_1 (t_k + l_k s_k^* r_k)^* l_k s_k^* a_3 a_4, \quad (8)$$

$$r_{k+1} = (a_1 \parallel a_3) + a_1 a_2 s_k^* r_k (t_k + l_k s_k^* r_k)^* a_4 a_3 \quad (9)$$

(здесь $k = 0, \dots, n-2$).

Доказательство. Рассмотрим автомат (1), (2). Аналогично примерам 1–3 удалением в каждом из $n + 1$ блоков рис. 2 левого нижнего и правого верхнего состояний (т.е. состояний q_i , $i \bmod 4 \in \{1, 2\}$, $1 \leq i \leq 4n + 2$) получаем обобщенный автомат

$$M_1 = \langle \{q_0, q_3, q_4, q_7, \dots, q_{4n}, q_{4n+3}\}, \{a_1, a_2, a_3, a_4\}, \Delta_1, \{q_0\}, \{q_0\} \rangle,$$

$$\Delta_1 = \{(q_i, a_1 a_2, q_{i+4}) : i = 0, 4, \dots, 4n - 4\} \cup \{(q_{i+4}, a_3 a_4, q_i) : i = 0, 4, \dots, 4n - 4\} \cup$$

$$\cup \{(q_i, a_2 a_1, q_{i+4}) : i = 3, 7, \dots, 4n - 1\} \cup \{(q_{i+4}, a_4 a_3, q_i) : i = 3, 7, \dots, 4n - 1\} \cup$$

$$\cup \{(q_i, a_1 \parallel a_3, q_{i-1}) : i = 4, 8, \dots, 4n\} \cup \{(q_{i-1}, a_2 \parallel a_4, q_i) : i = 4, 8, \dots, 4n\}.$$

Удаляем состояние q_{4n+3} , тем самым добавляя петлю $(q_{4n-1}, a_2 a_1 a_4 a_3, q_{4n-1})$, добавляем петлю $(q_{4n}, 0, q_{4n})$ и вводим обозначения (5). Множеству переходов полученного автомата принадлежат, в том числе, переходы

$$(q_{4n-4}, a_1 a_2, q_{4n}), (q_{4n}, a_3 a_4, q_{4n-4}), (q_{4n}, s_0, q_{4n}),$$

$$(q_{4n}, r_0, q_{4n-1}), (q_{4n-1}, l_0, q_{4n}), \tag{10}$$

$$(q_{4n-5}, a_2 a_1, q_{4n-1}), (q_{4n-1}, a_4 a_3, q_{4n-5}), (q_{4n-1}, t_0, q_{4n-1}). \tag{11}$$

Удаляем состояние q_{4n} . Тогда из множества переходов удаляются переходы (10), а добавляются переходы

$$(q_{4n-4}, a_1 a_2 s_0^* a_3 a_4, q_{4n-4}), (q_{4n-4}, a_1 a_2 s_0^* r_0, q_{4n-1}),$$

$$(q_{4n-1}, l_0 s_0^* a_3 a_4, q_{4n-4}), (q_{4n-1}, l_0 s_0^* r_0, q_{4n-1}); \tag{12}$$

объединение двух параллельных петель — третьего перехода из формулы (11) и четвертого перехода из формулы (12) — дает петлю $(q_{4n-1}, t_0 + l_0 s_0^* r_0, q_{4n-1})$. Аналогичными рассуждениями удаление состояния q_{4n-1} , объединение параллельных переходов и введение регулярных выражений s_1 , t_1 , l_1 и r_1 согласно формулам (6)–(9) приводит к добавлению переходов

$$(q_{4n-4}, s_1, q_{4n-4}), (q_{4n-5}, t_1, q_{4n-5}), (q_{4n-4}, r_1, q_{4n-5}), (q_{4n-5}, l_1, q_{4n-4}).$$

Выполняя далее подобную процедуру, т.е. удаляя состояния в порядке q_{4n-4} , q_{4n-5} , ..., q_7 , получаем рекурсивные формулы (6)–(9) и автомат

$$M_2 = \langle \{q_0, q_3, q_4\}, \{a_1, a_2, a_3, a_4\}, \{(q_0, a_1 a_2, q_4), (q_4, a_3 a_4, q_0),$$

$$(q_4, s_{n-1}, q_4), (q_3, t_{n-1}, q_3), (q_4, r_{n-1}, q_3), (q_3, l_{n-1}, q_4)\}, \{q_0\}, \{q_0\} \rangle.$$

Удаляя состояние q_3 , а затем q_4 , получаем регулярное выражение, доказывающее формулу (3). Заменой же автомата M_2 на эквивалентный с эпсилон-переходом (см. примеры 2, 3)

$$M'_2 = \langle \{q_0, q'_0, q_3, q_4\}, \{a_1, a_2, a_3, a_4\}, \{(q_0, a_1 a_2, q_4), (q_4, a_3 a_4, q'_0), (q_0, 1, q'_0),$$

$$(q_4, a_3 a_4 a_1 a_2 + s_{n-1}, q_4), (q_3, t_{n-1}, q_3), (q_4, r_{n-1}, q_3), (q_3, l_{n-1}, q_4)\}, \{q_0\}, \{q'_0\}\}$$

и последовательным удалением состояний q_3 и q_4 получаем регулярное выражение, доказывающее формулу (4). Теорема доказана.

Следствие 1. Удаление состояний автомата M_1 теоремы 1 в ином порядке $q_{4n+3}, q_{4n-1}, q_{4n}, q_{4n-5}, q_{4n-4}, \dots, q_8$ приводит к эквивалентным рекурсивным формулам:

$$\hat{s}_0 = s_0 = 0, \hat{l}_0 = l_0 = a_2 \parallel a_4, \hat{t}_0 = t_0 = a_2 a_1 a_4 a_3, \hat{r}_0 = r_0 = a_1 \parallel a_3,$$

$$\hat{s}_{k+1} = a_1 a_2 (\hat{s}_k + \hat{r}_k \hat{t}_k^* \hat{l}_k)^* a_3 a_4, \quad (13)$$

$$\hat{t}_{k+1} = a_2 a_1 \hat{t}_k^* a_4 a_3 + a_2 a_1 \hat{t}_k^* \hat{l}_k (\hat{s}_k + \hat{r}_k \hat{t}_k^* \hat{l}_k)^* \hat{r}_k \hat{t}_k^* a_4 a_3, \quad (14)$$

$$\hat{l}_{k+1} = (a_2 \parallel a_4) + a_2 a_1 \hat{t}_k^* \hat{l}_k (\hat{s}_k + \hat{r}_k \hat{t}_k^* \hat{l}_k)^* a_3 a_4, \quad (15)$$

$$\hat{r}_{k+1} = (a_1 \parallel a_3) + a_1 a_2 (\hat{s}_k + \hat{r}_k \hat{t}_k^* \hat{l}_k)^* \hat{r}_k \hat{t}_k^* a_4 a_3 \quad (16)$$

($k = 0, \dots, n-2$) при сохранении формул (3)–(5), т.е.

$$\begin{aligned} L_n &= L[(a_1 a_2 (\hat{s}_{n-1} + \hat{r}_{n-1} \hat{t}_{n-1}^* \hat{l}_{n-1})^* a_3 a_4)^*] = \\ &= L[1 + a_1 a_2 (a_3 a_4 a_1 a_2 + \hat{s}_{n-1} + \hat{r}_{n-1} \hat{t}_{n-1}^* \hat{l}_{n-1})^* a_3 a_4]. \end{aligned}$$

Следствие 2. Удаление состояний $q_{4n+3}, q_{4n}, q_{4n-1}, q_{4n-4}, q_{4n-5}, \dots, q_{12}, q_{11}$ автомата M_1 теоремы 1 с последующей заменой автомата на эквивалентный с эпсилон-переходом (см. примеры 2, 3) приводит к автомату

$$M_2'' = \langle \{q_0, q'_0, q_3, q_4, q_7, q_8\}, \{a_1, a_2, a_3, a_4\}, \Delta_2'', \{q_0\}, \{q'_0\} \rangle,$$

$$\Delta_2'' = \{(q_0, a_1 a_2, q_4), (q_4, a_1 a_2, q_8), (q_4, a_3 a_4, q'_0), (q_8, a_3 a_4, q_4), (q_0, 1, q'_0),$$

$$(q_3, a_2 a_1, q_7), (q_7, a_4 a_3, q_3), (q_4, a_1 \parallel a_3, q_3), (q_3, a_2 \parallel a_4, q_4), (q_4, a_3 a_4 a_1 a_2, q_4),$$

$$(q_8, s_{n-2}, q_8), (q_7, t_{n-2}, q_7), (q_8, r_{n-2}, q_7), (q_7, l_{n-2}, q_8)\};$$

дальнейшим удалением состояний q_3 и q_4 получаем автомат

$$M_3 = \langle \{q_0, q'_0, q_7, q_8\}, \{a_1, a_2, a_3, a_4\}, \Delta_3, \{q_0\}, \{q'_0\} \rangle,$$

$$\Delta_3 = \{(q_0, 1 + a_1 a_2 (a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4))^* a_3 a_4 = e_1, q'_0),$$

$$(q_0, a_1 a_2 (a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4))^* (a_1 \parallel a_3) a_2 a_1 = e_2, q_7),$$

$$(q_0, a_1 a_2 (a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4))^* a_1 a_2 = e_3, q_8),$$

$$(q_7, a_4 a_3 (a_2 \parallel a_4)(a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4))^* a_3 a_4 = e_4, q'_0),$$

$$(q_7, t_{n-2} + a_4 a_3 a_2 a_1 + a_4 a_3 (a_2 \parallel a_4)(a_3 a_4 a_1 a_2 +$$

$$\begin{aligned}
 &+ (a_1 \parallel a_3)(a_2 \parallel a_4))^* (a_1 \parallel a_3)a_2a_1 = e_5, q_7), \\
 &(q_7, l_{n-2} + a_4a_3(a_2 \parallel a_4)(a_3a_4a_1a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4))^* a_1a_2 = e_6, q_8), \\
 &(q_8, a_3a_4(a_3a_4a_1a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4))^* a_3a_4 = e_7, q'_0), \\
 &(q_8, r_{n-2} + a_3a_4(a_3a_4a_1a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4))^* (a_1 \parallel a_3)a_2a_1 = e_8, q_7), \\
 &(q_8, s_{n-2} + a_3a_4(a_3a_4a_1a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4))^* a_1a_2 = e_9, q_8)\}
 \end{aligned}$$

и соответствующее ему регулярное выражение

$$e_1 + e_2e_5^*e_4 + (e_3 + e_2e_5^*e_6)(e_9 + e_8e_5^*e_6)^*(e_7 + e_8e_5^*e_4).$$

Замечание 1. Оценим высоту итерации языка L_n . Поскольку $s_0 = 0$ согласно формуле (5), регулярные выражения (6)–(9) для случая $k = 0$ представимы в виде

$$\begin{aligned}
 s'_1 &= a_1a_2a_3a_4 + a_1a_2r_0(t_0 + l_0r_0)^* l_0a_3a_4, \\
 t'_1 &= a_2a_1(t_0 + l_0r_0)^* a_4a_3, \\
 l'_1 &= (a_2 \parallel a_4) + a_2a_1(t_0 + l_0r_0)^* l_0a_3a_4, \\
 r'_1 &= (a_1 \parallel a_3) + a_1a_2r_0(t_0 + l_0r_0)^* a_4a_3.
 \end{aligned}$$

Следовательно, $\text{sh } s'_1 = \text{sh } t'_1 = \text{sh } l'_1 = \text{sh } r'_1 = 1$, а из формул (6)–(9) и следствия 2 получаем

$$\begin{aligned}
 \text{sh } s_k &= \text{sh } t_k = \text{sh } l_k = \text{sh } r_k = 2k - 1 \quad (k \geq 1), \\
 \text{sh } e_1 &= \text{sh } e_2 = \text{sh } e_3 = \text{sh } e_4 = \text{sh } e_7 = 1, \\
 \text{sh } e_5 &= \text{sh } e_6 = \text{sh } e_8 = \text{sh } e_9 = 2n - 5,
 \end{aligned}$$

$$\text{sh } L_n \leq 2n - 3 \quad (n \geq 3). \tag{17}$$

Отметим, что ни формула (4) не улучшает полученную оценку (17), поскольку $\text{sh } s_{n-1} = \text{sh } t_{n-1} = \text{sh } l_{n-1} = \text{sh } r_{n-1} = 2n - 3$, $\text{sh } L_n \leq 2n - 1$, ни формулы (13)–(16), поскольку

$$\begin{aligned}
 \text{sh } \hat{s}_1 &= \text{sh } \hat{t}_1 = \text{sh } \hat{l}_1 = \text{sh } \hat{r}_1 = 2, \\
 \text{sh } \hat{s}_{n-1} &= \text{sh } \hat{t}_{n-1} = \text{sh } \hat{l}_{n-1} = \text{sh } \hat{r}_{n-1} = 2n - 2,
 \end{aligned}$$

$\text{sh } L_n \leq 2n$. В случаях $n = 1$ и $n = 2$ высота итерации языка $\text{sh } L_1 = \text{sh } L_2 = 2$ (см. примеры 1, 2), в случае $n = 3$ $\text{sh } L_3 \leq 3$ (см. пример 3). Вопрос улучшения оценки (17), равно как и вопрос нахождения $\text{sh } L_n$, для случаев $n \geq 3$ на данный момент остается для автора открытым.

НЕКОТОРЫЕ ОПЕРАЦИИ НАД ФОРМАЛЬНЫМИ ЯЗЫКАМИ L_n

Для двух языков L_n и L_m ($n > m$) выполняется очевидное вложение $L_n \supset L_m$, потому что $L_n \cup L_m = L_n$, $L_n \cap L_m = L_m$. Из формулы (3) следует равенство $(L_n)^* = L_n$.

Для конкатенации выполняется равенство $L_n L_m = L_m L_n = L_n$. Действительно, для $w \in L_n$ ввиду принадлежности $\varepsilon \in L_m$ для произвольного m имеем $w = w \cdot \varepsilon \in L_n L_m$, откуда следует вложение $L_n \subset L_n L_m$. Докажем обратное вложение. Пусть $w = w_1 w_2 \in L_n L_m$, где $w_1 \in L_n$, $w_2 \in L_m$. Сеть Петри с ограниченным буфером размера n , порождающая слово w_1 , может породить и слово w_2 , поскольку для его порождения требуется меньший буфер размера $m < n$. Потому она может породить и слово $w = w_1 w_2$, т.е. $w \in L_n$, $L_n L_m \subset L_n$. Равенство $L_m L_n = L_n$ доказывается аналогично.

Рассмотрим разность языков L_n и L_m для $n > m$. Язык $L_n \setminus L_m = L_n \cap \overline{L_m}$ регулярен, как пересечение регулярного языка L_n с регулярным языком $\overline{L_m}$ — дополнением к регулярному языку L_m (свойства замкнутости регулярных языков см., например, [4, 5, 7]).

Пример 4. Рассмотрим язык $L_2 \setminus L_1$. Построим конечный автомат M_1 , допускающий данный язык, согласно следующему принципу. Буфер (позиция p_5) имеет ограниченный размер 2 и буфер хотя бы в один момент времени содержит две фишки. Пусть t_0 — момент времени, когда впервые выполнилось условие $\mu(p_5) = 2$. До момента t_0 (формальнее: до соответствующего такта) M_1 функционирует как автомат M_1 примера 1, допускающий язык L_1 ; с момента t_0 — как автомат M_1 примера 2, допускающий язык L_2 . Автомат M_1 имеет вид, указанный на рис. 3:

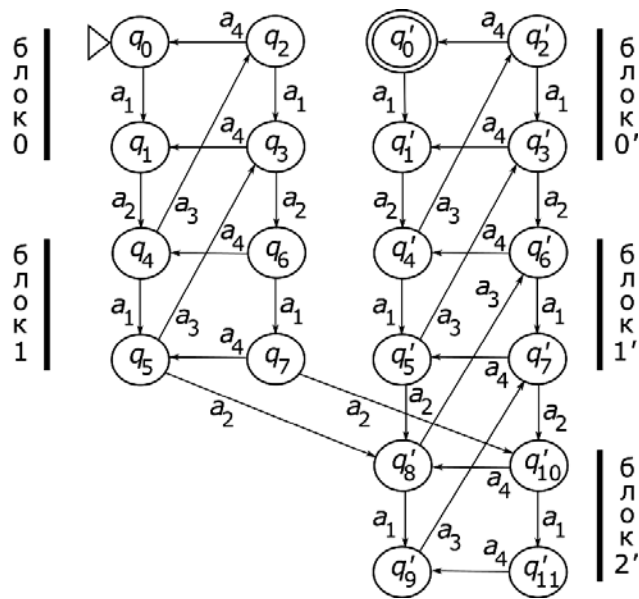


Рис. 3. Автомат, допускающий язык $L_2 \setminus L_1$

$$M_1 = \langle \{q_0, \dots, q_7\} \cup \{q'_0, \dots, q'_{11}\}, \{a_1, a_2, a_3, a_4\}, \Delta'_1 \cup \Delta''_1 \cup \Delta'''_1, \{q_0\}, \{q'_0\} \rangle,$$

$$\Delta'_1 = \{(q_0, a_1, q_1), (q_2, a_1, q_3), (q_4, a_1, q_5), (q_6, a_1, q_7), (q_1, a_2, q_4), (q_3, a_2, q_6),$$

$$(q_4, a_3, q_2), (q_5, a_3, q_3), (q_2, a_4, q_0), (q_3, a_4, q_1), (q_6, a_4, q_4), (q_7, a_4, q_5)\}, (18)$$

$$\Delta_1'' = \{(q_5, a_2, q_8'), (q_7, a_2, q_{10}')\}, (19)$$

$$\begin{aligned} \Delta_1''' = & \{(q'_i, a_1, q'_{i+1}) : i = 0, 2, 4, 6, 8, 10\} \cup \{(q'_i, a_2, q'_{i+3}) : i = 1, 3, 5, 7\} \cup \\ & \cup \{(q'_i, a_3, q'_{i-2}) : i = 4, 5, 8, 9\} \cup \{(q'_i, a_4, q'_{i-2}) : i = 2, 3, 6, 7, 10, 11\}. \end{aligned}$$

В построенном автомате состояния q_0, \dots, q_7 и переходы из множества Δ_1' соответствуют автомату M_1 примера 1, состояния q'_0, \dots, q'_{11} и переходы из множества Δ_1''' — автомату M_1 примера 2, переходы из множества Δ_1'' «связывают» две различные части. Построенный автомат является детерминированным.

Регулярное выражение находим следующим образом. В каждом из пяти блоков (двух блоков в левой колонке рис. 3 и трех в правой) удаляем левое нижнее и правое верхнее состояния аналогично примеру 2, т.е. состояния $q_1, q_2, q_5, q_6, q'_1, q'_2, q'_5, q'_6, q'_9$ и q'_{10} :

$$\begin{aligned} M_2 = & \langle \{q_0, q_3, q_4, q_7, q'_0, q'_3, q'_4, q'_7, q'_8, q'_{11}\}, \{a_1, a_2, a_3, a_4\}, \Delta_2' \cup \\ & \cup \Delta_2'' \cup \Delta_2''', \{q_0\}, \{q'_0\} \rangle, \end{aligned}$$

$$\begin{aligned} \Delta_2' = & \{(q_0, a_1 a_2, q_4), (q_4, a_3 a_4, q_0), (q_4, a_1 \parallel a_3, q_3), \\ & (q_3, a_2 \parallel a_4, q_4), (q_3, a_2 a_1, q_7), (q_7, a_4 a_3, q_3)\}, (20) \end{aligned}$$

$$\Delta_2'' = \{(q_4, a_1 a_2, q'_8), (q_7, a_2 a_1, q'_{11}), (q_7, a_2 \parallel a_4, q'_8)\}, (21)$$

$$\begin{aligned} \Delta_2''' = & \{(q'_0, a_1 a_2, q'_4), (q'_4, a_1 a_2, q'_8), (q'_4, a_3 a_4, q'_0), (q'_8, a_3 a_4, q'_4), \\ & (q'_3, a_2 a_1, q'_7), (q'_7, a_2 a_1, q'_{11}), (q'_7, a_4 a_3, q'_3), (q'_{11}, a_4 a_3, q'_7), \\ & (q'_4, a_1 \parallel a_3, q'_3), (q'_8, a_1 \parallel a_3, q'_7), (q'_3, a_2 \parallel a_4, q'_4), (q'_7, a_2 \parallel a_4, q'_8)\}. \end{aligned}$$

Заменяем автомат M_2 на эквивалентный, удаляя переходы $(q_4, a_3 a_4, q_0)$, $(q'_0, a_1 a_2, q'_4)$ и добавляя петли $(q_4, a_3 a_4 a_1 a_2, q_4)$, $(q'_4, a_3 a_4 a_1 a_2, q'_4)$. Последовательно удаляем состояния q'_{11} , q_3 и q'_3 , а также учитываем формулы (5):

$$M_3 = \langle \{q_0, q_4, q_7, q'_0, q'_4, q'_7, q'_8\}, \{a_1, a_2, a_3, a_4\}, \Delta_3' \cup \Delta_3'' \cup \Delta_3''', \{q_0\}, \{q'_0\} \rangle,$$

$$\Delta_3' = \{(q_0, a_1 a_2, q_4), (q_4, a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4), q_4),$$

$$(q_4, (a_1 \parallel a_3) a_2 a_1, q_7), (q_7, a_4 a_3 (a_2 \parallel a_4), q_4), (q_7, a_4 a_3 a_2 a_1, q_7)\},$$

$$\Delta_3'' = \{(q_4, a_1 a_2, q'_8), (q_7, a_2 a_1 a_4 a_3 = t_0, q'_7), (q_7, a_2 \parallel a_4 = l_0, q'_8)\},$$

$$\Delta_3''' = \{(q'_4, a_3 a_4, q'_0), (q'_4, a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4), q'_4), (q'_4, (a_1 \parallel a_3) a_2 a_1, q'_7)\},$$

$$(q'_7, a_4 a_3 (a_2 \parallel a_4), q'_4), (q'_7, a_2 a_1 a_4 a_3 + a_4 a_3 a_2 a_1 = t_0 + a_4 a_3 a_2 a_1, q'_7),$$

$$(q'_4, a_1 a_2, q'_8), (q'_8, a_3 a_4, q'_4), (q'_8, a_1 \parallel a_3 = r_0, q'_7), (q'_7, a_2 \parallel a_4 = l_0, q'_8)\}. \quad (22)$$

Последовательно удаляем состояния q'_8 , q_7 и q'_7 ; при удалении состояния q_7 объединение двух параллельных петель

$$(q_4, a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4), q_4),$$

$$(q_4, (a_1 \parallel a_3) a_2 a_1 (a_4 a_3 a_2 a_1)^* a_4 a_3 (a_2 \parallel a_4), q_4)$$

дает петлю $(q_4, a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 a_1 a_4 a_3)^* (a_2 \parallel a_4), q_4)$; с учетом этого получаем автомат

$$M_4 = \langle \{q_0, q_4, q'_0, q'_4\}, \{a_1, a_2, a_3, a_4\}, \Delta_4, \{q_0\}, \{q'_0\} \rangle,$$

$$\Delta_4 = \{(q_0, a_1 a_2, q_4), (q'_4, a_3 a_4, q'_0),$$

$$(q_4, a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 a_1 a_4 a_3)^* (a_2 \parallel a_4), q_4),$$

$$(q_4, a_1 a_2 a_3 a_4 + (a_1 \parallel a_3) a_2 a_1 (a_4 a_3 a_2 a_1)^* (a_2 \parallel a_4) a_3 a_4 +$$

$$+ (a_1 a_2 (a_1 \parallel a_3) + (a_1 \parallel a_3) a_2 a_1 (a_4 a_3 a_2 a_1)^* (a_2 a_1 a_4 a_3 + (a_2 \parallel a_4)(a_1 \parallel a_3))) \cdot$$

$$\cdot (a_2 a_1 a_4 a_3 + a_4 a_3 a_2 a_1 + (a_2 \parallel a_4)(a_1 \parallel a_3))^* (a_4 a_3 (a_2 \parallel a_4) + (a_2 \parallel a_4) a_3 a_4), q'_4),$$

$$(q'_4, a_1 a_2 a_3 a_4 + a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4) + (a_1 a_2 (a_1 \parallel a_3) + (a_1 \parallel a_3) a_2 a_1) \cdot$$

$$\cdot (a_2 a_1 a_4 a_3 + a_4 a_3 a_2 a_1 + (a_2 \parallel a_4)(a_1 \parallel a_3))^* (a_4 a_3 (a_2 \parallel a_4) + (a_2 \parallel a_4) a_3 a_4), q'_4)\}.$$

Удаляя состояние q_4 и q'_4 , получаем искомое регулярное выражение

$$r_4 = a_1 a_2 [a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 a_1 a_4 a_3)^* (a_2 \parallel a_4)]^* \cdot$$

$$\cdot [a_1 a_2 a_3 a_4 + (a_1 \parallel a_3) a_2 a_1 (a_4 a_3 a_2 a_1)^* (a_2 \parallel a_4) a_3 a_4 +$$

$$+ (a_1 a_2 (a_1 \parallel a_3) + (a_1 \parallel a_3) a_2 a_1 (a_4 a_3 a_2 a_1)^* (a_2 a_1 a_4 a_3 + (a_2 \parallel a_4)(a_1 \parallel a_3))) \cdot$$

$$\cdot (a_2 a_1 a_4 a_3 + a_4 a_3 a_2 a_1 + (a_2 \parallel a_4)(a_1 \parallel a_3))^* (a_4 a_3 (a_2 \parallel a_4) + (a_2 \parallel a_4) a_3 a_4) \cdot$$

$$\cdot [a_1 a_2 a_3 a_4 + a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4) + (a_1 a_2 (a_1 \parallel a_3) + (a_1 \parallel a_3) a_2 a_1) \cdot$$

$$\cdot (a_2 a_1 a_4 a_3 + a_4 a_3 a_2 a_1 + (a_2 \parallel a_4)(a_1 \parallel a_3))^* (a_4 a_3 (a_2 \parallel a_4) + (a_2 \parallel a_4) a_3 a_4)]^* a_3 a_4$$

высотой итерации $\text{sh } r_4 = 2$. Тогда высота итерации языка $\text{sh}(L_2 \setminus L_1) = 2$.

Отметим, что $\text{sh}(L_2 \setminus L_1) = \text{sh } L_2$ с учетом примера 2.

Регулярное выражение для $L_n \setminus L_1$ имеет следующий рекурсивный вид.

Теорема 2. Язык $L_n \setminus L_1$, $n \geq 2$ имеет вид $L_n \setminus L_1 = L[r_{n \setminus 1}]$, где

$$\begin{aligned}
 r_{n\setminus 1} = & a_1 a_2 [a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 a_1 a_4 a_3)^* (a_2 \parallel a_4)]^* \cdot \\
 & \cdot [(a_1 a_2 + (a_1 \parallel a_3) a_2 a_1 (a_4 a_3 a_2 a_1)^* l_{n-2}) s_{n-2}^* a_3 a_4 + \\
 & + (a_1 a_2 s_{n-2}^* r_{n-2} + (a_1 \parallel a_3) a_2 a_1 (a_4 a_3 a_2 a_1)^* (t_{n-2} + l_{n-2} s_{n-2}^* r_{n-2})) \cdot \\
 & \cdot (t_{n-2} + l_{n-2} s_{n-2}^* r_{n-2} + a_4 a_3 a_2 a_1)^* (a_4 a_3 (a_2 \parallel a_4) + l_{n-2} s_{n-2}^* a_3 a_4)] \cdot \\
 & \cdot [a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4) + a_1 a_2 s_{n-2}^* a_3 a_4 + ((a_1 \parallel a_3) a_2 a_1 + a_1 a_2 s_{n-2}^* r_{n-2}) \cdot \\
 & \cdot (t_{n-2} + l_{n-2} s_{n-2}^* r_{n-2} + a_4 a_3 a_2 a_1)^* (a_4 a_3 (a_2 \parallel a_4) + l_{n-2} s_{n-2}^* a_3 a_4)]^* a_3 a_4, \quad (23)
 \end{aligned}$$

а s_{n-2} , t_{n-2} , l_{n-2} и r_{n-2} определены рекурсивно формулами (5)–(9).

Доказательство. Сперва отметим, что для случая $n = 2$ непосредственная подстановка формул (5) в формулу (23) доказывает, что $r_{n\setminus 1}$ совпадает с регулярным выражением r_4 примера 4. Рассмотрим автомат

$$\begin{aligned}
 M_1 = & \langle \{q_0, \dots, q_7\} \cup \{q'_0, \dots, q'_{4n+3}\}, \{a_1, a_2, a_3, a_4\}, \Delta'_1 \cup \Delta''_1 \cup \Delta'''_1, \{q_0\}, \{q'_0\} \rangle, \\
 \Delta'_1 = & \{(q'_i, a_1, q'_{i+1}) : i \geq 2, 0 \leq i \leq 4n + 2\} \cup \{(q'_i, a_2, q'_{i+3}) : i \geq 2, 1 \leq i \leq 4n - 1\} \cup \\
 & \cup \{(q'_i, a_3, q'_{i-2}) : i \bmod 4 \in \{0, 1\}, 4 \leq i \leq 4n + 1\} \cup \\
 & \cup \{(q'_i, a_4, q'_{i-2}) : i \bmod 4 \in \{2, 3\}, 2 \leq i \leq 4n + 3\}
 \end{aligned}$$

(здесь Δ'_1 и Δ''_1 определены формулами (18), (19) соответственно, а Δ'''_1 аналогично формуле (2)). Рассматриваемый автомат является детерминированным. Аналогично теореме 1 удалением в каждом из $n + 3$ блоков левого нижнего и правого верхнего состояний (т.е. состояний q_1 , q_2 , q_5 , q_6 и q'_i , $i \bmod 4 \in \{1, 2\}$, $1 \leq i \leq 4n + 2$) получаем обобщенный автомат

$$\begin{aligned}
 M_2 = & \langle \{q_0, q_3, q_4, q_7, q'_0, q'_3, q'_4, q'_7, \dots, q'_{4n}, q'_{4n+3}\}, \{a_1, a_2, a_3, a_4\}, \Delta'_2 \cup \\
 & \cup \Delta''_2 \cup \Delta'''_2, \{q_0\}, \{q'_0\} \rangle, \\
 \Delta'_2 = & \{(q'_i, a_1 a_2, q'_{i+4}) : i = 0, 4, \dots, 4n - 4\} \cup \{(q'_{i+4}, a_3 a_4, q'_i) : i = 0, 4, \dots, 4n - 4\} \cup \\
 & \cup \{(q'_i, a_2 a_1, q'_{i+4}) : i = 3, 7, \dots, 4n - 1\} \cup \{(q'_{i+4}, a_4 a_3, q'_i) : i = 3, 7, \dots, 4n - 1\} \cup \\
 & \cup \{(q'_i, a_1 \parallel a_3, q'_{i-1}) : i = 4, 8, \dots, 4n\} \cup \{(q'_{i-1}, a_2 \parallel a_4, q'_i) : i = 4, 8, \dots, 4n\}
 \end{aligned}$$

(здесь Δ'_2 и Δ''_2 определены формулами (20), (21) соответственно, а Δ'''_2 аналогично множеству Δ_1 из доказательства теоремы 1). Последовательно удаляем состояния q'_{4n+3} , q'_{4n} , q'_{4n-1} , q'_{4n-4} , q'_{4n-5} , ..., q'_{15} и вводим рекурсивные формулы (6)–(9) по аналогии с рассуждениями доказательства теоремы 1. Отметим, что данные действия не изменяют переходы множеств Δ'_2 и Δ''_2 . Полученный автомат имеет вид

$$\tilde{M}_2 = \langle \{q_0, q_3, q_4, q_7, q'_0, q'_3, q'_4, q'_7, q'_8, q'_{11}, q'_{12}\}, \{a_1, a_2, a_3, a_4\}, \Delta'_2 \cup \cup \Delta''_2 \cup \tilde{\Delta}'''_2, \{q_0\}, \{q'_0\} \rangle,$$

$$\begin{aligned} \tilde{\Delta}'''_2 = & \{(q'_i, a_1 a_2, q'_{i+4}) : i = 0, 4, 8\} \cup \{(q'_{i+4}, a_3 a_4, q'_i) : i = 0, 4, 8\} \cup \\ & \cup \{(q'_3, a_2 a_1, q'_7), (q'_7, a_2 a_1, q'_{11}), (q'_7, a_4 a_3, q'_3), (q'_{11}, a_4 a_3, q'_7), \\ & (q'_4, a_1 \parallel a_3, q'_3), (q'_8, a_1 \parallel a_3, q'_7), (q'_3, a_2 \parallel a_4, q'_4), (q'_7, a_2 \parallel a_4, q'_8), \\ & (q'_{12}, s_{n-3}, q'_{12}), (q'_{11}, t_{n-3}, q'_{11}), (q'_{12}, r_{n-3}, q'_{11}), (q'_{11}, l_{n-3}, q'_{12})\}. \end{aligned}$$

Заменяем автомат \tilde{M}_2 на эквивалентный, удаляя переходы $(q_4, a_3 a_4, q_0)$, $(q'_0, a_1 a_2, q'_4)$ и добавляя петли $(q_4, a_3 a_4 a_1 a_2, q_4)$, $(q'_4, a_3 a_4 a_1 a_2, q'_4)$. Последовательно удаляем состояния q'_{12} и q'_{11} , используем рекурсивные формулы (6)–(9), затем удаляем состояния q_3 и q'_3 :

$$M_3 = \langle \{q_0, q_4, q_7, q'_0, q'_4, q'_7, q'_8\}, \{a_1, a_2, a_3, a_4\}, \Delta'_3 \cup \Delta''_3 \cup \Delta'''_3, \{q_0\}, \{q'_0\} \rangle,$$

$$\begin{aligned} \Delta'_3 = & \{(q_0, a_1 a_2, q_4), (q_4, a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4), q_4), \\ & (q_4, (a_1 \parallel a_3) a_2 a_1, q_7), (q_7, a_4 a_3 (a_2 \parallel a_4), q_4), (q_7, a_4 a_3 a_2 a_1, q_7)\}, \end{aligned} \quad (24)$$

$$\Delta''_3 = \{(q_4, a_1 a_2, q'_8), (q_7, t_{n-2}, q'_7), (q_7, l_{n-2}, q'_8)\}, \quad (25)$$

$$\begin{aligned} \Delta'''_3 = & \{(q'_4, a_3 a_4, q'_0), (q'_4, a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 \parallel a_4), q'_4), (q'_4, (a_1 \parallel a_3) a_2 a_1, q'_7), \\ & (q'_7, a_4 a_3 (a_2 \parallel a_4), q'_4), (q'_7, t_{n-2} + a_4 a_3 a_2 a_1, q'_7), (q'_4, a_1 a_2, q'_8), \\ & (q'_8, a_3 a_4, q'_4), (q'_8, r_{n-2}, q'_7), (q'_7, l_{n-2}, q'_8), (q'_8, s_{n-2}, q'_8)\}. \end{aligned} \quad (26)$$

Сравнение данного автомата с автоматом M_3 примера 4 показывает:

- 1) в множествах Δ'_3 переходы совпадают;
- 2) в множествах Δ''_3 второй и третий переходы различаются лишь индексами;
- 3) в множествах Δ'''_3 пятый, восьмой и девятый переходы различаются лишь индексами, а десятый переход множества (26) при $n = 2$ с учетом формул (5) является пустым переходом $(q'_8, s_0 = 0, q'_8)$ и его можно без потери общности добавить в множество (22).

Потому, выполняя предложенные в примере 4 дальнейшие действия, а именно последовательно удаляя состояния q'_8 , q_7 , q'_7 , q_4 и q'_4 , а также принимая во внимание оговоренную в п. 3) петлю (q'_8, s_{n-2}, q'_8) , получаем искомое регулярное выражение (23).

Теорема доказана.

Следствие 3. Язык $L_n \setminus L_1$ также допускает представление в виде $L_n \setminus L_1 = L[r'_{n \setminus 1}]$, где

$$r'_{n \setminus 1} = a_1 a_2 [a_3 a_4 a_1 a_2 + (a_1 \parallel a_3)(a_2 a_1 a_4 a_3)^* (a_2 \parallel a_4)]^*.$$

$$\begin{aligned} & \cdot [s_{n-1} + (a_1 \parallel a_3)a_2a_1(a_4a_3a_2a_1)^*(t_{n-2} + l_{n-2}s_{n-2}^*r_{n-2})^*l_{n-2}s_{n-2}^*a_3a_4 + \\ & + (a_1a_2s_{n-2}^*r_{n-2} + (a_1 \parallel a_3)a_2a_1(a_4a_3a_2a_1)^*(t_{n-2} + l_{n-2}s_{n-2}^*r_{n-2})) \cdot \\ & \cdot (t_{n-2} + l_{n-2}s_{n-2}^*r_{n-2})^*a_4a_3t_{n-1}l_{n-1}] \cdot [a_3a_4a_1a_2 + s_{n-1} + r_{n-1}t_{n-1}^*l_{n-1}]^*a_3a_4. \quad (27) \end{aligned}$$

Доказательство. В доказательстве теоремы 2 при построении автомата M_3 по автомату \tilde{M}_2 не удаляем (в отличие от доказательства теоремы 2) состояние q'_3 , получая следующий автомат:

$$\begin{aligned} \tilde{M}_3 = & \langle \{q_0, q_4, q_7, q'_0, q'_3, q'_4, q'_7, q'_8\}, \{a_1, a_2, a_3, a_4\}, \Delta'_3 \cup \Delta''_3 \cup \tilde{\Delta}''_3, \{q_0\}, \{q'_0\} \rangle, \\ \tilde{\Delta}''_3 = & \{(q'_4, a_3a_4, q'_0), (q'_4, a_3a_4a_1a_2, q'_4), (q'_4, a_1 \parallel a_3, q'_3), (q'_3, a_2 \parallel a_4, q'_4), \\ & (q'_3, a_2a_1, q'_7), (q'_7, a_4a_3, q'_3), (q'_7, t_{n-2}, q'_7), (q'_4, a_1a_2, q'_8), \\ & (q'_8, a_3a_4, q'_4), (q'_8, r_{n-2}, q'_7), (q'_7, l_{n-2}, q'_8), (q'_8, s_{n-2}, q'_8)\} \end{aligned}$$

(здесь Δ'_3 и Δ''_3 определены формулами (24), (25) соответственно). Последовательно удаляем состояния q'_8 и q'_7 , используем рекурсивные формулы (6)–(9):

$$\begin{aligned} \tilde{M}_4 = & \langle \{q_0, q_4, q_7, q'_0, q'_3, q'_4\}, \{a_1, a_2, a_3, a_4\}, \Delta'_3 \cup \tilde{\Delta}_4, \{q_0\}, \{q'_0\} \rangle, \\ \tilde{\Delta}_4 = & \{(q_4, a_1a_2s_{n-2}^*r_{n-2}(t_{n-2} + l_{n-2}s_{n-2}^*r_{n-2})^*a_4a_3 \neq r_{n-1}, q'_3), (q_4, s_{n-1}, q'_4), \\ & (q_7, (t_{n-2} + l_{n-2}s_{n-2}^*r_{n-2})^+a_4a_3, q'_3), (q_7, (t_{n-2} + l_{n-2}s_{n-2}^*r_{n-2})^*l_{n-2}s_{n-2}^*a_3a_4, q'_4), \\ & (q'_4, a_3a_4, q'_0), (q'_4, a_3a_4a_1a_2 + s_{n-1}, q'_4), (q'_4, r_{n-1}, q'_3), (q'_3, l_{n-1}, q'_4), (q'_3, t_{n-1}, q'_3)\} \end{aligned}$$

(в третьем переходе множества $\tilde{\Delta}_4$ использована запись $e^+ = e^*e$, см. например, [4, 5]). Последовательным удалением состояний q_7 и q'_3 получаем автомат

$$\begin{aligned} \tilde{M}_5 = & \langle \{q_0, q_4, q'_0, q'_4\}, \{a_1, a_2, a_3, a_4\}, \tilde{\Delta}_5, \{q_0\}, \{q'_0\} \rangle, \\ \tilde{\Delta}_5 = & \{(q_0, a_1a_2, q_4), (q'_4, a_3a_4, q'_0), \\ & (q_4, a_3a_4a_1a_2 + (a_1 \parallel a_3)(a_2a_1a_4a_3)^*(a_2 \parallel a_4), q_4), \\ & (q_4, s_{n-1} + (a_1 \parallel a_3)a_2a_1(a_4a_3a_2a_1)^*(t_{n-2} + l_{n-2}s_{n-2}^*r_{n-2})^*l_{n-2}s_{n-2}^*a_3a_4 + \\ & + (a_1a_2s_{n-2}^*r_{n-2} + (a_1 \parallel a_3)a_2a_1(a_4a_3a_2a_1)^*(t_{n-2} + l_{n-2}s_{n-2}^*r_{n-2})) \cdot \\ & \cdot (t_{n-2} + l_{n-2}s_{n-2}^*r_{n-2})^*a_4a_3t_{n-1}l_{n-1}, q'_4), (q'_4, a_3a_4a_1a_2 + s_{n-1} + r_{n-1}t_{n-1}^*l_{n-1}, q'_4)\} \end{aligned}$$

и искомое регулярное выражение (27). Следствие доказано.

Замечание 2. Формула (27), использующая индексы $n-2$ и $n-1$, выглядит более компактной по сравнению с формулой (23). По сути же регулярное выражение $r'_{n\setminus 1}$ из-за дополнительного шага рекурсии сложнее, чем $r_{n\setminus 1}$. Именно потому в примере 4 и в доказательстве теоремы 2 состояние q'_3 удаляется до состояний q'_8 и q'_7 , хотя шаг рекурсии предполагает обратное.

Замечание 3. Для высоты итерации языка $L_n \setminus L_m$ ($n > m$) выполняется равенство $\text{sh}(L_n \setminus L_m) = \text{sh} L_n$, поскольку высоту итерации языка определяет именно n (подобно тому, как на рис. 3 высоту итерации языка определяет именно правая, более высокая колонка).

Отметим, что в работе [15] предложен эвристический алгоритм построения по заданному конечному автомату регулярного выражения, высота итерации которого близка к высоте итерации соответствующего регулярного языка (такое регулярное выражение в работе [15] называется псевдооптимальным).

Замечание 4. Напомним, что языком сети Петри T -типа называют множество таких слов $\sigma(w)$, что после запуска последовательности переходов w ни один переход невозможно запустить [1]. Порождаемый сетью Петри, изображенной на рис. 1, язык T -типа не содержит ни одного слова $L = \emptyset$ как для варианта задачи с ограниченным буфером, так и с неограниченным.

ЛИТЕРАТУРА

1. Дж. Питерсон, *Теория сетей Петри и моделирование систем*. М.: Мир, 1984, 264 с.
2. В.Е. Котов, *Сети Петри*. М.: Наука. Гл. ред. физ.-мат. лит.-ры, 1984, 160 с.
3. Т. Мурата, "Сети Петри: Свойства, анализ, приложения", *ТИИЭР*, т. 77, № 4, с. 41–85, 1989.
4. А.Е. Пентус и М.Р. Пентус, *Теория формальных языков*. М.: Изд-во ЦПИ при механико-математическом ф-те МГУ, 2004, 80 с.
5. А.Е. Пентус и М.Р. Пентус, *Математическая теория формальных языков*. М.: «Бином», 2006, 247 с.
6. С. Гинзбург, *Математическая теория контекстно-свободных языков*. М.: Мир, 1970, 326 с.
7. Дж. Э. Хопкрофт, Р. Мотвани, и Дж. Д. Ульман, *Введение в теорию автоматов, языков и вычислений*. М.: Вильямс, 2002, 528 с.
8. В.М. Статкевич, "Регулярные выражения для языков сети Петри «производитель/потребитель» с ограниченным буфером размера 1 и 2", на *Системный анализ и информационные технологии, материалы 19-й Международной научно-технической конференции SAIT 2017*, Киев, 22–25 мая 2017, с. 123.
9. V. Statkevych, "On regular expressions for producer/consumer Petri net languages with bounded buffer", in *Nonlinear Analysis and Applications: Materials of 4th International scientific conference on memory of corresponding member of National Academy of Science of Ukraine V.S. Mel'nik*, Kyiv, 4–6 April, 2018, p. 68.
10. V. Mukhin and V. Statkevych, "On one context-free language for producer/consumer Petri net with the unbounded buffer", in *15th International conference on Development and Application Systems*, Suceava, Romania, 21–23 May, 2020, pp. 137–140.

11. М.З. Згуровский и В.А. Денисенко, *Дискретно-непрерывные системы с управляемой структурой. Теория, моделирование, применение*. К.: Наукова думка, 1998, 351 с.
12. А. Саломая, *Жемчужины теории формальных языков*. М.: Мир, 1986, 159 с.
13. R. Valk and G. Vidal-Naquet, "Petri nets and regular languages", *Journal of Computer and System Sciences*, **23**, pp. 299–325, 1981.
14. A. Gronewold and H. Fleischhack, "Computing Petri net languages by reductions", in *Fundamentals of computation theory: 10th International conference; proceedings / FCT'95*, Drezden, Germany, August 22–25, 1995, Springer, pp. 253–262.
15. С.В. Баумгертнер и Б.Ф. Мельников, "Мультиэвристический подход к проблеме звездно-высотной минимизации недетерминированных конечных автоматов", *Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии*, № 1, с. 5–7, 2010.

Поступила 11.08.2020

INFORMATION ON THE ARTICLE

Vitalii M. Statkevych, ORCID: 0000-0001-5210-9890, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: mstatkevich@yahoo.com

REGULAR EXPRESSIONS FOR SOME PETRI NET LANGUAGES FOR THE PRODUCER/CONSUMER PROBLEM / V.M. Statkevych

Abstract. We consider a Petri net for the producer/consumer problem (one of the classical synchronization problems) with the bounded buffer of size n and the regular formal languages L_n , generated by the net. We propose regular expressions denoting these languages in the recursive form, in case of the bounded buffer of size from 1 to 3 the explicit formulas are proposed. We transform a reachability graph into a finite automaton and use the state elimination method. We give an upper estimate for the star-height of the mentioned languages, in case of the bounded buffer of size 1 and 2 the exact values are calculated. We also consider union, intersection, Kleene closure, concatenation and set difference operations on mentioned languages. We give the finite automaton and propose regular expressions denoting the set difference of languages $L_n \setminus L_1$ in the recursive form, for $L_2 \setminus L_1$ the explicit formula is proposed.

Keywords: Petri net, producer/consumer problem, Petri net language, formal language, regular language, finite automaton, regular expression, star-height.

РЕГУЛЯРНІ ВИРАЗИ ДЛЯ ДЕЯКИХ МОВ МЕРЕЖ ПЕТРІ В ЗАДАЧІ ПРО ПОСТАЧАЛЬНИКА ТА СПОЖИВАЧА / В.М. Статкевич

Анотація. Розглянуто мережу Петрі в задачі про постачальника та споживача (одній з класичних задач синхронізації) з обмеженим буфером розміру n і регулярні формальні мови L_n , які вона породжує. Для цих мов знайдено регулярні вирази в рекурсивному вигляді, а у випадках обмеженого буфера розміру від 1 до 3 — у вигляді явних формул. За графом досяжності побудовано скінченний автомат, застосовано метод послідовного видалення вершин. Для висоти ітерації (зіркової висоти) вказаних мов надано оцінку зверху, а у випадках обмеженого буфера розміру 1 та 2 знайдено точні значення. Для вказаних мов розглянуто операції об'єднання, перетину, замикання Кліні, конкатенації та різниці. Для різниці мов $L_n \setminus L_1$ побудовано скінченний автомат і знайдено регулярні вирази в рекурсивному вигляді, а для різниці $L_2 \setminus L_1$ — у вигляді явної формули.

Ключові слова: мережа Петрі, задача про постачальника та споживача, мова мережі Петрі, формальна мова, регулярна мова, скінченний автомат, регулярний вираз, висота ітерації (зіркова висота).

REFERENCES

1. J. Peterson, *Petri net theory and modeling of systems*. Moscow: Mir, 1984.
2. V.E. Kotov, *Petri nets*. Moscow: Nauka, 1984.
3. T. Murata, "Petri nets: Properties, analysis and applications", *Trudy Instituta inzhenerov po elektrotekhnike i radioelektronike*, vol. 77, no. 4, pp. 41–85, 1989.
4. A.E. Pentus and M.R. Pentus, *Formal language theory*. Moscow: Mekhaniko-matematicheskii fakul'tet Moskovskogo Gosudarstvennogo Universiteta, 2004.
5. A.E. Pentus and M.R. Pentus, *The mathematical theory of formal languages*. Moscow: Binom, 2006.
6. S. Ginsburg, *The mathematical theory of context-free languages*. Moscow: Mir, 1970.
7. J.E. Hopcroft, R. Motwani and J.D. Ullman, *Introduction to automata theory, languages, and computation*, 2nd ed. Moscow: Williams Publishing House, 2002.
8. V.M. Statkevych, "Regular expressions for producer/consumer Petri net languages with bounded buffer of size 1 and 2", in *19th Int. Conf. on System Analysis and Information Technology (SAIT)*, Kyiv, Educational and Scientific Complex "Institute for Applied System Analysis" of National Technical University of Ukraine "Kyiv Polytechnic Institute", 22–25 May 2017, pp. 123.
9. V. Statkevych, "On regular expressions for producer/consumer Petri net languages with bounded buffer", in *4th Int. Scientific Conf. Nonlinear Analysis and Applications on memory of corresponding member of National Academy of Science of Ukraine V.S. Mel'nik*, Kyiv, National Technical University of Ukraine "Kyiv Polytechnic Institute", 4–6 Apr. 2018, pp. 68.
10. V. Mukhin and V. Statkevych, "On one context-free language for producer/consumer Petri net with the unbounded buffer", in *15th Int. Conf. on Development and Application Systems (DAS)*, Suceava, Romania, 21–23 May 2020, pp. 137–140. doi: 10.1109/DAS49615.2020.9108948
11. M.Z. Zgurovsky and V.A. Denisenko, *Discrete-continuous systems with controlled structure. Theory, modeling, applications*. Kiev: Naukova Dumka, 1998.
12. A. Salomaa, *Jewels of formal language theory*. Moscow: Mir, 1986.
13. R. Valk and G. Vidal-Naquet, "Petri nets and regular languages", *Journal of Computer and System Sciences*, vol. 23, issue 3, pp. 299–325, 1981.
14. A. Gronewold and H. Fleischhack, "Computing Petri net languages by reductions", in *10th Int. Conf. Fundamentals of computation theory (FCT)*, Drezden, Germany, Springer, 22–25 Aug. 1995, pp. 253–262. doi: 10.1007/3-540-60249-6_57
15. S.V. Baumgertner and B.F. Melnikov, "Multi-heuristic approach for the star-height minimization of non-deterministic finite automata", *Proceedings of Voronezh State University. Series: Systems analysis and information technologies*, no. 1, pp. 5–7, 2010.

SIMULATING THE ROTATION OF A BLACK HOLE AND ANTIGRAVITY

Y. MATSUKI, P.I. BIDYUK

Abstract. In this article we show that rotation of a black hole can create antigravity and anti-gravitational waves, given that there is a strong gravity in the black hole, which distorts time and space. At first, we derived the curvature tensors upon Einstein's field equation, using spherical polar coordinates, and then calculated the coefficients of the curvature tensors to simulate the strength of each component of the tensors. It is assumed that the stress-energy tensor, which is located outside of the black hole, can reflect the strength of the gravitational field and the gravitational waves. As the result, we concluded that, if the time and space are distorted in the black hole, the rotation can create antigravity and the anti-gravitational waves. In addition, the result of the simulation shows that the antigravity positively contributes to the stress-energy tensor, which may expand the size of the Universe.

Keywords: antigravity, curvature tensor, stress-energy tensor, Einstein's field equation.

INTRODUCTION (RESEARCH QUESTION)

In our previous two researches [1, 2], we reported as follows: the negative flow of gravitational waves (anti-gravitational waves) must be described by the expression: $-g^{\mu\nu}g_{\rho\sigma,\mu\nu} = 0$, while Dirac [3] predicted that, $g^{\mu\nu}g_{\rho\sigma,\mu\nu} = 0$, describes the gravitational waves. This means that the negative waves move backward from the direction of the positive flow of the waves. Usually the positive flow and the negative flow should be balanced; therefore, neither of the positive flow nor negative flow of gravitational waves is observable. However, when a star moves, the movement of the mass of the star breaks the balance; and then gravitational waves of both positive and negative flows appear [1]. Upon this conclusion, we investigated the curvature tensors of gravitational waves that are emitted from a black hole and found that the tensors of the gravitational waves from a black hole share the same mathematical forms with the tensors of gravitational field of the black hole [2].

And then, we made the next research to investigate the effect of rotation of the black hole, assuming that the rotation of the black hole breaks the balance of positive and negative flows so that anti-gravitational waves would appear. We also examined, whether or not, the antigravity appears when the black hole rotates and this antigravity creates the energy that may expand size of the Universe to larger scale. This article reports the results of these investigations.

CURVATURE TENSORS FOR SIMULATION

Gravitational field

According to Einstein and Dirac [3], the gravitational field is described by the curvature tensors:

$$R_{\mu\nu} = \Gamma_{\mu\alpha,\nu}^{\alpha} - \Gamma_{\mu\nu,\alpha}^{\alpha} - \Gamma_{\mu\nu}^{\alpha}\Gamma_{\alpha\beta}^{\beta} + \Gamma_{\mu\beta}^{\alpha}\Gamma_{\nu\alpha}^{\beta},$$

where

$$\Gamma_{\mu\nu}^{\lambda} = g^{\lambda\alpha}\Gamma_{\alpha\mu\nu} = \frac{1}{2}g^{\lambda\alpha}(g_{\alpha\mu,\nu} + g_{\alpha\nu,\mu} - g_{\mu\nu,\alpha}). \quad (1)$$

Here, $g^{\lambda\alpha}$, are the fundamental tensors that describe the curvature of the 4-dimensional space in spherical polar coordinates, which is diagonal and symmetric as shown below:

$$g^{\lambda\alpha} = \begin{bmatrix} g^{00} & g^{01} & g^{02} & g^{03} \\ g^{10} & g^{11} & g^{12} & g^{13} \\ g^{20} & g^{21} & g^{22} & g^{23} \\ g^{30} & g^{31} & g^{32} & g^{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -\frac{2m}{\mu(\rho-\tau)^{\frac{2}{3}}} & 0 & 0 \\ 0 & 0 & -\mu^2(\rho-\tau)^{\frac{4}{3}} & 0 \\ 0 & 0 & 0 & -\mu^2(\rho-\tau)^{\frac{4}{3}}\sin^2\theta \end{bmatrix}.$$

And it makes the geodesic of the kind:

$$\Gamma_{\mu\nu}^{\alpha} = g^{\alpha\alpha}\Gamma_{\alpha\mu\nu} = \frac{1}{2}g^{\alpha\alpha}(g_{\alpha\mu,\nu} + g_{\alpha\nu,\mu} - g_{\mu\nu,\alpha}).$$

Therefore, the equation (1) becomes like this: $\Gamma_{\mu\nu}^{\alpha} = g^{\alpha\alpha}\Gamma_{\alpha\mu\nu} = \frac{1}{2}g^{\alpha\alpha}(g_{\alpha\mu,\nu} + g_{\alpha\nu,\mu} - g_{\mu\nu,\alpha})$, where $g_{\alpha\mu,\nu} = \frac{\partial g_{\alpha\mu}}{\partial x_{\nu}}$, and x_{ν} , is the vector in ν -th coordinate.

Then, we derived all the components of $R_{\mu\nu}$, and then according to Einstein's rule ($R_{\mu\nu} = \sum_{\mu\nu} R_{\mu\nu}$), summated them to obtain:

$$R_{00} = \frac{-11}{9(\rho-\tau)^2}; R_{01} = R_{10} = \frac{-4}{9(\rho-\tau)^2}, R_{11} = \frac{20}{3(\rho-\tau)^2} + \frac{11\mu}{18m(\rho-\tau)^{4/3}};$$

$$R_{22} = \frac{28}{9\mu^2(\rho-\tau)^{10/3}} + \frac{140m}{9\mu^2(\rho-\tau)^4} + \frac{4}{\sin^2\theta} + \cot^2\theta;$$

$$R_{33} = \frac{-28}{9\mu^2(\rho - \tau)^{\frac{10}{3}} \sin^2 \theta} + \frac{140m}{9\mu^3(\rho - \tau)^4 \sin^2 \theta} + \frac{4}{\sin^2 \theta} + \frac{11 \cot^2 \theta}{\sin^2 \theta}.$$

Here, μ and m are constants, where, $\mu = (3/2\sqrt{2m})^{2/3}$. All other, $R_{\mu\nu} = 0$. The non-diagonal components, R_{01} and R_{10} , appear because τ and ρ are not independent. However, in this research only the space components of the curvature tensors, R_{11} , R_{22} , and R_{33} , are considered.

Gravitational waves

The curvature tensors of gravitational waves, which penetrate the boundary of a black hole [2], are:

$$\begin{aligned} &g^{\zeta\eta} g_{\rho\sigma, \zeta\eta} + g_{,\sigma}^{\zeta\eta} (g_{\rho\zeta, \eta} - \frac{1}{2} g_{\zeta\eta, \rho}) + g_{,\rho}^{\zeta\eta} (g_{\sigma\zeta, \eta} - \frac{1}{2} g_{\zeta\eta, \sigma}) + \\ &+ \frac{1}{2} g^{\rho\zeta} g_{\beta\beta} g_{\rho\zeta, \beta} g_{\eta\zeta, \beta\sigma} + \frac{1}{2} g^{\rho\zeta} g_{\beta\beta} g_{\rho\zeta, \beta\sigma} g_{\eta\zeta, \beta} + \\ &+ \frac{1}{2} g_{,\sigma}^{\rho\zeta} g_{\beta\beta} g_{\rho\zeta, \beta} g_{\eta\zeta, \beta} + \frac{1}{2} g^{\rho\zeta} g_{\beta\beta, \sigma} g_{\rho\zeta, \beta} g_{\eta\zeta, \beta} + \\ &+ \frac{1}{2} g^{\sigma\zeta} g_{\beta\beta} g_{\sigma\zeta, \beta} g_{\zeta\eta, \beta\rho} + \frac{1}{2} g^{\sigma\zeta} g_{\beta\beta} g_{\sigma\zeta, \beta\rho} g_{\zeta\eta, \beta} + \\ &+ \frac{1}{2} g_{,\rho}^{\sigma\zeta} g_{\beta\beta} g_{\sigma\zeta, \beta} g_{\eta\zeta, \beta} + \frac{1}{2} g^{\sigma\zeta} g_{\beta\beta, \rho} g_{\sigma\zeta, \beta} g_{\eta\zeta, \beta}. \end{aligned}$$

$$\text{When } \zeta = \eta = 0 : \frac{4\mu}{9m(\rho - \tau)^{4/3}} - \frac{112}{9\mu^2(\rho - \tau)^{10/3}} - \frac{112}{9\mu^2(\rho - \tau)^{10/3} \sin^2 \theta};$$

$$\begin{aligned} \text{When } \zeta = \eta = 1 : &\frac{-16}{9(\rho - \tau)^2} - \frac{2}{3(\rho - \tau)^2} - \frac{2\mu}{81m(\rho - \tau)^{10/3}} + \frac{\mu}{9m(\rho - \tau)^{7/3}} - \\ & - \frac{\mu}{9m(\rho - \tau)^{4/3}} - \frac{2}{(\rho - \tau)^2} + \frac{224m}{9\mu^3(\rho - \tau)^4} + \frac{224m}{9\mu^3(\rho - \tau)^4 \sin^2 \theta}; \end{aligned}$$

$$\text{When } \zeta = \eta = 2 : \frac{24 \cos \theta}{\sin^4 \theta} + \frac{8}{\sin^2 \theta}, \text{ when, } \zeta = \eta = 3 : \frac{64}{9(\rho - \tau)^2} - 32 \cot^2 \theta.$$

Distortion of time and space in strong gravity

Using the curvature tensors, Dirac [3] invented a coordinate system that describes the gravitational field from the center of strong gravity in a black hole, in which time and space are distorted by affecting each other. He suggested that if we travel toward the center of the strong gravity, it takes infinite time to reach the center. Upon this Dirac’s prescription, we assumed that the time and the distance between the center of the gravitational field and the edge of the Universe are as shown in Fig. 1 and Fig. 2.

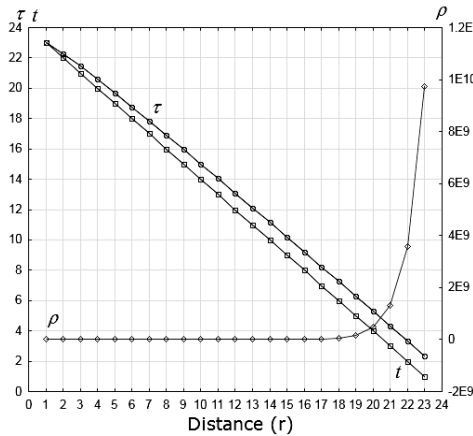


Fig. 1. Time and distance from the center of the gravitational field, Case-1 (non-linear distortion): $f(r) = \log r$ and $g(r) = e^r$

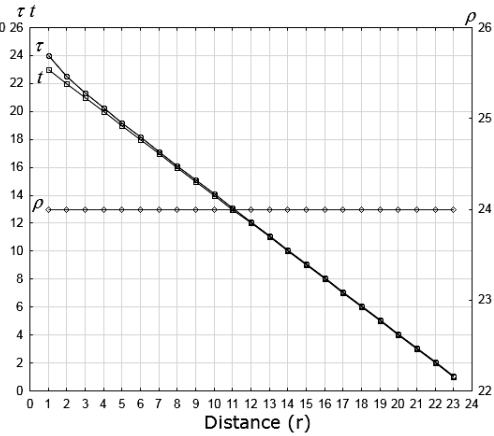


Fig. 2. Time and distance from the center of the gravitational field, Case-2 (linear distortion): $f(r) = (1/4)$ and $g(r) = r$

In these figures, τ is a relative time in the coordinate system, which expands and shrinks depending on the distance r , where $\tau = t + f(r)$; and ρ is the relative distance, which expands and shrinks depending on the time t , where $\rho = t + g(r)$; and $f(r)$, and $g(r)$ are functions of r . For the simulation, we assumed Case 1: $f(r) = \log r$, and, $g(r) = e^r$; and Case 2: $f(r) = \frac{1}{4}$, and $g(r) = r$.

Note: r is the distance from the center of strong gravity, t is the time to travel on the distance, f and g are given functions, and $\tau = t + f(r)$; and $\rho = t + g(r)$.

ALGORITHM

Einstein's field equation [3] that rules the motion of particles in the gravitational field is as follows: $(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R)_{,\nu} = 0$. Then, $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = kT$, where T is the stress-energy tensor and k is a constant [4]. Then, we propose the following algorithm to calculate the relative intensities of the components of curvature tensors:

$$H = kT - R_{\mu\nu} = kT - (c_1X_1 + c_2X_2 + \dots + c_lX_l),$$

and

$$H^2 = \{kT - (c_1X_1 + c_2X_2 + \dots + c_lX_l)\}^2,$$

where c_1, \dots, c_l are the coefficients, that create a column vector, c . And, $X = [X_1 \ X_2 \ \dots \ X_l]$, then $H = kT - Xc$. Then we set the constraint, $X'H = 0$, then $X'(kT - Xc) = 0$, where X' is transpose matrix of X .

Then, $X'Xc = X'kT$, $c = (X'X)^{-1}X'kT$, and $\Sigma = V(c) = \sigma^2(X'X)^{-1}$, where $V(c) = \sigma^2$ is the variance of the c , and $\sigma^2 = e'e/(n-l)$, where $e = MkT$, $M = I - X(X'X)^{-1}X'$; n is the number of rows of each column of X (in this

simulation $n = 23$); l is the number of columns of X ; I is a 23×23 unit matrix that holds 1 on all diagonal elements and 0 for the other elements; $(X'X)^{-1}$ is the inverse matrix of $X'X$; and e' is the transpose vector of e . By calculating c and $V(c)$, we estimated the relative strength of each component of, $R_{\mu\nu}$, to the stress-energy tensor in the system of spherical polar coordinates.

Rotation of the object that contains strong gravity

When an object rotates as shown in Fig. 3, its coordinate system will be transformed by the transformation matrix D of the Euler's angles [4]. For the rotation around one axis, the tensors of the object's coordinate system will be multiplied

$$\text{by: } D = \begin{bmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

And then the curvature tensor will be transformed to the following form:

$$DR_{\mu\nu} = \begin{bmatrix} \cos \varphi & \sin \varphi & 0 \\ -\sin \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} R_{11} & 0 & 0 \\ 0 & R_{22} & 0 \\ 0 & 0 & R_{33} \end{bmatrix} =$$

$$= \begin{bmatrix} \cos \varphi R_{11} & \sin \varphi R_{22} & 0 \\ -\sin \varphi R_{11} & \cos \varphi R_{22} & 0 \\ 0 & 0 & R_{33} \end{bmatrix}.$$

The components of $R_{\mu\nu}$ before and after the rotation are shown in Tables 1 and 2.

For the simulation, we used the components, $\cos \varphi R_{11}$, $\cos \varphi R_{22}$ and R_{33} , which correspond to the coordinates that describe the space coordinates, ρ , θ and φ . The components of R_{33} doesn't change by the rotation, under the operation of $DR_{\mu\nu}$, because $D_{33} = 1$. We selected these three diagonal components for calculating the coefficients of the curvature tensors with the algorithm mentioned above, which simulates the relative strength of each components of the curvature tensor to the stress-energy tensor. However, we

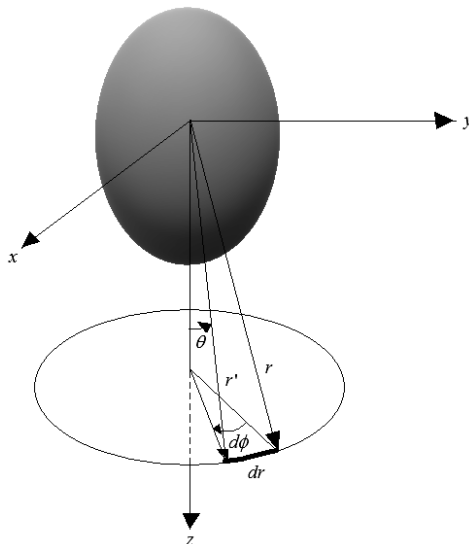


Fig. 3. Rotation of an object

didn't use the non-diagonal components, $-\sin \varphi R_{11}$ and $\sin \varphi R_{22}$, because these are perpendicular to the diagonal components, therefore do not contribute to the stress-energy tensor.

SIMULATION

Input data

Time t is set as shown in Fig. 1 for Case-1, and in Fig. 2 for Case-2, with which its slope to the distance, r , from the center of the gravitational field is a constant. For simulating the spatial expansion of the gravitational field, we assumed as if θ becomes larger in far distance. On the other hand, for simulating the flow of gravitational waves, we assumed that θ becomes smaller in far distance, as shown in Fig. 4. For simulating the rotation of the object, we set two cases, assuming φ_1 (Rotation1) and φ_2 (Rotation 2) also as shown in Fig.4. With these settings, $\sin \theta$, $\cos \theta$, $\cot \theta$, and $\cos \varphi$ of the gravitational field behave like as shown in Fig. 5.

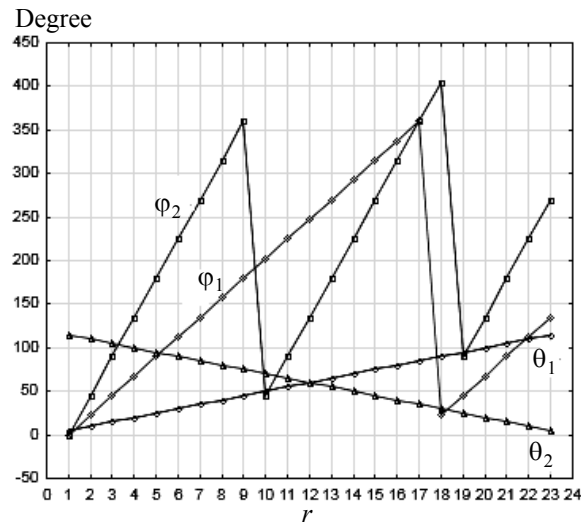


Fig. 4. Angles θ and φ for simulating gravitational field and gravitational waves

Note: θ_1 : for gravitational field, θ_2 : for gravitational waves, φ_1 : for the rotation 1, φ_2 : for the rotation 2

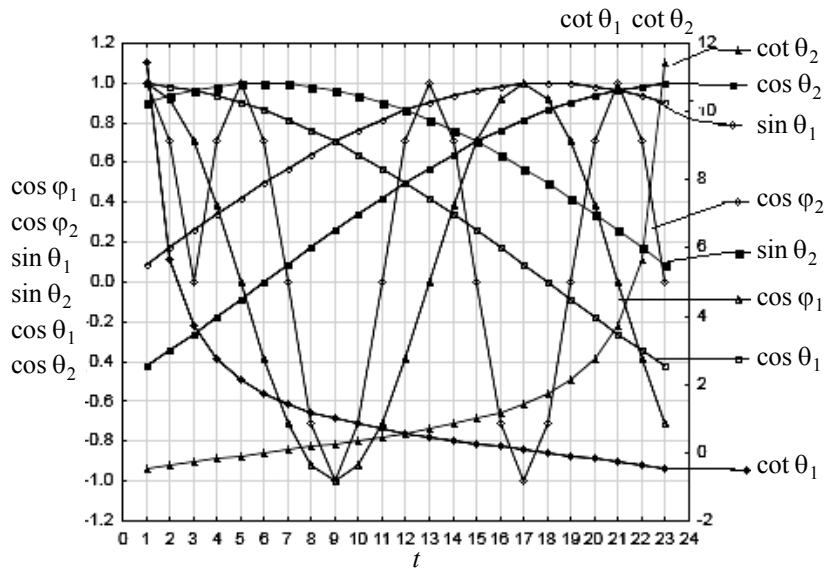


Fig. 5. $\sin \theta$, $\cos \theta$, $\cot \theta$, and $\cos \varphi$ of the simulated gravitational field

In addition, for this simulation, we set the stress-energy tensor kT to be 1; because, the purpose of this simulation is to measure the order of magnitude of the relative strength of each component of, $R_{\mu\nu}$, to the stress-energy tensor.

RESULTS

Gravitational field

The results of simulation for the gravitational field are shown in Table 1 for Case-1, and Table 2 for Case-2. The both Tables show the calculated coefficients of the simulation with no rotation, with the rotation 1 and the rotation 2.

Table 1. Results of the simulation of gravitational field, Case-1

Components of $R_{\mu\nu}$	c and $\sqrt{V(c)}$ of $R_{\mu\nu}$ before the rotation	Components of $DR_{\mu\nu}$	c and $\sqrt{V(c)}$ (Rotation 1)	c and $\sqrt{V(c)}$ (Rotation 2)
$\frac{1}{(\rho - \tau)^2}$	$2,902 \cdot 10^{-2}$ ($1,875 \cdot 10^4$)	$\frac{\cos \varphi}{(\rho - \tau)^2}$	$-4,255 \cdot 10^5$ ($5,078 i \cdot 10^5$)	$1,783 \cdot 10^5$ ($7,302 \cdot 10^5$)
$\frac{1}{(\rho - \tau)^{4/3}}$	$3,496 \cdot 10^{-3}$ ($2,573 \cdot 10^3$)	$\frac{\cos \varphi}{(\rho - \tau)^{4/3}}$	$1,390 \cdot 10^4$ ($9,997 \cdot 10^3$)	$-6,495 \cdot 10^3$ ($2,272 \cdot 10^4$)
$\frac{1}{(\rho - \tau)^{10/3}}$	$-1,488 \cdot 10^2$ ($1,064 \cdot 10^7$)	$\frac{\cos \varphi}{(\rho - \tau)^{10/3}}$	$-2,768 \cdot 10^7$ ($1,608i \cdot 10^8$)	$-2,996 \cdot 10^7$ ($1,004 \cdot 10^8$)
$\frac{1}{(\rho - \tau)^4}$	$-2,623 \cdot 10^3$ ($1,808 i \cdot 10^8$)	$\frac{\cos \varphi}{(\rho - \tau)^4}$	$-2,676 \cdot 10^8$ ($8,577 i \cdot 10^8$)	$1,351 \cdot 10^8$ ($4,202 \cdot 10^8$)
$\frac{1}{\sin^2 \theta}$	1,000 ($6,252 \cdot 10^{-2}$)	$\frac{\cos \varphi}{\sin^2 \theta}$	0,7787 (0,3866)	0,2913 (0,4413)
$\cot^2 \theta$	-1,000 (0,5245)	$\cos \varphi \cdot \cot^2 \theta$	-2,788 (0,7892)	-0,7126 (1,432)
$\frac{1}{(\rho - \tau)^4 \sin^2 \theta}$	-55,86 ($3,877 i \cdot 10^6$)	$\frac{1}{(\rho - \tau)^4 \sin^2 \theta} *$	$-6,454 \cdot 10^6$ ($2,594 i \cdot 10^7$)	$-6,129 \cdot 10^5$ ($2,190 \cdot 10^6$)
$-\frac{1}{(\rho - \tau)^{10/3} \sin^2 \theta}$	-38,63 ($2,688 i \cdot 10^6$)	$-\frac{1}{(\rho - \tau)^{10/3} \sin^2 \theta} *$	$-4,436 \cdot 10^6$ ($1,736 i \cdot 10^7$)	$-1,686 \cdot 10^4$ ($2,473 \cdot 10^5$)
$\frac{\cot^2 \theta}{\sin^2 \theta}$	$-2,773 \cdot 10^{-7}$ (0,2009)	$\frac{\cot^2 \theta}{\sin^2 \theta} *$	$2,057 \cdot 10^{-2}$ (0,2481 i)	0,2367 (0,1400)
The values in the brackets are $\sqrt{V(c)}$. For example, $1,808 i \cdot 10^8 = \sqrt{-3.27 \cdot 10^{16}}$. * This component corresponds to the coordinate of the axis of the rotation, therefore $\cos \varphi$ is not multiplied.				

In Case-1 (non-linear distortion of the time and space), the coefficient, c , of $\frac{\cot^2 \theta}{\sin^2 \theta}$ changes its sign from minus to plus after the rotation of φ_1 (the Rotation 1) and of φ_2 (the Rotation 2). The gravity must be negative, and it is so to the stress-energy tensor when it doesn't rotate, but it becomes positive to the stress-energy tensor after the rotations. This result means that the antigravity appears after the rotation.

Table 2. Results of the simulation of gravitational field, Case-2

Components of $R_{\mu\nu}$	c and $\sqrt{V(c)}$ of $R_{\mu\nu}$ before the rotation	Components of $DR_{\mu\nu}$	c and $\sqrt{V(c)}$ (Rotation 1)	c and $\sqrt{V(c)}$ (Rotation 2)
$\frac{1}{(\rho - \tau)^2}$	$-8,518 \cdot 10^{-3}$ ($1,896 \cdot 10^{-2}$)	$\frac{\cos \varphi}{(\rho - \tau)^2}$	$1,278 \cdot 10^4$ ($4,437 \cdot 10^3$)	$5,473 \cdot 10^3$ ($3,900 \cdot 10^3$)
$\frac{1}{(\rho - \tau)^{4/3}}$	$1,217 \cdot 10^{-3}$ ($2,820 \cdot 10^{-3}$)	$\frac{\cos \varphi}{(\rho - \tau)^{4/3}}$	$-2,182 \cdot 10^3$ ($7,573 \cdot 10^2$)	$-9,831 \cdot 10^2$ ($6,530 \cdot 10^2$)
$\frac{1}{(\rho - \tau)^{10/3}}$	$0,1086$ ($0,2162$)	$\frac{\cos \varphi}{(\rho - \tau)^{10/3}}$	$-6,724 \cdot 10^4$ ($2,833 \cdot 10^4$)	$-3,707 \cdot 10^4$ ($2,353 \cdot 10^4$)
$\frac{1}{(\rho - \tau)^4}$	$-0,2701$ ($0,5121$)	$\frac{\cos \varphi}{(\rho - \tau)^4}$	$9,317 \cdot 10^4$ ($4,145 \cdot 10^4$)	$4,968 \cdot 10^4$ ($3,173 \cdot 10^4$)
$\frac{1}{\sin^2 \theta}$	$1,000$ ($1,864 \cdot 10^{-5}$)	$\frac{\cos \varphi}{\sin^2 \theta}$	$16,76$ ($5,815$)	$8,595$ ($4,914$)
$\cot^2 \theta$	$-1,000$ ($2,679 \cdot 10^{-5}$)	$\cos \varphi \cdot \cot^2 \theta$	$-33,90$ ($9,405$)	$-9,779$ ($9,921$)
$\frac{1}{(\rho - \tau)^4 \sin^2 \theta}$	$-2,229 \cdot 10^{-4}$ ($4,083 \cdot 10^{-4}$)	$\frac{1}{(\rho - \tau)^4 \sin^2 \theta} *$	$-1,052 \cdot 10^{-3}$ ($5,370 \cdot 10^2$)	$-6,204 \cdot 10^2$ ($4,183 \cdot 10^2$)
$-\frac{1}{(\rho - \tau)^{10/3} \sin^2 \theta}$	$-3,371 \cdot 10^{-3}$ ($5,829 \cdot 10^{-3}$)	$-\frac{1}{(\rho - \tau)^{10/3} \sin^2 \theta} *$	$-2,856 \cdot 10^2$ ($1,743 \cdot 10^2$)	$-3,109 \cdot 10^2$ ($2,104 \cdot 10^2$)
$\frac{\cot^2 \theta}{\sin^2 \theta}$	$-1,707 \cdot 10^{-8}$ ($6,560 \cdot 10^{-8}$)	$\frac{\cot^2 \theta}{\sin^2 \theta} *$	$0,1293$ ($3,106 \cdot 10^2$)	$8,552 \cdot 10^{-3}$ ($4,746 \cdot 10^2$)
The values in the brackets are $\sqrt{V(c)}$. For example, $1,808 \cdot 10^8 = \sqrt{-3,27 \cdot 10^{16}}$. * This component corresponds to the coordinate of the axis of the rotation, therefore $\cos \varphi$ is not multiplied.				

In addition, the coefficients of $\frac{1}{(\rho - \tau)^4}$ and $\frac{\cot^2 \theta}{\sin^2 \theta}$ change these signs from minus to plus, only for the rotation of φ_2 in Case-1. And in Case-2 (linear distortion)

tion of time and space), only the coefficients of $\frac{1}{(\rho - \tau)^2}$, $\frac{1}{(\rho - \tau)^4}$, and $\frac{\cot^2 \theta}{\sin^2 \theta}$, change these signs from minus to plus after the both rotations of, φ_1 , and φ_2 .

The summation of each of the positive coefficients and the negative coefficients are shown in Table 5, and in Fig. 6 for Case-1, and Fig. 7 for Case-2. In case of non-linear distortion of time and space (Case-1), antigravity appears after the Rotation 2 in Case-1; while in case of linear distortion of time and space (Case-2), the antigravity appears after both of the Rotation 1 and 2.

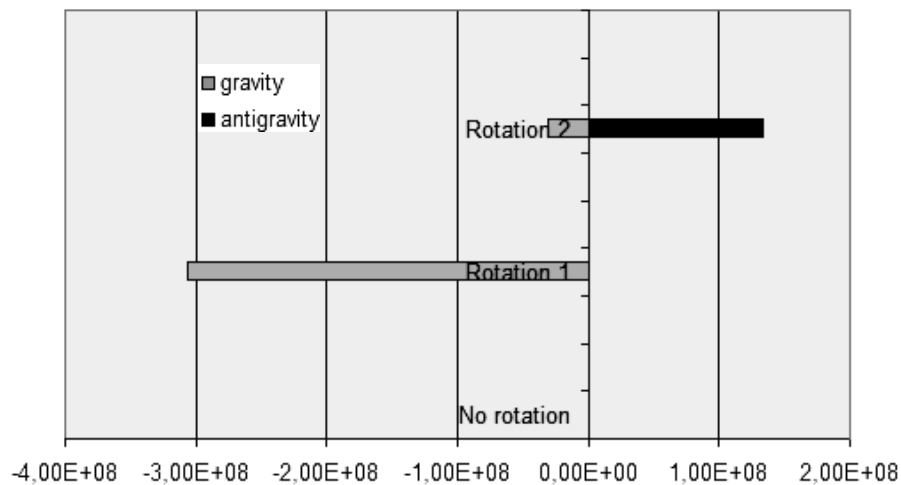


Fig. 6. Gravity and antigravity (Case – 1: non-linear distortion of time and space)

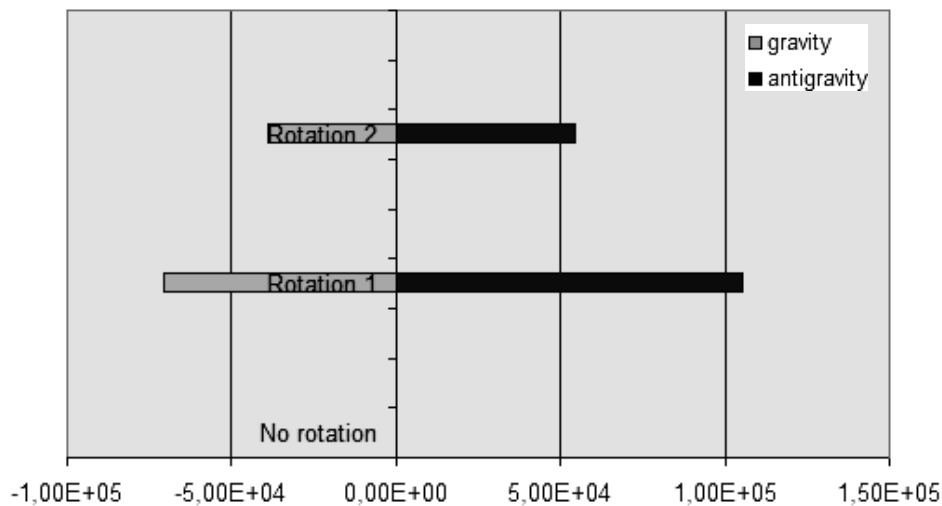


Fig. 7. Gravity and antigravity (Case – 2: linear distortion of time and space)

Gravitational waves

The results of simulation for the gravitational waves are shown in Table 3 for Case-1 and Table 4 for Case-2. For the gravitational waves the coefficients of the

curvature tensors are positive to the stress-energy tensor; while the negative coefficients represent the anti-gravitational waves.

Table 3. Results of the simulation of gravitational waves, Case-1

Components of gravitational waves before the rotation	c and $\sqrt{V(c)}$ of gravitational waves	Components of gravitational waves after the rotation	c and $\sqrt{V(c)}$ Rotation 1	c and $\sqrt{V(c)}$ Rotation 2
$\frac{1}{(\rho - \tau)^2}$	1,200 · 10 ⁴ (6,038 · 10 ⁻⁸)	$\frac{\cos \varphi}{(\rho - \tau)^2}$	1,185 · 10 ⁶ (3,619 · 10 ¹⁰)	-1,210 · 10 ⁶ (3,772 i · 10 ⁹)
$\frac{1}{(\rho - \tau)^{4/3}}$	-84,48 (5,520 · 10 ⁶)	$\frac{\cos \varphi}{(\rho - \tau)^{4/3}}$	-1,723 · 10 ⁴ (7,669)	9,443 · 10 ³ (8,600 i · 10 ⁶)
$\frac{1}{(\rho - \tau)^4}$	1,360 · 10 ⁷ (9,263 i · 10 ¹¹)	$\frac{\cos \varphi}{(\rho - \tau)^4}$	1,450 · 10 ⁹ (6,839 i · 10 ¹³)	9,839 · 10 ⁻⁸ (5,481 i · 10 ¹²)
$\frac{1}{\sin^2 \theta}$	1,001 (1,747 · 10 ²)	$\frac{\cos \varphi}{\sin^2 \theta}$	0,1496 (1,703 · 10 ³)	2,997 · 10 ⁻² (1,422 · 10 ²)
$-\cot^2 \theta$	-1,001 (1,877 · 10 ⁻²)	$-\cot^2 \theta^*$	8,722 · 10 ⁻² (6,337 · 10 ⁻²)	8,884 · 10 ⁻³ (11,21)
$\frac{1}{(\rho - \tau)^4 \sin^2 \theta}$	-6,752 · 10 ⁶ (4,939 i · 10 ¹¹)	$\frac{\cos \varphi}{(\rho - \tau)^4 \sin^2 \theta}$	-7,830 · 10 ⁻⁸ (4,291 i · 10 ¹³)	-8,148 · 10 ⁻⁸ (4,387 i · 10 ¹²)
$-\frac{1}{(\rho - \tau)^{10/3} \sin^2 \theta}$	1,476 · 10 ⁷ (7,651 i · 10 ¹¹)	$-\frac{\cos \varphi}{(\rho - \tau)^{10/3} \sin^2 \theta}$	1,293 · 10 ⁹ (2,133 i · 10 ¹³)	-3,424 · 10 ⁻⁸ (2,597 i · 10 ¹²)
$-\frac{1}{(\rho - \tau)^{10/3}}$	-1,544 · 10 ⁷ (7,783 i · 10 ¹¹)	$-\frac{\cos \varphi}{(\rho - \tau)^{10/3}}$	-1,335 · 10 ⁹ (1,668 i · 10 ¹³)	4,209 · 10 ⁻⁸ (2,913 i · 10 ¹²)
$\frac{1}{(\rho - \tau)^{7/3}}$	-4,509 · 10 ⁴ (1,953 · 10 ⁹)	$\frac{\cos \varphi}{(\rho - \tau)^{7/3}}$	-3,890 · 10 ⁶ (1,437 · 10 ¹¹)	4,601 · 10 ⁶ (1,723 i)
$\frac{\cos \theta}{\sin^4 \theta}$	2,996 · 10 ⁷ (0,1763)	$\frac{\cos \varphi \cdot \cos \theta}{\sin^4 \theta}$	-2,889 · 10 ⁻⁴ (11,36)	1,859 · 10 ⁴ (5,165)
-	-	$\frac{1}{(\rho - \tau)^2}^*$	1,876 · 10 ⁴ (6,619 · 10 ³)	3,842 · 10 ⁻² (5,644 · 10 ⁵)
The values in the brackets are $\sqrt{V(c)}$. For example, 1,808 i · 10 ⁸ = $\sqrt{-3.27 \cdot 10^{16}}$. * This component corresponds to the coordinate of the axis of the rotation, therefore $\cos \varphi$ is not multiplied.				

Table 4. Results of the simulation of gravitational waves, Case-2

Components of gravitational waves before the rotation	c and $\sqrt{V(c)}$ of gravitational waves	Components of gravitational waves after the rotation	c and $\sqrt{V(c)}$ Rotation 1	c and $\sqrt{V(c)}$ Rotation 2
$\frac{1}{(\rho - \tau)^2}$	$2,335 \cdot 10^{-2}$ (21,38)	$\frac{\cos \varphi}{(\rho - \tau)^2}$	$-2,080 \cdot 10^4$ ($7,251 \cdot 10^3$)	$3,964 \cdot 10^3$ ($1,428 \cdot 10^4$)
$\frac{1}{(\rho - \tau)^{4/3}}$	$-1,708 \cdot 10^{-3}$ (1,511)	$\frac{\cos \varphi}{(\rho - \tau)^{4/3}}$	$1,220 \cdot 10^3$ ($4,052 \cdot 10^{-3}$)	$-5,052 \cdot 10^2$ ($9,136 \cdot 10^2$)
$\frac{1}{(\rho - \tau)^4}$	1,217 ($1,140 \cdot 10^3$)	$\frac{\cos \varphi}{(\rho - \tau)^4}$	$-1,463 \cdot 10^6$ ($6,858 \cdot 10^5$)	$3,541 \cdot 10^6$ ($2,825 \cdot 10^6$)
$\frac{1}{\sin^2 \theta}$	1,000 ($9,834 \cdot 10^{-4}$)	$\frac{\cos \varphi}{\sin^2 \theta}$	0,8678 (0,3550)	0,8594 (0,7481)
$\cot^2 \theta$	-1,000 ($2,405 \cdot 10^{-3}$)	$-\cot^2 \theta^*$	$5,531 \cdot 10^{-3}$ ($2,318 \cdot 10^{-3}$)	$8,384 \cdot 10^{-3}$ ($6,984 \cdot 10^{-3}$)
$\frac{1}{(\rho - \tau)^4 \sin^2 \theta}$	-0,9864 ($9,137 \cdot 10^{-2}$)	$\frac{\cos \varphi}{(\rho - \tau)^4 \sin^2 \theta}$	$1,169 \cdot 10^6$ ($5,448 \cdot 10^5$)	$-2,773 \cdot 10^6$ ($2,218 \cdot 10^6$)
$\frac{1}{(\rho - \tau)^{10/3} \sin^2 \theta}$	-0,1682 ($1,523 \cdot 10^2$)	$\frac{\cos \varphi}{(\rho - \tau)^{10/3} \sin^2 \theta}$	$1,568 \cdot 10^5$ ($6,602 \cdot 10^4$)	$-3,016 \cdot 10^5$ ($2,612 \cdot 10^5$)
$-\frac{1}{(\rho - \tau)^{10/3}}$	0,2770 ($2,517 \cdot 10^2$)	$-\frac{\cos \varphi}{(\rho - \tau)^{10/3}}$	$-2,829 \cdot 10^5$ ($1,292 \cdot 10^5$)	$6,841 \cdot 10^5$ ($5,614 \cdot 10^5$)
$\frac{1}{(\rho - \tau)^{7/3}}$	$-2,439 \cdot 10^2$ (22,20)	$\frac{\cos \varphi}{(\rho - \tau)^{7/3}}$	$2,115 \cdot 10^4$ ($8,154 \cdot 10^3$)	$2,338 \cdot 10^4$ ($2,391 \cdot 10^4$)
$\frac{\cos \theta}{\sin^4 \theta}$	$8,679 \cdot 10^{-10}$ ($1,092 \cdot 10^{-6}$)	$\frac{\cos \varphi \cdot \cos \theta}{\sin^4 \theta}$	$3,683 \cdot 10^{-4}$ ($6,707 \cdot 10^{-4}$)	$-1,286 \cdot 10^{-2}$ ($8,520 \cdot 10^{-3}$)
-	-	$\frac{1}{(\rho - \tau)^2}^*$	30,60 (8,602)	11,74 (6,283)

The values in the brackets are $\sqrt{V(c)}$. For example, $1,808 i \cdot 10^8 = \sqrt{-3.27 \cdot 10^{16}}$.
 * This component corresponds to the coordinate of the axis of the rotation, therefore $\cos \varphi$ is not multiplied.

In Case-1 (non-linear distortion of time and space), the coefficient, c, of $\frac{\cos \theta}{\sin^4 \theta}$, changes its sign from plus to minus after the rotation of φ_1 (the Rotation 1), and, $\frac{1}{(\rho - \tau)^2}$, and $-\frac{1}{(\rho - \tau)^{10/3} \sin^2 \theta}$, change these signs from plus to minus after the rotation of φ_2 (the Rotation 2). And in Case-2 (linear distortion of time

and space), the coefficients of $\frac{1}{(\rho - \tau)^2}$, $\frac{1}{(\rho - \tau)^4}$ and $-\frac{1}{(\rho - \tau)^{10/3}}$ change these signs from plus to minus after the rotation of φ_1 (the Rotation 1), while the coefficient of, $\frac{\cos\theta}{\sin^4\theta}$, changes its sign from plus to minus after the rotation of φ_2 (the Rotation 2).

Table 5. Strengths of gravity and antigravity

Case	Case-1		Case-2	
	Gravity	Antigravity	Gravity	Antigravity
No rotation	-2,867·10 ³	1,033	-1,282	1,110
Rotation 1	-3,066·10 ⁸	1,390·10 ⁴	-7,079·10 ⁴	1,060·10 ⁵
Rotation 2	-3,060·10 ⁷	1,353·10 ⁸	-3,899·10 ⁴	5,516·10 ⁴

The summation of each of the positive coefficients and the negative coefficients are shown in Table 6, and in Fig. 8 for Case-1, and Fig. 9 for Case-2. In both cases, gravity and antigravity are balanced without the rotation of the black hole, but the balance is broken after the rotations, then antigravity appears with the Rotation 2 in Case-1, and with both of the Rotation 1 and Rotation 2 in Case-2. Also, gravitational waves and anti-gravitational waves are balanced without the rotation, but they appear when the black hole rotates.

Table 6. Strengths of gravitational waves and anti-gravitational waves

Case	Case - 1		Case - 2	
	Gravitational waves	Anti-gravitational waves	Gravitational waves	Anti-gravitational waves
No rotation	2,838·10 ⁷	-2,224·10 ⁷	2,517	-2,181
Rotation 1	2,744·10 ⁹	-2,122·10 ⁹	1,348·10 ⁶	-1,766·10 ⁶
Rotation 2	1,409·10 ⁹	-1,158·10 ⁹	4,252·10 ⁶	-3,075·10 ⁶

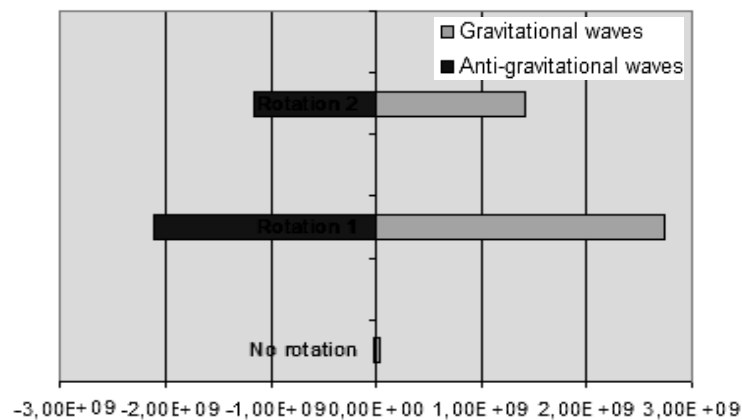


Fig. 8. Gravitational waves and anti-gravitational waves (Case-1: non-linear distortion of time and space)

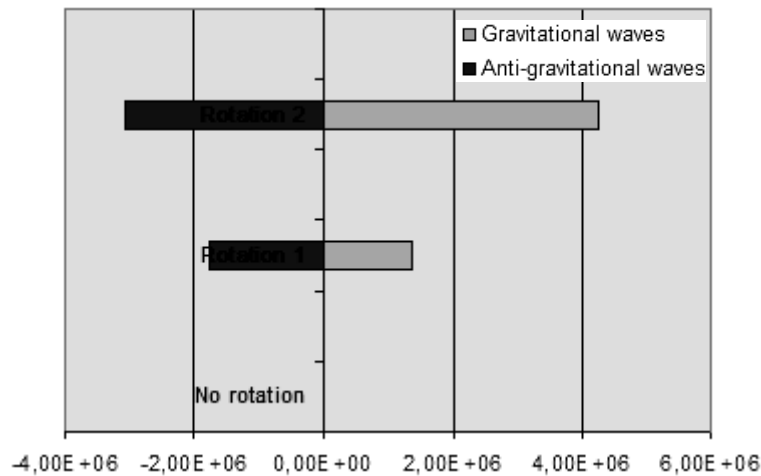


Fig. 9. Gravitational waves and anti-gravitational waves (Case-2: linear distortion of time and space)

PHYSICAL MEANING OF THE RESULT

Here, $R_{\mu\nu} - (1/2)g_{\mu\nu}R + \Lambda g = kT$, is the equation of gravitational field of the Universe [4], where $R_{\mu\nu}$ are curvature tensors, named Ricci tensors, $g_{\mu\nu}$ are fundamental tensors, $R_{\mu\nu} = R$, and, $g_{\mu\nu} = g$, where $\mu = \nu$, T is the stress-energy tensor and k is a constant, and $\Lambda < 0$ is the cosmological constant named “dark energy”, which is a positive contribution to kT . The above result of our simulation shows that the rotation of the black hole makes positive contribution to the stress-energy tensor, which may expand the size of the Universe, however it is unknown if the antigravity is related to the dark energy.

Fig. 6 and Fig.7 show that the gravity and the antigravity are balanced without the rotation, but balance is broken when the black hole rotates. Also, Fig. 8 and Fig. 9 show that the gravitational waves and the anti-gravitational waves are balanced without the rotation, but the balance is broken when the black hole rotates. This finding is consistent with our previous report [1].

CONCLUSIONS AND RECOMMENDATIONS

In this simulation, assuming that the coordinates of time and space can be distorted in the strong gravity in a black hole, we investigated whether a rotation of a black hole can produce antigravity and anti-gravitational waves, or not, by calculating the relative strengths of the components of the curvature tensors of the black hole, which are measured by the stress-energy tensor that is placed outside of the black hole, upon Einstein’s field equation. In order to simulate the curvature in the strong gravity, we used the system of the spherical polar coordinates so that we could simulate rotation of the black hole with Euler’s angles.

The results of the simulation show that the rotating black hole can produce the antigravity and anti-gravitational waves, if time and space are distorted linearly and non-linearly. Also, the results suggest a possible explanation about the expansion of the Universe.

Further investigations are needed about the process of the time-space distortions and of the angular momentum of the rotation.

REFERENCES

1. Y. Matsuki, P.I. Bidyuk, “Analysis of negative flow of gravitational waves (Part 5)”, *System Research & Information Technology*, no. 4, pp. 7–18, 2019.
2. Y. Matsuki, P.I. Bidyuk, “Numerical Simulation of Gravitational Waves from a Black hole, using Curvature Tensors (Part 6)”, *System Research & Information Technology*, no.1, pp. 54–67, 2020.
3. P.A.M. Dirac, *General Theory of Relativity*, Florida University, A Wiley-Interscience Publication, John Wiley & Sons, New York, 1975, pp. 69 .
4. H. Goldstein, C.P. Poole, J.L. Safko, *Classical Mechanics*, 3rd Edition, published by Pearson Education, Inc., 2002, pp. 646; [especially Chapter 4 “The Kinematics of Rigid Body Motion”, pp.134–183, Chapter 7.11 “Introduction to the general theory of relativity”, pp. 324–328].

Received 02.07.2020

From the Editorial Board: the article corresponds completely to submitted manuscript.

INFORMATION ON THE ARTICLE

Petro I. Bidyuk, ORCID: 0000-0002-7421-3565, Educational and Scientific Complex “Institute for Applied System Analysis” of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: pbidyuke_00ukr.net

Yoshio Matsuki, ORCID: 0000-0002-5917-8263, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: matsuki@wdc.org.ua

ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ ОБЕРТАННЯ ЧОРНОЇ ДІРИ ТА АНТИГРАВІТАЦІЇ
/ Й. Мацуки, П.І. Бідюк

Анотація. Показано, що обертання чорної діри може створити антигравітацію та антигравітаційні хвилі за умови, що у чорній дірі існує сильна гравітація, яка викривлює час і простір. Отримано тензори кривизни на підставі рівняння поля Ейнштейна з використанням сферичних полярних координат, розраховано коефіцієнти тензорів для моделювання сили кожного компонента тензорів. Зроблено припущення, що тензор енергії-імпульсу, розміщений за межами чорної діри, може відображати силу гравітаційного поля і гравітаційних хвиль. У результаті сформовано такий висновок: якщо час і простір викривляються у чорній дірі, то обертання може створити антигравітацію та антигравітаційні хвилі. Результат моделювання показав, що антигравітація робить позитивний внесок у тензор енергії імпульсу, що може розширити розмір Всесвіту.

Ключові слова: антигравітація, тензор кривизни, тензор енергії напруження, рівняння Ейнштейна для поля.

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ВРАЩЕНИЯ ЧЕРНОЙ ДЫРЫ И АНТИГРАВИТАЦИИ / Й. Мацуки, П.И. Бидюк

Аннотация. Показано, что вращение черной дыры может создать антигравитацию и антигравитационные волны при условии, что в черной дыре существует сильная гравитация, которая искажает время и пространство. Получены тензоры кривизны на основании уравнения поля Эйнштейна с использованием сферических полярных координат, рассчитаны коэффициенты тензоров для моделирования силы каждого компонента тензоров. Предполагается, что тензор энергии импульса, расположенный за пределами черной дыры, может отражать силу гравитационного поля и гравитационных волн. В результате сформулирован вывод: если время и пространство искривляются в черной дыре, вращение может создать антигравитацию и антигравитационные волны. Результат моделирования показал, что антигравитация делает позитивный вклад в тензор энергии импульса, что может расширить размер Вселенной.

Ключевые слова: антигравитация, тензор кривизны, тензор энергии напряжения, уравнение Эйнштейна для поля.

MODELING OF CONTACT INTERACTION OF A HEATED PLANE RIGID ELLIPTICAL PUNCH WITH A TRANSVERSALLY ISOTROPIC ELASTIC HALF-SPACE

V.S. KIRILYUK, O.I. LEVCHUK, V.V. GAVRILENKO, M.B. VITER

Abstract. On the base of a rigorous mathematical model, the problem of the contact interaction of a heated flat punch of an elliptical section with a transversely isotropic elastic half-space is investigated. It is assumed that the half-space surface is the isotropy plane of a transversely isotropic material, and also that there is a smooth (without friction) contact. Expressions of contact stresses and displacements of a heated flat elliptical punch are found explicitly. In the form of a simple inequality, a condition for separating the elastic material from the surface of a flat elliptical punch is obtained. Numerical calculations are carried out. Contact interaction of a heated flat punch is studied taking into account the separation of material from the punch.

Keywords: mathematical model, contact interaction, elastic half-space, transversally-isotropic material, plane elliptical punch, heating, stress distribution, domain of material separation.

INTRODUCTION

Currently, methods for solving spatial problems of contact interaction for isotropic elastic bodies are quite well developed. Among the papers on this topic, classical monographs [1–5], as well as articles [6–8], can be noted. However, the solution of spatial contact problems for transversely isotropic bodies is associated with significant mathematical difficulties, since the initial system of equations for determining the stress state has a more complex structure. Contact problems of thermoelasticity for a transversely isotropic half-space were studied in [8–10] and others. An approach was used in [9, 10] that allows one to investigate problems only for a circular contact region. In [8], the contact problem of thermoelasticity is studied with a special distribution of the temperature field on the surface of the punch, which is proportional to the contact pressure under the paraboloidal punch. In the papers [11–15] and [16–21], spatial problems for transversely-isotropic elastic and electroelastic bodies respectively were considered. At the same time, analytical solutions of spatial contact problems for transversely isotropic elastic bodies were not obtained when hard punch heated in an arbitrary manner.

In this paper, the problem of thermoelasticity on the indentation of a heated plane hard punch of elliptical cross-section into a transversely isotropic elastic half-space is considered. Expressions of contact stresses and displacements of a heated flat elliptical punch are found explicitly. In the form of inequality, the relationship between the values of the indentation force, the heating temperature, and the thermoelastic properties of a transversely isotropic material is obtained, which makes it possible to predict the appearance of a material separation zone under a

flat elliptical punch (for given force and temperature influences). The influence of material properties, heating temperature, and indentation force on the distribution of contact pressure is investigated. It is shown that the appearance of separation (peeling) of the material significantly affects the type of the distribution of contact stresses under the punch.

Formulation of the problem. Let us consider a transversely isotropic half-space that occupies a region $z \leq 0$ and into which a heated flat hard punch of elliptical section is pressed without friction. We assume that the axis Oz coincides with the axis of symmetry of the transversely isotropic material. The boundary conditions on the surface of the half-space have the following form:

$$\begin{aligned} \sigma_{xz} = \sigma_{yz} = 0, \quad z = 0; \quad \sigma_{zz} = 0, \quad (x, y) \notin \Omega; \\ T(x, y, 0) = T_0(x, y), \quad (x, y) \in \Omega; \quad T_0(x, y)|_{\partial\Omega} = 0; \\ T(x, y, 0) = 0, \quad (x, y) \in R^2 \setminus \Omega; \\ u_z(x, y, 0) = \delta, \quad (x, y) \in \Omega, \end{aligned} \quad (1)$$

where $\Omega: x^2/a^2 + y^2/b^2 \leq 1$; $T(x, y, 0) > 0$ — punch heating temperature; δ — unknown displacement value. The indentation force applied in the center of the punch is related to the contact pressure by the ratio $P = \iint_{\Omega} p(x, y) dx dy$, where

$p(x, y)$ is the unknown contact pressure.

Basic relations. The equations of stationary thermoelasticity for an elastic transversely isotropic medium in the absence of body forces and heat sources in the body according to [8] can be written as

$$\begin{aligned} c_{11} \frac{\partial^2 u_x}{\partial x^2} + \frac{1}{2}(c_{11} - c_{12}) \frac{\partial^2 u_x}{\partial y^2} + c_{44} \frac{\partial^2 u_x}{\partial z^2} + \\ + \frac{\partial}{\partial x} \left[\frac{1}{2}(c_{11} + c_{12}) \frac{\partial u_y}{\partial y} + (c_{13} + c_{44}) \frac{\partial u_z}{\partial z} \right] = \beta \frac{\partial T}{\partial x}; \\ \frac{1}{2}(c_{11} - c_{12}) \frac{\partial^2 u_y}{\partial x^2} + c_{11} \frac{\partial^2 u_y}{\partial y^2} + c_{44} \frac{\partial^2 u_y}{\partial z^2} + \\ + \frac{\partial}{\partial y} \left[\frac{1}{2}(c_{11} + c_{12}) \frac{\partial u_x}{\partial x} + (c_{13} + c_{44}) \frac{\partial u_z}{\partial z} \right] = \beta \frac{\partial T}{\partial y}; \\ c_{44} \left(\frac{\partial^2 u_z}{\partial y^2} + \frac{\partial^2 u_z}{\partial x^2} \right) + c_{33} \frac{\partial^2 u_z}{\partial z^2} + (c_{13} + c_{44}) \frac{\partial}{\partial z} \left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} \right) = \beta_1 \frac{\partial T}{\partial z}; \quad (2) \\ \partial^2 T / \partial x^2 + \partial^2 T / \partial y^2 + n_4 \partial^2 T / \partial z^2 = 0. \end{aligned}$$

In the above expressions, c_{ij} are elastic constants; β, β_1, n_4 — constants depending on the thermophysical properties (thermal conductivity and thermal linear expansion coefficients) of the material. The solution of the system of equa-

tions (2) can be represented by means of four potential functions Φ_i ($i = 1, 2, 3, 4$) in according to [8] in this way:

$$\begin{aligned} u_x &= \partial\Phi_1 / \partial x + \partial\Phi_2 / \partial x + \partial\Phi_3 / \partial y + \partial\Phi_4 / \partial x; \\ u_y &= \partial\Phi_1 / \partial y + \partial\Phi_2 / \partial y - \partial\Phi_3 / \partial x + \partial\Phi_4 / \partial y; \\ u_z &= m_1\partial\Phi_1 / \partial z + m_2\partial\Phi_2 / \partial z + m_4\partial\Phi_4 / \partial z, \end{aligned} \tag{3}$$

where Φ_1, Φ_2, Φ_3 are functions satisfying the equations

$$(\partial^2 / \partial x^2 + \partial^2 / \partial y^2 + n_j \partial^2 / \partial z^2) \Phi_j = 0,$$

also $n_3 = 2c_{44} / (c_{11} - c_{12})$; n_1, n_2 are the roots of the quadratic equation

$$c_{11}c_{44}n^2 - [c_{44}^2 + c_{33}c_{11} - (c_{13} + c_{44})^2]n + c_{33}c_{44} = 0; \tag{4}$$

$$m_j = \frac{c_{11}n_j - c_{44}}{c_{13} + c_{44}} = \frac{n_j(c_{13} + c_{44})}{c_{33} - n_jc_{44}} \quad (j = 1, 2).$$

The function Φ_4 simultaneously satisfies two equations

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + n_4 \frac{\partial^2}{\partial z^2} \right) \Phi_4 = 0; \quad \frac{\partial^2 \Phi_4}{\partial z^2} = m_3 T.$$

We use the notation $z_j = zn_j^{-1/2}$ ($j = 1, 2, 3, 4$). Functions $\Phi_1(x, y, z_1)$, $\Phi_2(x, y, z_2)$, $\Phi_3(x, y, z_3)$, $\Phi_4(x, y, z_4)$ will be harmonic functions in the corresponding coordinate system. The constants m_3, m_4 included in relations (3) depend on the elastic and thermophysical properties of a transversely isotropic medium and are written as follows:

$$m_3 = \frac{\beta}{c_{44} + (c_{13} + c_{44})m_4 - c_{11}n_4}; \quad m_4 = \frac{\beta_1(c_{44} - n_4c_{11}) + \beta n_4(c_{13} + c_{44})}{\beta(c_{33} - n_4c_{44}) - \beta_1(c_{13} + c_{44})}.$$

Solution method. We write the temperature field in the form of the harmonic potential of the double layer

$$T(x, y, z_4) = \frac{\partial}{\partial z_4} \left(-\frac{1}{2\pi} \iint_{\Omega} \frac{T_0(\xi, \eta) d\xi d\eta}{\sqrt{(\xi - x)^2 + (\eta - y)^2 + z_4^2}} \right).$$

It follows from the properties of the derivative of the potential of a simple layer [3] that

$$T(x, y, z_4)|_{z_4=z=0} = \begin{cases} T_0(x, y), & (x, y) \in \Omega; \\ 0, & (x, y) \notin \Omega. \end{cases}$$

Harmonic function is a solution to the Dirichlet problem for the stationary heat conduction equation for a half-space (for a given distribution of the temperature field inside a flat region and zero temperature outside this region on the surface of the half-space [3]). Note that contact problems of thermoelasticity with a known temperature distribution in the contact area and the absence of a temperature field outside it (in the contact plane) were also considered in [8, 24].

Next, we present the solution of problem in the form of a superposition of states, for the first of which we take the function Φ_4 in the form of one of the Boussinesq potentials [2]

$$\begin{aligned} \Phi_4^{(1)}(x, y, z_4) &= F(x, y, z_4) = \\ &= -\frac{1}{2\pi} \iint_{\Omega} T_0(\xi, \eta) \ln\left(\sqrt{(\xi-x)^2 + (\eta-y)^2 + z_4^2} + z_4\right) d\xi d\eta. \end{aligned}$$

For the first state, we also set

$$\Phi_i^{(1)}(x, y, z_i) = \alpha_i F(x, y, z_i) \quad (i=1,2); \quad \Phi_3^{(1)} = 0,$$

where α_1, α_2 are the unknown constants.

Constants a_1, a_2 , we define by means of this way:

$$a_1 = -\frac{n_1^{1/2} (m_2 - m_4)}{n_4^{1/2} (m_2 - m_1)}; \quad a_2 = \frac{n_2^{1/2} (m_1 - m_4)}{n_4^{1/2} (m_2 - m_1)}.$$

As a result, for the first state we get

$$\begin{aligned} \sigma_{xz}^{(1)} &= \sigma_{yz}^{(1)} = 0, \quad z = 0; \\ u_z^{(1)}|_{z=0} &= 0, \quad (x, y) \in \Omega; \\ \sigma_{zz}^{(1)}|_{z=0} &= \begin{cases} -\gamma^{\text{Trans}} T_0(x, y) & (x, y) \in \Omega; \\ 0, & (x, y) \notin \Omega. \end{cases} \end{aligned}$$

We find the value γ^{Trans} in the form

$$\begin{aligned} \gamma^{\text{Trans}} &= \beta_1 - m_3(c_{33}m_4 - n_4c_{13}) + \\ &+ c_{44}m_3n_4^{1/2} \left[\frac{(m_1 - m_4)}{(m_1 - m_2)}(1 + m_2)n_2^{1/2} - \frac{(m_2 - m_4)}{(m_1 - m_2)}(1 + m_1)n_1^{1/2} \right]; \end{aligned} \quad (5)$$

$$n_4 = k_1/k; \quad \beta = (c_{11} + c_{12})\alpha + c_{13}\alpha_1; \quad \beta_1 = 2c_{13}\alpha + c_{33}\alpha_1,$$

where k_1/k is the ratio of the coefficient of thermal conductivity in the direction Oz to the coefficient of thermal conductivity in the direction Ox (or Oy); α, α_1 are the coefficients of linear thermal expansion of the material in the direction of Ox (or Oy) and Oz . In the transition from a transversely isotropic material to the isotropic material, we obtain $\gamma^{\text{Trans}} \rightarrow \mu\alpha(1+\nu)/(1-\nu)$ (ν is the Poisson's ratio, μ is the shear modulus), which fully corresponds to the result [5] for an isotropic material. Note that expression (5) for γ^{Trans} it was also used in the papers [22, 23] to find the thermo-stressed state of a transversely isotropic material with an elliptical crack.

For the second state of superposition, we choose the functions Φ_j ($j=1,2,3,4$) as follows:

$$\Phi_1^{(2)}(x, y, z_1) = \frac{1}{2\pi} \frac{-n_1^{1/2}}{(1+m_1)c_{44}(n_1^{1/2} - n_2^{1/2})} \iint_{\Omega} p(\xi, \eta) \ln(\rho_1 + z_1) d\xi d\eta;$$

$$\Phi_2^{(2)}(x, y, z_2) = \frac{1}{2\pi} \frac{n_2^{1/2}}{(1+m_2)c_{44}(n_1^{1/2} - n_2^{1/2})} \iint_{\Omega} p(\xi, \eta) \ln(\rho_2 + z_2) d\xi d\eta,$$

where

$$\rho_j = \sqrt{(x - \xi)^2 + (y - \eta)^2 + z_j^2} \quad (j = 1, 2).$$

In this case we obtain

$$\frac{\partial}{\partial z_i} \Phi_i^{(2)}(x, y, z_i) = \frac{1}{2\pi} \iint_{\Omega} \frac{p(\xi, \eta) d\xi d\eta}{\sqrt{(x - \xi)^2 + (y - \eta)^2 + z_i^2}} \quad (j = 1, 2).$$

Take also $\Phi_4^{(2)} = \Phi_3^{(2)} = 0$.

As a result of superposition of states, we obtain

$$\sigma_{xz}^{(1)} + \sigma_{xz}^{(2)} = \sigma_{yz}^{(1)} + \sigma_{yz}^{(2)} = 0 \quad \text{for } z = 0;$$

$$(u_z^{(1)} + u_z^{(2)})|_{z=0} = \frac{1}{2\pi} A^{\text{Trans}} \iint_{\Omega} \frac{p(\xi, \eta) d\xi d\eta}{\sqrt{(x - \xi)^2 + (y - \eta)^2}};$$

$$(\sigma_{zz}^{(1)} + \sigma_{zz}^{(2)})|_{z=0} = \begin{cases} -p(x, y) - \gamma^{\text{Trans}} T_0(x, y), & (x, y) \in \Omega; \\ 0, & (x, y) \notin \Omega. \end{cases}$$

The value A^{Trans} is determined as

$$A^{\text{Trans}} = \frac{1}{c_{44}(n_1^{1/2} - n_2^{1/2})} \left[\frac{m_2 - m_1}{(1+m_1)(1+m_2)} \right] =$$

$$= \frac{c_{11}}{c_{44}} \frac{(n_1^{1/2} + n_2^{1/2})(c_{13} + c_{44})}{(c_{11}n_1 + c_{13})(c_{11}n_2 + c_{13})}. \quad (6)$$

From the obtained expression (6) for a transversely isotropic material, one can easily obtain the case of an isotropic material. Let's put

$$n_1 = n_2 = 1; \quad c_{11} = \lambda + 2\mu; \quad c_{13} = \lambda; \quad c_{44} = \mu.$$

Then from formula (6) it follows

$$A^{\text{Trans}} \rightarrow \frac{\lambda + 2\mu}{2\mu(\lambda + \mu)} = \frac{1 - \nu}{\mu}.$$

Thus, for an isotropic material we obtain a coincidence of the results with the known data [5].

Note that expressions A^{Trans} (6) can be converted to a more convenient form. Using Vieta's theorem for the roots of the quadratic equation, from (6) we obtain

$$A^{\text{Trans}} = \frac{\sqrt{c_{11}}}{(c_{11}c_{33} - c_{13}^2)\sqrt{c_{44}}} \left[\sqrt{c_{11}c_{33} - c_{13}^2 - 2c_{44}c_{13} + 2c_{44}\sqrt{c_{11}c_{33}}} \right].$$

The obtained expression allows to find the desired value by directly substituting the elastic constants of the material into it without first determining the roots of the quadratic equation (4), as in the case of formulas (6).

Correspondence between solutions of contact problems for isotropic and transversely isotropic elastic half-spaces (in contact with heated plane rigid punch of elliptical section). According to the results of [5], the solution of the contact problem of thermoelasticity for an isotropic elastic half-space with boundary conditions (1) can lead to a search for an unknown potential density of a simple layer. It remains to satisfy the boundary condition

$$u_z(x, y, 0) = \delta = \frac{(1-\nu)}{2\pi\mu} \iint_{\Omega} \frac{p(\xi, \eta) d\xi d\eta}{\sqrt{(x-\xi)^2 + (y-\eta)^2}}. \quad (7)$$

Stress distribution has such form under the plane punch

$$\sigma_{zz}|_{z=0} = -p(x, y) - \frac{(1+\nu)}{(1-\nu)} \alpha \mu T_0(x, y), \quad (x, y) \in \Omega. \quad (8)$$

As a result of a superposition of states for a transversely isotropic half-space, we obtain

$$u_z(x, y, 0) = \delta = \frac{1}{2\pi} A^{\text{Trans}} \iint_{\Omega} \frac{p(\xi, \eta) d\xi d\eta}{\sqrt{(x-\xi)^2 + (y-\eta)^2}}. \quad (9)$$

The normal stresses under the plane punch in this case have the form

$$\sigma_{zz}|_{z=0} = -p(x, y) - \gamma^{\text{Trans}} T_0(x, y), \quad (x, y) \in \Omega. \quad (10)$$

All other boundary conditions (1) are satisfied. Comparing expressions (7), (8) and (9), (10), we conclude that such contact characteristics as contact pressure and displacement under the plane punch for a transversely isotropic half-space can be calculated from the corresponding expressions for an isotropic half-space by replacing the values $(1-\nu)/\mu$ with A^{Trans} and $(1+\nu)\alpha\mu/(1-\nu)$ by γ^{Trans} .

Solutions of new contact problems. When pressing a flat elliptical punch (in the absence of the rotations around the axes $0x$ and $0y$) according to the found correspondence of expressions (7), (8) and (9), (10) and results [4, 5] we obtain the values of contact pressure and displacement under the plane punch:

$$p(x, y) = \frac{P - Q_1}{2\pi ab} \left(1 - \frac{x^2}{a^2} - \frac{y^2}{b^2} \right)^{-1/2} + \gamma^{\text{Trans}} T_0(x, y);$$

$$\delta = \frac{P - Q_1}{2\pi a} A^{\text{Trans}} K(e), \quad (11)$$

where a is the semimajor axis of the ellipse, e is its eccentricity;

$$P \geq Q_1 = \gamma^{\text{Trans}} \iint_{\Omega} T_0(x, y) dx dy. \quad (12)$$

When inequality (12) is fulfilled, the contact stress under the flat punch are compressive and have a root singularity when approaching the punch boundary, which is determined by the first term in formulas (11).

Consider the distribution of the temperature field under the plane punch in the form

$$T_0(x, y) = T_q(1 - x^2/a^2 - y^2/b^2)^q, \quad q > 0.$$

Then, for a heated plane hard elliptical punch, the values of stresses and displacements under the punch we obtain in the form

$$-\sigma_{zz}(x, y) = \frac{P - Q_1}{2\pi ab} (1 - x^2/a^2 - y^2/b^2)^{-1/2} + \gamma^{\text{Trans}} T_q (1 - x^2/a^2 - y^2/b^2)^q;$$

$$\delta = \frac{P - Q_1}{2\pi a} B_1^{\text{Piezo}} K(e);$$

$$Q_1 = \gamma^{\text{Trans}} \iint_{\Omega} T_q (1 - x^2/a^2 - y^2/b^2)^q dx dy; \quad P \geq Q_1. \quad (13)$$

Note that formulas (13) have the following physical meaning. When the inequality $P \geq Q_1$ is fulfilled, the plane punch is pressed to the material over the entire contact area and under it there is no area of separation (delamination) of the material. After integration, we obtain the inequality

$$P \geq Q_1 = \frac{1}{q+1} T_q \gamma^{\text{Trans}} \pi ab. \quad (14)$$

If the opposite inequality holds

$$P < \frac{1}{q+1} T_q \gamma^{\text{Trans}} \pi ab$$

tensile stresses arise when approaching the punch edge (due to the first term in the stress components in formulas (13)), i.e. a material separation zone appears.

In [7], for the problem of the contact of a heated plane circular punch with an elastic isotropic half-space, it was proposed to search for a new contact zone, which is smaller than the size of the punch itself, from the problem for a non-planar punch, directing $R \rightarrow \infty$. The two-dimensional contact problem of thermoelasticity was considered in a similar way in [24]. Using this approach, and considering for this the problem of a heated paraboloidal punch of elliptical cross section, which is pressed into an elastic transversely isotropic half-space without friction, we obtain

$$\frac{a^3}{R_1} = [(P - Q_1) A^{\text{Trans}}]^{1/3} \left[\frac{3}{2\pi} \left(\frac{K(e) - E(e)}{e^2} \right) \right]^{1/3},$$

where e is the eccentricity of the elliptical base of the punch. Directing $R_1 \rightarrow \infty$, we get

$$P = Q_1 = \frac{1}{q+1} \gamma^{\text{Trans}} T_q \pi a_* b_*,$$

where a_*, b_* are the semi-axes of the new contact area under the punch. They are smaller than the corresponding semi-axes of the plane punch. However, contact zone remains elliptical and the relation remains $b_*/a_* = b/a$. Using the expression (14) we find further

$$a_* = \sqrt{\frac{P(q+1)}{\gamma^{\text{Trans}} T_q \pi \sqrt{1-e^2}}}; \quad b_* = a_* \sqrt{1-e^2}.$$

Contact stress under a heated plane punch in the case of separation of material near the edge of the punch takes the form

$$-\sigma_{zz}(x, y) = \gamma^{\text{Trans}} T_q (1 - x^2/a_*^2 - y^2/b_*^2)^q, \quad (15)$$

since the singular term disappears in formulas (13) at $P = Q_1$. Therefore, an increase in punch heating when a certain threshold value is exceeded, which depends on the strength P and thermoelastic properties of the transversely isotropic material, leads to the appearance of separation zone of the material under the plane punch.

Note that for a plane circular punch ($e = 0$), the radius of the contact area when the material is separated from the punch according to formulas (15) takes the form

$$a_* = \sqrt{\frac{P(q+1)}{\gamma^{\text{Trans}} T_q \pi}}, \quad b_* = a_*.$$

Analysis of the results of numerical investigations. Consider the case of the distribution of the temperature field under a plane circular punch of radius a in the form $T_0(1 - x^2/a^2 - y^2/a^2)^{1/4}$, where $T_0 > 0$. We investigate next three cases of punch heating: 1) $T_0 = \frac{1,2P}{\gamma^{\text{Trans}} \pi a^2}$; 2) $T_0 = \frac{1,3P}{\gamma^{\text{Trans}} \pi a^2}$ and

3) $T_0 = \frac{1,5P}{\gamma^{\text{Trans}} \pi a^2}$. First, we verify the fulfillment of inequality (14) to find out

whether the material is peeling off. As a result, we obtain that for the first case inequality (14) is satisfied, i.e. material separation under the punch does not occur. At the same time, with increasing heating (cases 2 and 3), such a separation of the material takes place.

After simple calculations for the first case of heating, the pressure expression under the punch takes the form

$$-\frac{\sigma_{zz}}{T_0 \gamma^{\text{Trans}}} = \frac{1}{60} (1 - r^2/a^2)^{-1/2} + (1 - r^2/a^2)^{1/4}.$$

For the second and third cases, the contact pressure has the same expression

$$-\frac{\sigma_{zz}}{T_0 \gamma^{\text{Trans}}} = (1 - r^2/a_*^2)^{1/4}.$$

However, the radii of the new contact area for these cases are different. For the second case of heating $a_* = 5a/\sqrt{26}$, at the same time for the third case we find that $a_* = a\sqrt{5/6}$.

Fig. 1 shows the change in contact pressure under the punch, while the pressure curve for case 1 (without separating the material under the punch) is shown by line 1, and for the second and third cases – by means of the lines with corresponding numbers. It is seen that with increasing heating of the flat punch, the contact area with the half-space decreases.

Consider a flat elliptical punch with the distribution of the temperature field under the punch in the form $T_0(1 - x^2/a^2 - y^2/b^2)^{1/4}$ of $T_0 > 0$. Put $T_0 = P/[(\pi ab)\gamma^{\text{Trans}} \beta^*]$. We study cases of punch heating, assuming that it β^*

takes the following values: 1) $\beta^* = 1000$; 2) $\beta^* = 4$; 3) $\beta^* = 2$; 4) $\beta^* = 1$. Since the value T_0 is inversely proportional to the value β^* , in the latter case we get the highest punch heating. For the selected parameter values β^* , the material does not separate under the punch, and the stress expression under the punch takes the form:

$$-\frac{\sigma_{zz} \alpha^*}{[P/(\pi ab)]} = \frac{1}{2} \left[\alpha^* - \frac{4}{5} \right] (1 - x^2/a^2 - x^2/b^2)^{-1/2} + (1 - x^2/a^2 - x^2/b^2)^{1/4}.$$

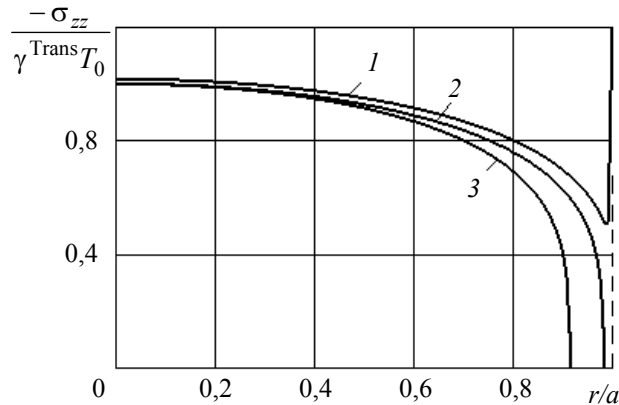


Fig. 1. The stress distribution under a heated flat punch, taking into account the separation of the material

In fig. 2 shows the distribution of the contact pressure under a flat punch with an elliptical section at different values of the heating of the punch.

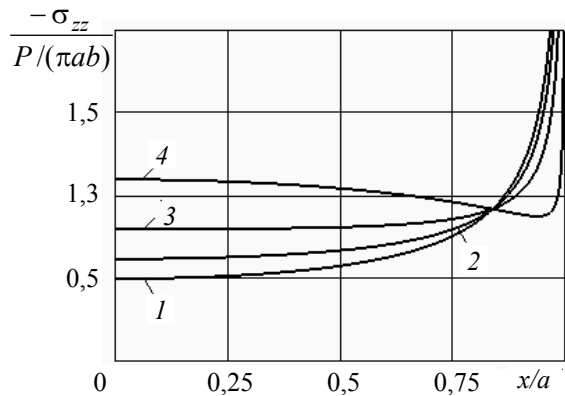


Fig. 2. Stress distribution under a plane elliptical punch

Lines 1–4 in Fig. 2 correspond to punch heating options noted above. It can be seen an increase of the temperature heating of a plane punch lead to the stresses increase in the center of the punch and decrease when approaching its boundary.

Thus, in the paper expressions of contact stresses and displacements of a heated flat elliptical punch are found explicitly. By means of the inequality, the condition for the occurrence of material separation under a flat heated elliptical punch is obtained, which is pressed without friction into the transversely isotropic elastic half-space. This inequality includes the values of the indentation force, temperature heating, and thermoelastic properties of a transversely isotropic

material. It is shown that with increasing heating, the region of complete contact (with separation of the material) decreases. The famous results for an elastic isotropic material follow from the obtained data as a special case. The influence of heating on the distribution of contact stresses, as well as the appearance of a region of separation (delamination) of a transversely isotropic material under a punch, is investigated.

REFERENCES

1. J.R. Barber, *Contact Mechanics*. New York, USA: Springer, 2018.
2. K.L. Johnson, *Contact Mechanics*. Cambridge, Great Britain: Cambridge Univ. Press, 1985.
3. A.I. Lurie, *Theory of Elasticity*. Berlin, Germany: Springer, 2005.
4. L.A. Galin and G.M.L. Gladwell (Editor) *Contact Problems*. Dordrecht, Germany: Springer, 2008.
5. L.A. Galin, *Development of the Theory of Contact Problems in the USSR* [in Russian]. Moscow, USSR: Nauka, 1976.
6. N.M. Borodachev, "On solving the contact problem of thermoelasticity in the case of axial symmetry", *Izv. AN SSSR, Otd. Tekhn. Nauk Mekh. Mashinost*, no. 5, pp. 86–90, 1962.
7. J.R. Barber, "Indentation of an elastic half space by a cooled flat punch", *Q.J. Mech. Appl. Math.*, vol. 35, no.1, pp. 141–154, 1982.
8. Yu.N. Podil'chuk, V.F. Tkachenko, Ya.I. Sokolovskii, "Thermoelastic contact problem on the penetration of a transversely isotropic half-space by a heated die elliptical in plan", *Int. Appl. Mech.*, vol. 32, no. 11, pp. 851–857, 1996.
9. D.V. Grilitsky and Ya.M. Kizyma, *Axisymmetric contact problems of the theory of elasticity and thermoelasticity* [in Russian]. Lvov, Ukraine: Vyscha Shkola, 1981.
10. B.G. Shelestovskii and G.V. Gabrusev, "Thermoelastic state of transversely isotropic layer between two annular punches", *Int. Appl. Mech.*, vol. 40, no. 4, pp. 417–425, 2004.
11. Y.S. Chai and I.I. Argatov, "Local tangential contact of elastically similar, transversely isotropic elastic bodies", *Meccanica*, vol. 53, no. 11–12, pp. 3137–3143, 2018.
12. V.I. Fabrikant, "Contact problem for an arbitrarily oriented transversely isotropic half-space", *Acta Mechanica*, vol. 228, no. 4, pp. 1541–1560, 2017.
13. P.F. Hou, W.H. Zhang and J.-Y. Chen, "Three-dimensional exact solutions of homogeneous transversely isotropic coated structures under spherical contact", *Int. J. Solids Structures*, vol. 161, no. 5, pp. 136–173, 2019.
14. F. Marmo, F. Toraldo and L. Rosati, "Analytical formulas and design charts for transversely isotropic half-spaces subject to linearly distributed pressures", *Meccanica*, vol. 51, no. 11, pp. 2909–2928, 2016.
15. Yu.V. Tokovyy and C.C. Ma, "Three-dimensional elastic analysis of transversely-isotropic composites", *Journal of Mechanics*, vol. 33, no. 6, pp. 821–830, 2018.
16. V.S. Kirilyuk, "On the relationship between the solutions of static contact problems of elasticity and electroelasticity for a half-space", *Int. Appl. Mech.*, vol. 42, no. 11, pp. 1266–1269, 2006.
17. V.S. Kirilyuk and O.I. Levchuk, "Indentation of punches into a piezoceramic body: Two-dimensional contact problem of electroelasticity", *Int. Appl. Mech.*, vol. 44, no. 11, pp. 1244–1257, 2008.
18. V.S. Kirilyuk, "Stress state of a piezoceramic body with a plane crack opened by a rigid inclusion", *Int. Appl. Mech.*, vol. 44, no. 7, pp. 757–768, 2008.
19. V.S. Kirilyuk and O.I. Levchuk, "Stress state of an orthotropic piezoelectric material with an elliptic crack", *Int. Appl. Mech.*, vol. 53, no. 3, pp. 305–312, 2017.
20. V.S. Kirilyuk, O.I. Levchuk, and H. Altenbach, "Calculation of stress intensity factors for an arbitrary oriented penny-shaped crack under inner pressure in an orthotropic electroelastic material", *Advanced Structured Materials*, vol. 108, pp. 211–222, 2019.
21. Yu.N. Podil'chuk, "Exact analytical solutions of static electroelastic and thermoelectroelastic problems for a transversely isotropic body in curvilinear coordinate systems", *Int. Appl. Mech.*, vol. 39, no. 2, pp. 132–170, 2003.

22. V.S. Kirilyuk, "The thermoelastic equilibrium of a transversally isotropic medium with an elliptic crack under symmetric loading", *Int. Appl. Mech.*, vol. 36, no. 4, pp. 509–517, 2000.
23. V.S. Kirilyuk, "Equilibrium of a transversally isotropic body with an elliptic crack under thermal action", *Int. Appl. Mech.*, vol. 37, no. 10, pp. 1304–1310, 2001.
24. V. Pauk, "Plane contact of hot flat-ended punch and thermoelastic half-space involving finite friction", *J. Appl. Mech.*, vol. 74, no. 6, pp. 1172–1177, 2007.

Received 11.09.2020

INFORMATION ON THE ARTICLE

Vitaly S. Kirilyuk, ORCID: 0000-0002-8513-0378, S.P. Timoshenko Institute of mechanics of NAS of Ukraine, e-mail: kirilyuk_v@ukr.net.

Olga I. Levchuk, ORCID: 0000-0002-6514-6225, S.P. Timoshenko Institute of mechanics of NAS of Ukraine, e-mail: 2013levchuk@gmail.com.

Olena V. Gavrilenko, ORCID: 0000-0003-0413-6274, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: iem.gavrilenko@meta.ua.

Mykhailo B. Viter, ORCID: 0000-0003-4109-005X, National Transport University of Ukraine, e-mail: mbviter@gmail.com

МОДЕЛЮВАННЯ КОНТАКТНОЇ ВЗАЄМОДІЇ НАГРІТОГО ЖОРСТКОГО ЕЛІПТИЧНОГО ШТАМПА З ТРАНСВЕРСАЛЬНО-ІЗОТРОПНИМ ПРУЖНИМ ПІВПРОСТОРОМ / В.С. Кирилюк, О.І. Левчук, В.В. Гавриленко, М.Б. Вітер

Анотація. На основі строгої математичної моделі досліджено задачу контактної взаємодії нагрітого плоского штампа еліптичного перерізу з трансверсально-ізотропним пружним півпростором. Припускається, що поверхня півпростору є площиною ізоотропії трансверсально-ізотропного матеріалу, а також має гладкий (без тертя) контакт. У явному вигляді знайдено вирази контактних напружень і переміщення нагрітого плоского еліптичного штампа. У вигляді простої нерівності отримано умову відділення пружного матеріалу від поверхні плоского еліптичного штампа. Виконано числові розрахунки. Вивчено контактну взаємодію нагрітого плоского штампа з урахуванням відділення матеріалу від штампа.

Ключові слова: математична модель, контактна взаємодія, пружний півпростір, трансверсально-ізотропний матеріал, плоский еліптичний штамп, нагрівання, розподіл напружень, ділянка відділення матеріалу.

МОДЕЛИРОВАНИЕ КОНТАКТНОГО ВЗАИМОДЕЙСТВИЯ НАГРЕТОГО ПЛОСКОГО ЖЕСТКОГО ЭЛЛИПТИЧЕСКОГО ШТАМПА С ТРАНСВЕРСАЛЬНО-ИЗОТРОПНЫМ УПРУГИМ ПОЛУПРОСТРАНСТВОМ / В.С. Кирилюк, О.И. Левчук, В.В. Гавриленко, М.Б. Витер

Аннотация. На основе строгой математической модели исследована задача контактного взаимодействия нагретого плоского штампа эллиптического сечения с трансверсально-изотропным упругим полупространством. Предполагается, что поверхность полупространства является плоскостью изотропии трансверсально-изотропного материала, а также имеет гладкий (без трения) контакт. В явном виде найдены выражения контактных напряжений и перемещения нагретого плоского эллиптического штампа. В виде простого неравенства получено условие отделения упругого материала от поверхности плоского эллиптического штампа. Выполнены числовые расчеты. Изучено контактное взаимодействие нагретого плоского штампа с учетом отделения материала от штампа.

Ключевые слова: математическая модель, контактное взаимодействие, упругое полупространство, трансверсально-изотропный материал, плоский эллиптический штамп, нагрев, распределение напряжений, область отделения материала.

ВІДОМОСТІ ПРО АВТОРІВ

Бідюк Петро Іванович,

професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Бодянський Євгеній Володимирович,

професор, доктор технічних наук, професор кафедри штучного інтелекту Харківського національного університету радіоелектроніки, Україна, Харків

Вігер Михайло Богданович,

доцент, кандидат фізико-математичних наук, доцент кафедри інформаційних систем та технологій факультету транспортних та інформаційних технологій Національного транспортного університету, Україна, Київ

Гавриленко Олена Валеріївна,

доцент, кандидат фізико-математичних наук, доцент кафедри автоматизованих систем обробки інформації та управління факультету інформатики та обчислювальної техніки КПІ ім. Ігоря Сікорського, Україна, Київ

Гамідов Галіб,

доктор філософії, керівник департаменту компанії «Азеришиг», Азербайджан, Баку

Дрогобицький Іван Миколайович,

професор, доктор економічних наук, професор кафедри системного аналізу в економіці Фінансового університету при уряді Російської Федерації, Росія, Москва

Зайченко Юрій Петрович,

професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Крамов Артем Андрійович,

аспірант Київського національного університету імені Тараса Шевченка, Україна, Київ

Кривий Сергій Лук'янович,

професор, доктор фізико-математичних наук, професор кафедри інтелектуальних програмних систем Київського національного університету імені Тараса Шевченка, Україна, Київ

Кулішова Нонна Євгенівна,

доцент, кандидат технічних наук, професор кафедри медіасистем і технологій Харківського національного університету радіоелектроніки, Україна, Харків

Курило Олена Олександрівна,

інженер-програміст 3-ї категорії відділу інформаційних технологій АТ «ВНДІАЕН», Україна, Суми

Кирилюк Віталій Семенович,

старший науковий співробітник, доктор фізико-математичних наук, провідний науковий співробітник відділу механіки стохастично-неоднорідних середовищ Інституту механіки імені С.П.Тимошенка НАН України, Київ

Левчук Ольга Іванівна,

старший науковий співробітник, кандидат фізико-математичних наук, старший науковий співробітник відділу механіки стохастично-неоднорідних середовищ Інституту механіки імені С.П. Тимошенка НАН України, Київ

Мацукі Йошіо,

доктор технічних наук, завідувач лабораторії економетрики і прогнозування Світового центру даних з геоінформатики і сталого розвитку КПІ ім. Ігоря Сікорського, Україна, Київ

Мельник Ігор Віталійович,

професор, доктор технічних наук, професор кафедри електронних приладів та пристроїв факультету електроніки КПІ ім. Ігоря Сікорського, Україна, Київ

Мухойд Олександр Вікторович,

завідувач сектору САПР відділу інформаційних технологій АТ «ВНДІАЕН», Україна, Суми

Погорілий Сергій Дем'янович,

професор, доктор технічних наук, завідувач кафедри комп'ютерної інженерії Київського національного університету імені Тараса Шевченка, Україна, Київ

Починок Аліна Володимирівна,

кандидат технічних наук, доцент кафедри інформаційних систем і технологій Навчально-наукового інституту інформаційних технологій та менеджменту Університету державної фіскальної служби України, Київ

Рудзевич Анна-Марія Павлівна,

студентка ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Рязановський Кирило Денисович,

магістр кафедри технічної кібернетики факультету інформатики та обчислювальної техніки КПІ ім. Ігоря Сікорського, Україна, Київ

Синєглазов Віктор Михайлович,

професор, доктор технічних наук, завідувач кафедри авіаційних комп'ютерно-інтегрованих комплексів Національного авіаційного університету, Україна, Київ

Слинько Максим Сергійович,

аспірант Київського національного університету імені Тараса Шевченка, Україна, Київ

Статкевич Віталій Михайлович,

кандидат фізико-математичних наук, науковий співробітник відділу прикладного нелінійного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Чумаченко Олена Іллівна,

доцент, доктор технічних наук, доцент кафедри технічної кібернетики факультету інформатики та обчислювальної техніки КПІ ім. Ігоря Сікорського, Україна, Київ