



№4-2022

ISSN 1999-9429

ИЗВЕСТИЯ ЮФУ

ТЕХНИЧЕСКИЕ НАУКИ

- Анализ данных и моделирование
- Алгоритмы обработки информации
- Электроника, связь и навигация

ИЗВЕСТИЯ ЮФУ. ТЕХНИЧЕСКИЕ НАУКИ IZVESTIYA SFedU. ENGINEERING SCIENCES

Свидетельство о регистрации средства массовой информации
ПИ № ФС77-28889 от 12.07.2007

Федеральная служба по надзору в сфере связи, информационных технологий
и массовых коммуникаций

Научно-технический и прикладной журнал

Издается с 1995 года, до середины 2007 года под названием «Известия ТРТУ»

Подписной индекс 41970

№ 4 (228). 2022 г.

Тематический выпуск

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Журнал включен в «Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук».

Редакционный совет

Курейчик В.В. (гл. редактор); Кравченко Ю.А. (зам. гл. редактора); Курейчик В.М. (зам. гл. редактора); Бородянский И.М. (ученый секретарь); Абрамов С.М.; Агеев О.А.; Бабенко Л.К.; Боженюк А.В.; Веселов Г.Е.; Гайдук А.Р.; Горбанёва О.И.; Каляев И.А.; Коноплев Б.Г.; Касьянов А.О.; Левин И.И.; Куповых Г.В.; Лысенко И.Е.; Малюков С.П.; Медведев М.Ю.; Мельник Э.В.; Никитов С.А.; Обуховец В.А.; Панич А.Е.; Петров В.В.; Пшихопов В.Х.; Редько В.Г.; Румянцев К.Е.; Семенихина Д.В.; Сергеев Н.Е.; Стемпковский А.Л.; Сухинов А.И.; Тополов В.Ю.; Турулин И.И.; Угольницкий Г.А.; Усов А.Б.; Федосов В.П.; Целых А.Н.; Юханов Ю.В.

Учредитель Южный федеральный университет.

Издатель Южный федеральный университет.

Ответственный за выпуск Курейчик В.В.

Технический редактор Ярошевич Н.В.

Оригинал-макет выполнен Ярошевич Н.В.

Дата выхода в свет 31.10.2022 г. Формат 70×108 $\frac{1}{16}$. Бумага офсетная.

Офсетная печать. Усл. печ. л. – 22,3. Уч.-изд. л. – 19,7.

Заказ № 8687. Тираж 250 экз.

Адрес издателя: 344091, г. Ростов-на-Дону, пр. Стачки, 200/1. Тел. 8(863)2478051.

Адрес типографии: Отпечатано в отделе полиграфической, корпоративной и сувенирной продукции Издательско-полиграфического комплекса КИБИ МЕДИА ЦЕНТРА ЮФУ. 344090, г. Ростов-на-Дону, пр. Стачки, 200/1, тел (863) 247-80-51.

Адрес редакции: 347922, г. Таганрог, ул. Чехова, 22, ЮФУ, тел. +7 (928) 909-57-82, e-mail: iborodyanskiy@sfedu.ru, <http://izv-tn.tti.sfedu.ru/>.

16+

Цена свободная

ISSN 1999-9429 (Print)

ISSN 2311-3103 (Online)

© Южный федеральный университет, 2022

СОДЕРЖАНИЕ

РАЗДЕЛ I. АНАЛИЗ ДАННЫХ И МОДЕЛИРОВАНИЕ

П.А. Воронин, А.М. Белевцев, Ф.Г. Садреев МЕТОДИКА ОЦЕНКИ И ВЫБОРА PLM-СИСТЕМ ПРИ ЦИФРОВОЙ ТРАНСФОРМАЦИИ МАШИНОСТРОИТЕЛЬНОГО ПРЕДПРИЯТИЯ	6
Е.М. Герасименко, В.В. Курейчик, С.И. Родзин, А.П. Кухаренко ПРИМЕНЕНИЕ НЕЧЕТКОЙ ЛОГИКИ ДЛЯ ПРИНЯТИЯ РЕШЕНИЙ ОБ ЭВАКУАЦИИ ПРИ НАВОДНЕНИИ.....	15
Э.В. Кулиев, А.В. Котельва, М.М. Семенова, С.В. Игнатьева, А.П. Кухаренко ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ В УПРАВЛЕНИИ ПРЕДПРИЯТИЕМ НА ОСНОВЕ АЛГОРИТМА ИМИТАЦИИ ОТЖИГА	29
Д.В. Загуменнов, В.В. Мкртчян ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ НАДЕЖНОСТИ СХЕМ ШИРОКОВЕЩАТЕЛЬНОГО ШИФРОВАНИЯ С АЛГЕБРОГЕОМЕТРИЧЕСКИМИ КОДАМИ МАЛОЙ МОЩНОСТИ.....	40
А.К. Мельников, И.И. Левин, А.И. Дордопуло, Л.М. Сластен ОЦЕНКА ВОЗМОЖНОСТЕЙ ПЕРСПЕКТИВНЫХ ВЫЧИСЛИТЕЛЬНЫХ ТЕХНОЛОГИЙ ДЛЯ РАСЧЕТА ТОЧНЫХ ПРИБЛИЖЕНИЙ РАСПРЕДЕЛЕНИЙ ВЕРОЯТНОСТЕЙ ЗНАЧЕНИЙ СТАТИСТИК	50
К.Е. Румянцев, П.Д. Миронова, Шакир Хайдер Хуссейн ОЦЕНКА ВЛИЯНИЯ ФУНКЦИОНАЛЬНЫХ ЭЛЕМЕНТОВ НА ПАРАМЕТРЫ СИСТЕМЫ КВАНТОВОГО РАСПРЕДЕЛЕНИЯ КЛЮЧА НА ОСНОВЕ ПРОТОКОЛА B92.....	63
А.С. Черепанцев ЗАКОНОМЕРНОСТИ ПЕРЕХОДНОГО РЕЖИМА В ДИССИПАТИВНОЙ КЛЕТОЧНОЙ МОДЕЛИ ЗЕМЛЕТРЯСЕНИЙ	77
В.А. Павский, К.В. Павский ОЦЕНКА ОСУЩЕСТВИМОСТИ РЕШЕНИЯ ЗАДАЧ НА ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ ПРИ ГРУППОВОМ ОБСЛУЖИВАНИИ.....	95
И.Д. Русаловский, Л.К. Бабенко, О.Б. Макаревич РАЗРАБОТКА МЕТОДОВ ГОМОМОРФНОГО ДЕЛЕНИЯ	103

РАЗДЕЛ II. АЛГОРИТМЫ ОБРАБОТКИ ИНФОРМАЦИИ

О.Б. Лебедев, А.А. Жиглатый БИОИНСПИРИРОВАННЫЙ АЛГОРИТМ РЕШЕНИЯ ИНВАРИАНТНЫХ ГРАФОВЫХ ЗАДАЧ	113
В.В. Бова, Ю.А. Кравченко, С.И. Родзин МЕТОДЫ И АЛГОРИТМЫ КЛАСТЕРИЗАЦИИ ТЕКСТОВЫХ ДАННЫХ (ОБЗОР).....	122
Б.К. Лебедев, О.Б. Лебедев, Е.О. Лебедева ЭВОЛЮЦИОННЫЙ ПОПУЛЯЦИОННЫЙ МЕТОД РЕШЕНИЯ ТРАНСПОРТНОЙ ЗАДАЧИ.....	143
Т.Н. Кондратьева, Е.Р. Мунтян, И.Ф. Развеева РЕАЛИЗАЦИЯ ТОРГОВОГО СОВЕТНИКА ДЛЯ МУЛЬТИРЫНОЧНОЙ ПЛАТФОРМЫ METATRADER 5	157
И.И. Турулин, Аль-Карави Хуссейн Шукор Мукер МЕТОДИКА И АЛГОРИТМ СИНТЕЗА УПРАВЛЯЕМЫХ ЦИФРОВЫХ ФИЛЬТРОВ ЧЕБЫШЕВА I РОДА НИЖНИХ ЧАСТОТ НА БАЗЕ МЕТОДА БИЛИНЕЙНОГО ПРЕОБРАЗОВАНИЯ	170

А.А. Чусов, М.А. Копаева АЛГОРИТМ ПОТОКОВОЙ КОМПРЕССИИ ДАННЫХ С ПЛАВАЮЩЕЙ ЗАПЯТОЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ ОБЕСПЕЧЕНИЯ НАУЧНЫХ ЭКСПЕРИМЕНТОВ	182
В.А. Частикова, С.А. Жерлицын НЕЙРОСЕТЕВАЯ МЕТОДИКА ИДЕНТИФИКАЦИИ ЛИЧНОСТИ ПО РИСУНКУ ВЕН ЛАДОНИ	192
К.Н. Алексеев, Д.А. Сорокин, А.Л. Леонтьев МЕТОДИКА СОЗДАНИЯ ТОПОЛОГИЧЕСКИХ ОГРАНИЧЕНИЙ ПРИ ВЫСОКОЙ УТИЛИЗАЦИИ РЕСУРСОВ ПЛИС	200

РАЗДЕЛ III. ЭЛЕКТРОНИКА, СВЯЗЬ И НАВИГАЦИЯ

Н.М. Богатов, В.С. Володин, Л.Р. Григорьян, А.И. Коваленко, М.С. Коваленко ИМПУЛЬСНЫЕ ХАРАКТЕРИСТИКИ КРЕМНИЕВЫХ СТРУКТУР С N-P ПЕРЕХОДОМ, ОБЛУЧЕННЫХ ПРОТОНАМИ	213
Р.Р. Ибадов, В.П. Федосов, С.Р. Ибадов АНАЛИЗ ПОДСТИЛАЮЩЕЙ ПОВЕРХНОСТИ ПРИ ФОРМИРОВАНИИ ИЗОБРАЖЕНИЯ В РЕЖИМЕ ДОПЛЕРОВСКОГО ОБУЖЕНИЯ ЛУЧА	222
П.В. Луферчик, А.А. Комаров, П.В. Штро, А.Н. Конев РАЗРАБОТКА АДАПТИВНОЙ СИСТЕМЫ СВЯЗИ НА ОСНОВЕ OFDM ДЛЯ ТРОПОСФЕРНОГО И РАДИОРЕЛЕЙНОГО КАНАЛА	231
П.В. Луферчик, П.В. Штро, А.Н. Конев, А.А. Комаров РАЗРАБОТКА ЭНЕРГЕТИЧЕСКИ ЭФФЕКТИВНОЙ СИСТЕМЫ СВЯЗИ В ТРОПОСФЕРНОМ РАДИОКАНАЛЕ НА БАЗЕ OFDM-СИГНАЛОВ	238
О.В. Белоусова, Г.В. Куповых, А.Г. Клово, В.В. Гривцов РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ ЭЛЕКТРОДИНАМИЧЕСКОЙ СТРУКТУРЫ ТУРБУЛЕНТНОГО ПРИЗЕМНОГО СЛОЯ	245

CONTENT

SECTION I. DATA ANALYSIS AND MODELING

P.A. Voronin, A.M. Belevtsev, F.G. Sadreev METHODOLOGY FOR EVALUATION AND SELECTION OF PLM-SYSTEMS IN THE DIGITAL TRANSFORMATION OF A MACHINE-BUILDING ENTERPRISE	6
E.M. Gerasimenko, V.V. Kureichik, S.I. Rodzin, A.P. Kukhareno APPLICATION OF FUZZY LOGIC FOR MAKING DECISIONS ABOUT EVACUATION IN CASE OF FLOODING	16
E.V. Kuliev, A.V. Kotelva, M.M. Semenova, S.V. Ignateva, A.P. Kukhareno INTELLIGENT DATA ANALYSIS IN ENTERPRISE MANAGEMENT BASED ON THE ANNEALING SIMULATION ALGORITHM.....	30
D.V. Zagumennov, V.V. Mkrtichyan EXPERIMENTAL STUDY OF THE RELIABILITY OF BROADCAST ENCRYPTION SCHEMES WITH LOW-POWER ALGEBRAIC GEOMETRIC CODES	40
A.K. Melnikov, I.I. Levin, A.I. Dordopulo, L.M. Slasten ANALYSIS OF ADVANCED COMPUTER TECHNOLOGIES FOR CALCULATION OF EXACT APPROXIMATIONS OF STATISTICS PROBABILITY DISTRIBUTIONS	51
K.E. Rumyantsev, P.D. Mironova, Shakir Hayder Hussein EVALUATION OF THE FUNCTIONAL ELEMENTS INFLUENCE ON THE PARAMETERS OF THE QKD SYSTEM BASED ON B92 PROTOCOL	63
A.S. Cherepantsev THE TRANSIENT REGIME PATTERNS IN THE DISSIPATIVE CELL MODEL OF EARTHQUAKES	77
V.A. Pavsky, K.V. Pavsky ESTIMATION OF REALIZABILITY OF SOLVING TASKS ON COMPUTER SYSTEMS IN GROUP MAINTENANCE	95
I.D. Rusalovsky, L.K. Babenko, O.B. Makarevich DEVELOPMENT OF HOMOMORPHIC DIVISION METHODS.....	104

SECTION II. INFORMATION PROCESSING ALGORITHMS

O.B. Lebedev, A.A. Zhiglatiy BIOINSPIRED ALGORITHM FOR SOLVING INVARIANT GRAPH PROBLEMS ..	113
V.V. Bova, Y.A. Kravchenko, S.I. Rodzin METHODS AND ALGORITHMS FOR TEXT DATA CLUSTERING (REVIEW).....	122
B.K. Lebedev, O.B. Lebedev, E.O. Lebedeva EVOLUTIONARY POPULATION METHOD FOR SOLVING THE TRANSPORT PROBLEM.....	144
T.N. Kondratieva, E.R. Muntyan, I.F. Razveeva IMPLEMENTATION OF A TRADING ADVISOR FOR THE METATRADER 5 MULTI-MARKET PLATFORM.....	158
I.I. Turulin, Al-Karawi Hussein Shakor Mogheer METHOD AND ALGORITHM FOR SYNTHESIS OF CONTROLLED CHEBYSHEV DIGITAL FILTERS OF THE FIRST KIND OF LOW FREQUENCIES BASED ON THE BILINEAR TRANSFORMATION METHOD	170
A.A. Chusov, M.A. Kopaeva ALGORITHM FOR COMPRESSION OF FLOATING-POINT DATA IN SCIENTIFIC RESEARCH SUPPORT SYSTEMS	183
V.A. Chastikova, S.A. Zherlitsyn DEVELOPMENT OF A METHOD FOR PERSONAL IDENTIFICATION BASED ON THE PATTERN OF PALM VEINS.....	193
K.N. Alekseyev, D.A. Sorokin, A.L. Leontyev METHODOLOGY OF TOPOLOGICAL RESTRICTIONS FOR INTENSIVELY USED FPGA RESOURCE	201

SECTION III. ELECTRONICS, COMMUNICATIONS AND NAVIGATION

N.M. Bogatov, V.S. Volodin, L.R. Grigoryan, A.I. Kovalenko, M.S. Kovalenko	
IMPULSE CHARACTERISTICS OF SILICON STRUCTURES WITH N-P JUNCTION IRRADIATED BY PROTONS	213
R.R. Ibadov, V.P. Fedosov, S.R. Ibadov	
ANALYSIS OF UNDERLYING SURFACE IN IMAGE FORMATION IN DOPPLER BEAM SHARPENING MODE.....	222
P.V. Luferschik, A.A. Komarov, P.V. Shtro, A.N. Konev	
DEVELOPMENT OF ADAPTIVE OFDM-BASED COMMUNICATION SYSTEM FOR TROPOSPHERE AND RADIO RELAY CHANNEL	231
P.V. Luferschik, P.V. Shtro, A.N. Konev, A.A. Komarov	
DEVELOPMENT OF ENERGY EFFICIENT COMMUNICATION SYSTEM IN THE TROPOSPHERE RADIO CHANNEL BASED ON OFDM SIGNALS	239
O.V. Belousova, G.V. Kupovykh, A.G. Klovo, V.V. Gritsov	
MODELING RESULTS OF THE TURBULENT SURFACE LAYER ELECTRODYNAMIC STRUCTURE	245

Раздел I. Анализ данных и моделирование

УДК 004

DOI 10.18522/2311-3103-2022-4-6-15

П.А. Воронин, А.М. Белевцев, Ф.Г. Садреев

МЕТОДИКА ОЦЕНКИ И ВЫБОРА PLM-СИСТЕМ ПРИ ЦИФРОВОЙ ТРАНСФОРМАЦИИ МАШИНОСТРОИТЕЛЬНОГО ПРЕДПРИЯТИЯ

В рамках перехода на шестой технологический уклад перед предприятиями промышленности остро встает вопрос эффективного проведения цифровой трансформации предприятия. Для этого предприятиям необходимо обеспечить переход к процессному управлению, автоматизировать бизнес-процессы и интегрировать все процессы, приложения и данные на единой платформе. Возникает проблема выбора способа объединения информационных потоков между различными программными средствами на предприятии, одним из решений является использование PLM-системы. При дискретном типе производства их использование затрудняется из-за большого количества программных решений в основных, организационных, обеспечивающих и бизнес-процессах развития. В этой связи необходимо определить оптимальную систему, которая отвечала бы всем необходимым критериям для построения единой платформы для «бесшовного» процесса производства. На рынке представлено большое количество PLM-систем, которые реализуют определенное количество функциональных назначений. Выбор PLM-системы должен основываться на удовлетворении взаимосвязанного набора требований предприятия, которые определяются на основе бизнес-процессов. Это позволит отсеять большинство неподходящих программных решений. Вместе с тем, PLM-система, которая имеет самый широкий функционал, может не отвечать требованиям по созданию системы обеспечения интеграции с системами автоматизации бизнес-процессов, а также экономическим, социальным, политическим и иным требованиям. Поэтому для каждого предприятия составляется индивидуальный перечень критериев в соответствии со сферой деятельности, типом производства, наличием различного программного обеспечения, существующим уровнем автоматизации и другими параметрами. В статье предлагается методика выбора PLM-систем для машиностроительного предприятия радиоэлектронного профиля, основанная на анализе бизнес-процессов и определении требований предприятия, проведении мониторинга и определении функциональных назначений PLM-систем и определении оптимального варианта на основе метода анализа иерархий. Получение интегральной оценки определенных критериев и вариантов программных решений позволит произвести объективный выбор оптимальной PLM-системы для конкретного предприятия. Использование методики позволит ускорить и повысить качество процесса выбора PLM-системы для определенного предприятия в условиях перехода к шестому технологическому укладу и индустрии 4.0.

Цифровая трансформация; PLM-система; процессное управление; метод анализа иерархий; машиностроительное предприятие.

P.A. Voronin, A.M. Belevtsev, F.G. Sadreev

METHODOLOGY FOR EVALUATION AND SELECTION OF PLM-SYSTEMS IN THE DIGITAL TRANSFORMATION OF A MACHINE-BUILDING ENTERPRISE

As part of the transition to the sixth technological order, the question of effective implementation of the digital transformation of the enterprise is acute for industrial enterprises. To do this, enterprises need to ensure the transition to process management, automate business processes and integrate all processes, applications and data on a single platform. There is a problem of choosing a way to combine information flows between various software tools in the enterprise, one of the solutions is the use of a PLM system. With a discrete type of production, their use is difficult due to the large number of software

solutions in the main, organizational, supporting and business development processes. In this regard, it is necessary to determine the optimal system that would meet all the necessary criteria for building a single platform for a "seamless" production process. There are a large number of PLM systems on the market that implement a certain number of functional appointments. The choice of a PLM system should be based on the satisfaction of an interconnected set of enterprise requirements, which are determined based on business processes. This will eliminate most unsuitable software solutions. At the same time, a PLM system that has the widest functionality may not meet the requirements for creating a system for ensuring integration with business process automation systems, as well as economic, social, political and other requirements. Therefore, an individual list of criteria is compiled for each enterprise in accordance with the field of activity, the type of production, the availability of various software, the existing level of automation and other parameters. The article proposes a methodology for choosing PLM systems for a radio-electronic engineering enterprise, based on the analysis of business processes and determining the requirements of the enterprise, monitoring and determining the functional purposes of PLM systems and determining the optimal option based on the hierarchy analysis method. Obtaining an integrated assessment of certain criteria and software solutions options will allow an objective choice of the optimal PLM system for a particular enterprise. The use of the methodology will speed up and improve the quality of the PLM system selection process for a certain enterprise in the conditions of transition to the sixth technological order and industry 4.0.

Digital transformation; PLM system; process management; hierarchy analysis method; machine-building enterprise.

Введение. В настоящее время существует тренд цифровой трансформации высокотехнологичных предприятий. Она требует фундаментального преобразования бизнес-моделей и бизнес-возможностей [1]. Одним из этапов цифровой трансформации является внедрение PLM-систем [2], при выборе которых, появляются следующие вопросы:

- ◆ выбор типов технологических процессов;
- ◆ выбор типа производства (непрерывное, дискретное);
- ◆ оценка существующего уровня автоматизации и информационной поддержки изделий;
- ◆ обеспечение интеграции архитектуры PLM-систем.

Решение этих проблем усложняются тем, что на предприятиях не осуществлен переход к процессному управлению и, в настоящее время, автоматизация бизнес-процессов и внедрение low-code платформ находятся на стадии формирования и разработки [3].

В этой связи методика выбора PLM-систем для заданного типа предприятий является актуальной.

Основная часть. Для решения проблемных вопросов предлагается методика, которая позволит более обоснованно подойти к выбору PLM-системы для заданного предприятия (рис. 1)

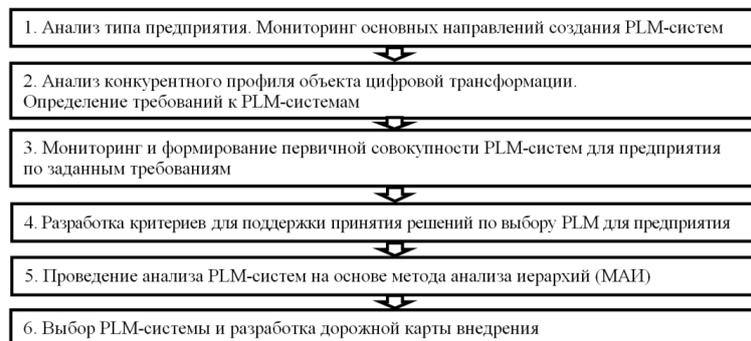


Рис. 1. Методика выбора PLM-системы для предприятий

Существует два типа производств: дискретное и непрерывное. Производство предприятия радиоэлектронного профиля является дискретным [4]. Т.е. в процесс изготовления изделий включены цеха химической обработки, обработки металла, проектирования и изготовления печатных плат, лакокрасочного покрытия, штамповки, литья, а также отделы проектирования, конструирования, опытного производства, пост гарантийного обслуживания и прочие.

Исходя из этого, для PLM-системы можно выделить следующие технические требования:

1) Возможность интеграции с наибольшим количеством программного обеспечения (ПО), имеющегося на предприятии [5]. PLM-система должна быть связывающим звеном между другими системами, уже встроенными в бизнес-процессы предприятия;

2) Возможность развертывания в распределенном режиме. Необходимо для организации информационных потоков между различными филиалами предприятия;

3) Наличие системы работы с клиентами (CRM). Необходимо для организации оперативной обратной связи с клиентами [6];

4) Наличие Workflow. Необходимо для моделирования и управления потоками работ, процессов [7];

5) Организация системы единого хранения данных. Необходимо для хранения данных в систематизированном виде;

6) Защищенность передачи данных. Необходимо для сохранности данных во время хранения и передачи информации [8];

7) Возможность развертывания в различных операционных системах. Необходимо для гибкого развертывания в условиях импортозамещения, разработки шлюзов;

8) Открытость интерфейса и встроенные языки программирования. Необходимо для разработки шлюзов между PLM-системой и другим программным обеспечением, доработки интерфейсных форм [9];

9) Полнота функциональных назначений. Обобщенное требование для проведения цифровой трансформации [10];

Помимо технических требований необходимо выделить и экономические требования:

1) Стоимость покупки и эксплуатации. Необходимо установить затраты на внедрение и поддержку во время эксплуатации PLM-системы;

2) Зависимость от зарубежного ПО. Важно в современных условиях санкций [11];

3) Сопровождение в эксплуатации. Необходимо для настройки информационных потоков на этапе введения PLM-системы, а также во время работы, например, при изменении какого-либо бизнес-процесса [12].

Для предприятий ОПК в современных условиях использование целого ряда PLM систем (Siemens, Dassault System, Oracle), разрабатываемых странами, подержавшими санкции против России, затруднено и в ходе исследований не рассматривалось. В рамках исследования были проанализированы следующие PLM-системы: Союз PLM, Intermech Professional Solutions (IPS), T-Flex, «Цифровое предприятие», ЛОЦМАН PLM, Крок КСЭД, Appius-PLM. Результаты анализа по критериям для выбора PLM-систем представлены в табл. 1 [13–20].

Таблица 1

Анализ PLM-систем

Наименование PLM / Критерии	Разработчик	Основные модули *Указаны модули, которые имеют не все PLM-системы	Интеграция с другими системами	Языки программирования	ОС
Союз PLM	Программсоюз (РФ)	-Workflow -Модуль статистики	SolidWorks, КОМПАС-3D, AutoCAD, Autodesk Inventor, T-FLEX CAD, MS Word, Excel Altium Designer IC, SAP R3	C#	Windows
Intermech Professional Solutions (IPS)	ИНТЕРМЕХ (Республика Беларусь)	-Workflow - CRM -Модуль статистики	AutoCAD, BricsCAD, Inventor, NX, Pro/Engineer, Creo, Solid Edge, Solidworks, КОМПАС-3D, Mentor Graphics, Altium Designer IC	C++, C#, Visual Basic, Delphi, F#	Windows
T-Flex PLM	ЗАО «Топ Системы» (РФ)	-Workflow -Собственная САПР - CRM	T-FLEX CAD, Solid Works, Autodesk Inventor, AutoCAD, Компас, Pro/E, Creo, Catia, Siemens NX; IC, Infor ERP, Галактика ERP и Microsoft Dynamics AX	C# (.NET Framework)	Windows, Linux
Система «Цифровое предприятие»	«РосАтом», «Ростех» и РФЯЦ-ВНИИЭФ (РФ)	-Workflow - CRM	Web-браузеры, Pentaho, СУБД «Синергия», Синергия 1.0 (CAD/CAE/PDM),	Java, Python	Astra Linux, Синергия, Windows
ЛЮЦМАН PLM	АСКОН (РФ)	-Workflow -Собственная САПР -Модуль статистики	PostgreSQL, КОМПАС-3D, ВЕРТИКАЛЬ, ПОЛИНОМ:MDM, SolidWorks, Autodesk Inventor, CATIA, Creo, NX, Altium Designer, Delta Design	Visual C++, Visual C#	Windows
КРОК КСЭД	КРОК (РФ)	- CRM -Модуль статистики	PostgreSQL, Microsoft SQL Server, Oracle, IBM DB2 SAP, IC	Java	Windows, Unix/Linux
Аппиус-PLM	ГК АРПИУС (РФ)	-Workflow -Модуль статистики	КОМПАС 3D, SOLIDWORKS, Solid Edge, Inventor и T-FLEX CAD Altium Designer, EPLAN,P-CAD, OrCAD.,IC	JavaScript	Windows

Мониторинг PLM-систем по полноте функциональных назначений приведен в табл. 2. Однако, данный способ не позволяет дать окончательный вариант PLM-системы. Необходима интегральная оценка для сопоставления всех групп критериев и самих критериев. В данной статье для оценки и выбора PLM-системы был выбран метод анализа иерархий.

Таблица 2

Мониторинг PLM-систем по полноте функциональных назначений

Наименование PLM/ Функциональные назначения	Интеграция с другими системами	Workflow	Работа в распределенном режиме	Организация хранения данных	Защищенность передачи данных	Независимость от зарубежного ПО
Союз PLM						
IPS						
T-Flex PLM						
Система «ЦП»						
ЛОЦМАН PLM						
КРОК КСЭД						
Аппиус-PLM						

Критерии для оценки приоритетной PLM-системы для предприятия радиоэлектронного профиля представлены на рис. 2.

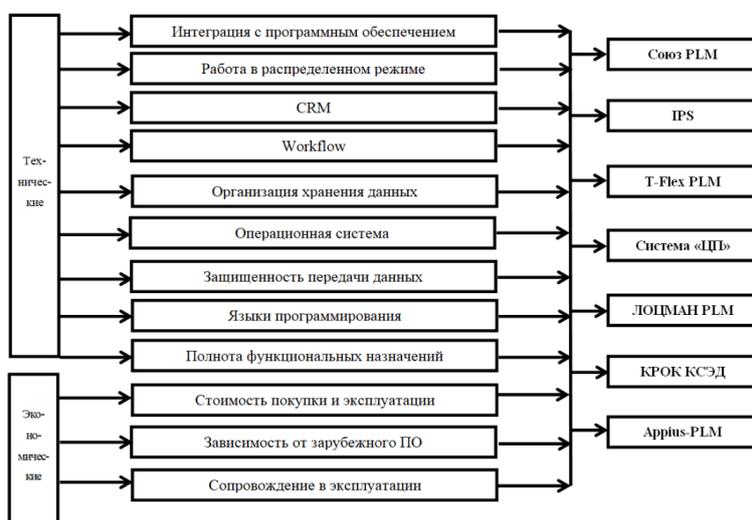


Рис. 2. Критерии для оценки приоритетной PLM-системы

Процедура проведения анализа PLM-систем на основе МАИ приведена на рис. 3 [21].

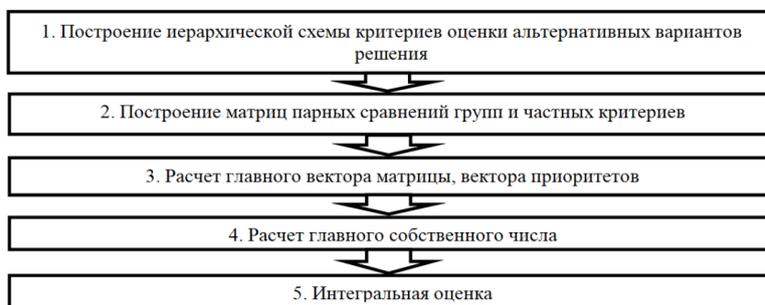


Рис. 3. Процедура МАИ

Исходя из проведенного анализа существует 7 вариантов выбора PLM-систем.

- V1 – Союз PLM
- V2 – IPS
- V3 – T-Flex PLM
- V4 – Система «Цифровое Предприятие»
- V5 – Лоцман PLM
- V6 – Крок КСЭД
- V7 – Appius-PLM

Необходимо установить множество критериев для предприятия радиоэлектронного профиля:

- K1 - Интеграция с программным обеспечением;
- K2 - Работа в распределенном режиме;
- K3 – Наличие системы CRM;
- K4 – Наличие Workflow;
- K5 – Организация хранения данных;
- K6 – Операционная система;
- K7 – Защищенность передачи данных;
- K8 – Языки программирования;
- K9 – Полнота функциональных назначений;
- K10 – Стоимость покупки и эксплуатации;
- K11 – Зависимость от зарубежного ПО;
- K12 – Сопровождение в эксплуатации.

Матрица парных сравнений критериев приведена в табл. 3, матрица веса альтернатив – в табл. 4. Интегральные оценки приоритетов PLM-систем приведены на рис. 4.

Таблица 3

Матрица парных сравнений критериев

	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12
K1	1,000	3,000	7,000	5,000	1,000	3,000	1,000	3,000	1,000	5,000	3,000	3,000
K2	0,333	1,000	5,000	3,000	0,333	1,000	0,333	1,000	0,333	3,000	1,000	1,000
K3	0,143	0,200	1,000	0,333	0,143	0,200	0,143	0,200	0,143	0,333	0,200	0,200
K4	0,200	0,333	3,000	1,000	0,200	0,333	0,200	0,333	0,200	1,000	0,333	0,333
K5	1,000	3,000	7,000	5,000	1,000	3,000	1,000	3,000	1,000	5,000	3,000	3,000
K6	0,333	1,000	5,000	3,000	0,333	1,000	0,333	1,000	0,333	3,000	1,000	1,000
K7	1,000	3,000	7,000	5,000	1,000	3,000	1,000	3,000	1,000	5,000	3,000	3,000
K8	0,333	1,000	5,000	3,000	0,333	1,000	0,333	1,000	0,333	3,000	1,000	1,000
K9	1,000	3,000	7,000	5,000	1,000	3,000	1,000	3,000	1,000	5,000	3,000	3,000
K10	0,200	0,333	3,000	1,000	0,200	0,333	0,200	0,333	0,200	1,000	0,333	0,333
K11	0,333	1,000	5,000	3,000	0,333	1,000	0,333	1,000	0,333	3,000	1,000	1,000
K12	0,333	1,000	5,000	3,000	0,333	1,000	0,333	1,000	0,333	3,000	1,000	1,000

Таблица 4

Матрица веса альтернатив

	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12
B1	0,133	0,059	0,067	0,159	0,065	0,065	0,077	0,057	0,050	0,200	0,067	0,067
B2	0,133	0,176	0,200	0,159	0,065	0,065	0,077	0,364	0,137	0,200	0,067	0,200
B3	0,342	0,176	0,200	0,159	0,065	0,159	0,077	0,061	0,351	0,067	0,333	0,200
B4	0,049	0,176	0,200	0,159	0,384	0,385	0,231	0,252	0,137	0,067	0,333	0,067
B5	0,176	0,176	0,067	0,159	0,179	0,065	0,231	0,150	0,137	0,067	0,067	0,200
B6	0,046	0,059	0,200	0,048	0,179	0,181	0,231	0,057	0,050	0,200	0,067	0,067
B7	0,119	0,176	0,067	0,159	0,065	0,079	0,077	0,057	0,137	0,200	0,067	0,200



Рис. 4. Интегральные оценки приоритетов PLM-систем

На основе анализа делаем вывод, что наиболее предпочтительными PLM-системами для предприятий радиоэлектронного профиля будут:

- 1) Система «Цифровое Предприятие»;
- 2) T-Flex PLM;
- 3) ЛОЦМАН PLM.

Выводы. Представленные системы дают возможность «запустить» процесс цифровой трансформации предприятия, охватывая весь цикл изготовления изделия. Модули PLM-систем позволяют обрабатывать всю информацию согласно типовым бизнес-процессам. Но шаблоны и инструменты для их настройки в PLM-системах не позволяют в полной мере автоматизировать бизнес-процессы в режиме «как надо», так как не имеют на данный момент достаточно гибкого функционала для создания прикладных приложений (модулей) для автоматизации требуемых бизнес-процессов. На рынке ПО одним из трендов разработки становится создание low-code платформ-конструкторов, которые дополняют PLM-системы, позволяющие решить данную проблему.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Томас Сибел. Цифровая трансформация. Как выжить и преуспеть в новую эпоху. – М.: Изд-во «МИФ», 2020.
2. Ераскин Н.Г., Иванова М.А., Салтыкова Н.В. Внедрение концепции PLM на предприятии // Молодой ученый. – 2021. – № 2. – С. 14-17.
3. Воронин П.А., Белевцев А.М., Епанешникова И.К., Садреев Ф.Г. Разработка общих процедур интеграции информационного взаимодействия plm-систем и low-code платформ-конструкторов // Всероссийская научно-практическая конференция молодых специалистов «Системы управления полным жизненным циклом высокотехнологичной продукции в машиностроении: новые источники роста». – 2021.

4. *Бутусов Н.О., Буза М.Н., Алехин А.Г.* Разработка автоматизированной системы управления производством дискретного типа // Известия Волгоградского государственного технического университета. – 2022.
5. *Капустина А.И.* Интеграция систем программного и математического обеспечения для моделирования и исследования автоматизированных систем управления // ЭНЕРГИЯ-2021: Шестнадцатая всероссийская (восьмая международная) научно-техническая конференция студентов, аспирантов и молодых ученых. В 6 т. Иваново, 2021.
6. *Якубова В.А.* Исследование процессов взаимодействия с потребителями в рамках CRM-системы предприятия по производству и реализации мебельных изделий // Студенческий вестник. – 2022.
7. *Ифанов А.В., Полтавцев А.А.* Технологии workflow в создании распределенных систем обработки заказов // Проблемы информатики в образовании, управлении, экономике и технике: Сб. статей XXI Международной научно-технической конференции. – Пенза: Пензенский государственный университет, 2021.
8. *Кулаченко И.А.* Системы информационной защиты при передаче данных. – Научно-исследовательский центр "Вектор развития", 2022.
9. *Власенко А.Н., Александрова С.С., Садреев Ф.Г., Воронин П.А.* Автоматизация создания инженерных учебных webgl приложений // Тр. Международной конференции по компьютерной графике и зрению "Графикон". – 2021.
10. *Логинов И.В.* Диагностирование функциональной пригодности развивающихся многофункциональных автоматизированных систем на основе перестраиваемой модели // Программные продукты и системы. – 2022. – Т. 35, № 2. – С. 195-207.
11. *Синицына Е.В., Гусев А.А.* Проблема импортозамещения в сфере микроэлектроники и создания российского аппаратного обеспечения // Вестник Института мировых цивилизаций. – 2022.
12. *Хан В.Л.* Сопровождение систем автоматизации программного обеспечения // Перспективы развития информационных технологий. – 2011. – № 3-2. – С. 152-156.
13. ЛОЦМАН: PLM – Аскон. – URL: <https://ascon.ru/>.
14. T-Flex PLM – «ТОП Системы». – URL: <https://www.tflex.ru/plm/>.
15. Союз-PLM – Програмсоюз. – URL: <http://www.programsoyuz.ru/products/system-soyuz-plm/>.
16. IPS – Интермех. – URL: https://intermech.ru/ips_product.html.
17. КСЭД – КРОК. – URL: <https://www.croc.ru/business-solutions/content-management-and-workflow/>.
18. Аппиус – ГК Аппиус. – URL: <https://www.appius.ru/>.
19. Система полного жизненного цикла. – URL: <https://nnov.hse.ru/mirror/pubs/share/226534016>.
20. Система «Цифровое предприятие». – URL: <https://цифровоепредприятие.рф>.
21. *Балыбердин В.А., Белевцев А.М., Бендерский Г.П.* Прикладные методы оценки и выбора решений в стратегических задачах инновационного менеджмента. – М.: Дашков и К, 2014. – 240 с.

REFERENCES

1. *Tomas Sibel.* Tsifrovaya transformatsiya. Kak vyzhit' i preuspet' v novuyu epokhu [Digital transformation. How to survive and succeed in the new era]. Moscow: Izd-vo «MIF», 2020.
2. *Eraskin N.G., Ivanova M.A., Saltykova N.V.* Vnedrenie kontseptsii PLM na predpriyatii [Introduction of the PLM concept at the enterprise], *Molodoy uchenyy* [Young scientist], 2021, No. 2, pp. 14-17.
3. *Voronin P.A., Belevtsev A.M., Epaneshnikova I.K., Sadreev F.G.* Razrabotka obshchikh protsedur integratsii informatsionnogo vzaimodeystviya plm-sistem i low-code platform-konstruktorov [Development of common procedures for the integration of information interaction between PLM systems and low-code design platforms], *Vserossiyskaya nauchno-prakticheskaya konferentsiya molodykh spetsialistov «Sistemy upravleniya polnym zhiznennym tsiklom vysokotekhnologichnoy produktsii v mashinostroenii: novye istochniki rosta»* [All-Russian Scientific and Practical conference of young specialists "Management systems for the full life cycle of high-tech products in mechanical engineering: new sources of growth"], 2021,
4. *Butusov N.O., Buza M.N., Alekhin A.G.* Razrabotka avtomatizirovannoy sistemy upravleniya proizvodstvom diskretnogo tipa [Development of an automated production management system of discrete type], *Izvestiya Volgogradskogo gosudarstvennogo tekhnicheskogo universiteta* [Izvestiya Volgograd State Technical University], 2022.

5. *Kapustina A.I.* Integratsiya sistem programmno i matematicheskogo obespecheniya dlya modelirovaniya i issledovaniya avtomatizirovannykh sistem upravleniya [Integration of software and mathematical support systems for modeling and research of automated control systems], *ENERGIYA-2021: S Hestnadsataya vserossiyskaya (vos'maya mezhdunarodnaya) nauchno-tekhnicheskaya konferentsiya studentov, aspirantov i molodykh uchenykh* [ENERGIA-2021: The sixteenth All-Russian (eighth International) Scientific and Technical Conference of students, postgraduates and young scientists]. In 6 vol. Ivanovo, 2021.
6. *Yakubova V.A.* Issledovanie protsessov vzaimodeystviya s potrebitelyami v ramkakh CRM-sistemy predpriyatiya po proizvodstvu i realizatsii mebel'nykh izdeliy [Research of the processes of interaction with consumers within the CRM system of the enterprise for the production and sale of furniture products], *Studencheskiy vestnik* [Student Bulletin], 2022.
7. *Ifanov A.V., Poltavtsev A.A.* Tekhnologii workflow v sozdanii raspredelennykh sistem obrabotki zakazov [Workflow technologies in the creation of distributed order processing systems], *Problemy informatiki v obrazovanii, upravlenii, ekonomike i tekhnike: Sb. statey XXI Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii* [Problems of informatics in education, management, economics and technology: Collection of articles of the XXI International Scientific and Technical Conference]. Penza: Penzenskiy gosudarstvennyy universitet, 2021.
8. *Kulachenko I.A.* Sistemy informatsionnoy zashchity pri peredache dannykh [Information security systems for data transmission]. Nauchno-issledovatel'skiy tsentr "Vektor razvitiya", 2022.
9. *Vlasenko A.N., Aleksandrova S.S., Sadreev F.G., Voronin P.A.* Avtomatizatsiya sozdaniya inzhenernykh uchebnykh webgl prilozheniy [Automating the creation of engineering educational webgl applications], *Tr. Mezhdunarodnoy konferentsii po komp'yuternoy grafike i zreniyu "Grafikon"* [Proceedings of the International Conference on Computer Graphics and Vision "Graphicon"], 2021.
10. *Loginov I.V.* Diagnostirovanie funktsional'noy prigodnosti razvivayushchikhsya mnogo-funktsional'nykh avtomatizirovannykh sistem na osnove perestraivaemoy modeli [Diagnosing the functional suitability of developing multi-functional automated systems based on a tunable model], *Programmnye produkty i sistemy* [Software products and systems], 2022, Vol. 35, No. 2, pp. 195-207.
11. *Sinitsyna E.V., Gusev A.A.* Problema importozameshcheniya v sfere mikroelektroniki i sozdaniya rossiyskogo apparatnogo obespecheniya [The problem of import substitution in the field of microelectronics and the creation of Russian hardware], *Vestnik Instituta mirovykh tsivilizatsiy* [Bulletin of the Institute of World Civilizations], 2022.
12. *Khan V.L.* Soprovozhdenie sistem avtomatizatsii programmno obespecheniya [Maintenance of software automation systems], *Perspektivy razvitiya informatsionnykh tekhnologiy* [Prospects for the development of information technologies], 2011, No. 3-2, pp. 152-156.
13. LOTSMAN: PLM – Askon [PILOT: PLM – Askon]. Available at: <https://ascon.ru/>.
14. T-Flex PLM – «TOP Sistemy» [T-Flex PLM – "TOP Systems"]. Available at: <https://www.tfex.ru/plm/>.
15. Soyuz-PLM – Programsoyuz [Union-PLM – Program Union]. Available at: <http://www.programsoyuz.ru/products/system-soyuz-plm/>.
16. IPS – Intermech [IPS – Intermech]. Available at: https://intermech.ru/ips_product.html.
17. KSED – KROK [CSED – CROC]. Available at: <https://www.croc.ru/business-solutions/content-management-and-workflow/>.
18. Appius – GK Appius [Appius – GC Appius]. Available at: <https://www.appius.ru/>.
19. Sistema polnogo zhiznennogo tsikla [A full life cycle system]. Available at: <https://nnov.hse.ru/mirror/pubs/share/226534016>.
20. Sistema «Tsifrovoe predpriyatie» [The Digital Enterprise system]. Available at: <https://tsifrovoepredpriyatie.rf>.
21. *Balyberdin V.A., Belevtsev A.M., Benderskiy G.P.* Prikladnye metody otsenki i vybora resheniy v strategicheskikh zadachakh innovatsionnogo menedzhmenta [Applied methods of evaluation and choice of solutions in strategic tasks of innovation management]. Moscow: Dashkov i K, 2014, 240 p.

Статью рекомендовал к опубликованию д.т.н. К.Н. Климов.

Воронин Павел Александрович – Московский авиационный институт (национальный исследовательский университет); e-mail: heyholyrampage@yandex.ru; г. Москва, Россия; тел.: +79857048284; ассистент.

Белевцев Андрей Михайлович – e-mail: ambelevtsev@yandex.ru; тел.: +79037691788; д.т.н.; профессор.

Садреев Фарид Гайнутдинович – e-mail: sadreevfg@yandex.ru; тел.: +79773635011; к.т.н.; доцент.

Voronin Pavel Aleksandrovich – Moscow Aviation Institute (National Research University); e-mail: heyholyrampage@yandex.ru; Moscow, Russia; phone: +79857048284; assistant;

Belevtsev Andrey Michailovitch – e-mail: ambelevtsev@yandex.ru; phone: +79037691788; dr. of eng. sc.; professor.

Sadreev Farid Gainutdinovich – e-mail: sadreevfg@yandex.ru; phone: +79773635011; cand. of eng. sc.; associate professor.

УДК 007.519.816

DOI 10.18522/2311-3103-2022-4-15-29

Е.М. Герасименко, В.В. Курейчик, С.И. Родзин, А.П. Кухаренко
ПРИМЕНЕНИЕ НЕЧЕТКОЙ ЛОГИКИ ДЛЯ ПРИНЯТИЯ РЕШЕНИЙ
ОБ ЭВАКУАЦИИ ПРИ НАВОДНЕНИИ*

Речь идет о стихийных бедствиях, таких как наводнение, которые можно спрогнозировать за несколько часов до того, как они произойдут, чтобы можно было организовать эвакуацию населения. Эвакуация означает, что люди в районах бедствия должны покинуть эти районы и добраться до укрытий или убежищ. Представлен анализ процесса принятия решения об эвакуации, основные критерии, определяющие решение и основные этапы применения нечеткой логики для принятия решения об эвакуации на основе качественных и количественных значений критериев принятия решения. Эти этапы включают выбор критериев, определение качественных входных и выходных переменных, фаззификацию переменных, определение базы нечетких правил, построение нечеткого вывода, визуализацию результатов и анализ чувствительности. При моделировании учитывались следующие критерии: прогнозируемый уровень наводнения, уровень опасности, уязвимость района предполагаемого наводнения и возможность безопасной эвакуации. Прогнозируемый уровень наводнения основывался на параметрах максимального уровня и скорости подъема воды. Уровень опасности отражал физические характеристики наводнения и его потенциальное воздействие на безопасность людей в районе наводнения. Уязвимость района предполагаемого наводнения определялась как неспособность на местном уровне предотвратить непосредственный контакт людей с наводковыми водами во время события. Возможность безопасной эвакуации определялась как множество ограничений и потенциальных негативных аспектов, которые могут задержать или помешать успешному проведению эвакуации. Представлено описание качественных переменных критериев принятия решения о необходимости эвакуации, примеры определения базы нечетких правил. Нечеткая модель реализована с помощью Matlab Fuzzy Logic Toolbox. Описана процедура нечеткого вывода и интерпретации решения и модель нескольких сценариев и ситуаций наводнения. Рассмотрен способ, с помощью которого нечеткая модель принятия решения об эвакуации может быть применена в сочетании с геоинформационной системой. Представлены действия, связанные с необходимостью эвакуации для различных сценариев и обстоятельств.

Принятие решений, эвакуация, чрезвычайная ситуация, факторы неопределенности, нечеткая логика, «мягкие» вычисления.

* Исследование выполнено за счет гранта Российского научного фонда № 22-71-10121, <https://rscf.ru/project/22-71-10121/> в Южном федеральном университете.

E.M. Gerasimenko, V.V. Kureichik, S.I. Rodzin, A.P. Kukhareenko

APPLICATION OF FUZZY LOGIC FOR MAKING DECISIONS ABOUT EVACUATION IN CASE OF FLOODING

We are talking about natural disasters, such as flooding, which can be predicted a few hours before they occur so that evacuation of the population can be organized. Evacuation means that people in disaster areas must leave these areas and reach shelters. The article presents an analysis of the decision-making process on evacuation, the main criteria determining the decision and the main stages of using fuzzy logic to make a decision on evacuation based on qualitative and quantitative values of the decision-making criteria. These stages include selection of criteria, determination of qualitative input and output variables, fuzzification of variables, definition of the base of fuzzy rules, construction of fuzzy inference, visualization of results and sensitivity analysis. When modeling, the following criteria were taken into account: the predicted flood level, the level of danger, the vulnerability of the area of the expected flood and the possibility of safe evacuation. The predicted flood level was based on the parameters of the maximum level and the rate of water rise. The hazard level reflected the physical characteristics of the flood and its potential impact on the safety of people in the flood area. The vulnerability of the area of the expected flood was defined as the inability at the local level to prevent people from direct contact with flood waters during the event. The possibility of safe evacuation was defined as a set of limitations and potential negative aspects that could delay or hinder the successful evacuation. The description of qualitative variable criteria for making a decision on the need for evacuation, examples of determining the base of fuzzy rules are presented. The fuzzy model is implemented using Matlab Fuzzy Logic Toolbox. The procedure of fuzzy inference and interpretation of the solution and a model of several scenarios and flood situations are described. The method by which a fuzzy model of decision-making on evacuation can be applied in combination with a geoinformation system is considered. The actions related to the need for evacuation for various scenarios and circumstances are presented.

Decision-making; evacuation; emergency; uncertainty factors; fuzzy logic; "soft" calculations.

Введение. Для эффективного управления чрезвычайной ситуацией при наводнении необходимо сочетание экспертных знаний и информации с местной ситуации [1]. Нечеткая логика позволяет интегрировать экспертные знания и информацию с мест бедствия [2].

Интерес представляет исследование многокритериальной методологии принятия решений с использованием нечеткой логики для оценки необходимости эвакуации путем объединения качественных и количественных значений критериев принятия решения [3]. Термин "необходимость эвакуации" (*Necessity to Evacuate, NTE*) получил международное признание и определяется как число в диапазоне от 0 до 1, которое указывает на уровень потенциальной потребности в эвакуации в районе стихийного бедствия или техногенной катастрофы. Чем выше значение *NTE*, например при наводнении, тем более необходимой является эвакуация с целью защиты местного населения. *NTE* представляет синтетическую информацию, основанную на прогнозе потенциальной опасности для людей, а также возможности безопасной эвакуации в случае чрезвычайной ситуации. С помощью геоинформационной системы (ГИС) *NTE* может быть отображен на картах, которые используются в качестве инструмента поддержки принятия решений для организации и управления эвакуацией.

В статье представлен анализ процесса принятия решения об эвакуации, основные критерии, определяющие решение и процедура применения нечеткой логики для принятия решения об эвакуации на основе *NTE*.

Проблемы принятия решения об эвакуации и нечеткая логика. Речь идет о стихийных бедствиях, которые можно спрогнозировать за несколько часов до того, как они произойдут в подверженных риску районах, чтобы можно было ор-

ганизовать превентивную эвакуацию населения. Эвакуация означает, что люди в районах бедствия должны покинуть эти районы и добраться до укрытий или убежищ. В некоторых случаях укрытия могут находиться внутри зоны риска (вертикальная эвакуация). Организация антикризисного управления не может быть импровизированной, особенно в случае массовой эвакуации, для которой должны быть заранее подготовлены конкретные планы задолго до кризиса. Эти планы могут использоваться в качестве поддержки принятия решений на начальном этапе, но их недостаточно для учета обновляемых данных в режиме реального времени.

Большинство существующих методов и инструментов для подготовки и управления эвакуацией были разработаны для речных наводнений и ураганов [4]. Если прогнозируется наводнение, то власти в режиме реального времени должны ответить на основные вопросы относительно решения об эвакуации:

- ◆ каков реальный уровень опасности для жизни людей в подверженных риску районах и требуется ли эвакуация из этих районов?
- ◆ насколько безопасно может быть проведена эвакуация?
- ◆ способны ли местные органы власти с точки зрения ресурсов и времени эвакуировать часть или все население, находящееся в зоне риска, за время до наступления события?

Необходимость принятия решения об эвакуации вызвана потенциально катастрофической угрозой в виде сильного наводнения. Гидрографическая информация об угрозе наводнения сравнивается с заранее определенными пороговыми значениями за определенный период наблюдений. В системах прогнозирования и предупреждения используются следующие уровни предупреждения: красный, оранжевый, желтый и т.д., которые указывают на прогнозируемый уровень риска (высокий, умеренный, незначительный).

Требуется анализ факторов риска наводнения и его потенциальных последствий для ответа на вопрос об уровне риска, при котором эвакуация является оправданной с точки зрения безопасности. Эвакуация уменьшает риски угрозы жизни и движимому имуществу. Поэтому понимание риска наводнения и опасности для жизни является основным вопросом при принятии решения об эвакуации. Угроза жизни людей вызвана сочетанием уязвимости места нахождения людей (внутри/снаружи здания, характер жилья), характеристик наводнения (уровень подъема воды, его скорость) и характеристик населения (возраст, состояние здоровья).

Ответ на вопрос о безопасности эвакуации в основном связан с погодными условиями, временным фактором и безопасностью инфраструктуры. Погодные условия (дождь, туман, снег, ветер) влияют на безопасность дорожного движения. Время, как и состояние дорожной сети, возможности местных органов власти, также являются важными факторами при эвакуации.

Принятие решения об эвакуации зависит от множества определенных и неопределенных факторов:

- ◆ прогнозные и имитационные модели, основанные на местных уровнях воды и картировании рисков;
- ◆ фактическая доступность эвакуационных ресурсов;
- ◆ вероятность отказа сетей, аварий, нехватки топлива.

Для работы с множеством факторов принятия решений об эвакуации в сочетании экспертными знаниями целесообразно использовать нечеткую логику, позволяющую моделировать процесс принятия решения об эвакуации с учетом неопределенности.

Необходимо различать стохастическую неопределенность и субъективную неопределенность. Стохастическая неопределенность связана с тем, что информация по отдельным аспектам решаемой задачи и используемому инструментарию

носит вероятностный характер. Субъективная неопределенность означает нечеткость (размытость) некоторых представлений о проблеме и способах ее решения. Этот тип неопределенности возникает из-за лингвистической неточности, когда люди используют неточные понятия или выражения для описания свойств концепта. Разница в том, что субъективная неопределенность связана с человеческими определениями, а не с объективным фактом природы. Однако утверждения, использующие субъективные категории, играют важную роль в процессах принятия решений человеком [5].

Нечеткая логика дает соответствующие описания субъективной неопределенности с помощью нечетких множеств. Под нечётким множеством A понимается совокупность упорядоченных пар, составленных из элементов x универсального множества X и соответствующих степеней принадлежности $\mu_A(x)$:

$$A = \{(x, \mu_A(x)) | x \in X, \mu_A(x) \in [0,1],$$

Функция принадлежности $\mu_A(x)$ – это непрерывная кривая, которая определяет степень принадлежности любой числовой переменной к лингвистической переменной. Предложено множество различных типов функций принадлежности (*MFS*) для нечеткой логики: z -, λ -, π - и s -типа [6]. Для практических реализаций чаще всего используются функции принадлежности с линейной или сплайновой формами, как более простые и достаточно точные функции для моделирования большинства переменных в системах поддержки принятия решений. Согласно психолингвистическим исследованиям человеческой классификации качественных переменных, базовые функции принадлежности сплайнов неплохо моделируют лингвистические отношения для сложных приложений анализа данных и поддержки принятия решений [7].

В области планирования и моделирования эвакуации принадлежность к λ -типу и π -типу (треугольные и трапециевидные) функции используются для проектирования систем нечеткой логики. Например, в [8] использовалась функция π -типа для разработки системы нечеткой логики, которая моделировала поведение человека во время эвакуации во время урагана.

Предметом нечёткой логики является исследование рассуждений в условиях нечёткости, размытости. В некоторых обстоятельствах нечеткое высказывание, например “осторожно”, является более четким и эффективным, нежели попытки количественного описания ситуаций для принятия решения человеком, находящимся в опасности. Принципы нечеткой логики основаны на известных навыках человеческого мышления путем синтеза и упрощения сложных задач принятия решений в условиях неопределенности [8].

В контексте многокритериального анализа для принятия решения об эвакуации необходимо определить, как различные нечеткие переменные будут “объединены”, чтобы сделать выводы или предложить решения, в зависимости от типа проблемы. Подход нечеткой логики предлагает определить эвристические или экспертные продукционные правила типа “ЕСЛИ-ТО” для выражения взаимосвязей между комбинациями нечетких входных переменных и выводами, которые следуют из их комбинаций. Основная проблема при использовании нечетких продукционных правил заключается в том, чтобы на основе некоторых нечетких высказываний с известной степенью истинности, которые являются условиями нечетких правил, оценить степень истинности других нечетких высказываний, являющимися выводами нечетких правил. Чтобы иметь возможность решить эту проблему, необходимо ответить на более частный вопрос: чему должна быть равна степень истинности вывода нечеткого правила, если известна степень истинности условия этого правила? Таким образом, в системах нечетких продукций центральное место занимает метод определения истинности заключений [9].

Взаимосвязь между условием и заключением в нечетком правиле в общем случае представляет собой бинарное нечеткое отношение на декартовом произведении универсумов соответствующих нечетких высказываний. Этот подход используется для определения различных методов нечеткого вывода в продукционных нечетких системах.

Перед созданием базы правил нечеткой модели необходимо определить нечеткие множества и функции принадлежности, чтобы согласовать значение нечетких переменных, исходя из экспертных знаний и опыта. Нечеткая логика содержит широкий спектр нечетких операций, которые могут быть выполнены с нечеткими отношениями и нечеткими множествами:

- ◆ $\mu(x) = \max(\mu_i)$ – нечеткое ИЛИ определяется как максимальное из значений *MFS*;
- ◆ $\mu(x) = \min(\mu_i)$ – нечеткое И определяется как минимальное из значений *MFS*;
- ◆ $\mu(x) = 1 - (\mu_i)$ – нечеткое НЕ определяется как аддитивное дополнение значений *MFS*;
- ◆ $\mu(x) = \text{product}(\mu_i)$ – нечеткое произведение определяется как алгебраическое произведение значений *MFS*;
- ◆ $\mu(x) = \sum \mu_i - \text{product}(\mu_i)$ – нечеткая сумма определяется как разность между алгебраической суммой *MFS* и алгебраическим произведением.

Нечеткие операции должны применяться к правилам вывода. Необходимо оперировать нечеткими входными данными и нечеткими отношениями для получения выходных данных, которые также представляют решение нечетким образом. Эта операция нечеткой импликации или нечеткого вывода ($A \rightarrow B$). Она не так проста и интуитивно понятна, как в классической логике. Ниже приведены некоторые наиболее часто используемые операторы нечеткой импликации:

- ◆ $\mu_{A \rightarrow B}(x, y) = \max(\min(\mu_A(x), \mu_B(x)), 1 - \mu_A(x))$ – импликация по Заде;
- ◆ $\mu_{A \rightarrow B}(x, y) = \min(\mu_A(x), \mu_B(x))$ – импликация по Мамдани;
- ◆ $\mu_{A \rightarrow B}(x, y) = \min(1, 1 - \mu_A(x), \mu_B(x))$ – импликация по Лукасевичу;
- ◆ $\mu_{A \rightarrow B}(x, y) = \begin{cases} 1, & \text{if } (\mu_A(x) \leq \mu_B(y)) \\ \mu_B(y) & \text{else} \end{cases}$ – импликация по Геделю;
- ◆ $\mu_{A \rightarrow B}(x, y) = \max(1 - \mu_A(x), \mu_B(x))$ – импликация по Дайнесу-Решеру;
- ◆ $\mu_{A \rightarrow B}(x, y) = \mu_A(x) \cdot \mu_B(x)$ – импликация по формуле произведения.

Основанные на правилах нечеткой импликации системы могут быть использованы для прогнозирования наводнений, управление рисками наводнений, моделирование поведения людей при эвакуации.

Нечеткая система, основанная на экспертных знаниях, включает в себя базу нечетких правил и механизм вывода. База нечетких правил "ЕСЛИ-ТО" – это знания и ноу-хау экспертов о конкретной проблеме, например, об эвакуации районов, подверженных риску, в случае сильного наводнения [10]. Механизм вывода интерпретирует базу правил с помощью *MFS* входных и выходных нечетких переменных, операторов импликации и методов дефаззификации. Система вывода на основе правил реализует отображение входных нечетких множеств в выходные нечеткие множества. Процесс фаззификации и дефаззификации используется для того, чтобы обеспечить возможность управления системой с нечеткими входами и нечеткими выходами.

Фаззификация преобразует входные параметры системы в «нечёткую» область, описываемую лингвистическими переменными в базе знаний с использованием *MFS*. Фаззификация не является строгой процедурой и выполняется экспертами на основе опыта и знаний.

Дефаззификация – это преобразование нечеткого множества в четкое число. Перед дефаззификацией выходных данных все нечеткие выходные данные системы объединяются с помощью оператора нечеткого максимума. Существуют различные методы дефаззификации: центр области под кривой *MFS* нечеткого выходного множества, биссектриса площади под кривой *MFS*, среднее значение нечеткого выходного множества среди максимальных значений *MFS*, наибольшее значение нечеткого выходного множества среди максимальных значений *MFS*, минимальное значение нечеткого выходного множества среди максимальных значений *MFS*. Выбор метода дефаззификации зависит от контекста проблемы принятия решения.

Применение нечеткой логики для принятия решений об эвакуации при наводнении. Окончательное решение о массовой эвакуации при наводнении сталкивается с двумя трудностями: (1) необходимо учитывать многочисленные и разнородные критерии при неточной и неполной информации; (2) большинство значений этих критериев имеют неопределенную границу принятия решения.

Примером трудности (1) может быть выбор маршрута эвакуации с учетом затоплений, условий дорожного движения (“легкие”, “тяжелые”, “бампер к бамперу”). Качественная информация, которая представляет собой субъективное знание, использующее лингвистические описания, должна обрабатываться при принятии решений. Однако такую информацию трудно обработать с помощью классической математической логики. Между тем, качественная информация не должна игнорироваться при принятии решений. Очевидно, что качественная информация о наводнении от незначительного до крупного, поскольку она была определена на основе существующих и подтвержденных пороговых значений, имеет больше смысла для ЛПР. Такую неопределенность в человеческом восприятии можно легко смоделировать с помощью нечеткой логики.

Для трудности (2) значения критериев могут включать некоторую неопределенность границ критериев, что может повлиять на принятие решения. Например, прогнозируемый уровень воды (например, от 4,85 до 5,05 м), варианты решения об эвакуации для конкретного района (“без эвакуации”, “бдительность”, “рекомендация”, “мягкий приказ”, “экстренная эвакуация”) с помощью нечеткой логики могут быть представлены через *MFS*.

Рассмотрим, как применяется нечеткая логика путем объединения качественной и количественной информации для оценки и принятия решения о необходимости эвакуации (*NTE*) в случае чрезвычайной ситуации из-за наводнения.

В нечеткой системе для принятия решения об эвакуации при наводнении критерии принятия решения выбираются в качестве входных данных, а *NTE* определяется как единственная выходная переменная, которая колеблется в интервале [0, 1].

Основными этапами проектирования нечеткой системы являются (1) выбор критериев; (2) определение качественных переменных критериев и *NTE*; (3) построение *MFS*; (4) определение нечетких правил; (5) определение нечеткого вывода; (6) визуализация и анализ чувствительности.

Вначале путем консультаций с экспертами, заинтересованными сторонами, анализом плана эвакуации определяются исходные данные (критериев принятия решения) с учетом прогноза наводнения, опасности для населения, расчетного времени эвакуации и т.д.

Затем необходимо уточнить количественные и качественные переменные для каждого критерия и *NTE*. Все диапазоны значений количественных переменных затем классифицируются для определения качественных переменных и множества их возможных значений. Эти переменные определяются с использованием экспертных знаний, статистических данных, исторических данных о наводнениях в данной местности и имеющихся планов эвакуации.

Далее, строятся *MFS* каждого критерия и *NTE*. Определяется экспертная база правил, описывающая нечеткие отношения между критериями принятия решений и *NTE*, которая интерпретируется с использованием процедуры нечеткого вывода для установления количественных отношений между критериями принятия решения и *NTE*. Чтобы облегчить анализ чувствительности *NTE* относительно входных данных рекомендуется визуализировать 3D-поверхности относительное влияние двух входных данных на *NTE*, чтобы принять окончательное решение об эвакуации.

Пространственное распределение конечного значения *NTE* может быть визуализировано на картах для получения более детальных решений, таких как определение приоритетов районов эвакуации. Проиллюстрируем подробнее каждый из этапов нечеткой системы принятия решения об эвакуации при наводнении.

Этапы принятия решения об эвакуации при наводнении

Этап 1: выбор критериев.

Принятие решения об эвакуации из-за наводнения зависит от результатов оценки риска наводнения, а также возможностей безопасной эвакуации. Риск наводнения отражает потенциальную угрозу населению, в то время как возможность безопасной эвакуации отражает потенциальные аспекты, которые могут помешать успешной эвакуации. Таким образом критерии принятия решения должны включать потенциальную опасность наводнения, угрозу для жизни населения, а также пропускную способность эвакуационного транспорта, время для эвакуации.

В данном исследовании с целью иллюстрации применения нечеткой логики при принятии решений об эвакуации при наводнении выбрано конкретное критериев: (1) прогнозируемый уровень наводнения, (2) уровень опасности, (3) уязвимость района предполагаемого наводнения и (4) возможность безопасной эвакуации.

Этап 1.1. Прогнозируемый уровень наводнения основывается на нескольких параметрах, из которых наиболее важной является информация о максимальном уровне и скорости подъема воды.

Этап 1.2. Уровень опасности. Прогнозируемый уровень наводнения дает представление о наводнении в целом, но его недостаточно для оценки опасности для населения в каждом районе, подверженном наводнению. Уровень опасности отражает физические характеристики наводнения и его потенциальное воздействие на безопасность людей в каждом районе, на каждой улице. Это детальное знание определяет решение об эвакуации отдельных районов в приоритетном порядке. Будем рассматривать уровень подъема воды на локальных участках в качестве количественной переменной уровня опасности.

Этап 1.3. Уязвимость района предполагаемого наводнения определяется как неспособность предотвратить непосредственный контакт людей с паводковыми водами во время события. Поэтому возможности защиты данного района от наводнения необходимо учитывать до принятия решения об эвакуации. Переменные уязвимости района от наводнения включают характер проживания, высоту зданий и т.п.

Этап 1.4. Возможность безопасной эвакуации может быть определена как набор ограничений (например, пропускная способность дорог) и потенциальных негативных аспектов (плохие погодные условия, аварии, пробки на дорогах), которые могут задержать или помешать успешному проведению эвакуации. Этот критерий включает множество переменных, таких как количество людей, подлежащих эвакуации, доступное время для эвакуации, доступный транспорт, пропускная способность дорог, точки выхода из районов, подверженных наводнениям.

Время является наиболее важной переменной во время эвакуации. В последние десятилетия были разработаны модели эвакуации, включая моделирование дорожного движения, чтобы помочь оценить время эвакуации, загруженность перекрестков маршрутов, расстояния эвакуации, уровни заторов [11].

Различные параметры, отражающие способность и безопасность эвакуации, должны быть объединены в единый показатель, представляющий расчетное соотношение эвакуируемых. Расчетное соотношение эвакуированных определяется как процент людей, успевших эвакуироваться до наступления наводнения.

Этап 2: определение качественных переменных критериев принятия решений и NTE.

Первоначально различные критерии принятия решения об эвакуации имеют либо количественные, либо качественные значения. Перед построением нечеткой системы для принятия решения об эвакуации необходимо связать каждый критерий с качественными переменными в соответствии со значениями его количественной переменной. Качественные переменные представляют некоторый диапазон количественных значений, определяемых обычно экспертами.

Этап 2.1. Качественную переменную для прогнозируемого уровня наводнения и связанного с этим уровня риска определим через цветовую кодировку; “красный” (крупное наводнение, создающее прямую угрозу безопасности людей и имущества), “оранжевый” (риск возникновения значительного уровня затопления, которое может оказать значительное влияние на безопасность имущества и людей), “желтый” (риск наводнения, которое не влечет за собой значительного ущерба, но требует бдительности) и “зеленый” (нет риска наводнения). Пороговые значения уровня воды, которые вызывают соответствующее оповещение, зависят от конкретного водного бассейна. Например, для прогнозируемого уровня наводнения примем следующие пороговые значения уровня воды (м), которые вызывают соответствующее оповещение: зеленый (менее 6,5), желтый ([6,5; 6,7]), оранжевый ([6,7; 7,0]), красный (более 7,0).

Этап 2.2. Качественная переменная уровня опасности. На практике во многих исследованиях учитывается только высота подъема воды. Соответственно области наводнения в зависимости от высоты подъема воды (h) принято делить на зоны. Например, зона (a) определяется произведением h на скорость потока воды (v) и составляет более $7 \text{ м}^2/\text{с}$ (прорыв дамбы), зона (b) с максимальной высотой подъема воды 1 м, зона (c) с максимальной высотой подъема воды $0,5 \text{ м} < h < 1 \text{ м}$, зона (d) с максимальной высотой подъема воды $0,25 \text{ м} < h < 0,5 \text{ м}$, зона (e) с максимальной высотой подъема воды $h < 0,5 \text{ м}$. Иными словами, термины зон (a), (b), (c), (d), (e) представляют значения качественной переменной уровня опасности, связанного с уровнем затопления.

Этап 2.3. Качественная переменная уязвимости района предполагаемого наводнения, например, может характеризоваться следующими уровнями: высокий, средний и низкий, в основном в зависимости от возможности укрытия на месте. Высокий уровень уязвимости означает небольшое количество убежищ, находящихся в районе с паводковыми водами, средний уровень уязвимости – имеются убежища и высокие дома, низкий уровень уязвимости – большое число укрытий, однако при сильном наводнении целостность этих укрытий может быть нарушена. Количественных оценок этих уровней нет, однако, чтобы построить непрерывную функцию принадлежности можно назначить низкому, среднему и высокому уровням номер, например 1, 2 или 3, соответственно.

Этап 2.4. Качественная переменная, отражающая возможность безопасной эвакуации. Будем характеризовать эту переменную пятью качественными значениями: очень низкая, низкая, средняя, высокая, очень высокая, в зависимости от значений процента людей, решивших эвакуироваться. Например, очень низкая – [0; 27,5], низкая – [27,5; 37,5], средняя – [37,5; 67,5], высокая – [67,5; 80], очень высокая – более 80.

Этап 2.5. Определение качественной переменной *NTE*. В некоторых из существующих подходов предлагается принятие решения об эвакуации с учетом затрат [12]. Однако если главной целью является спасение жизни людей, то для оценки необходимости эвакуации целесообразно использовать несколько иной подход. Переменная *NTE* определяется как уровень потенциальной потребности эвакуации относительно приоритетной цели обеспечения безопасности людей. *NTE* варьируется от 0 до 100% и является синтетическим показателем для ЛПП при нечетком многокритериальном анализе, основанном на входных критериях: прогнозируемый уровень наводнения, уровень опасности, уязвимость района предполагаемого наводнения и возможность безопасной эвакуации в случае чрезвычайной ситуации с наводнением [13]. Пусть *NTE* определяется пятью качественными значениями со следующими пороговыми значениями: “очень низкий” – [0; 27,5], “низкий” – [27,5; 37,5], “средний” – [37,5; 67,5], “высокий” – [67,5; 80], “очень высокий” более 80.

Этап 3: фаззификация входных и выходных данных.

После того, как все качественные и количественные переменные критериев принятия решения и *NTE* определены, входные и выходные данные нечеткой системы для принятия решения об эвакуации “размываются” путем проектирования *MFS*. В качестве примера будем использовать треугольные и трапециевидные формы *MFS*, потому что их легко интерпретировать (линейная экстраполяция) и достаточно для проектирования системы нечеткой логики.

Этап 4: определение базы нечетких правил.

Нечеткие правила в базе правил ЕСЛИ-ТО указывают на логические связи между качественными переменными и отражают различные виды ситуаций и экспертные выводы, которые могут быть сделаны в соответствии с имеющимися знаниями и опытом [14]. Например, *IF* (прогнозируемый уровень наводнения – красный *AND* уровень опасности – зона *(a)* *AND* уязвимость района предполагаемого наводнения – высокая *AND* возможность безопасной эвакуации – очень высокая) *THEN* (необходимость эвакуации – очень высокая).

Доступные знания на самом деле ограничены, особенно при анализе новых катастроф. Поэтому только часть всех потенциальных правил может быть определена в нечеткой системе. Однако база правил может дополняться новыми правилами с добавлением новой информации и опыта, что повышает точность нечеткой модели в долгосрочной перспективе. Теоретически, общее число возможных правил, которые могут быть определены в нечеткой системе, равно числу всех комбинаций качественных значений входных данных [15, 16]. Например, в представленной выше нечеткой системе общее количество правил равно $4 \times 5 \times 3 \times 5 = 300$. Чтобы упростить обработку правил или ограничить количество входных данных в нечеткой системе, можно формировать синтетические показатели в качестве входных данных нечеткой системы. Другое решение заключается в разработке многослойной нечеткой системы.

Этап 5: процедура нечеткого вывода и интерпретации решения.

Этап 5.1: применение функций принадлежности для получения нечетких входных данных из реальных значений входных переменных;

Этап 5.2: применение нечетких операторов для получения одного нечеткого ввода;

Этап 5.3: применение операторов импликации для получения нечетких выходных данных;

Этап 5.4: применение нечетких операторов для агрегирования нечеткого выходного сигнала;

Этап 5.5: применение методов дефаззификации для получения четкого значения выходной переменной.

Проиллюстрируем процедуру нечеткого вывода в нечеткой системе на примере двух правил:

Правило 1. *IF* (прогнозируемый уровень наводнения – оранжевый *AND* уровень опасности – зона (b) *AND* уязвимость района предполагаемого наводнения – высокая *AND* возможность безопасной эвакуации – высокая) *THEN* (*NTE* – высокая);

Правило 2. *IF* (прогнозируемый уровень наводнения – красный *AND* уровень опасности – зона (a) *AND* уязвимость района предполагаемого наводнения – высокая *AND* возможность безопасной эвакуации – очень высокая) *THEN* (*NTE* – очень высокая).

Пусть реальные количественные значения четырех входных переменных следующие: прогнозируемый уровень наводнения – 7,12 м; уровень опасности – 2,13 м; уязвимость района – 3; возможность безопасной эвакуации – 75 %.

Применим треугольные и трапециевидных *MFS* для получения нечетких значений из реальных значений входных переменных. Например, для размытия реального значения 7,12 м прогнозируемого уровня, связанного с нечеткой переменной “красный” в Правиле 1, получим степень принадлежности 0,8.

В примере четыре входных переменных и два правила, поэтому результатом фаззификации будут четыре нечетких входных переменных для каждого правила.

Объединяем все нечетких входных данных каждого правила для получения единого нечеткого входного условия с использованием нечетких операторов, соответствующих логическим операторам *AND* в условной части правил. К логическому отношению *AND* можно применить два стандартных нечетких оператора: нечеткое *AND* и нечеткое произведение: *Fuzzy AND*: $\mu(x) = \min(\mu_i)$; *Fuzzy product*: $\mu(x) = \text{product}(\mu_i)$.

Выберем нечеткий оператор *AND*. Операция минимизации четырех нечетких входных данных в *IF*-части Правила 1 (степени принадлежности: 0,2; 0,4; 1,0; 0,75) дает одну нечеткую входную переменную со степенью принадлежности, равной 0,2. Операция минимизации четырех нечетких входных данных в *IF*-части Правила 2 (максимальные степени принадлежности: 0,8; 0,6; 1,0; 0,25) дает один нечеткий вход со степенью принадлежности равной 0,25.

Применим оператор импликации по Мамдани для получения нечеткого вывода для каждого правила: $\mu_{A \rightarrow B}(x, y) = \min(\mu_A(x), \mu_B(y))$

Для Правила 1 применение оператора импликации Мамдани ($\min(0,2; 1)$) устанавливает значение *NTE* “высокий” на уровне 0,2. Аналогично, MF “очень высокий” *NTE* в Правиле 2 определяется на уровне 0,25.

Объединяем все нечеткие выходные переменные в базе правил. Для агрегирования можно использовать два стандартных нечетких оператора (нечеткий ИЛИ, нечеткая сумма): $\mu(x) = \max(\mu_i)$; $\mu(x) = \sum \mu_i - \text{product}(\mu_i)$,

Выберем нечеткий оператор ИЛИ. В правиле 1 максимальная степень принадлежности составляет 0,2, в то время как в Правиле 2 максимальная степень принадлежности составляет 0,25. Оператор нечеткого ИЛИ дает максимальную степень принадлежности конечного нечеткого вывода, равную 0,25.

В качестве метода дефаззификации для получения конечного значения *NTE*, например, выберем метод центра площади. Тогда значение *NTE*, основанное на двух правилах, составляет 79,7 %.

Нечеткая система для принятия решения об эвакуации при наводнении была реализована с помощью *Matlab Fuzzy Logic Toolbox*. Этот инструмент также позволяет визуализировать 3D нечеткие поверхности, которые представляют отношения между двумя входными переменными (остальные фиксированы) и одним выходом. Такого рода представление помогает понять, как система будет вести себя для всего диапазона значений во входном пространстве. Кроме того, его можно использовать для проверки и калибровки правил и *MFS*, а также для определения необходимости внесения изменений в базу правил для нечетких множеств.

Был проведен анализ чувствительности предлагаемого метода путем моделирования нескольких сценариев и ситуаций наводнения.

В частности, первый сценарий предполагал, что если уязвимость района наводнения фиксируется на низком уровне (значение 1), а возможность безопасной эвакуации – на очень высоком уровне (значение 95%), то 3D-поверхность позволяет понять, как ведет себя выходная переменная *NTE* при изменении прогнозируемого уровня наводнения и уровня опасности, обнаружить пороговые эффекты комбинации входных данных. Например, *NTE* становится чувствительным, когда прогнозируемый уровень приближается к 7 м, а уровень опасности одновременно приближается к 2 м. *NTE* значительно увеличивается более чем на 50% при уровне опасности выше 1,25 м.

Второй сценарий представлял ситуацию, когда прогнозируемый уровень наводнения достигает 6,9 м (оранжевый уровень), а уязвимость района низкая (значение 1). В этом сценарии *NTE*, в основном, зависит от местного уровня опасности. Однако, когда расчетное соотношение эвакуированных падает ниже 40%, то независимо от повышения местного уровня опасности, *NTE* составляет не более 50%. Это обстоятельство указывает на существование минимальных пороговых значений доступных ресурсов для начала эвакуации. Эту информацию некоторые ЛПР могут забыть в критической ситуации.

Третий пример сценария представляет ситуацию, когда уровень опасности зафиксирован как очень опасно (2 м), а уязвимость района низкая (значение 1). Когда предполагаемое количество эвакуированных составляет менее 40%, то независимо от увеличения прогнозируемого уровня *NTE* составляет не более 50%, как в предыдущем сценарии. Низкое расчетное соотношение эвакуируемых ограничивает *NTE*, поскольку эффективность эвакуации ограничена.

Четвертый сценарий представляет ситуацию, когда местный уровень опасности зафиксирован на уровне 2 м (очень опасно), а возможности эвакуации и безопасность очень высоки (значение 80%). В этом случае *NTE* существенно не меняется в зависимости от уязвимости области, поскольку возможности района по предотвращению наводнений очень ограничены, особенно в случае катастрофического события.

В целом, анализ чувствительности помогает откалибровать нечеткую нелинейную модель, а также изучить и сравнить различные сценарии и стратегии эвакуации. С этой точки зрения нечеткая модель принятия решений также может быть использована в качестве инструмента для подготовки плана эвакуации.

Применение нечеткой системы в сочетании с геоинформационной системой. Геоинформационные системы (ГИС), являясь универсальным программным средством, предлагают набор сервисов для пространственного анализа и поддержки принятия решений: графическую визуализацию географических объектов района в виде карты, пространственную базу данных с географической привязкой, статистику и анализ данных [17]. При интеграции в ГИС карты, база данных и статистика становятся очень мощными инструментами, особенно для управления чрезвычайными ситуациями. Например, рассмотрим способ, с помощью которого нечеткая логическая модель принятия решения об эвакуации может быть применена в сочетании с ГИС.

Пусть пространственные данные обрабатываются с помощью ГИС. Вначале необходимо подготовить данные о количественных значениях входных переменных в данной местности (местные уровни воды, уязвимость района и т.п.) перед обработкой в нечеткой системе [18]. Далее, необходимо импортировать пространственные данные из ГИС в *Matlab*. Перед выполнением импорта в *Matlab* таблица атрибутов должна быть преобразована в формат файла *Excel*. Затем нечеткий алгоритм, реализованный с помощью *Matlab*, может быть применен к каждой строке файла, соответствующей геометрическому объекту.

Наоборот, полученные с помощью *Matlab* выходные данные *NTE* могут быть импортированы в ГИС для создания карты пространственного распределения *NTE*. Рис. 1 иллюстрирует процесс применения модели нечеткой логики в сочетании с ГИС.

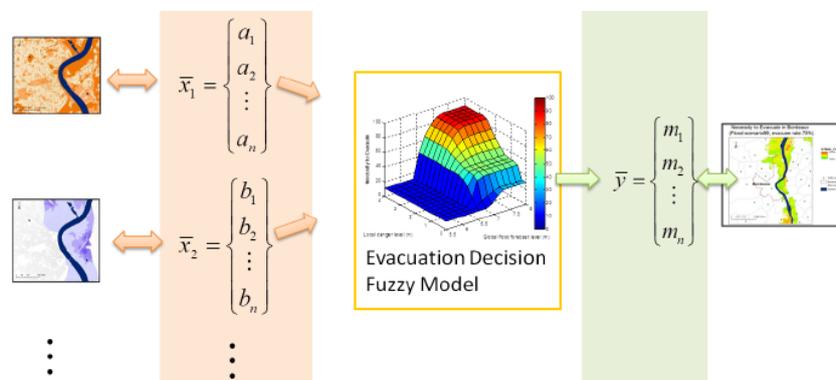


Рис. 1. Применение модели нечеткой логики в сочетании с ГИС

Это помогает ЛПР понять контекст ситуации, определить основную информацию, касающуюся возможной необходимости эвакуации, и принять решение на основе многокритериального анализа [19, 20].

Попробуем установить связь между числовым значением *NTE*, выраженным в процентах (%), и множеством качественных значений входных переменных. Одним из возможных решений для соответствующих предлагаемых действий, связанных с необходимостью эвакуации, может быть решение, указанное в табл. 1.

Таблица 1

Предлагаемые действия, связанные с необходимостью эвакуации

Вариант эвакуации	Описание	Необходимо эвакуировать (%)	
		уровень	диапазон
Никакой эвакуации	Никаких дополнительных действий не требуется	очень низкий	(0, 15)
Бдительность	Меры предосторожности при риске наводнения, включая подготовку к эвакуации	низкий	(15, 40)
Консультативная эвакуация	Частичная эвакуация, некоторым людям из группы риска, рекомендуется эвакуироваться	средний	(40, 65)
Мягкий приказ об эвакуации	Полная эвакуация людей из группы риска	высокий	(65, 80)
Приказ о срочной эвакуации	Приказ о полной эвакуации, населению, находящемуся в опасности, рекомендуется покинуть район наводнения	очень высокий	(80, 100)

Заключение. Предложена нечеткая модель принятия решения об эвакуации на основе таких критериев принятия решения как уровень прогнозируемого наводнения, уровень опасности, уязвимость района и возможность безопасной эвакуации. Нечеткая модель реализована с помощью *Matlab*. Данные пространственного измерения критериев принятия решения предварительно обрабатываются с

помощью платформы ГИС, а затем импортируются в нечеткую модель принятия решения об эвакуации для оценки *NTE* конкретной области. Пространственные значения *NTE* можно импортировать в платформу ГИС для отображения на картах, которые могут использоваться в качестве поддержки принятия решений для управления эвакуацией. Нечеткая модель принятия решения об эвакуации определяет нелинейную связь между критериями принятия решения и необходимостью эвакуации, основанную на эвристических правилах с использованием нечеткой логики. Это обеспечит оценку *NTE* для должностных лиц в конкретной области для различных сценариев и обстоятельств.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Lumbroso D., Vinet F.* Tools to Improve the Production of Emergency Plans for Floods: Are They Being Used by the People that Need Them? // *Jour. of Contingencies and Crisis Management.* – 2012. – Vol. 20. – P. 149-165.
2. *Gerasimenko E., Rozenberg I.* Earliest arrival dynamic flow model for emergency evacuation in fuzzy conditions // *IOP Conference Series: Materials Science and Engineering, Krasnoyarsk Science and Technology.* – 2020. – Vol. 734 (1). – P. 012147.
3. *Родзин С.И.* Вычислительный интеллект: немонотонные логики и графическое представление знаний // *Программные продукты и системы.* – 2002. – № 1, – С. 20-25.
4. *Lindell M.K., Prater C.S.* A hurricane evacuation management decision support system (EMDSS) // *Natural Hazards.* – 2007. – Vol. 40. – P. 627-634.
5. *Kacprzyk J., Zadrozny S., Nurmi H., Bozhenyuk A.* Towards Innovation Focused Fuzzy Decision Making // *IEEE Int. Conf. on Fuzzy Systems.* – 2021. – P. 256-268.
6. *Родзина О.Н.* Проблемно-ориентированные алгоритмы мягких вычислений. – Чебоксары: ИД "Среда", 2020. – 96 с. – DOI: 10.31483/a-200.
7. *Sivanandam S., Sumathi S., Deepa S.* Introduction to Fuzzy Logic using MATLAB. – Verlag Berlin Heidelberg: Springer, 2006.
8. *Герасименко Е.М.* Нечеткая модель нахождения максимального динамического потока для решения задачи эвакуации зданий // *Известия ЮФУ. Технические науки.* – 2019. – № 4 (206). – С. 15-26.
9. *Родзин С.И., Скобцов Ю.А., Эль-Хатиб С.А.* Биоэвристики: теория, алгоритмы и приложения. – Чебоксары: Изд. дом «Среда», 2019. – 224 с.
10. *Герасименко Е.М., Нужнов Е.В.* Решение задачи нахождения максимального потока в задачах эвакуации на основе нечетких колеблющихся операторов агрегирования // *Известия ЮФУ. Технические науки.* – 2021. – № 4 (221), – С. 145-154.
11. *Tiglioglu T.S.* Modeling hurricane evacuation using transportation models, fuzzy set, and possibility theory: Ph.D. dissertation, Texas Tech University, 2001.
12. *Raymond A.M.* Modeling hurricane evacuation vulnerability: a case study of Pinellas County. Department of Geography, University of South Florida, 2005.
13. *Rodzin S., Rodzina O.* New computational models for big data and optimization // *Proc. 9th Int. Conf. on Application of Information and Communication Technologies (AICT'2015).* – 2015. – P. 3-7.
14. *Gerasimenko E., Kureichik V.V.* Minimum cost lexicographic evacuation flow finding in intuitionistic fuzzy networks // *Jour. of Intelligent and Fuzzy Systems.* – 2022. – Vol. 42, No. 1. – P. 251-263.
15. *Курейчик В.В., Курейчик Вл.Вл., Саак А.Э.* Эволюционный алгоритм для решения задачи диспетчеризации // *Известия ЮФУ. Технические науки.* – 2021. – № 2 (219). – С. 50-59.
16. *Курейчик В.В., Курейчик Вл.Вл., Саак А.Э.* Основные стратегии и принципы при решении задач диспетчерования в сборнике // *Тр. межд. конгресса "Интеллектуальные системы и информационные технологии".* – 2020. – С. 235-241.
17. *Родзин С.И., Родзина Л.С.* Биоинспирированный поиск решений: теория и приложения для обработки проблемно-ориентированных знаний в геоинформатике // *Известия ЮФУ. Технические науки.* – 2015. – № 4 (165). – С. 203-216.
18. *Родзин С.И.* Биоэвристики многокритериальной оптимизации: проектирование и гибридизация // *Тр. межд. конгресса «IS&IT'18».* – Таганрог: Изд-во Ступина С.А., 2018. – Т. 1. – С. 33-47.

19. Петровский А.Б. Снижение размерности признакового пространства: метод СОКПАТ // Искусственный интеллект и принятие решений. – 2020. – № 2. – С. 63-77.
20. Kailiponi P. Analyzing evacuation decision using Multi-Attribute Utility Theory (MAUT) // Procedia Engineering. – 2010. – Vol. 3. – P. 163-174.

REFERENCES

1. Lumbroso D., Vinet F. Tools to Improve the Production of Emergency Plans for Floods: Are They Being Used by the People that Need Them?, *Jour. of Contingencies and Crisis Management*, 2012, Vol. 20, pp.149-165.
2. Gerasimenko E., Rozenberg I. Earliest arrival dynamic flow model for emergency evacuation in fuzzy conditions, *IOP Conference Series: Materials Science and Engineering*. , 2020, Vol. 734 (1), pp. 012147.
3. Rodzin S.I. Vychislitel'nyy intellekt: nemonotonnye logiki i graficheskoe predstavlenie znaniy [Computational intelligence: non-monotonic logics and graphical representation of knowledge], *Programmnye produkty i sistemy* [Software products and systems], 2002, No. 1, pp. 2-25.
4. Lindell M.K., Prater C.S. A hurricane evacuation management decision support system (EMDSS), *Natural Hazards*, 2007, Vol. 40, pp. 627-634.
5. Kacprzyk J., Zadrozny S., Nurmi H., Bozhenyuk A. Towards Innovation Focused Fuzzy Decision Making. *IEEE Int. Conf. on Fuzzy Systems*, 2021, pp. 256-268.
6. Rodzina O.N. Problemno-orientirovannyye algoritmy myagkikh vychisleniy [Problem-oriented algorithms of soft computing]. *Cheboksary: ID "Sreda"*, 2020, 96 p. DOI: 10.31483/a-200.
7. Sivanandam S., Sumathi S., Deepa S. Introduction to Fuzzy Logic using MATLAB. Verlag Berlin Heidelberg: Springer, 2006.
8. Gerasimenko E.M. Nechetkaya model' nakhozhdeniya maksimal'nogo dinamicheskogo potoka dlya resheniya zadachi evakuatsii zdaniy [Fuzzy model of finding the maximum dynamic flow for solving the problem of evacuation of buildings], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2019, No. 4, pp. 15-26.
9. Rodzin S.I., Skobtsov Yu.A., El-Khatib S.A. Bioevristiki: teoriya, algoritmy i prilozheniya [Bioheuristics: theory, algorithms and applications]. *Cheboksary: ID "Sreda"*, 2019, 224 p.
10. Gerasimenko E.M., Nuzhmov E.V. Reshenie zadachi nakhozhdeniya maksimal'nogo potoka v zadachakh evakuatsii na osnove nechetkikh koleblyushchikhsya operatorov agregirovaniya [Solving the problem of finding the maximum flow in evacuation problems based on fuzzy oscillating aggregation operators], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, no. 4, pp. 145-154.
11. Tiglioglu T.S. Modeling hurricane evacuation using transportation models, fuzzy set, and possibility theory: Ph.D. dissertation, Texas Tech University, 2001.
12. Raymond A.M. Modeling hurricane evacuation vulnerability: a case study of Pinellas County. Department of Geography, University of South Florida, 2005.
13. Rodzin S., Rodzina O. New computational models for big data and optimization, *Proc. 9th Int. Conf. on Application of Information and Communication Technologies (AICT'2015)*, 2015, pp. 3-7.
14. Gerasimenko E., Kureichik V.V. Minimum cost lexicographic evacuation flow finding in intuitionistic fuzzy networks, *Jour. of Intelligent and Fuzzy Systems*, 2022, Vol. 42, No. 1, pp. 251-263.
15. Kureychik V.V., Kureychik V.V., Saak A.E. Evolyucionnyy algoritim dlya resheniya zadachi dispatcherizatsii [An evolutionary algorithm for solving the dispatching problem], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 2, pp. 50-59.
16. Kureychik V.V., Kureychik V.V., Saak A.E. Osnovnye strategii i printsipy pri reshenii zadach dispatcherirovaniya v sbornike [Basic strategies and principles in solving dispatching tasks in the collection]. *Tr. mezhd. kongressa "Intellektual'nye sistemy i informacionnye tekhnologii"* [Proceedings of the International Congress "Intelligent Systems and Information Technologies"], 2020, pp. 235-241.
17. Rodzin S.I., Rodzina L.S. Bioinspirirovannyy poisk resheniy: teoriya i prilozheniya dlya obrabotki problemno-orientirovannykh znaniy v geoinformatike [Bioinspired search for solutions: theory and applications for processing problem-oriented knowledge in geoinformatics], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2015, No. 4, pp. 203-216.

18. Rodzin S.I. Bioevristiki mnogokriterial'noy optimizatsii: proektirovanie i gibridizatsiya [Bio heuristics of multicriteria optimization: design and hybridization], *Tr. mezhd. kongressa "Intellectual'nye sistemy i informatsionnye tekhnologii"* [Proceedings of the International Congress "IS&IT'18"]. Taganrog: Izd-vo Stupina S.A., 2018, Vol. 1. pp. 33-47.
19. Petrovsky A.B. Snizhenie razmernosti priznakovogo prostranstva: metod SOKRAT [Reducing the dimension of the feature space: the SOCRATES method], *Iskusstvennyj intellekt i prinyatie reshenij* [Artificial intelligence and decision-making], 2020, No. 2, pp. 63-77.
20. Kailiponi P. Analyzing evacuation decision using Multi-Attribute Utility Theory (MAUT), *Procedia Engineering*, 2010, Vol. 3, pp. 163-174.

Статью рекомендовала к опубликованию д.т.н., профессор Л.С. Лисицына.

Герасименко Евгения Михайловна – Южный федеральный университет; e-mail: vkur@sfedu.ru; г. Таганрог, Россия; тел.: 88634371651; кафедра систем автоматизированного проектирования; к.т.н.; доцент.

Курейчик Владимир Викторович – e-mail: vkur@sfedu.ru; кафедра систем автоматизированного проектирования; зав. кафедрой САПР; д.т.н.; профессор.

Родзин Сергей Иванович – e-mail: srodzin@sfedu.ru; тел.: 88634371673; кафедра математического обеспечения и применения ЭВМ; к.т.н.; профессор.

Кухаренко Анатолий Павлович – e-mail: apkuharenko@sfedu.ru; кафедра интеллектуальных и многопроцессорных систем; к.т.н.; доцент.

Gerasimenko Evgeniya Michailovna – Southern Federal University; e-mail: vkur@sfedu.ru; Taganrog, Russia; phone: +78634371651; the department of computer aided design; cand. of eng. sc.; associate professor.

Kureichik Vladimir Victorovich – e-mail: vkur@sfedu.ru; the department of computer aided design; head of CAD department; dr. of eng. sc.; professor.

Rodzin Sergey Ivanovich – e-mail: srodzin@sfedu.ru; phone: +78634371673; the department of software engineering; cand. of eng. sc.; professor.

Kukharenko Anatoliy Pavlovich – e-mail: apkuharenko@sfedu.ru; the department of intelligent and multiprocessor systems; cand. of eng. sc.; associate professor.

УДК 004.89

DOI 10.18522/2311-3103-2022-4-29-39

Э.В. Кулиев, А.В. Котельва, М.М. Семенова, С.В. Игнатьева, А.П. Кухаренко

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ В УПРАВЛЕНИИ ПРЕДПРИЯТИЕМ НА ОСНОВЕ АЛГОРИТМА ИМИТАЦИИ ОТЖИГА *

Рассмотрен аналитический обзор алгоритма имитации отжига для задачи эффективного управления предприятием. Проведена оптимизация алгоритма имитации отжига для задачи эффективного управления предприятием. Для анализа случаев использовалась оптимизация графика работы рабочих в организации. Установлена модель планирования рабочих с сильными и слабыми ограничениями. Смоделированный алгоритм отжига используется для оптимизации стратегии решения модели планирования рабочего графика персонала. Алгоритм имитации отжига представляет собой алгоритм, пригодный для решения крупномасштабных задач комбинаторной оптимизации. Он также оценивает и получает оптимальную стратегию планирования. Алгоритм имитации отжига хорошо влияет на интеллектуальный анализ данных управления человеческими ресурсами. Интеллектуальный анализ больших данных может помочь компаниям проводить динамический

* Исследование выполнено за счет гранта Российского научного фонда № 22–21–00316, <https://rscf.ru/project/22-21-00316/> в Южном федеральном университете.

анализ при наборе талантов, а план набора талантов выполняется качественно и стандартно, чтобы проанализировать характеристики различных талантов со многих сторон и повысить уровень управления человеческими ресурсами. Разработан алгоритм реализующий процесс работы алгоритма имитации отжига. Алгоритм имитации отжига принимает новые решения по критерию Метрополиса, поэтому помимо принятия оптимизированного решения он также принимает ослабленное решение в ограниченном диапазоне. Алгоритм Метрополиса – алгоритм семплирования, использующийся, в основном, для сложных функций распределения. Он отчасти похож на алгоритм выборки с отклонением, однако здесь вспомогательная функция распределения меняется со временем. Проведены экспериментальные исследования, которые показывают, что модель планирования рабочих, основанная на сильных и слабых ограничениях, значительно лучше, чем модель ручного планирования, достигая эффективного баланса между контролем затрат на зарплату в организации и повышением удовлетворенности персонала. Успешное применение модели планирования персонала, основанной на моделируемом алгоритме отжига, приносит новые идеи и идеи для решения крупномасштабных задач планирования рабочих. Приведенные результаты могут служить отправной точкой для изучения систем управления персоналом, основанных на технологии интеллектуального анализа данных.

Интеллектуальный анализ данных; алгоритм имитации отжига; управление человеческими ресурсами.

E.V. Kuliev, A.V. Kotelva, M.M. Semenova, S.V. Ignateva, A.P. Kukhareenko

INTELLIGENT DATA ANALYSIS IN ENTERPRISE MANAGEMENT BASED ON THE ANNEALING SIMULATION ALGORITHM

The article considers an analytical review of the annealing simulation algorithm for the problem of efficient enterprise management. The optimization of the annealing simulation algorithm for the problem of efficient enterprise management has been carried out. For the analysis of cases, the optimization of the work schedule of workers in the organization was used. Established worker scheduling model with strong and weak constraints. The simulated annealing algorithm is used to optimize the strategy for solving the staff scheduling model. The simulated annealing algorithm is an algorithm suitable for solving large-scale combinatorial optimization problems. It also evaluates and obtains the optimal scheduling strategy. The simulated annealing algorithm has a good effect on the data mining of human resource management. Big data mining can help companies conduct dynamic analysis in talent recruitment, and the talent recruitment plan is carried out in a quality and standard way to analyze the characteristics of various talents from many angles and improve the level of human resource management. An algorithm has been developed that implements the operation of the annealing simulation algorithm. The simulated annealing algorithm makes new decisions based on the Metropolis criterion, so in addition to making an optimized decision, it also makes a reduced decision in a limited range. The Metropolis algorithm is a sampling algorithm mainly used for complex distribution functions. It is somewhat similar to the variance sampling algorithm, but here the auxiliary distribution function changes over time. Experimental studies have been carried out that show that a worker scheduling model based on strong and weak constraints is significantly better than a manual scheduling model, achieving an effective balance between controlling wage costs in an organization and increasing employee satisfaction. The successful application of a workforce scheduling model based on a simulated annealing algorithm brings new insights and insights to solve large-scale worker scheduling problems. The results presented can serve as a starting point for studying personnel management systems based on data mining technology.

Data mining; annealing simulation algorithm; human resource management.

Введение. Развитие социальной экономики создает возможности и проблемы для развития предприятий. Предприятия должны внедрять эффективные методы управления, особенно методы управления человеческими ресурсами, чтобы повысить свою конкурентоспособность [1]. На данном этапе, с наступлением эры больших данных, управление человеческими ресурсами предприятия постепенно

использует методы управления «поиска» больших данных. Адаптация к изменениям времени и внедрение инноваций в средства и методы управления человеческими ресурсами – самая большая проблема, с которой сталкиваются менеджеры по персоналу предприятий в эпоху больших данных. Алгоритм имитации отжига (SA) представляет собой алгоритм, пригодный для решения крупномасштабных задач комбинаторной оптимизации. Алгоритм имитации отжига подходит для интеллектуального анализа данных управления человеческими ресурсами предприятия [2]. Чтобы реализовать оптимизацию управления оптимальным распределением человеческих ресурсов предприятия и повысить эффективность управления человеческими ресурсами предприятия, можно использовать модель оптимального распределения человеческих ресурсов предприятия, основанную на алгоритме имитации отжига.

Постановка задачи. Цель задач комбинаторной оптимизации состоит в том, чтобы найти оптимальное решение из допустимого пространства решений комбинаторной задачи. Как правило, он содержит три основных элемента: переменные, ограничения и целевые функции. Основные параметры, выбранные в процессе решения, называются переменными. Различные ограничения на значение переменных называются ограничениями. Функция, которая представляет эталон измерения допустимого решения, называется целевой функцией. Решение задач комбинаторной оптимизации заключается в поиске наиболее подходящего решения в множестве решений целевой функции, что неизбежно требует применения определенных алгоритмов для снижения временной и пространственной сложности процесса решения. Алгоритм имитации отжига – это алгоритм для решения задач комбинаторной оптимизации [3]. Он использует критерии приемлемости Метрополиса, чтобы алгоритм избежал ловушки локального «оптимума», использует «график охлаждения» для управления всем процессом реализации алгоритма и, наконец, позволяет алгоритму получить приближенное оптимальное решение за полиномиальное время.

Шаги алгоритма имитации отжига. Предположим, что целевая функция $f(i)$ решения i задачи комбинаторной оптимизации эквивалентна энергии E_i микроскопического состояния i твердого тела. Пусть управляющий параметр t , уменьшающий свое значение по мере развития алгоритма, играет роль температуры T в процессе отжига твердого тела, а затем принимает значение для каждого управляющего параметра t . Алгоритм продолжает процесс «генерации новых решений, принятия и отбрасывания», то есть однократного выполнения алгоритма Метрополиса [1, 2]. Когда управляющий параметр t стремится к 0, окончательно может быть получено общее оптимальное решение задачи комбинаторной оптимизации. Алгоритм имитации отжига использует алгоритм Метрополиса для генерации последовательности решений задачи комбинаторной оптимизации и определяется вероятностью перехода, соответствующей критерию Метрополиса [3]:

$$p'_i(i \Rightarrow j) = \begin{cases} i, & \text{если } f(i) \leq f(j) \\ \exp\left(\frac{f(i)-f(j)}{t}\right) & \end{cases} \quad (1)$$

Определим, следует ли принять переход от текущего решения i к новому решению j , $t \in R$ представляет параметр управления. Начнем с большего значения t (соответствующего температуре растворения твердого вещества). После того, как будет сделано достаточное количество передач, медленно уменьшайте значение t . Если это повторяется, алгоритм завершается, когда выполняется определенный критерий остановки. Предполагая наличие доменных структур и генераторов, пусть tk обозначает значение управляющего параметра t на k -й итерации алгоритма Метрополиса, а Lk обозначают количество преобразований, сгенерированных в течение k -й итерации алгоритма Метрополиса.

Задача оптимизации может быть описана следующим образом: где S – дискретное конечное пространство состояний, а i представляет состояние. Для такой задачи оптимизации этапы расчета алгоритма SA можно описать следующим образом: $\min f(i), i \in S$. Упомянутый выше алгоритм имитации отжига можно визуальнo описать блок-схемой. Процесс работы алгоритма имитации отжига показан на рис. 1 [4].

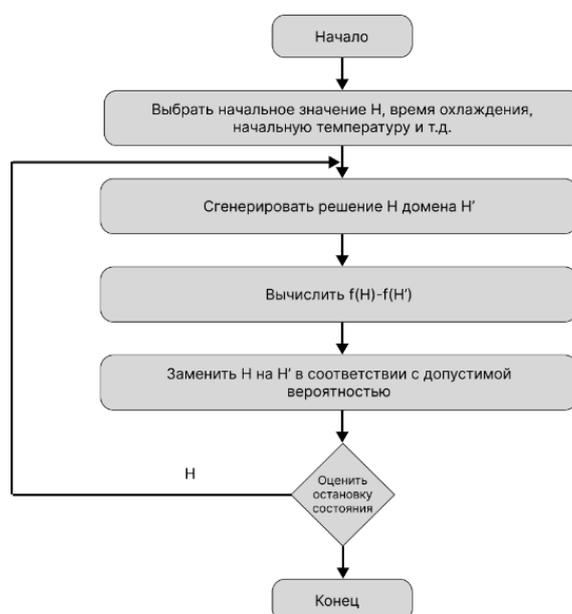


Рис. 1. Процесс работы алгоритма имитации отжига

Из потока алгоритма видно, что функция генерации нового состояния, функция принятия нового состояния, функция детемпературы, критерий стабилизации выборки, критерий окончания детемпературы и начальная температура являются основными звеньями и факторами, которые влияют на результаты оптимизации алгоритма.

Экспериментальная производительность алгоритма имитации отжига имеет преимущества высокого качества, сильной устойчивости к начальным значениям и простоты реализации. Алгоритм имитации отжига принимает новые решения по критерию Метрополиса, поэтому помимо принятия оптимизированного решения он также принимает ослабленное решение в ограниченном диапазоне [5–8]. В этом существенное отличие алгоритма имитации отжига от алгоритма локального поиска. В начале, если значение t велико, можно принять худшие ухудшающиеся решения; по мере уменьшения значения могут быть приняты только лучшие ухудшающиеся решения; наконец, когда значение t стремится к 0, ухудшающиеся решения больше не принимаются. Это позволяет алгоритму имитации отжига вырваться из «ловушки» локальной оптимизации, и он с большей вероятностью найдет общее оптимальное решение задачи комбинаторной оптимизации, но не потеряв при этом своей простоты и универсальности.

Интеллектуальный анализ данных управления человеческими ресурсами предприятия. С повышением уровня информации и возможностей принятия решений на основе данных традиционное управление человеческими ресурсами в эпоху больших данных постоянно меняется. Основываясь на корпоративном виде-

нии и стратегии, эффективное применение технологии интеллектуального анализа данных в корпоративном управлении человеческими ресурсами способствует разумному подбору персонала и должностей, полному раскрытию рабочих способностей и потенциала сотрудников, повышению эффективности организации и сотрудников, а также достижению устойчивого корпоративного развития [9–11].

Человеческие ресурсы предприятия могут узнавать о талантах друг от друга посредством управления интеллектуальным анализом больших данных, таких как производительность личного сайта в социальной сети и личная оценка в «круге друзей», а затем определять, подходит ли талант для требований компании по найму, обеспечивать качество набора корпоративных талантов, а также способствовать улучшению управления человеческими ресурсами. Например, при назначении должностей бизнес-менеджеры могут судить о том, какой человек подходит для какой должности на основе результатов анализа, и какой человек может создать высокие преимущества. При управлении человеческими ресурсами можно не только принимать эффективные решения, но также можно создать базу данных, чтобы можно было отслеживать конкретную ситуацию с талантами в режиме реального времени. Оценка высокоэффективных талантов требует четырех аспектов принятия решений, а именно данных резюме, данных о производительности таланта в первый год, использования талантом рабочего времени и эффективности работы, а также динамики таланта в социальной сфере.

Ключевой частью управления человеческими ресурсами предприятия является полное понимание характеристик различных талантов, а затем назначение подходящих для них должностей. Все больше и больше компаний полагаются на интеллектуальный анализ больших данных для создания точных моделей человеческого потенциала, чтобы анализировать характеристики различных талантов со многих сторон и повышать уровень управления человеческими ресурсами [5, 12]. Интеллектуальный анализ больших данных может помочь компаниям проводить динамический анализ при наборе талантов, а план набора талантов выполняется качественно и стандартно, чтобы проанализировать характеристики различных талантов со многих сторон и повысить уровень управления человеческими ресурсами.

Общий процесс интеллектуального анализа данных выглядит следующим образом:

- ◆ Предварительно обработать данные. Собирать и очищать информацию из источников данных и хранить ее в хранилище данных.
- ◆ Поиск модели. Использовать инструменты интеллектуального анализа данных, чтобы найти модели в данных. Этот процесс поиска может выполняться системой автоматически.
- ◆ Анализ результатов. Процесс поиска интеллектуального анализа данных обычно необходимо повторять много раз, потому что после того, как аналитик оценивает выходные результаты, могут образоваться некоторые новые проблемы и новые входные данные.
- ◆ Усвоение знаний. Интерпретация отчета о результатах, интерпретация результатов и принятие соответствующих мер на основе результатов – это ручной процесс.

Используя эту модель, можно обнаружить типы талантов, которые существуют в организации, а также определить, к какому из этих типов относится сотрудник.

На основе всех записей данных в хранилище данных создается выборка, подлежащая классификации. Объекты, подлежащие классификации, называются образцами, в качестве пробных наборов. Чтобы получить разумную классификацию выборки, конкретные атрибуты должны быть определены количественно. Количественные атрибуты становятся индикаторами выборки [6].

$$h_i = \{h_{i1}, h_{i2}, \dots, h_{im}\}, (i = 1, 2, \dots, n). \quad (2)$$

Поскольку фактические данные едины, собранные данные часто не являются номерами закрытых интервалов, поэтому эти необработанные данные должны быть стандартизированы. Сначала нужно найти среднее значение. Для определенного индекса выборки данные можно получить, где представляет собой данные, полученные по i -й выборке для k -го показателя. Их среднее значение рассчитывается по следующей формуле:

$$h'_k = \frac{\{h_{1k}, h_{2k}, \dots, h_{nk}\}}{n} = \sum_{i=1}^n h_{ik}, \quad k = 1, 2, \dots, m. \quad (3)$$

Затем рассчитать стандартное отклонение этих исходных данных по следующей формуле:

$$S_k = \sqrt{\frac{\sum_{i=1}^n (h'_{ik} - h'_k)^2}{n}}. \quad (4)$$

Рассчитать стандартизированное значение

$$h''_{ik} = \left| \frac{h'_{ik} - h'_k}{S_k} \right|. \quad (5)$$

Если стандартизированные данные, полученные в это время, не находятся в замкнутом интервале, то используется следующая нормированная формула экстремального значения [7]:

$$h_{ik} = \frac{h''_{ik} - h''_{min k}}{h''_{max k} - h''_{min k}}. \quad (6)$$

В формуле $h''_{max k}$ и $h''_{min k}$ представляют максимальное и минимальное значения в $h''_{1k}, h''_{2k}, \dots, h''_{nk}$ соответственно. Общая форма установления отношения сходства модулей R выглядит следующим образом

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \dots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{bmatrix}, \quad 0 \leq r_{ij} \leq 1, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, n. \quad (7)$$

Используем метод максимума и минимума для расчета r_{ij}

$$r_{ij} = \frac{\sum_{k=1}^m \min(h_{ik}, h_{kj})}{\sum_{k=1}^m \max(h_{ik}, h_{kj})}, \quad (i, j \leq n). \quad (8)$$

Принят метод максимального дерева, то есть строится специальный граф, вершинами которого являются все классифицированные объекты. Конкретный метод заключается в том, чтобы сначала нарисовать определенное i в наборе вершин, а затем соединить ребра в порядке r_{ij} от наибольшего к наименьшему и не требовать циклов, пока все вершины не будут соединены так, чтобы было получено максимальное дерево. Каждому ребру можно присвоить определенный вес, а именно r_{ij} . Однако из-за разных способов подключения самое большое дерево не может быть уникальным. Затем необходимо взять λ и вырезать набор на максимальное количество; то есть удалить ребра с весом $r_{ij} < \lambda$, $\lambda \in [0, 1]$. Таким образом, дерево разрезается на несколько поддеревьев, не связанных друг с другом [8, 13]. Хотя самое большое дерево не является уникальным, после взятия набора разрезов полученные поддерева будут одинаковыми, и эти подданные являются шаблонами, найденными по индукции в хранилище данных.

Экспериментальные исследования. Данные о персонале в статье взяты из открытых источников о режиме работы в организации X. Предположим, что период планирования составляет одну неделю, а ежедневное рабочее время делится на три класса: класс А (8:00–16:00), класс Б (16:00–0:00) и класс В (0:00–8:00).

Максимальное время непрерывной работы каждого рабочего составляет 4 смены, а самая продолжительная непрерывная ночная смена – 2 смены. В календарном цикле самая длинная рабочая смена каждого рабочего составляет не более 6 смен, а самая короткая рабочая смена - не менее одной смены. Общее количество рабочих часов в календарном цикле составляет около 40 часов, а фактическое количество персонала, необходимых для каждой смены в день, дается старшим менеджером

В организации действует система «гибкого графика». Недельный график составляется старшим менеджером вручную в зависимости от спроса на персонал в компании, семейных условий и условий жизни рабочего.

Пространство решений и выбор кода. Комбинаторная оптимизация заключается в поиске оптимального решения, так что, $x^* \forall x_i \in \Omega, C(x^*) = \min C(x_i)$, где $\Omega = \{x_1, x_2, \dots, x_n\}$ – это пространство решений, образованное всеми состояниями, и представляет собой значение целевой функции, соответствующее состоянию [3, 15–18].

Стратегии кодирования модели планирования рабочего в основном включают кодирование ближайшего соседа, кодирование порядка, двоичное кодирование и матричное кодирование. Последовательное кодирование не способствует глобальной оптимизации. Двоичное кодирование неестественно и требует дополнительных операторов коррекции для обеспечения легитимности решения; матричное кодирование имеет большой объем памяти и влияет на эффективность оптимизации генетических операторов. Исходя из этого, кодирование ближайшего соседа является широко используемой стратегией для описания проблем планирования рабочих.

Так называемое кодирование ближайшего соседа напрямую использует решение для построения оптимизированной формы, например решение 010101...0110. Соответствующий код ближайшего соседа (0 1 0 1 0 1...0 1 1 0). Этот метод кодирования соответствует характеристикам решения задачи целочисленного программирования 0–1, а также подходит для проектирования операций оптимизации.

Разработка функции приемлемости состояния алгоритма имитации отжига. Функция приемлемости состояния является ключом к способности алгоритма генерировать вероятностные скачки и может избегать локальных минимумов под руководством механизма распределения [2, 19]. В сочетании с функцией генерации состояния, основанной на случайной операции, чтобы процесс поиска имел возможность преодолевать локальный минимум и удовлетворять условию симметрии алгоритма имитации отжига. Критерий является наиболее часто используемой схемой принятия нового состояния, где представляет собой разность целевых значений между старым и новым состоянием, а представляет собой температуру.

Начальная температура и начальное состояние. Наиболее часто используемая и понятная схема определения начальной и температуры состоит в том, чтобы сначала случайным образом сгенерировать набор состояний, определить максимальную целевую разницу между двумя состояниями, а затем использовать $|\Delta_{max}|t_0 = -\frac{|\Delta_{max}|}{\ln p_r}$. Среди них – начальная вероятность принятия (теоретически она должна быть близка к 1, а в реальной конструкции может быть равна 0,1), а начальное состояние генерируется случайной функцией 0–1 [4, 20].

Конструкция функции понижения температуры. Теоретически температура должна уменьшаться очень медленно, как обратная величина логарифма. Однако, чтобы избежать слишком длительного процесса поиска и найти хороший компромисс между качеством оптимизации и временными характеристиками, экспоненциальная функция инверсии является наиболее часто используемой стратегией снижения температуры.

Разработка критерия изменения температуры и критерия завершения алгоритма. Чтобы адаптироваться к динамическим изменениям производительности алгоритма и лучше сбалансировать производительность оптимизации и производительность алгоритма по времени, можно принять два критерия «температурная модификация» и «завершение алгоритма», разработанные пороговым методом. То есть, если наилучшее оптимизированное значение, полученное в процессе оптимизации, остается неизменным в течение 20 последовательных поколений, то температура снижается. Если оптимальное значение остается неизменным в течение 20 последовательных, процесс поиска прекращается, и оптимальное значение является результатом оптимизации алгоритма.

Результаты расчетов подробной модели расписания рабочих показаны на рис. 2

	Фактическое ручное расписание	Модель расписания	Отклонение
Целевое значение	252,5	143	-43,4%
Стоимость заработной платы	575	513	-10,8%
Удовлетворение на этот раз	715	967	35,24%
Время работы алгоритма (с)	-	5.08	

Рис. 2. Результаты расчетов модели рабочих

Подробная сходимость смоделированной модели планирования рабочего с использованием смоделированного алгоритма отжига показана на рис. 3.

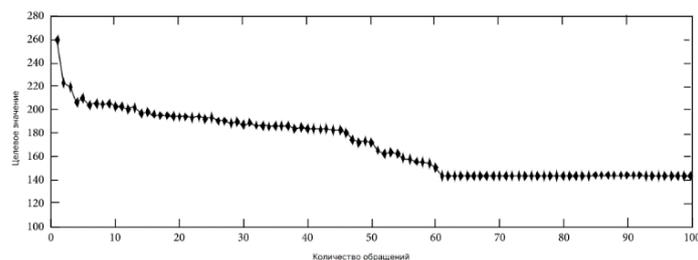


Рис. 3. График сходимости на основе алгоритма имитации отжига

График рабочих оптимизируется алгоритмом имитации отжига. По результатам моделирования видно следующее:

1) Из рис. 2 видно, что целевое значение модели планирования рабочего на 43,43% ниже целевого значения фактического ручного планирования, из которых затраты на заработную плату снижены на 10,8%, но удовлетворенность рабочего сменой увеличилась на 35,24%. Это показывает, что модель планирования рабочих, основанная на сильных и слабых ограничениях, значительно лучше, чем модель ручного планирования, достигая эффективного баланса между контролем затрат на зарплату в организации и повышением удовлетворенности персонала. Успешное применение модели планирования персонала, основанной на моделируемом алгоритме отжига, приносит новые идеи и идеи для решения крупномасштабных задач планирования рабочих.

2) Можно обнаружить, что после нескольких итераций на основе алгоритма имитации отжига оптимальное значение решения алгоритма имитации отжига остается неизменным на 60-м поколении. Затем процесс поиска останавливается при достижении 100-го поколения, а решением в это время является оптимальное значение оптимизации, найденное алгоритмом.

Заключение. В статье рассмотрена задача составления расписания рабочих, основанная на моделируемом алгоритме отжига, в основном с двух сторон: совершенствование модели планирования работы рабочих и изучение задачи составления расписания персонала. Большое количество организаций по-прежнему полагаются на старших менеджеров, которые составляют расписание вручную с многолетним опытом составления расписания, и часто наблюдается хаотичное расписание и неудовлетворенность персоналом. Это серьезно тормозит развитие современного информационного менеджмента и усовершенствованного менеджмента в компаниях. В течение длительного времени в зарубежных исследованиях разрабатывались различные алгоритмы математического планирования и эвристики. Тем не менее, существует большой разрыв между трудовыми нормами и потребностями рабочими и ограничениями по сменам. Устанавливается модель планирования рабочего с сильными и слабыми ограничениями, и, наконец, смоделированный алгоритм отжига используется для оптимизации стратегии для решения модели планирования рабочего. Он также оценивает алгоритм решения и получает оптимальную стратегию планирования. Можно видеть, что алгоритм имитации отжига хорошо влияет на интеллектуальный анализ данных управления человеческими ресурсами.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Курейчик В.В., Курейчик В.М., Родзин С.И. Теория эволюционных вычислений. – М.: Физматлит, 2012. – 260 с.
2. Курейчик В.М., Курейчик В.В. Генетический алгоритм разбиения графа // Известия Российской академии наук. Теория и системы управления. – 1999. – № 4. – С. 79-87.
3. Курейчик В.М., Лебедев Б.К., Лебедев О.К. Поисковая адаптация: теория и практика. – М.: Физматлит, 2006. – 272 с. – ISBN 5-9221-0749-6.
4. Кошевой, Н.Д., Кошевая И.И., Бельмега А.В. Оптимизация композиционных планов второго порядка с использованием жадного алгоритма и алгоритма имитации отжига // International Scientific and Practical Conference World science. – 2016. – Т. 1. – № 10 (14). – С. 11-16.
5. Сухов И.В. Анализ алгоритма имитации отжига в решении задачи коммивояжера // Академия педагогических идей Новация. Серия: Научный поиск. – 2019. – № 6. – С. 16-19.
6. Борознов В.О. Исследование решения задачи коммивояжера // Вестник Астраханского государственного технического университета. Управление, вычислительная техника и информатика. – 2009. – С. 147-151.
7. Курейчик В.М., Курейчик В.В. Генетический алгоритм разбиения графа // Известия Российской академии наук. Теория и системы управления. – 1999. – № 4. – С. 79-87.
8. Кравченко Ю.А., Мансур А.М., Мохаммад Ж.Х. Векторизация текста с использованием методов интеллектуального анализа данных // Известия ЮФУ. Технические науки. – 2021. – № 2 (219). – С. 154-167.
9. Wang W., Liu L., Liu J., and Chen Z. Energy management and optimization of vehicle-to-grid systems for wind power integration // CSEE Journal of Power and Energy Systems. – 2020. – Vol. 7, No. 1. – P. 172-180.
10. Wu W., Wu W., and Wang S. Thermal management optimization of a prismatic battery with shape-stabilized phase change material // International Journal of Heat and Mass Transfer. – 2018. – Vol. 121. – P. 967-977.
11. Курейчик В.М., Лебедев Б.К., Лебедев О.К. Поисковая адаптация: теория и практика. – М.: Физматлит, 2006. – 272 с. – ISBN 5-9221-0749-6.
12. Кулиев Э.В., Лежебоков А.А., Кравченко Ю.А. Роевой алгоритм поисковой оптимизации на основе моделирования поведения летучих мышей // Известия ЮФУ. Технические науки. – 2016. – №7(180). – С. 53-62.
13. Крипак Е.М., Тимофеев Д.Н. Оптимизация использования человеческих ресурсов предприятия на основе решения задачи о назначении с применением нечетко-множественного моделирования // Экономика и предпринимательство. – 2014. – № 4-2 (45). – С. 743-747.
14. Курейчик В.М., Курейчик В.В., Родзин С.И. Модели параллелизма эволюционных вычислений // Вестник Ростовского государственного университета путей сообщения. – 2011. – № 3 (43). – С. 93-97.

15. Кулиев Э.В., Запорожец Д.Ю., Кравченко Ю.А., Семенова М.М. Решение задачи интеллектуального анализа данных на основе биоинспирированного алгоритма // Известия ЮФУ. Технические науки. – 2021. – № 6 (223). – С. 89-99.
16. Кравченко Ю.А., Мансур Али Махмуд, Мохаммад Жуман Хуссейн. Модифицированный метод устранения неоднозначности смысла слов, основанный на методах распределенного представления // Известия ЮФУ. Технические науки. – 2021. – № 3 (220). – С. 92-101.
17. Bray M., Wang W., Rees M.A. et al. KPDGUI: an interactive application for optimization and management of a virtual kidney paired donation program // Computers in Biology and Medicine. – 2019. – Vol. 108. – P. 345-353.
18. Byrne R.H., Nguyen T.A., Copp D.A., Chalamala B.R., and Gyuk I. Energy management and optimization methods for grid energy storage systems // IEEE Access. – 2018. – Vol. 6. – P. 13231-13260.
19. Bova V.V., Kravchenko Y.A., Rodzin S.I., Kuliev E.V. Simulation of the semantic network of knowledge representation in intelligent assistant systems based on ontological approach // Communications in Computer and Information Science (см. в книгах). – 2021. – Vol. 1396 CCIS. – P. 241-252.
20. Кулиев Э.В., Кривенко М.П., Семенова М.М., Игнатьева С.В. Интеллектуальная подсистема поддержки принятия решений на основе биологически правдоподобных алгоритмов самоорганизации // Известия ЮФУ. Технические науки. – 2021. – № 4 (221). – С. 105-116.

REFERENCES

1. Kureychik V.V., Kureychik V.M., Rodzin S.I. Teoriya evolyutsionnykh vychisleniy [Theory of evolutionary calculations]. Moscow: Fizmatlit, 2012, 260 p.
2. Kureychik V.M., Kureychik V.V. Geneticheskiy algoritm razbiyeniya grafa [Genetic algorithm of graph partitioning], *Izvestiya Rossiyskoy akademii nauk. Teoriya i sistemy upravleniya* [Proceedings of the Russian Academy of Sciences. Theory and control systems], 1999, No. 4, pp. 79-87.
3. Kureychik V.M., Lebedev B.K., Lebedev O.K. Poiskovaya adaptatsiya: teoriya i praktika [Search adaptation: theory and practice]. Moscow: Fizmatlit, 2006, 272 p. ISBN 5-9221-0749-6.
4. Koshevoy N.D., Koshevaya I.I., Bel'mega A.V. Optimizatsiya kompozitsionnykh planov vtorogo poriyadka s ispol'zovaniyem zhadnogo algoritma i algoritma imitatsii otzhiga [Optimization of second-order composite plans using a greedy algorithm and an annealing simulation algorithm], *International Scientific and Practical Conference World science*, 2016, Vol. 1, No. 10 (14), pp. 11-16
5. Sukhov I.V. Analiz algoritma imitatsii otzhiga v reshenii zadachi kommvoyazhora [Analysis of the algorithm of simulated annealing in solving the traveling salesman problem], *Akademiya pedagogicheskikh idey Novatsiya. Seriya: Nauchnyy poisk* [Academy of Pedagogical Ideas Novation. Series: Scientific search], 2019, No. 6, pp. 16-19.
6. Boroznov V.O. Issledovaniye resheniya zadachi kommvoyazhera [Investigation of the solution of the traveling salesman problem], *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Upravleniye, vychislitel'naya tekhnika i informatika* [Bulletin of the Astrakhan State Technical University. Management, computer engineering and computer science], 2009, pp. 147-151.
7. Kureychik V.M., Kureychik V.V. Geneticheskiy algoritm razbiyeniya grafa [Genetic algorithm of graph partitioning], *Izvestiya Rossiyskoy akademii nauk. Teoriya i sistemy upravleniya* [Proceedings of the Russian Academy of Sciences. Theory and control systems], 1999, No. 4, pp. 79-87.
8. Kravchenko Yu.A., Mansur A.M., Mokhammad Zh.Kh. Vektorizatsiya teksta s ispol'zovaniyem metodov intellektual'nogo analiza dannykh [Text vectorization using data mining methods], *Izvestiya YuFU. Tekhnicheskiye nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 2 (219), pp. 154-167.
9. Wang W., Liu L., Liu J., and Chen Z. Energy management and optimization of vehicle-to-grid systems for wind power integration, *CSEE Journal of Power and Energy Systems*, 2020, Vol. 7, No. 1, pp. 172-180,
10. Wu W., Wu W., and Wang S. Thermal management optimization of a prismatic battery with shape-stabilized phase change material, *International Journal of Heat and Mass Transfer*, 2018, Vol. 121, pp. 967-977.
11. Kureychik V.M., Lebedev B.K., Lebedev O.K. Poiskovaya adaptatsiya: teoriya i praktika [Search adaptation: theory and practice]. Moscow: Fizmatlit, 2006, 272 p. ISBN 5-9221-0749-6.

12. *Kuliev E.V., Lezhebokov A.A., Kravchenko Yu.A.* Royevoy algoritm poiskovoy optimizatsii na osnove modelirovaniya povedeniya letuchikh myshey [Swarm search engine optimization algorithm based on bat behavior modeling], *Izvestiya YuFU. Tekhnicheskiye nauki* [Izvestiya SFedU. Engineering Sciences], 2016, No. 7 (180), pp. 53-62.
13. *Kripak Ye.M., Timofeyev D.N.* Optimizatsiya ispol'zovaniya chelovecheskikh resursov predpriyatiya na osnove resheniya zadachi o naznachenii s primeneniym nechetko-mnozhestvennogo modelirovaniya [Optimization of the use of human resources of the enterprise based on the solution of the assignment problem using fuzzy-multiple modeling], *Ekonomika i predprinimatel'stvo* [Economics and entrepreneurship], 2014, No. 4-2 (45), pp. 743-747
14. *Kureychik V.M., Kureychik V.V., Rodzin S.I.* Modeli parallelizma evolyutsionnykh vychisleniy [Models of parallelism of evolutionary computations], *Vestnik Rostovskogo gosudarstvennogo universiteta putey soobshcheniya* [Bulletin of the Rostov State University of Railways], 2011, No. 3 (43), pp. 93-97.
15. *Kuliev E.V., Zaporozhets D.Yu., Kravchenko Yu.A., Semenova M.M.* Resheniye zadachi intellektual'nogo analiza dannykh na osnove bioinspirirovannogo algoritma [Solving the problem of data mining based on a bioinspired algorithm], *Izvestiya YuFU. Tekhnicheskiye nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 6 (223), pp. 89-99.
16. *Kravchenko Yu.A., Mansur Ali Makhmud, Mokhammad Zhuman Khussayn.* Modifitsirovannyi metod ustraneniya neodnoznachnosti smysla slov, osnovannyi na metodakh raspredelenogo predstavleniya [A modified method for eliminating the ambiguity of the meaning of words based on distributed representation methods], *Izvestiya YuFU. Tekhnicheskiye nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 3 (220), pp. 92-101.
17. *Bray M., Wang W., Rees M.A. et al.* KPDGUI: an interactive application for optimization and management of a virtual kidney paired donation program, *Computers in Biology and Medicine*, 2019, Vol. 108, pp. 345-353.
18. *Byrne R.H., Nguyen T.A., Copp D.A., Chalamala B.R., and Gyuk I.* Energy management and optimization methods for grid energy storage systems, *IEEE Access*, 2018, Vol. 6, pp. 13231-13260.
19. *Bova V.V., Kravchenko Yu.A., Rodzin S.I., Kuliev E.V.* Simulation of the semantic network of knowledge representation in intelligent assistant systems based on ontological approach, *Communications in Computer and Information Science (sm. v knigakh)*, 2021, Vol. 1396 CCIS, pp. 241-252.
20. *Kuliev E.V., Krivenko M.P., Semenova M.M., Ignat'yeva S.V.* Intellektual'naya podsistema podderzhki prinyatiya resheniy na osnove biologicheskii pravdopodobnykh algoritmov samoorganizatsii [Intelligent decision support subsystem based on biologically plausible self-organization algorithms], *Izvestiya YuFU. Tekhnicheskiye nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 4 (221), pp. 105-116.

Статью рекомендовал к опубликованию к.т.н., доцент С.Г. Буланов.

Кулиев Эльмар Валерьевич – Южный федеральный университет; e-mail: ekuliev@sfedu.ru; г. Таганрог, Россия, тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Котельва Ангелина Васильевна – e-mail: kotelva@sfedu.ru; тел.: 88634371651; студент.

Семенова Марина Максимовна – e-mail: bogomolova@sfedu.ru; тел.: 89185512464; студент.

Игнатьева Сабина Валерьевна – e-mail: signateva@sfedu.ru; тел.: 88634371651; студент.

Кухаренко Анатолий Павлович – e-mail: apkuharenko@sfedu.ru; кафедра интеллектуальных и многопроцессорных систем; к.т.н.; доцент.

Kuliev Elmar Valerievich – Southern Federal University; e-mail: ekuliev@sfedu.ru; Taganrog, Russia; phone: +78634371651; the department of computer aided design; associate professor.

Kotelva Angelina Vasilevna – e-mail: kotelva@sfedu.ru; phone: +78634371651; student

Semenova Marina Maksimovna – e-mail: bogomolova@sfedu.ru; phone: +79185512464; student.

Ignateva Sabina Valerievna – e-mail: signateva@sfedu.ru; phone: +78634371651; student.

Kukharenko Anatoliy Pavlovich – e-mail: apkuharenko@sfedu.ru; the department of intelligent and multiprocessor systems; cand. of eng. sc.; associate professor.

Д.В. Загуменнов, В.В. Мкртчян

**ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ НАДЕЖНОСТИ СХЕМ
ШИРОКОВЕЩАТЕЛЬНОГО ШИФРОВАНИЯ
С АЛГЕБРОГЕОМЕТРИЧЕСКИМИ КОДАМИ МАЛОЙ МОЩНОСТИ***

Рассматриваются схемы специального широковещательного шифрования – криптографический протокол, решающий задачу распространения цифровой продукции среди авторизованных пользователей. Широковещательное шифрование находит применение в различных областях, например, защита данных в компьютерных сетях, кабельное и спутниковое цифровое телевидение, распределенное хранение информации. В схемах широковещательного шифрования данные распространяются свободно, но в зашифрованном виде, и каждому легальному пользователю выдается уникальный набор ключей для их расшифрования. В схемах специального широковещательного шифрования возможны атаки со стороны коалиций злоумышленников из числа авторизованных пользователей, пытающихся создать “пиратские” ключи и получить несанкционированный доступ к распространяемым данным. Эффективный способ борьбы с такими атаками найден в использовании линейных кодов, обладающих специальными идентифицирующими свойствами, в частности, так называемыми “frameproof” (FP) и “traceability” (TA) свойствами. Ранее получены теоретические границы мощности коалиции злоумышленников, в пределах которой применимы схемы, основанные на использовании идентифицирующих алгеброгеометрических кодов. В работе представлена информационная система для проведения экспериментальных исследований надежности схем, основанных на использовании идентифицирующих алгеброгеометрических кодов малой мощности, в частности, для вычисления вероятностей нарушения идентифицирующих свойств таких кодов, в том числе при превышении известных теоретических границ. В качестве примера использования представленной системы приведены и проанализированы результаты вычислительного эксперимента для двух алгеброгеометрических кодов. В заключение рассмотрены открытые вопросы, представляющие интерес для дальнейших исследований, в частности, возможность расширения экспериментальных исследований до кодов произвольной мощности.

Математические методы защиты информации; широковещательное шифрование; алгеброгеометрические коды; идентифицирующие коды.

D.V. Zagumennov, V.V. Mkrtichyan

**EXPERIMENTAL STUDY OF THE RELIABILITY OF BROADCAST
ENCRYPTION SCHEMES WITH LOW-POWER ALGEBRAIC GEOMETRIC
CODES**

Broadcast encryption is a data distribution protocol that solve the problem of distributing digital products to authorized users and prevent unauthorized parties from accessing the data. It is widely used in computer networks data protection, digital television and distributed storage. In broadcast encryption schemes, data is distributed freely, but in encrypted form, and each legal user is given a unique set of keys to decrypt it. However, broadcast encryption schemes are vulnerable to attacks from coalitions of malicious users from among authorized users who are trying to create “pirated” keys and gain unauthorized access to distributed data. Attacks of this kind can be handled in broadcast encryption schemes by using error-correction codes that have special identifying properties, in particular, frameproof (FP) and traceability (TA) properties. Previously, theoretical limits were obtained for the power of a coalition of attackers, within which schemes based on identifying algebraic geometric codes are applicable. The paper presents an information system for conducting experimental studies of schemes reliability based on low-power identifying algebraic geometric codes, in

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-31-90098.

particular, for calculating identifying properties violation probabilities, including when exceeding known theoretical limits. As an example of using the presented system, the results of a computational experiment for two algebraic geometric codes are presented and analyzed. In conclusion, some open questions are considered that are of interest for further research, in particular, the possibility of expanding experimental studies to codes of arbitrary power.

Copy protection; broadcast encryption; algebraic geometric codes; identifying codes.

Введение. В работе рассматриваются схемы защиты легально тиражируемой цифровой продукции от несанкционированного копирования [1–5], называемые схемами специального широковещательного шифрования (ССШШ). В ССШШ распространитель тиражирует данные свободно, но в зашифрованном виде, и каждому легальному пользователю для расшифрования выдаёт уникальный набор ключей и векторов из некоторого линейного кода. Далее пользователи применяют эту информацию при выполнении легального доступа к данным. В случае обнаружения нелегального использования вектора и ключа их владелец может быть идентифицирован контролёром. В ССШШ допускаются атаки следующего вида: некоторые недобросовестные легальные пользователи могут объединяться в коалиции злоумышленников мощности $c \geq 2$ с целью создания пиратских вектора и ключа, которые можно использовать для выполнения нелегального доступа к данным. Исследования в этой области продолжены в других работах, например в [6–10], и в настоящее время представляются актуальными в связи с ростом популярности широковещательных служб и увеличения многообразия атак.

Для борьбы с коалиционными атаками в [3] предложен метод обнаружения членов коалиций, основанный на использовании линейных кодов, обладающих так называемыми идентифицирующими свойствами, например, свойствами c -FP и c -TA.

Интересной представляется задача исследования применимости алгеброгеометрических кодов [11–13] в ССШШ. Ранее проведены теоретические исследования наличия идентифицирующих свойств у алгеброгеометрических кодов [14–15], получены теоретические границы мощности коалиций злоумышленников, в пределах которых некоторые классы алгеброгеометрических кодов обладают c -FP и c -ТА свойствами.

Актуальной представляется задача экспериментального исследования надёжности ССШШ в случае, когда мощность атакующей коалиции превышает теоретические границы: интересно, как часто нарушаются идентифицирующие свойства у алгеброгеометрических кодов в данных условиях. Для проведения такого исследования нужно, во-первых, формализовать алгеброгеометрический код, представить его в удобном для необходимых вычислений виде: в работе коды представляются в виде порождающей и проверочной матриц кода, конструируемых с помощью пакета компьютерной алгебры Magma. Во-вторых, необходимо спланировать вычислительный эксперимент и написать программу, которая его реализует: в ходе работы такая программа разработана, но алгоритм, используемый в ней, имеет экспоненциальную сложность, что накладывает вычислительные ограничения на мощность рассматриваемых кодов. С целью иллюстрации работы алгоритма выбраны два алгеброгеометрических кода, для которых проведены соответствующие эксперименты. Результаты этих экспериментов и их анализ представлены в заключительном разделе.

Идентифицирующие коды. Пусть $F = F_q$ – конечное поле, и C – линейное подпространство размерности k в F^n . Подпространство C называют линейным кодом, k называют размерностью кода, а n – длиной кода. Также рассмотрим в F^n метрику $\rho(x, y) = |\{1 \leq i \leq n : x_i \neq y_i\}|$, то есть число координат, в которых вектора не совпадают. Эта метрика называется расстоянием

Хемминга. Минимальное из значений $\rho(x, y)$ для всех различных векторов x и y из кода C называется минимальным кодовым расстоянием кода и обозначается как d . Рассмотрим также множество $I(x, y)$ координат, совпадающих в векторах x и y , ясно, что $|I(x, y)| = n - \rho(x, y)$.

Коалицией кода C мощностью $c \in \mathbb{N} \setminus \{1\}$ называется набор из c различных кодовых векторов, множество всех коалиций кода C мощностью не более c обозначим как $coal_c(C)$. Множеством потомков коалиции C_0 называется следующее множество:

$$desc(C_0) = \{(y_1, \dots, y_n) \in F^n : \exists 1 \leq i \leq c, \exists 1 \leq j \leq n, y_i = u_{i,j}\}.$$

Код C называется c -FP кодом ([3], определение 1.1.1), если:

$$\forall C_0 \in coal_c(C) \forall z \in C \setminus C_0 z \notin desc(C_0).$$

Множеством FP-компрометации кода C называется множество таких значений c , для которых код C не является c -FP кодом. Очевидно, что это множество является целочисленным лучом, началом которого является некоторое натуральное число $R_{FP}(C)$. Это число называется рубежом FP-компрометации для кода C .

Код C называется c -ТА кодом ([3], определение 1.1.4), если:

$$\forall C_0 \in coal_c(C) \forall v \in C \setminus C_0 \forall u \in desc(C_0) \exists \omega \in C_0: \rho(\omega, u) < \rho(v, u).$$

Множеством ТА-компрометации кода C называется множество таких значений c , для которых код C не является c -ТА кодом. Аналогично, как и для свойства FP, что множество ТА-компрометации является целочисленным лучом, началом которого является некоторое натуральное число $R_{TA}(C)$, называемое рубежом ТА-компрометации для кода C .

Алгеброгеометрические коды. Пусть $F^{hom}[X, Y, Z]$ – множество однородных многочленов от трех переменных X, Y, Z над полем F . Пусть $G \in F^{hom}[X, Y, Z]$ – неприводим над полем и всеми его расширениями. Пусть $\chi = \chi(G, F)$ – проективная кривая в двумерном проективном пространстве $P^2(F)$, заданная нулями G , т.е. точки двумерного пространства $(a:b:c)$, такие, что $G(a, b, c) = 0$.

Рассмотрим теперь поле рациональных функций над кривой χ и обозначим его $F(\chi)$. Это поле состоит из рациональных функций, числители и знаменатели которых являются многочленами из $F^{hom}[X, Y, Z]$ одной и той же степени, при этом две функции считаются равными, если они отличаются на аддитивную константу, делящуюся на многочлен G ([12], 2.5.4). Согласно [12], 2.5.2, для каждой функции $H \in F(\chi)$ в каждой точке M кривой специальным образом задают целочисленный параметр, называемый порядком функции в точке и обозначаемый как $ord_M(H)$. В самом простом случае это значение равно разности кратности точки M как корня числителя и кратности точки M как корня знаменателя.

Дивизором D на кривой называют формальную сумму вида: $D = \sum_{M \in \chi} a_M M, a_M \in \mathbb{Z}$. Сумма всех таких a_M называются степенью дивизора и обозначается $deg(D)$, а множество точек M , для которых $a_M \neq 0$, называется носителем дивизора и обозначается как $supp(D)$. Дивизор называют эффективным и пишут $D \geq 0$, если все a_M неотрицательны. Таким образом, на множестве дивизоров вводится частичный порядок. Для каждой рациональной функции $H \in F(\chi)$ можно задать собственный дивизор: $(H) = \sum_{M \in \chi} ord_M(H)M$. Также дивизоры можно поточечно складывать и вычитать:

$$D_1 + D_2 = \sum_{M \in \chi} a_M M + \sum_{M \in \chi} b_M M = \sum_{M \in \chi} (a_M + b_M)M.$$

Пусть D – дивизор на гладкой проективной кривой χ , тогда множество

$$L(D) = \{H \in F(\chi) : (H) + D \geq 0\}$$

является конечномерным линейным пространством над полем F и называется пространством Римана-Роха.

Пусть $P = \{P_1, \dots, P_n\} \subset \chi, \text{supp}(D) \cap P = \emptyset, \text{deg}(D) = \alpha < n$. Определим алгеброгеометрический код как образ линейного отображения:

$$Ev: L(D) \rightarrow F^n, Ev(H) = (H(P_1), \dots, H(P_n))$$

и обозначим его как $C = C(\chi, P, D)$.

При работе с алгеброгеометрическими кодами удобно иметь порождающую и проверочную матрицы этого кода: это позволяет абстрагироваться от объектов алгебраической геометрии и проводить стандартные вычисления в терминах линейной алгебры. Напомним, что порождающей матрицей A алгеброгеометрического кода C является матрица отображения Ev , т.е. матрица размера $\dim(L(D)) \times n$, строками которой являются элементы базиса пространства Римана-Роха, отображенные в F^n с помощью Ev . Проверочной матрицей B является матрица кода, ортогональная к коду C как к векторному пространству. Таким образом, $A \times B^T = 0$.

Для получения порождающей и проверочной матриц кода использован пакет компьютерной алгебры *Magma* [16]. В листинге 1 приведем текст программы на языке программирования *Magma*, выводящую на экран порождающую и проверочную матрицы кода $C = C(\chi(X^5 + Y^4Z + YZ^4, F), 9 \cdot (0:1:0), P)$ над полем $F = F_{16}$, где $P = \chi \setminus \{(0:1:0)\}$, т.е. множество всех точек кривой, кроме точки $(0:1:0)$. На 1 строке мы строим конечное поле F_{16} , на второй – проективное пространство $P^2(F)$ над этим полем. Затем на строках 3-5 мы задаем многочлен G , строим кривую и вычисляем ее точки. Убедившись, что в массиве точек *places* необходимая нам точка имеет индекс 1, далее на строке 6 мы строим дивизор $D = 9 \cdot (0:1:0)$. Затем на строке 7 удаляем первую точку из массива *places*. На 8 строке мы получаем код. Далее остается вывести на экран порождающую и проверочную матрицы кода.

Схема эксперимента по исследованию надежности ССШШ с алгеброгеометрическими кодами. Пусть $s \in \mathbb{N} \setminus \{1\}$, C – алгеброгеометрический код. Пусть C_0 – случайно выбранная коалиция из $\text{coal}_c(C)$, а $\omega \in \text{desc}(C_0)$ – случайно выбранный потомок этой коалиции.

Листинг 1.

```

1 F<w> := GF(16);
2 P2<x,y,z> := ProjectiveSpace(F, 2);
3 G := x^5 + y^4*z + y*z^4;
4 X := Curve(P2, G);
5 places := Places(X, 1);
6 D := 9 * places[1];
7 Exclude(~places, places[1]);
8 C := AlgebraicGeometricCode(places, D);
9 GeneratorMatrix(C);
10 ParityCheckMatrix(C).
```

Рассмотрим следующие события: 1) A_1 : потомок ω является кодовым вектором. Это означает, что при наступлении события A_1 можно сделать вывод, что код C не является с-FP-кодом. 2) A_2 : потомок ω относительно метрики ρ лежит ближе к $C \setminus C_0$, чем к коалиции. Т.е., после вычисления d – минимального из расстояний

между ω и векторами коалиции – найден кодовый вектор из $C \setminus C_0$, расстояние от которого до ω не больше d . Это означает, что при наступлении события A_2 можно сделать вывод, что код C не является с-ТА кодом.

Рассмотрим схему проведения вычислительных экспериментов:

♦ Выбрать алгеброгеометрический код C над полем $F = F_q$, построить его порождающую и проверочную матрицы. Пусть k – размерность кода, а n – его длина.

♦ Положить $s = 2$, выбрать s случайных элементов пространства F^k , отобразить их с помощью отображения $E\nu$ в F^k , умножив на порождающую матрицу A . Таким образом, мы получим случайную коалицию C_0 .

♦ Выбрать случайным образом потомка ω коалиции C_0 , не совпадающего ни с одним вектором из коалиции, т.е. $\omega \in desc(C_0) \setminus C_0$.

♦ Проверить условие $\omega \in C \setminus C_0$, вычислив синдром: $s = \omega B^T$, где B – проверочная матрица кода. Если $s = 0$, то выполняется, что $\omega \in C \setminus C_0$, тогда необходимо зафиксировать события A_1 и A_2 , а в противном случае продолжить.

♦ Проверить, лежит ли ω ближе к C_0 , чем к $C \setminus C_0$. Для этого вычислить расстояние d – минимальное из расстояний между ω и векторами C_0 , а затем, перебрав все кодовые векторы из $C \setminus C_0$ и вычислив расстояния от ω до них, проверить, существует ли кодовый вектор из $C \setminus C_0$, расстояние от которого до ω не больше d . Если есть, то зафиксировать событие A_2 .

Эксперименты для данной величины s повторяются заданное количество раз. Далее значение s увеличивается, и эксперименты повторяются, пока $s < n$.

Вычислим, сколько раз нужно провести эксперимент при заданном s . Будем считать, что случайная величина, определенная на пространстве исходов событий A_1 и A_2 , имеет распределение Стьюдента. Выберем для события доверительную вероятность $p_\alpha = 0.99$ и точность оценки $\delta = 0.005$. Согласно [17], глава 16, параграф 15, $\delta = t^{-1}(p_\alpha) \sqrt{\frac{p(1-p)}{N}}$, где $t(x)$ – распределение Стьюдента, p – частота появления события, N – число экспериментов. Так как p неизвестно, мы можем полагать, что $p = 0.5$. Тогда в нашем случае $N = 66564$. В работе мы провели 70000 экспериментов для каждого из s .

Алгеброгеометрические коды, выбранные для проведения экспериментов. В качестве примера были проведены эксперименты для двух алгеброгеометрических кодов:

$$C_1 = (\chi_1 = \chi_1(Y^2Z + XYZ + YZ^2 - X^3 - Z^3, F), 3 \cdot (0: 1: 0), P_1),$$

$$C_2 = (\chi_1(X^3Y + Y^3Z + Z^3X, F), 5 \cdot (0: 1: 0), P_2),$$

где $P_1 = \chi_1 \setminus \{(0: 1: 0)\}$, $P_2 = \chi_2 \setminus \{(0: 1: 0), (1: 0: 0)\}$, $F = F_8$. Для первого кода длина $n = 13$, а для второго $n = 22$.

Приведем теорему о рубежах компрометации FR-свойства и ТА-свойства и затем вычислим теоретические оценки этих рубежей для кодов C_1 и C_2 .

Теорема 1 ([14], теоремы 3, 4). Пусть $C = C(\chi(G, F), D, \{P_1, \dots, P_n\})$ – алгеброгеометрический код над полем F . Тогда

$$R_{FP}(C) \geq \lceil \frac{n}{deg(D)} \rceil, R_{TA}(C) \geq \lceil \sqrt{\frac{n}{deg(D)}} \rceil$$

Если Q – единственная точка на кривой, и $D = \alpha Q$, а $|P| > 1$, то

$$R_{FP}(C) \leq \lceil \frac{n}{deg(D)/deg(G)} \rceil, R_{TA}(C) \leq \lceil \frac{n+\alpha}{2deg(D)/deg(G)} \rceil.$$

Согласно теореме 1,

$$\begin{aligned} \lceil \frac{13}{3} \rceil = 5 &\leq R_{FP}(C_1) \leq \lceil \frac{13}{\lfloor 3/3 \rfloor} \rceil = 13, \\ \lceil \sqrt{\frac{13}{3}} \rceil = 3 &\leq R_{TA}(C_1) \leq \lceil \frac{13+3}{2\lfloor 3/3 \rfloor} \rceil = 8, \\ R_{FP}(C_2) &\geq \lceil \frac{22}{5} \rceil = 5, R_{TA}(C_2) \geq \lceil \sqrt{\frac{22}{5}} \rceil = 3. \end{aligned}$$

Покажем, что для кода C_1 значение R_{TA} равно ровно 3, а значение R_{FP} равно ровно 5. Для этого предъявим коалицию C_0^1 мощности 3 и такого ее потомка, который лежит ближе к множеству $C \setminus C_0^1$, чем к коалиции C_0^1 , относительно метрики ρ . Это будет означать, что

$R_{TA}(C_1) \leq 3$, и тогда, учитывая оценки выше, это также означает, что $R_{TA}(C_1) = 3$. Далее предъявим коалицию C_0^2 мощности 5 и ее потомка, являющегося кодовым вектором. Аналогично, это будет означать, что $R_{FP}(C_1) = 5$. Коалиции будем строить согласно принципу, описанному в [5], лемме 1. Пусть $F = F_8 = F_2[\xi]/(\xi^3 + \xi + 1)$, а множество P_1 упорядочено следующим образом:

$$P_1 = \{(1:0:1), (\xi^4:1:1), (\xi^2:1:1), (\xi:1:1), (\xi:\xi:1), (\xi^5:\xi:1), (\xi^6:\xi:1), (\xi^2:\xi^2:1), (\xi^5:\xi^2:1), (\xi^3:\xi^2:1), (\xi^3:\xi^4:1), (\xi^4:\xi^4:1), (\xi^6:\xi^4:1)\}.$$

Тогда коалиция C_0^1 состоит из векторов:

$$\begin{aligned} u_1 &= (\xi, \xi^3, \xi^3, \xi^3, 0,0,0, \xi^4, \xi^4, \xi^4, \xi^2, \xi^2, \xi^2), \\ u_2 &= (\xi^2, \xi^6, \xi^6, \xi^6, \xi^4, \xi^4, \xi^4, 0,0,0, \xi, \xi, \xi), \\ u_3 &= (\xi^4, \xi^5, \xi^5, \xi^5, \xi^2, \xi^2, \xi^2, \xi, \xi, \xi, 0,0,0). \end{aligned}$$

Очевидно, что можно построить потомка ω этой коалиции:

$$\omega = (\xi, \xi^3, \xi^3, \xi^3, 0,0,0,0,0,0,0,0,0),$$

расстояние ρ от которого до нулевого вектора равно 4.

При этом $\rho(u_1, \omega) = 6, \rho(u_2, \omega) = 9, \rho(u_3, \omega) = 9$, т.е. ω лежит ближе к множеству $C \setminus C_0^1$, чем к коалиции C_0^1 , относительно метрики ρ , значит, $R_{TA}(C_1) = 3$. В коалицию C_0^2 поместим u_1, u_2, u_3 , а также следующие кодовые векторы:

$$\begin{aligned} u_4 &= (1,0,0,0, \xi^3, \xi^3, \xi^3, \xi^6, \xi^6, \xi^6, \xi^5, \xi^5, \xi^5), \\ u_5 &= (0, 1,1,1, \xi, \xi, \xi, \xi^2, \xi^2, \xi^2, \xi^4, \xi^4, \xi^4). \end{aligned}$$

Одним из потомков коалиции C_0^2 является нулевой вектор, это означает, что $R_{FP}(C_1) = 5$. Таким образом, теоретические исследования показывают, что $R_{FP}(C_1) = 5, R_{TA}(C_1) = 3, R_{FP}(C_2) \geq 5, R_{TA}(C_2) \geq 3$.

Результаты эксперимента. Вычисления проведены с помощью программы, реализованной на языке программирования Python с использованием библиотек *numpy* и *galois*. Использован компьютер с 12 ядрами мощностью 3.2 ГГц и ОЗУ объемом 32 Гб. Кроме того, для каждого из значений c эксперименты запущены параллельно на разных ядрах процессора с помощью библиотеки *concurrent.futures*.

В табл. 1 приведены полученные значения вероятности $p(A_1, c), p(A_2, c)$ появления событий A_1 и A_2 для кода C_1 , а в табл. 2 – значения $p(A_1, c), p(A_2, c)$ для кода C_2 . На рис. 1 показан график зависимости оценок вероятности p события A_2 для кодов C_1 и C_2 от величины c .

Таблица 1

Значения вероятностей наступления событий A_1 и A_2 для кода C_1

c	2	3	4	5	6	7
$p(A_1, c)$	0	0	0	0	0	
$p(A_2, c)$	0	0,052	0,188	0,315	0,406	0,478
c	8	9	10	11	12	13
$p(A_1, c)$	0	0	0	0	0	
$p(A_2, c)$	0,529	0,570	0,595	0,618	0,634	0,649

Проведенный вычислительный эксперимент подтверждает справедливость теоретических исследований, согласно которым значения R_{FP} и R_{TA} для кода C_1 равны 5 и 3 соответственно, так как оценки вероятности события A_1 для $c < R_{FP}(C_1) = 5$ и события A_2 для $c < R_{TA}(C_1) = 3$ на практике равны нулю. Аналогично, подтверждается, что $R_{TA}(C_2) \geq 3$, $R_{FP}(C_2) \geq 5$. Более того, на основании эксперимента можно утверждать, что для C_2 получено точное значение рубежа $R_{TA}(C_2)$, так как для C_2 выполняется, что $p(A_2, 3) > 0$, т.е., $R_{TA}(C_2) \leq 3$, и тогда, учитывая теоретическую границу $R_{TA}(C_2) \geq 3$, получаем, что $R_{TA}(C_2) = 3$.

Таблица 2

Значения вероятностей наступления событий A_1 и A_2 для кода C_2

c	2	3	4	5	6	7	8
$p(A_1, c)$	0	0	0	0	0		
$p(A_2, c)$	0	0,002	0,033	0,110	0,202	0,289	0,360
c	9	10	11	12	13	14	15
$p(A_1, c)$	0	0	0	0	0	0	0
$p(A_2, c)$	0,419	0,470	0,503	0,534	0,560	0,584	0,597
c	16	17	18	19	20	21	22
$p(A_1, c)$	0	0	0	0	0	0	0
$p(A_2, c)$	0,613	0,622	0,632	0,642	0,647	0,653	0,660

Обратим теперь внимание на оценки частоты нарушения идентифицирующих свойств при превышении теоретически допустимой мощности коалиции злоумышленников. Мы видим, что C_2 имеет лучшие показатели надежности при превышении границ. Однако отметим, что это достигается за счет большей длины кодового вектора, а обработка векторов большей длины требует больших вычислительных затрат и затрат памяти.

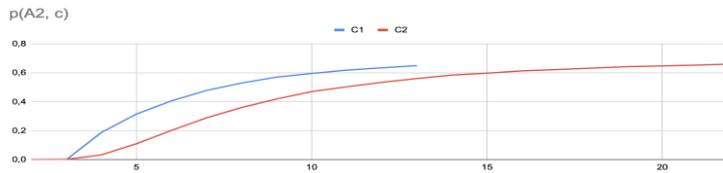


Рис. 1. График зависимости вероятности p события A_2 для кодов C_1 и C_2 от величины c

Для кода C_1 вероятность нарушения ТА-свойства менее 10% при $c = 3$, менее 50% при значениях c вплоть до 7; для кода C_2 вероятность нарушения ТА-свойства менее 10% при $c = 3, 4$, и менее 50% при значениях c вплоть до 10. Эти данные могут быть использованы при проектировании схем широкополосного шифрования: полученные значения означают, что при незначительном превышении рубежа R_{TA} для рассматриваемых кодов вероятность успешного нахождения нелегального пользователя из коалиции злоумышленников все еще достаточно велика.

Рассмотрим открытые вопросы, представляющие интерес для дальнейших исследований. Первым вопросом является переход от экспоненциальной сложности проверки события A_2 к полиномиальной. Возможным вариантом реализации такого перехода является использование полиномиального списочного декодера. Списочным декодером называют алгоритм, который по заданному коду, произвольному вектору из пространства (необязательно кодовому) и расстоянию d получает список всех кодовых векторов, лежащих не более, чем на расстоянии d от заданного вектора. Таким образом, при проверке события A_2 можно использовать списочный декодер, на вход которого необходимо подать код, случайного потомка случайной коалиции ω и минимальное из расстояний от ω до векторов из коалиции. Если списочный декодер будет иметь полиномиальную сложность, то и вся проверка будет иметь полиномиальную сложность. Полиномиальные списочные декодеры для алгеброгеометрических кодов существуют, например, [18–20]. Вторым вопросом является построение более точного события – индикатора нарушения FP-свойства. Предложенное в работе событие A_1 ни разу не наступило ни для кода C_1 , ни для кода C_2 . Это говорит как и об устойчивости FP-свойства для рассмотренных алгеброгеометрических кодов, так и о несовершенстве эксперимента.

Заключение. В работе представлен вычислительный эксперимент и программные инструменты для его проведения, которые могут быть использованы с целью оценки надежности схем специального широкополосного шифрования, основанных на использовании алгеброгеометрических кодов малой мощности, в том числе в условиях превышения допустимых теоретических рубежей мощности злоумышленников. Представленная схема может быть использована, во-первых, для уточнения теоретических рубежей R_{TA} и R_{FP} , и во-вторых, для анализа поведения схем специального широкополосного шифрования при различных условиях.

В качестве примера вычислительный эксперимент проведен для конкретных алгеброгеометрических кодов C_1 и C_2 . Для них получены оценки надежности схем специального широкополосного шифрования, в частности, частота нарушения идентифицирующих FP-свойства и ТА-свойства при превышении теоретически допустимой мощности коалиции злоумышленников. Эти данные проанализированы, в том числе в сравнении с полученными ранее теоретическими границами. С помощью эксперимента удалось вычислить неизвестную ранее точную оценки R_{TA} для кода C_2 . Очевидно, что представленный вычислительный эксперимент может быть использован для анализа других алгеброгеометрических кодов малой мощности.

Рассмотрены открытые вопросы, представляющие интерес для дальнейших исследований.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Fiat A., Naor M. Broadcast Encryption // Advances in cryptology. Lecture Notes in Computer Science 773. – SpringerVerlag, 1994. – P. 480-491.
2. Chor B., Fiat A., Naor M. Tracing traitors // Advances in cryptology – CRYPTO'94. – Springer Berlin Heidelberg, 1994. – P. 257-270.
3. Staddon J.N., Stinson D.R., Wei R. Combinatorial properties of frameproof and traceability codes // Information Theory, IEEE Transactions on. – 2001. – Vol. 47, No. 3. – P. 1042-1049.
4. Stinson D.R., Wei R. Combinatorial properties and constructions of traceability schemes and frameproof codes // Information Theory, IEEE Transactions on. – 2001. – Vol. 47, No. 3. – P. 1042-1049.
5. Silverberg A., Staddon J., Walker J.L. Applications of list decoding to tracing traitors // Information Theory, IEEE Transactions on. – 2003. – Vol. 49, No. 5. – P. 1312-1318.
6. Fernandez M., Cotrina J., Soriano M., Domingo N. A Note about the Traceability Properties of Linear Codes // Proc. 10th Int. Conf. on Information Security and Cryptology (ICISC'2007). Seoul, Korea. November 29–30, 2007. Lecture Notes in Comp. Science. – Vol. 4817. – Berlin: Springer, 2007. – P. 251-258.
7. Moreira J., Fernandez M. and Soriano M. A note on the equivalence of the traceability properties of Reed-Solomon codes for certain coalition sizes // 2009 First IEEE International Workshop on Information Forensics and Security (WIFS), London, 2009. – P. 36-40.
8. Кабатянский Г.А. Идентифицирующие коды и их обобщения // Проблемы передачи информации. – 2019. – Т. 55, № 3. – С. 90-111.
9. Егорова Е.Е., Фернандес М., Кабатянский Г.А., Мяо И. Существование и конструкции мультимедийных кодов, способных находить полную коалицию при атаке усреднения и шуме // Проблемы передачи информации. – 2020. – Т. 56, № 4. – С. 97-108.
10. Егорова Е.Е., Кабатянский Г.А. Разделимые коды для защиты мультимедиа от нелегального копирования коалициями // Проблемы передачи информации. – 2021. – Т. 57, № 2. – С. 178-198.
11. Гонна В.Д. Алгебраико-геометрические коды // Известия Российской академии наук. Серия математическая. – 1982. – Т. 46, № 4. – С. 762-781.
12. Влэдуц С.Г., Нозин Д.Ю., Цфасман М.А. Алгеброгеометрические коды. Основные понятия. – М.: МЦНМО, 2003.
13. Hoholdt T., van Lint J. H., Pellikaan R. Algebraic geometry codes // Handbook of coding theory. – 1998. – Vol. 1, No. Part 1. – P. 871-961.
14. Deundyak V.M. and Zagumennov D.V. On the Properties of Algebraic Geometric Codes as Copy Protection Codes // Automatic Control and Computer Sciences. – 2021. – Vol. 55, No. 7. – P. 795-808.
15. Деундяк В.М., Загуменнов Д.В. О границах мощности злоумышленников для идентифицирующих алгеброгеометрических кодов на специальных кривых // Прикладная дискретная математика. – 2021. – № 53. – С. 55-74.
16. Magma Computational Algebra System. Адрес доступа: <http://magma.maths.usyd.edu.au/magma/> (дата обращения: 06.08.2022).
17. Гмурман В.Е. Теория вероятностей и математическая статистика. – 8-е изд. – М.: Высшая школа, 2002. – 479 с.
18. Guruswami V, Sudan M. Improved decoding of Reed-Solomon and algebraic-geometric codes // Foundations of Computer Science. – Palo Alto: IEEE, 1998. – P. 28-37.
19. Shokrollahi A., Wasserman H. List Decoding of Algebraic-Geometric Codes // IEEE Transactions on Information Theory. – 1999. – Vol. 45, No. 2. – P. 432-437.
20. Fernandez M., Soriano M. Identification of Traitors in Algebraic-Geometric Traceability Codes // IEEE Transactions on Signal Processing. – 2004. – Vol. 52, No. 10. – P. 3073-3077.

REFERENCES

1. Fiat A., Naor M. Broadcast Encryption, Advances in cryptology. Lecture Notes in Computer Science. – SpringerVerlag, 1994. Vol. 773. – P. 480-491.
2. Chor B., Fiat A., Naor M. Tracing traitors, Advances in cryptology – CRYPTO'94. Springer Berlin Heidelberg, 1994, pp. 257-270.
3. Staddon J.N., Stinson D.R., Wei R. Combinatorial properties of frameproof and traceability codes, Information Theory, IEEE Transactions on, 2001, Vol. 47, No. 3, pp. 1042-1049.

4. Stinson D.R., Wei R. Combinatorial properties and constructions of traceability schemes and frameproof codes, *Information Theory, IEEE Transactions on*, 2001, Vol. 47, No. 3, pp. 1042--1049.
5. Silverberg A., Staddon J., Walker J.L. Applications of list decoding to tracing traitors // *Information Theory, IEEE Transactions on*, 2003, Vol. 49, No. 5, pp. 1312-1318.
6. Fernandez M., Cotrina J., Soriano M., Domingo N. A Note about the Traceability Properties of Linear Codes, *Proc. 10th Int. Conf. on Information Security and Cryptology (ICISC'2007). Seoul, Korea. November 29–30, 2007. Lecture Notes in Comp. Science*, Vol. 4817. Berlin: Springer, 2007, pp. 251–258.
7. Moreira J., Fernandez M. and Soriano M. A note on the equivalence of the traceability properties of Reed-Solomon codes for certain coalition sizes, *2009 First IEEE International Workshop on Information Forensics and Security (WIFS), London, 2009*, pp. 36-40.
8. Kabatyanskiy G.A. Identifitsiruyushchie kody i ikh obobshcheniya [Identifying codes and their generalizations], *Problemy peredachi informatsii* [Problems of information transmission], 2019, Vol. 55, No. 3, pp. 90-111.
9. Egorova E.E., Fernandes M., Kabatyanskiy G.A., Myao I. Sushchestvovanie i konstruktssii mul'timednykh kodov, sposobnykh nakhodit' polnuyu koalitsiyu pri atake usredneniya i shume [The existence and construction of multimedia codes capable of finding a complete coalition in the averaging attack and noise], *Problemy peredachi informatsii* [Problems of information transmission], 2020, Vol. 56, No. 4, pp. 97-108.
10. Egorova E.E., Kabatyanskiy G.A. Razdelimye kody dlya zashchity mul'timedia ot nelegal'nogo kopirovaniya koalitsiyami [Separable codes to protect multimedia from illegal copying by coalitions], *Problemy peredachi informatsii* [Problems of information transmission], 2021, Vol. 57, No. 2, pp. 178-198.
11. Goppa V.D. Algebraiko-geometricheskie kody [Algebraic-geometric codes], *Izvestiya Rossiyskoy akademii nauk. Seriya matematicheskaya* [Proceedings of the Russian Academy of Sciences. The series is mathematical], 1982, Vol. 46, No. 4, pp. 762-781.
12. Vleduts S.G., Nogin D.Yu., TSfasman M.A. Algebrogeometricheskie kody. Osnovnye ponyatiya [Algebraic-geometric codes. Basic concepts]. Moscow: MTSNMO, 2003.
13. Hoholdt T., van Lint J. H., Pellikaan R. Algebraic geometry codes, *Handbook of coding theory*, 1998, Vol. 1, No. Part 1, pp. 871-961.
14. Deundyak V.M. and Zagumennov D.V. On the Properties of Algebraic Geometric Codes as Copy Protection Codes, *Automatic Control and Computer Sciences*, 2021, Vol. 55, No. 7, pp. 795-808.
15. Deundyak V.M., Zagumennov D.V. O granitsakh moshchnosti zloumyshlennikov dlya identifitsiruyushchikh algebrogeometricheskikh kodov na spetsial'nykh krivykh [On the limits of the power of intruders for identifying algebra-geometric codes on special curves], *Prikladnaya diskretnaya matematika* [Applied Discrete Mathematics], 2021, No. 53, pp. 55-74.
16. Magma Computational Algebra System. Available at: <http://magma.maths.usyd.edu.au/magma/> (accessed 06 August 2022).
17. Gmurman V.E. Teoriya veroyatnostey i matematicheskaya statistika [Probability theory and mathematical statistics]. 8th ed. Moscow: Vysshaya shkola, 2002, 479 p.
18. Guruswami V, Sudan M. Improved decoding of Reed-Solomon and algebraic-geometric codes, *Foundations of Computer Science*. Palo Alto: IEEE, 1998, pp. 28-37.
19. Shokrollahi A., Wasserman H. List Decoding of Algebraic-Geometric Codes, *IEEE Transactions on Information Theory*, 1999, Vol. 45, No. 2, pp. 432-437.
20. Fernandez M., Soriano M. Identification of Traitors in Algebraic-Geometric Traceability Codes, *IEEE Transactions on Signal Processing*, 2004, Vol. 52, No. 10, pp. 3073-3077.

Статью рекомендовал к опубликованию к.ф.-м.н. А.В. Криворучко.

Загуменнов Денис Владимирович – Южный федеральный университет; e-mail: zagumen.denis@gmail.com; г. Ростов-на-Дону, Россия; тел.: +79185820982; кафедра алгебры и дискретной математики; м.н.с.; аспирант.

Мкртчян Вячеслав Витальевич – e-mail: mkrтчян@list.ru; тел.: +79034310555; кафедра алгебры и дискретной математики; к.т.н.; доцент.

Zagumennov Denis Vladimirovich – Southern Federal University; e-mail: zagumen.denis@gmail.com; Rostov-on-Don, Russia; phone: +79185820982; the department of algebra and discrete mathematics; junior researcher; graduate student.

Mkrtichyan Vyacheslav Vital'evich – e-mail: mkrtichan@list.ru; phone: +79034310555; the department of algebra and discrete mathematics; cand. of eng. sc.; associate professor.

УДК 519.224.22

DOI 10.18522/2311-3103-2022-4-50-62

А.К. Мельников, И.И. Левин, А.И. Дордопуло, Л.М. Сластен**ОЦЕНКА ВОЗМОЖНОСТЕЙ ПЕРСПЕКТИВНЫХ ВЫЧИСЛИТЕЛЬНЫХ ТЕХНОЛОГИЙ ДЛЯ РАСЧЕТА ТОЧНЫХ ПРИБЛИЖЕНИЙ РАСПРЕДЕЛЕНИЙ ВЕРОЯТНОСТЕЙ ЗНАЧЕНИЙ СТАТИСТИК**

Статья посвящена оценке аппаратного ресурса вычислительных систем для решения вычислительно-трудоемкой задачи – расчета распределений вероятностей значений статистик методом второй кратности на основе Δ -точных приближений для выборок объемом от 320 до 1280 знаков при мощности алфавита от 128 до 256 символов с точностью $\Delta=10^{-5}$. Общее время решения не должно превышать 30 дней или $2,592 \cdot 10^6$ секунд при круглосуточном режиме вычислений. Использование свойств метода второй кратности позволяет привести вычислительную сложность расчета к диапазону $9,68 \cdot 10^{22} - 1,60 \cdot 10^{52}$ операций с числом проверяемых векторов – от $6,50 \cdot 10^{23}$ до $1,39 \cdot 10^{50}$. Решение этой задачи для указанных параметров выборок в заданное время с помощью современных вычислительных средств (процессоров, графических ускорителей, программируемых логических интегральных схем) требует недостижимого на практике аппаратного ресурса. Поэтому в статье анализируются возможности перспективных квантовых и фотонных технологий для решения задачи с заданными параметрами. Основным преимуществом квантовых вычислительных систем является высокая скорость вычислений для всех возможных значений параметров. Однако, для расчета распределений вероятностей значений статистик квантовое ускорение не будет достигнуто из-за необходимости проверки всех полученных решений, число которых соответствует размерности задачи. Кроме того, текущий уровень развития элементной базы не позволяет создавать и использовать квантовые вычислители с разрядностью 120 кубитов, необходимой для решения рассматриваемой задачи. Фотонные вычислители могут обеспечить высокую скорость вычислений при низком энергопотреблении и для решения рассматриваемой задачи требуют наименьшее число узлов. Однако, нерешенные проблемы с физической реализацией элементов оперативного хранения данных и отсутствием доступной элементной базы не позволяют в обозримой перспективе (5–7 лет) использовать фотонные вычислительные технологии для расчета распределений вероятностей значений статистик, поэтому наиболее целесообразно применение гибридных вычислительных систем, содержащих узлы различных архитектур. Для реализации задачи на различных аппаратных платформах (универсальные процессоры, графические ускорители, программируемые логические интегральные схемы) и конфигурациях гибридных вычислительных систем предложено использование архитектурно-независимого языка программирования высокого уровня SET@L, объединяющего представление вычислений в виде множеств и совокупностей с помощью альтернативной теории множеств П. Вopenка с абсолютным параллелизмом информационного графа и парадигмами аспектно-ориентированного программирования.

Вероятность; статистика; точное распределение; точное приближение; алгоритмическая сложность; квантовые вычисления; фотонные технологии; архитектурно-независимое программирование; язык Set@L.

A.K. Melnikov, I.I. Levin, A.I. Dordopulo, L.M. Slasten

**ANALYSIS OF ADVANCED COMPUTER TECHNOLOGIES
FOR CALCULATION OF EXACT APPROXIMATIONS OF STATISTICS
PROBABILITY DISTRIBUTIONS**

The paper is devoted to the evaluation of the hardware resource of computer systems for solving a computational-expensive problem such as calculation of the probability distributions of statistics by the second multiplicity method based on Δ -exact approximations for samples with a size of 320-1280 characters and an alphabet power of 128-256 characters, and with an accuracy of $\Delta=10^{-5}$. The total solution time should not exceed 30 days or $2.592 \cdot 10^6$ seconds for 24/7 computing. Owing to the use of the properties of the second multiplicity method, the computational complexity of the calculations can be brought to the range of $9.68 \cdot 10^{22}$ - $1.60 \cdot 10^{52}$ operations with the number of tested vectors of $6.50 \cdot 10^{23}$ - $1.39 \cdot 10^{50}$. The solution of this problem for the specified parameters of samples during the given time requires the hardware resource which cannot be provided by modern computer means such as processors, graphics accelerators, programmable logic integrated circuits. Therefore, in the paper we analyze the possibilities of promising quantum and photon technologies for solving the problem with the given parameters. The main advantage of quantum computer systems is the high speed of calculations for all possible parameter values. However, quantum acceleration will not be achieved to calculate the probability distributions of statistics due to the need to check all the obtained solutions. Here, the number of obtained solutions corresponds to the dimension of the problem. In addition, due to the current development level of the quantum hardware components, it is impossible to create and use the 120-qubit quantum computers for the solution of the considered problem. Photon computers can provide high computation speed at low power consumption and require the smallest number of nodes to solve the considered problem. However, unsolved problems with the physical implementation of efficient memory elements and the lack of available hardware components make the use of photon computer technologies impossible for calculation of the probability distributions of statistics in the near future (5-7 years). Therefore, it is most reasonable to use hybrid computer systems containing nodes of different architectures. To solve the problem on various hardware platforms (general-purpose processors, GPUs, FPGAs) and configurations of hybrid computer systems, we suggest to use an architecture independent high-level programming language SET@L. The language combines the representation of calculations as sets and collections (based on the alternative set theory of P. Vopenka), the absolutely parallel form of the problem represented as an information graph, and the paradigm of aspect-oriented programming.

Probability; statistic; exact distribution; exact approximation; algorithmic complexity; quantum calculations; photonic technologies; architecture independent programming; Set@L language.

Введение. Для решения прикладных задач, требующих статистической обработки текстов как осмысленных символьных последовательностей, используются критерии, минимизирующие количество ложных решений о справедливости проверяемых гипотез. Наибольшей относительной эффективностью обладают критерии на основе точных распределений эталонных статистик [1], но расчет точных распределений – это вычислительно трудоемкая задача [1, 2], зависящая от мощности алфавита N и объема выборки n (длины текстовой последовательности).

Вычислительную сложность задачи можно сократить, если вместо точных распределений использовать их точные *приближения* (предельные распределения), которые минимально снижают эффективность используемых критериев [1, 2]. В качестве точных приближений распределений эталонных статистик используются Δ -точные распределения [2], отличающиеся от исходных на заранее задаваемую сколь угодно малую величину Δ . Для расчета точных распределений существует ряд методов, таких как метод вычисления точного распределения статистик типа Колмогорова-Смирнова [3–4], известный классический метод Монте-Карло [5] и другие. Наиболее предпочтительным является метод второй кратности [2], позволяющий при одинаковом ресурсе рассчитывать точные приближения распределе-

ний для максимальных значений параметров выборок. Метод второй кратности, основанный на решении систем линейных уравнений, имеет полиномиальную сложность, но его вычислительная трудоемкость для реальных прикладных задач все равно достаточно велика [2], поэтому расчет точных приближений в оперативно-приемлемое время с помощью современных вычислительных средств затруднен. Оценки необходимого вычислительного ресурса и времени решения задачи с помощью современных процессоров, графических ускорителей и ПЛИС для требуемых на практике параметров распределений приведены в [2]. В данной статье рассматриваются оценки необходимого ресурса и возможного времени решения задачи расчета точных приближений распределений вероятностей значений статистик с помощью перспективных вычислительных технологий – квантовых и фотонных вычислительных систем.

Постановка задачи. Распределения вероятностей значений статистик рассматриваются для алфавита $A_N = \{a_1, \dots, a_N\}$ мощностью $|A_N| = N$ и его выборки объемом n . Для расчета точных приближений распределений вероятностей значений статистики $P_\Delta \{S_{N,n} \geq c\}$, отличающихся от точных распределений $P_T \{S_{N,n} \geq c\}$ на задаваемую заранее сколь угодно малую величину Δ

$$|P_T \{S_{N,n} \geq c\} - P_\Delta \{S_{N,n} \geq c\}| \leq \Delta \quad (1)$$

применяется метод второй кратности (МВК), основанный на решении системы линейных уравнений

$$\begin{cases} \mu_0^{(v)} + \mu_1^{(v)} + \dots + \mu_n^{(v)} = N \\ 1 \cdot \mu_1^{(v)} + 2 \cdot \mu_2^{(v)} + \dots + n \cdot \mu_n^{(v)} = n \end{cases}, \quad (2)$$

где $\mu_j^{(v)}$ – это количества знаков алфавита A_N , встретившихся в v -й выборке объема n ровно j раз.

Метод МВК [2] основан на последовательном переборе векторов $\mu^{(v)} = \{\mu_0^{(v)}, \mu_1^{(v)}, \dots, \mu_n^{(v)}\}$ и проверке каждого $\mu^{(v)}$ на принадлежность к решениям системы линейных уравнений (2). Подробный теоретический обзор применимости и особенностей реализации МВК рассмотрен в [2]. Алгоритмическая сложность метода МВК определяется выражением

$$\begin{aligned} C_{MBK}(P_\Delta \{S_{N,n} \geq c\}) &= L_{\mu(N,n,r)} \times 5 \cdot r + \\ &+ K_\mu(N,n,r) \cdot (5 \cdot (r+1) + 2(N+r) + 3) + \\ &+ 2 \cdot K_\mu(N,n,r) \cdot \log_2 K_\mu(N,n,r) + 2 \cdot K_\mu(N,n,r), \end{aligned} \quad (3)$$

где $L_{\mu(N,n,r)}$ – это сокращенное количество проверяемых векторов возможных решений с.л.у. (2) с ограничениями $\{r \leq n \mid \forall i = \overline{r+1, n}, \mu_i^{(v)} = 0\}$, а $K_\mu(N,n,r)$ – число неотрицательных целочисленных решений (2) с ограничениями r . Основную трудоемкость выражения (3) определяет первый член многочлена [2]

$$L_{\mu(N,n,r)} = (N+1)^{\min([n/N], r)+1} \cdot \frac{(\min([n/N], r))!}{(n + \min([n/N], r))!} \cdot \frac{(n+r)!}{r!}. \quad (4)$$

Из формул (3) и (4) следует, что алгоритмическая сложность расчета точных приближений $C_{MBK}(P_\Delta \{S_{N,n} \geq c\})$ – это многочлен, зависящий не только от параметров выборки N, n , но и от параметра ограничения r , который, в свою очередь, функционально зависит от параметров выборки и точности вычислений Δ , то есть $r = r(N, n, \Delta)$.

Наиболее трудоемкая часть метода МВК – это процедура получения векторов потенциальных решений и проверки их принадлежности множеству решений. Для практических задач мощность алфавита N находится в диапазоне от 128 до 256, а объемы выборки – от 320 до 1280 при требуемом общем времени вычислений не более 30 дней. Поэтому для оценки алгоритмической сложности в [2] были определены заданные в формате (мощность алфавита, объем выборки) следующие выборки: (256, 1280), (128, 640), (128, 320) и (192, 320) с точностью вычислений $\Delta=10^{-5}$. Алгоритмическая сложность и требуемая производительность задачи для этих параметров выборок приведены в табл. 1.

Таблица 1

Характеристики метода расчета точных приближений для различных параметров выборок

№ п/п	Параметры выборки N/n	Параметр ограничения объема выборки r	Число проверок сокращенное $L_{\mu(N,n,r)}$	Число решений $K_{\mu(N,n,r)}$	Общая сложность $C_{МВК}(P_{\Delta}\{S_{N,n} \geq c\}$ (операций)	Необходимая производит. для расчета за 30 дней (оп/сек)
1	256 / 1280	23	$1,39 \times 10^{50}$	$5,60 \times 10^{25}$	$1,60 \times 10^{52}$	$6,15 \times 10^{45}$
2	128 / 640	22	$2,67 \times 10^{27}$	$1,76 \times 10^{20}$	$2,94 \times 10^{29}$	$1,13 \times 10^{23}$
3	192 / 320	14	$6,50 \times 10^{23}$	$4,67 \times 10^{12}$	$4,55 \times 10^{25}$	$1,75 \times 10^{19}$
4	128 / 320	16	$1,21 \times 10^{21}$	$2,23 \times 10^{13}$	$9,68 \times 10^{22}$	$3,72 \times 10^{16}$

Из табл. 1 следует, что вычислительная сложность расчета варьируется от $9,68 \times 10^{22}$ до $1,60 \times 10^{52}$ операций, средняя – порядка $4,55 \times 10^{25}$ операций, при этом нужно проверить от $6,50 \times 10^{23}$ до $1,39 \times 10^{50}$ векторов и получить от $4,67 \times 10^{12}$ до $5,60 \times 10^{25}$ решений.

Количество проверяемых переменных в (2) не превышает $(r+1)$, т.е. для приведенных в табл. 1 параметров не превышает 24: $\mu^{(v)} = \{\mu_0^{(v)}, \mu_1^{(v)}, \dots, \mu_{23}^{(v)}\}$, а остальные переменные равны нулю: $\{\mu_j^{(v)} = 0 \mid j = 24, \dots, n\}$. Значения $\mu_i^{(v)}$ целые и лежат в пределах $\{0 \leq \mu_i^{(v)} \leq 23 \mid i = \overline{0,23}\}$. Для размещения одного $\mu_i^{(v)}$ требуется не менее 5 бит ($2^4 < 23 < 2^5$), а для размещения всего вектора $\mu^{(v)}$, состоящего из 24-х компонентов $\mu_i^{(v)}$, требуется не менее $5 \times 24 = 120$ бит.

Важным свойством метода является возможность его распараллеливания по данным, так как проверка любых $\mu^{(i)}$ и $\mu^{(j)}$ на принадлежность к решениям системы при $i \neq j$ может проводиться независимо или параллельно.

Рассмотрим возможность применения перспективных технологий - квантовых и фотонных вычислительных систем – для решения задачи расчета точных приближений распределений.

Применение квантовых вычислительных технологий для расчета точных приближений распределений. Квантовые вычислительные технологии были предложены в 80-х годах 20 века целым рядом известных ученых, таких как Ричард Ф. Фейнман [6], Пол Бениофф [7], К.А. Валиев и А.А. Кокин [8], Ю.И. Манин [9]. Основная идея квантовых вычислений состоит в том, что квантовая система из q кубитов, работающих когерентно, имеет 2^q линейно независимых состояний. Согласно принципу квантовой суперпозиции, пространство состояний такого квантового регистра является 2^q -мерным гильбертовым пространством [10]. Одна операция над группой из q кубитов вычисляется сразу над всеми возможными её значениями, в отличие от группы классических битов, когда существует только одно текущее значение. Так обеспечивается максимально возможное распараллел-

ливание вычислений по данным и рост производительности до 2^q , которое называют «квантовым ускорением». Физическими системами, реализующими кубиты, могут быть любые объекты, имеющие два квантовых состояния: поляризационные состояния фотонов, электронные состояния изолированных атомов или ионов, спиновые состояния ядер атомов, и т.д.

Структурная схема квантового компьютера, предложенная русскими учеными К.А. Валиевым и А.А. Кокиным [8], представлена на рис. 1.

Р. Фейнман сформулировал принцип [6], что для любого алгоритма можно получить реализацию на квантовой системе, которая будет не хуже его реализации на классической фон-неймановской системе. Однако одновременно показано [9], что «квантовое ускорение» возможно не для всякого алгоритма и для произвольного алгоритма возможность получения квантового ускорения не гарантируется. Особенностью квантовых вычислений является вероятностный характер результата вычислений: получаемый результат верен только с некоторой вероятностью.

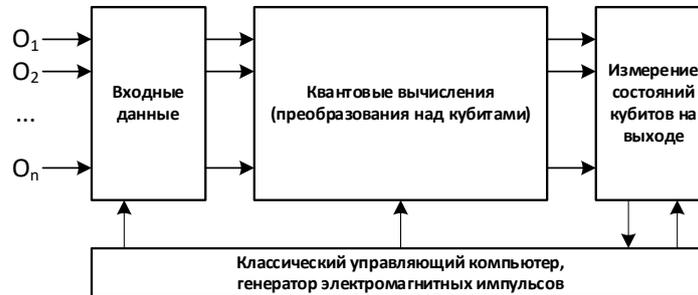


Рис. 1. Структурная схема квантового компьютера

Квантовые вычислители созданы в Гарвардском университете (система из 51 кубита) [11] и в Политехнической школе Университета Париж-Сакле (система из 70 кубит) [12]. Американская компания IonQ [13] анонсировала первый коммерческий квантовый компьютер на ионных ловушках, который содержит 160 кубит, но для квантовых операций используются 79 кубитов, а произвольные квантовые алгоритмы реализуются всего на 11 кубитах. В 2020 году фирма ИВМ [14] представила самый мощный квантовый компьютер на 64 кубитах. В России создание универсального квантового вычислителя из 100 кубитов ожидается к 2024 году [12], а квантового вычислителя, способного оперировать с данными в 120 бит, в настоящее время в мире не существует.

Оценим возможность решения задачи расчета точных приближений распределений на квантовом компьютере. Структурная схема процедуры проверки векторов в задаче расчета точных приближений распределений [2] соответствует структурной схеме квантового компьютера (рис. 1), что позволяет ожидать значительного эффекта при решении задачи на квантовых компьютерах. По минимальной оценке, для размещения вектора проверки решений $\mu^{(v)}$ требуется 120 бит, поэтому в квантовом компьютере нужно иметь 120 кубит, работающих совместно (когерентно). Предположим, что квантовый вычислитель, оперирующий с 120 кубитами (назовем его ГКВ-120), существует, и для него решена проблема достоверности результатов. При расчете точных приближений на ГКВ-120, для вектора проверок $\mu^{(v)} = \{\mu_0^{(v)}, \mu_1^{(v)}, \dots, \mu_{23}^{(v)}\}$ длиной 120 бит, получим все $\{\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(2^{120})}\}$, т.е. ($2^{120} \cong 1,3292 \times 10^{36}$) возможных решений, из которых нужно выделить $K_\mu(N, n, r)$ истинных решений. Для $N=256$ и $n=1280$ из табли-

цы 1 $K_{\mu}(256,1280,23)=5,60075 \cdot 10^{25}$. Таким образом, при измерении значения i -го возможного решения получаем состояние системы $\underbrace{\{\mu^{(i)}, \mu^{(i)}, \dots, \mu^{(i)}\}}_{2^{120}}$. Для полу-

чения $(i+1)$ -го возможного решения получаем состояние системы $\{\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(2^{120})}\}$ и считываем (измеряем) $(i+1)$ -е возможное решение. Для проверки всех возможных решений и получения $K_{\mu}(N, n, r)$ истинных решений необходимо выполнить 2^{120} обращений к ГКВ-120, что кардинально сокращает эффект от одновременного расчета 2^{120} возможных решений.

Необходимость проверки всех полученных решений, число которых соответствует размерности задачи, не позволяет достичь квантового ускорения и является существенным и принципиальным ограничением применения перспективных квантовых вычислительных технологий для расчета точных приближений распределений. Отсутствие технической и технологической базы не только в Российской Федерации, но и в мире, для создания квантовой вычислительной системы, оперирующей 120 кубитами, необходимыми для решения задачи, является дополнительным, технологическим ограничением. Поэтому перспективы применения квантовых вычислительных систем на существующем уровне развития науки и техники для решения задачи расчета точных приближений распределений вероятностей значений статистик достаточно скромны, несмотря на потенциально высокую производительность. Возможно, с развитием работ в области квантовых вычислений, приведенные оценки для рассматриваемой задачи могут быть пересмотрены.

Применение фотонных компьютеров для расчета точных приближений распределений. Еще одним актуальным направлением развития перспективных вычислительных средств, основанных на новых физических принципах, является создание вычислительных систем, использующих эффекты взаимодействия когерентных световых волн, порождаемых лазерным излучением и его носителями - фотонами [15, 16].

Структурная схема фотонного компьютера, разработанная во Всероссийском научно-исследовательском институте экспериментальной физики (г. Саров) [15], представлена на рис. 2. Фотонный компьютер (рис. 2) состоит из четырех крупных блоков: устройства преобразования задачи в программу для фотонного вычислителя (УПЗ), источника лазерного излучения (ИЛИ), устройства ввода-вывода (УВВ), а также фотонного процессора (ФП). В свою очередь, фотонный процессор содержит четыре процессорных элемента, которые содержат арифметико-логические устройства (АЛУ), устройства управления (УУ) и коммутаторы (К). Блоки фотонного компьютера соединены между собой электронными и оптическими каналами, а компоненты процессорного элемента – только оптическими каналами.

Блок УПЗ передает фотонную программу в блок ИЛИ. Лазерное излучение генерируется блоком ИЛИ и поступает в УВВ, где делится на световые лучи по количеству разрядов, одновременно подаваемым в ПЭ. Модуль УВВ формирует фотонную программу и передает ее в фотонный процессор, в котором вычисления выполняются процессорными элементами. Световые лучи взаимодействуют в пределах фотонного процессора, а синхронизацию выполняют оптические линии задержки [16]. В пределах фотонного процессора ПЭ можно соединить оптическими каналами в многопроцессорную среду любой топологии [16]. Одним из немаловажных достоинств фотонных вычислительных систем является низкое энергопотребление при высокой вычислительной производительности.

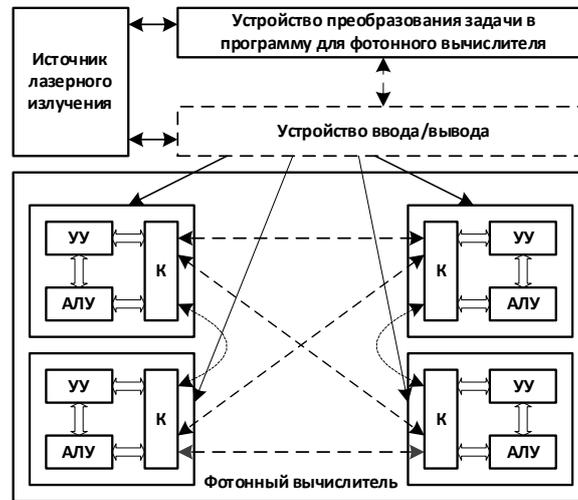


Рис. 2. Структурная схема фотонного компьютера

Показано, что производительность фотонного вычислительного узла может достигать до 10^{18} оп./сек. на 100 Вт потребляемой мощности [17].

Для достижения требуемой производительности фотонного компьютера P_{PHOTON} , нужно использовать N_{PH_UZ} узлов производительностью $P_{PH_UZ}=10^{18}$ оп/сек каждый:

$$P_{PHOTON} = N_{PH_UZ} \times P_{PH_UZ}. \quad (5)$$

Уровень производительности $P_{PHOTON} \geq 1,75 \times 10^{19}$ для расчетов точных приближений (параметры выборки № 3 в табл. 1) обеспечат

$$N_{PH_UZ} \geq 1,75 \times 10^{19} / P_{PH_UZ} = 1,75 \times 10^{19} / 1,0 \times 10^{18} \cong 17,5$$

узлов, то есть не менее 18 фотонных вычислительных узлов. Для достижения производительности, позволяющей вычислять точные приближения для всего спектра параметров рассматриваемых выборок $\{N = 2, 256, n = 1, 5N\}$, необходимо иметь не менее

$$N_{PH_UZ} \geq 6,15 \times 10^{45} / P_{PH_UZ} = 6,15 \times 10^{45} / 1,0 \times 10^{18} = 6,15 \times 10^{27}$$

вычислительных узлов, что существенно ниже, чем во всех рассмотренных ранее [2] технологиях.

В отличие от квантовых вычислительных систем, данных о действующих образцах цифровых фотонных вычислительных систем на настоящий момент не приводится. В большинстве работ [15–17] моделируется функционирование отдельных узлов и оцениваются возможные параметры вычислителя. Известны работы академика В.А. Соифера [18–20] и ряда других исследователей в области аналоговой фотоники, но каждый вычислительный узел создается для задачи с определенными параметрами, а технология разработки и создания аналогового фотонного вычислительного узла для произвольной задачи требует целого цикла научно-исследовательских и опытно-конструкторских работ. Поэтому действующих образцов фотонных вычислительных систем, технологической базы и/или коммерчески доступных компонентов на сегодняшний день не существует. Текущий уровень развития науки и техники не позволяет в ближайшие 5–7 лет рассматривать применение фотонных вычислительных систем для решения задачи

расчета точных приближений распределений вероятностей значений статистик, хотя потенциально такие вычислительные системы обладают очень высокой производительностью и требуют наименьшее число вычислительных узлов с низким энергопотреблением. Возможно, развитие технологической элементной базы для фотонных вычислительных технологий позволит пересмотреть и улучшить представленные оценки для рассматриваемой задачи в ближайшие несколько лет.

Архитектурно-независимое представление задачи расчета точных приближений для гибридных вычислительных систем. Одним из перспективных направлений расчета точных приближений является создание специализированной гибридной вычислительной системы (ГВС) с конвергентной архитектурой, которая может содержать вычислительные узлы современных архитектур: универсальные процессоры, графические ускорители, программируемые логические интегральные схемы. Поскольку вычисление распределений производится методом, единым для всех параметров выборок, то для такой системы особенно важна возможность программирования различных вычислительных архитектур в едином контуре, которая обеспечивается в архитектурно-независимой парадигме программирования [21]. С учетом возможного использования перспективных архитектур квантовых вычислителей и/или фотонных вычислительных узлов, архитектурно-независимое представление задачи расчета точных приближений для гибридных вычислительных систем становится особо значимым.

Такие возможности обеспечивает архитектурно-независимый аспектно-ориентированный язык SET@L, позволяющий сосредоточить внимание разработчика на методах параллелизма при решении задачи, а не на их технической реализации в выбранной вычислительной архитектуре. Язык Set@l позволяет свести проблему переноса программного обеспечения между разными по конфигурации и архитектуре ГВС к формированию аспектов (описаний) технических средств, в которых описываются ключевые моменты разбиения метода по параллелизму на данных технических средствах. При этом исходная программа, реализующая метод обработки данных, остается неизменной [22].

Язык Set@l основан на парадигме аспектно-ориентированного программирования (АОП) [23], согласно которой алгоритм решения прикладной задачи и особенности его реализации описываются в виде отдельных модулей программы. Программа на языке Set@l представляет информационный граф решаемой вычислительной задачи в виде множеств, разбиение и типизация которых задают разные варианты распараллеливания и другие аспекты реализации алгоритма. В отличие от других языков программирования на основе теории множеств, например, языков SETL, SETL2 и SETLX, Set@l использует типизацию множеств по различным критериям и позволяет оперировать нечеткими совокупностями в соответствии с альтернативной теорией множеств [24].

Разделение описаний алгоритма и особенностей его реализации на ВС в языке Set@l обеспечивается использованием парадигмы АОП, согласно которой сквозная функциональность программы, обуславливающая негативные эффекты спутывания и разбрасывания кода, представляется в виде отдельных программных модулей – аспектов. Исходный код, реализующий основную функциональность программы, содержит создаваемую пользователем разметку, которая определяет его взаимодействие с аспектами при трансляции или исполнении. Анализируя разметку и исходный код, транслятор-препроцессор формирует новую исполняемую программу с вплетенной сквозной функциональностью. Применение технологии АОП, как правило, упрощает разработку и дальнейшую поддержку ПО и повышает адаптируемость программ к различным изменениям.

Для неявного описания различных способов распараллеливания алгоритмов в языке архитектурно-независимого программирования ВС Set@1 введена классификация совокупностей по типу параллелизма их элементов при обработке. В тех случаях, когда характер параллелизма элементов совокупности четко определен, используются типы «множество» (par – параллельно-независимая обработка), «кортеж» (seq – последовательная обработка), «кортеж-конвейер» (pipe – конвейерная обработка) и «множество обработки по итерациям» (comp – параллельно-зависимая обработка). Однако в некоторых аспектах программы тип параллелизма ряда совокупностей нельзя определить однозначно, так как неизвестна архитектура ВС, на которой будет реализован алгоритм. В данном случае используется особый тип imp (неопределенный), а типизация совокупностей уточняется в других аспектах программы с помощью специальных синтаксических конструкций. Если аспекты программы на языке Set@1 не изменяют алгоритм в процессе его адаптации к архитектуре ВС, то решение задачи может быть представлено в рамках классической теории множеств Кантора–Больцано. Однако в ряде случаев целесообразно модернизировать алгоритм в соответствии с особенностями его реализации на ВС с определенной архитектурой. В таких случаях некоторые совокупности выделены нечетко и не являются множествами, поэтому их невозможно задать как объекты классической теории множеств. Set@1 позволяет описать разные способы реализации одного и того же алгоритма в единой аспектно-ориентированной программе. Для этого вводится классификация совокупностей по четкости выделения их элементов, предложенная в рамках альтернативной теории множеств П. Вopenка. Тип «множество» (set) соответствует классической четко выделенной совокупности элементов, тип «полумножество» (sm) – нечеткой совокупности, а тип «класс» (cls) – совокупности объектов, тип и разбиение которой нельзя однозначно определить на данном уровне абстракции. Примером вычислительной задачи, для описания которой на языке программирования Set@1 необходимы нечеткие совокупности, является алгоритм решения СЛАУ методом Якоби [25].

В отличие от других подходов к параллельному программированию высокопроизводительных ВС, язык программирования Set@1 задает не только граничные случаи реализации алгоритма, но и семейство промежуточных вариантов, которые неразличимы с точки зрения процедурного программирования, но обеспечивают непрерывность описания модели вычислений. Использование языка Set@1 позволяет синтезировать множество вариантов решения задачи и осуществлять переход между его элементами в зависимости от архитектуры и конфигурации ВС. Указанная возможность проиллюстрирована на примере программы быстрого преобразования Фурье: в зависимости от доступного объема памяти ВС комплексные поворотные коэффициенты вычисляются заранее и загружаются из памяти или рассчитываются через базовые и вспомогательные составляющие в процессе вычислений [26].

Таким образом, алгоритм решения задачи в языке Set@1 представляется в виде архитектурно-независимого исходного кода с помощью аспектно-ориентированного программирования, теоретико-множественного представления кода и реляционного исчисления. Особенности реализации алгоритма выделяются в отдельные аспекты, задающие разбиение на подмножества и типизацию основных совокупностей программы. Программа может оперативно портироваться между вычислительными системами с разными архитектурами и адаптируется к любым изменениям аппаратных ресурсов с помощью аспектов метода обработки, архитектуры и конфигурации. Благодаря предлагаемому подходу к параллельному программированию высокопроизводительных вычислительных систем на языке Set@1, открываются новые перспективы для разработки архитектурно- и ресурснезависимого программного обеспечения.

Язык Set@1 успешно применялся для решения алгоритмически сложных и ресурсоемких задач, таких как решение систем линейных алгебраических уравнений методом Гаусса [22] и Якоби [25], реализации алгоритма быстрого преобразования Фурье [26] и других, которые структурно не отличаются от решаемой задачи. Поэтому возможность использования языка Set@1 при расчете точных приближений распределений на ГВС позволит существенно облегчить сложность программирования вычислительных узлов различных архитектур: универсальных процессоров, графических ускорителей, программируемых логических интегральных схем, квантовых вычислителей и/или фотонных вычислительных узлов.

Заключение и выводы. В статье проанализирована возможность использования перспективных вычислительных технологий для решения вычислительно трудоемкой задачи расчета Δ -точных приближений распределений вероятностей значений статистик методом второй кратности, основанным на решении системы линейных уравнений второй кратности. Метод имеет полиномиальную сложность и допускает распараллеливание на фрагменты по данным.

Оценены возможности перспективных вычислительных технологий на основе квантовых и фотонных компьютеров для расчета распределений методом второй кратности. Анализ возможностей квантовых и фотонных вычислительных технологий показал, что они обладают большим потенциалом для решения задачи, но в настоящее время не позволяют решать задачу вычисления точных приближений распределений выборок с мощностью алфавита не более 256, объемом до 1280 знаков и ограничениями точности вычислений $\Delta=10^{-5}$. Проведен теоретический анализ квантовых вычислительных систем для решения задачи, который показал, что квантовое ускорение невозможно. Так, текущий уровень развития квантовых компьютеров является недостаточным, а также не существует алгоритма и критерия выбора подходящего решения из огромного множества получаемых решений задачи, а уровень развития фотонных компьютеров не позволяет создать вычислитель с требуемым числом вычислительных узлов.

Поддержка. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-07-00545.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Мельников А.К. Сложность расчета точных распределений вероятности симметричных аддитивно разделяемых статистик и область применения предельных распределений // Доклады ТУСУР. – Томск, 2017. – Т. 20, № 4, – С. 126-130. – ISSN 1818-0442.
2. Мельников А.К., Левин И.И., Дордопуло А.И., Писаренко И.В. Анализ возможностей современных вычислительных технологий для расчета точных приближений распределений вероятностей значений статистик // Известия ЮФУ. Технические науки. – 2021. – № 7 (224). – С. 6-19. – ISSN 1999-9429. – DOI: 10.18522/2311-3103-2021-7-6-19.
3. Тимонин В.И., Черномордик О.М. Метод вычисления точного распределения статистик типа Колмогорова-Смирнова при альтернативах Лемана // Теория вероятностей и ее применение. – 1985. – Т. 30. – Вып. 3. – С. 572-573.
4. Тяникова Н.Д., Тимонин В.И. Метод вычисления точных распределений статистик типа Колмогорова-Смирнова в случае нарушения однородности и независимости анализируемых выборок // Наука и образование. – 2014. – № 11. – С. 227-237. – DOI: 10.7463/1114.0740251 ISSN 1994-0448.
5. Расчет распределений методом Монте-Карло.
6. Feynman R.P. Simulating physics with computers (англ.) // International Journal of Theoretical Physics. – 1982. – Vol. 21, Iss. 6. – P. 467-488. – DOI: 10.1007/BF02650179.
7. Benioff P. Quantum mechanical hamiltonian models of turing machines (англ.) // Journal of Statistical Physics (англ.) русск.: journal. – 1982. – Vol. 29, No. 3. – P. 515-546. – DOI: 10.1007/BF01342185. – Bibcode: 1982JSP....29..515B.
8. Валев К.А., Кокин А.А. Квантовые компьютеры: надежды и реальность. – Ижевск: РХД, 2004. – 320 с.

9. Манин Ю.И. Вычислимое и невычислимое. – М.: Сов. радио, 1980. – 128 с. – (Кибернетика).
10. Морен К. Методы гильбертова пространства. – М.: Мир, 1965. – 570 с.
11. Физики из России и США создали первый 51-кубитный квантовый компьютер. – <https://ria.ru/20170714/1498476410.html>? (дата обращения: 2 мая 2022).
12. Как Россия потратит 15 млрд. на создание квантового компьютера. – https://gov.cnews.ru/articles/2020-01-9_kak_rossiya_potratit_15_mlrld_na_sozdanie? (дата обращения: 2 мая 2022).
13. IonQ заявила о создании самого мощного в мире квантового компьютера. – <https://3dnews.ru/1060908/kompaniya-ion-q-soobshchila-o-sozdanii-samogo-moshchnogo-v-mire-kvantovogo-kompyutera?> (дата обращения: 2 мая 2022).
14. ИВМ продемонстрировали квантовый компьютер с «квантовым объемом» 64. – <https://habr.com/ru/company/raiffeisenbank/news/t516062/?> (дата обращения: 2 мая 2022).
15. Степаненко С.А. Фотонная вычислительная машина. Принципы реализации. Оценка параметров // Вычислительная техника. Доклады Академии наук. – М., 2017. – Т. 476, № 4. – С. 389-394. – DOI: 10.1134/S1064562417050234.
16. Степаненко С.А. Фотонный компьютер: структура и алгоритмы, оценка параметров // Фотоника. – 2017. – № 7 / 67. – С. 72-83. – DOI: 10.22184/1993-7296.2017.67.7.72.83.
17. Степаненко С.А. Интерференционные логические элементы / представлено академиком РАН Ю.А. Трутневым. Доклады Академии наук. Математика, информатика, процессы управления. – М., 2020. – Т. 493. – С. 64-69. – DOI: 10.31857/S2686954320040189.
18. Soifer V.A., Kharitonov S.I., Khonina S.N. etc. Spiral caustics of vortex beams // Photonics. – 2021. – Vol. 8, Issue 1. – P. 1-20. – DOI: 10.3390/photonics8010024.
19. Soifer V.A., Doskolovich L.L., Bezus E.A. etc. Spatial differentiation of optical beams using a resonant metal-dielectric-metal structure // Journal of Optics. – 2021. – Vol. 23, Issue 2. – DOI: 10.1088/2040-8986/abe63b.
20. Soifer V.A., Doskolovich L.L., Bezus E.A. etc. An Optical Differentiator Based on a Three-Layer Structure with a W-Shaped Refractive Index Profile // Journal of Experimental and Theoretical Physics. – 2018. – Vol. 127, Issue 2. – P. 202-209. – DOI: 10.1134/S1063776118080174.
21. Левин И.И., Мельников А.К. Управление гибридными вычислительными системами на языке SET@L // Materiály XI mezinárodní vědecko-praktická konference «Aktuální vymoženosti vědy – 2015» Díl 7. Moderní informační technologie. Praha. Publishing House «Education and Science», 2015. – 96 с. – P. 23-28. – ISSN 978-966-8736-05-6.
22. Левин И.И., Дордопуло А.И., Писаренко И.В., Мельников А.К. Подход к архитектурно-независимому программированию вычислительных систем на основе аспектно-ориентированного языка Set@l // Известия ЮФУ. Технические науки. – 2018. – № 3. – С. 46-58.
23. Safonov V.O. Using Aspect-Oriented Programming for Trustworthy Software Development. – New York: John Wiley & Sons, 2008. – 352 p.
24. Вopenка П. Альтернативная теория множеств: Новый взгляд на бесконечность: пер. со славац. – Новосибирск: Изд-во Института математики, 2004. – 612 с.
25. Левин И.И., Дордопуло А.И., Писаренко И.В., Мельников А.К. Описание алгоритма решения систем линейных алгебраических уравнений методом Якоби на языке архитектурно-независимого программирования Set@l // Известия ЮФУ. Технические науки. – 2018. – № 5. – С. 34-48.
26. Левин И.И., Дордопуло А.И., Писаренко И.В., Мельников А.К. Архитектурно-независимая программа быстрого преобразования Фурье на языке программирования SET@L // Вестник РГРТУ. – 2019. – № 68. – С. 28-36. – ISSN 1995-4565. – DOI: 10.21667/1995-4565-2019-68-2-28-36. – <http://vestnik.rsreu.ru/ru/archive/2019/vypusk-68/805-1995-4565-2019-68-2-28-36>.

REFERENCES

1. Mel'nikov A.K. Slozhnost' rascheta tochnykh raspredeleniy veroyatnosti simmetrichnykh additivno razdelyaemykh statistik i oblast' primeneniya predel'nykh raspredeleniy [The complexity of calculating the exact probability distributions of symmetric additively separable statistics and the scope of marginal distributions], *Doklady TUSUR* [TUSUR reports]. Tomsk, 2017, Vol. 20, No. 4, pp. 126-130. ISSN 1818-0442.

2. Mel'nikov A.K., Levin I.I., Dordopulo A.I., Pisarenko I.V. Analiz vozmozhnostey sovremennykh vychislitel'nykh tekhnologiy dlya rascheta tochnykh priblizheniy raspredeleniy veroyatnostey znacheniy statistik [Analysis of the possibilities of modern computing technologies for calculating accurate approximations of probability distributions of statistical values], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 7 (224), pp. 6-19. ISSN 1999-9429. DOI: 10.18522/2311-3103-2021-7-6-19.
3. Timonin V.I., Chernomordik O.M. Metod vychisleniya tochnogo raspredeleniya statistik tipa Kolmogorova-Smirnova pri al'ternativakh Lemana [A method for calculating the exact distribution of Kolmogorov-Smirnov type statistics with Lehman alternatives], *Teoriya veroyatnostey i ee primeneniye* [Probability theory and its application], 1985, Vol. 30, Issue 3, pp. 572-573.
4. Tyannikova N.D., Timonin V.I. Metod vychisleniya tochnykh raspredeleniy statistik tipa Kolmogorova-Smirnova v sluchae narusheniya odnorodnosti i nezavisimosti analiziruemykh vyborok [Method for calculating exact distributions of Kolmogorov-Smirnov type statistics in case of violation of uniformity and independence of the analyzed samples], *Nauka i obrazovanie* [Science and Education], 2014, No. 11, pp. 227-237. DOI: 10.7463/1114.0740251 ISSN 1994-0448.
5. Raschet raspredeleniy metodom Monte-Karlo [Calculation of distributions by the Monte Carlo method].
6. Feynman R.P. Simulating physics with computers, *International Journal of Theoretical Physics*, 1982, Vol. 21, Iss. 6, pp. 467-488. DOI: 10.1007/BF02650179.
7. Benioff P. Quantum mechanical hamiltonian models of turing machines, *Journal of Statistical Physics*, 1982, Vol. 29, No. 3, pp. 515-546. DOI: 10.1007/BF01342185. Bibcode: 1982JSP....29..515B.
8. Valiev K.A., Kokin A.A. Kvantovye komp'yutery: nadezhdy i real'nost' [Quantum computers: Hopes and Reality]. Izhevsk: RKHD, 2004, 320 p.
9. Manin Yu.I. Vychislimoe i nevychislimoe [Computable and non-computable]. Moscow: Sov. radio, 1980, 128 p. (Cybernetics).
10. Moren K. Metody gil'bertova prostranstva [Methods of Hilbert space]. Moscow: Mir, 1965, 570 p.
11. Fiziki iz Rossii i SSHA sozdali pervyy 51-kubitnyy kvantovyy komp'yuter [Physicists from Russia and the USA have created the first 51-qubit quantum computer]. Available at: <https://ria.ru/20170714/1498476410.html?> (accessed 2 May 2022).
12. Kak Rossiya potratit 15 mlrd. na sozдание kvantovogo komp'yutera [How will Russia spend 15 billion? to create a quantum computer]. Available at: https://gov.news.ru/articles/2020-01-9_kak_rossiya_potratit_15_mlrd_na_sozdanie? (accessed 2 May 2022).
13. IonQ zayavila o sozdanii samogo moshchnogo v mire kvantovogo komp'yutera [IonQ has announced the creation of the world's most powerful quantum computer]. Available at: <https://3dnews.ru/1060908/kompaniya-ion-q-soobshchila-o-sozdanii-samogo-moshchnogo-v-mire-kvantovogo-kompyutera?> (accessed 2 May 2022).
14. IBM prodemonstirovali kvantovyy komp'yuter s «kvantovym ob'emom» 64. Available at: <https://habr.com/ru/company/raiffeisenbank/news/t/516062/?> (accessed 2 May 2022).
15. Stepanenko S.A. Fotonnaya vychislitel'naya mashina. Printsipy realizatsii. Otsenka parametrov [Photonic computing machine. Principles of implementation. Estimation of parameters], *Vychislitel'naya tekhnika. Doklady Akademii nauk* [Computer engineering. Reports of the Academy of Sciences]. Moscow, 2017, Vol. 476, No. 4, pp. 389-394. DOI: 10.1134/S1064562417050234.
16. Stepanenko S.A. Fotonnyy komp'yuter: struktura i algoritmy, otsenka parametrov [Photonic computer: structure and algorithms, estimation of parameters], *Fotonika* [Photonics], 2017, No. 7 / 67, pp. 72-83. DOI: 10.22184/1993-7296.2017.67.7.72.83.
17. Stepanenko S.A. Interferentsionnye logicheskie elementy [Interference logic elements], presented by Academician of the Russian Academy of Sciences Yu.A. Trutnev. *Doklady Akademii nauk. Matematika, informatika, protsessy upravleniya* [Reports of the Academy of Sciences. Mathematics, Computer Science, Management processes]. Moscow, 2020, Vol. 493, pp. 64-69. DOI: 10.31857/S2686954320040189.
18. Soifer V.A., Kharitonov S.I., Khonina S.N. etc. Spiral caustics of vortex beams, *Photonics*, 2021, Vol. 8, Issue 1, pp. 1-20. DOI: 10.3390/photonics8010024.
19. Soifer V.A., Doskolovich L.L., Bezus E.A. etc. Spatial differentiation of optical beams using a resonant metal-dielectric-metal structure, *Journal of Optics*, 2021, Vol. 23, Issue 2. DOI: 10.1088/2040-8986/abe63b.

20. *Soifer V.A., Daskolovich L.L., Bezus E.A. etc.* An Optical Differentiator Based on a Three-Layer Structure with a W-Shaped Refractive Index Profile, *Journal of Experimental and Theoretical Physics*, 2018, Vol. 127, Issue 2, pp. 202-209. DOI: 10.1134/S1063776118080174.
21. *Levin I.I., Mel'nikov A.K.* Upravlenie gibridnymi vychislitel'nymi sistemami na yazyke SET@L [Management of hybrid computing systems in the SET@L language], *Materiály XI mezinárodní vědecko-praktická konference «Aktuální vymoženosti vědy – 2015» Díl 7. Moderní informační technologie* [Materials XI international scientific and practical conference "current achievements of science-2015" Part 7. Modern information technology"]. Praha. Publishing House «Education and Science», 2015. – 96 s. – P. 23-28. – ISSN 978-966-8736-05-6.
22. *Levin I.I., Dordopulo A.I., Pisarenko I.V., Mel'nikov A.K.* Podkhod k arkhitekturno-nezavisimomu programirovaniyu vychislitel'nykh sistem na osnove aspektno-orientirovannogo yazyka Set@l [Approach to architecture-independent programming of computing systems based on aspect-oriented language Set@l], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2018, No. 3, pp. 46-58.
23. *Safonov V.O.* Using Aspect-Oriented Programming for Trustworthy Software Development. New York: John Wiley & Sons, 2008, 352 p.
24. *Vopenka P.* Alternativnaya teoriya mnozhestv: Novyy vzglyad na beskonechnost' [Alternative Set Theory: A New View of Infinity: Transl. from Slavack. Novosibirsk: Izd-vo Instituta matematiki, 2004, 612 p.
25. *Levin I.I., Dordopulo A.I., Pisarenko I.V., Mel'nikov A.K.* Opisanie algoritma resheniya sistem lineynykh algebraicheskikh uravneniy metodom Yakobi na yazyke arkhitekturno-nezavisimogo programirovaniya Set@l [Description of the algorithm for solving systems of linear algebraic equations by the Jacobi method in the language of architecture-independent programming Set@l], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2018, No. 5, pp. 34-48.
26. *Levin I.I., Dordopulo A.I., Pisarenko I.V., Mel'nikov A.K.* Arkhitekturno-nezavisimaya programma bystrogo preobrazovaniya Fur'e na yazyke programirovaniya SET@L [An architecture-independent fast Fourier transform program in the SET@L programming language], *Vestnik RGRU* [Vestnik of RSREU], 2019, No. 68, pp. 28-36. ISSN 1995-4565. DOI: 10.21667/1995-4565-2019-68-2-28-36. Available at: <http://vestnik.rsreu.ru/ru/archive/2019/vypusk-68/805-1995-4565-2019-68-2-28-36>.

Статью рекомендовал к опубликованию д.т.н. Э.В. Мельник.

Мельников Андрей Кимович – АО «Вычислительные решения»; e-mail: anta-mak@umail.ru; ak@comp-sol.ru; г. Москва, Россия; г.н.с.; к.т.н.; доцент ВАК.

Левин Илья Израилевич – Южный федеральный университет; e-mail: iilevin@sfedu.ru; г. Таганрог, Россия; зав. кафедрой ИМС; д.т.н.; профессор.

Дордопуло Алексей Игоревич – НИЦ супер-ЭВМ и нейрокомпьютеров; e-mail: dordopulo@superevm.ru; г. Таганрог, Россия; начальник отдела; к.т.н.

Сластен Любовь Михайловна – e-mail: lmslasten@yandex.ru; научный сотрудник; к.т.н.

Melnikov Andrey Kimovich – JSC "Computing Solutions"; e-mail: anta-mak@umail.ru; ak@comp-sol.ru; Moscow, Russia; chief researcher; cand. of eng. sc.; associate professor of the Higher Attestation Commission.

Levin Ilya Izrailevich – Southern Federal University; e-mail: iilevin@sfedu.ru; Taganrog, Russia; head of department of intelligent and multiprocessor systems; dr. of eng. sc.; professor.

Dordopulo Alexey Igorevich – Supercomputers and Neurocomputers Research Center; e-mail: dordopulo@superevm.ru; Taganrog, Russia; head of department; cand. of eng. sc.

Slasten Lyubov' Mikhaylovna – e-mail: lmslasten@yandex.ru; cand. of eng. sc.

К.Е. Румянцев, П.Д. Миронова, Шакир Хайдер Хуссейн

**ОЦЕНКА ВЛИЯНИЯ ФУНКЦИОНАЛЬНЫХ ЭЛЕМЕНТОВ
НА ПАРАМЕТРЫ СИСТЕМЫ КВАНТОВОГО РАСПРЕДЕЛЕНИЯ КЛЮЧА
НА ОСНОВЕ ПРОТОКОЛА B92**

Исследовано влияние параметров функциональных элементов на энергетические, временные и вероятностные характеристики системы квантового распределения ключа (КРК) на основе протокола B92. Построены зависимости вероятности записи правильного и ошибочного битов в сырую квантовую ключевую последовательность от длины волоконно-оптической линии связи (ВОЛС) и использования 4-х типов лазеров (EML, DFB, VCSEL, FP) и фотоприёмных модулей (id201; id210; id220; id230). Установлено, что, изменения вероятности записи правильного бита в сырую квантовую ключевую последовательность значительно более весомы, чем изменения вероятности записи ошибочного бита (50,9 раза против 3,3 раза при ширине спектра лазера 80 нм и изменении протяжённости ВОЛС с 10 до 100 км). Это связано с тем, что с ростом протяжённости ВОЛС резко растёт вероятность отсутствия регистрации на приёмной станции фотонов или импульсов темнового тока (ИТТ). Числовой материал указывает на прямую пропорциональную зависимость вероятности записи ошибочного бита от частоты генерации шумовых импульсов однофотонных лавинных фотодиодов (ОЛФД). Так, при увеличении частоты появления ИТТ в 60 раз (с 100 до 6000 Гц) вероятность записи ошибочного бита также увеличивается в 60 раз (например, при длине ВОЛС 100 км – 6,39 против 383,3). Установлено, что среднеквадратичное отклонение (СКО) времени задержки фотона прямо пропорционально длине ВОЛС и ширине спектра лазера. При ширине спектра лазера 10 нм и увеличении длины ВОЛС с 10 до 100 км (в 10 раз) среднеквадратичное отклонение времени задержки фотона также увеличивается в 10 раз (с 4,16 до 41,6 пс). Для достижения наилучших характеристик системы КРК в целом целесообразно использование лазера с минимальной шириной спектра излучения, например, EML-лазера. Однако EML-лазеры считаются самыми сложными и дорогостоящими из рассмотренных типов лазеров, поэтому использование EML-лазеров значительно повышает стоимость всей системы КРК.

Квантовое распределение ключей; протокол B92; энергетические характеристики; временные характеристики; вероятностные характеристики.

K.E. Rumyantsev, P.D. Mironova, Shakir Hayder Hussein

**EVALUATION OF THE FUNCTIONAL ELEMENTS INFLUENCE ON THE
PARAMETERS OF THE QKD SYSTEM BASED ON B92 PROTOCOL**

The influence of the parameters of functional elements on the energy, time and probabilistic characteristics of the quantum key distribution system (QKD) based on the B92 protocol is studied. The dependences of the probability of writing correct and erroneous bits into a raw quantum key sequence are plotted for various lengths of a fiber-optic communication line (FOCL) and the use of various lasers (EML-laser; DFB-laser; VCSEL-laser and FP-laser) and photodetector modules (id201; id210; id220; id230). Thus, changes in the probability of writing a correct bit into a raw quantum key sequence are much more significant than changes in the probability of writing an erroneous bit (50.9 times versus 3.3 times with FWHM=80 nm and a change in the length of the FOCL from 10 to 100 km). This is due to the fact that with an increase in the length of the FOCL, the probability of the absence of registration at the receiving station of photons or dark current pulses (DCP) sharply increases. Numerical material indicates a direct proportional dependence of the probability of writing an erroneous bit on the frequency of generation of noise pulses of single-photon avalanche photodiodes (SAPD). So, with an increase in the frequency of occurrence of DCP by 60 times (from 100 to 6000 Hz), the probability of recording an erroneous bit also increases by 60 times (for example, with a FOCL length of 100 km – 6.39 versus 383.3). It has been established that the root-mean-square deviation of the photon delay time is directly

proportional to the length of the FOCL and the width of the laser spectrum. With a spectrum width of FWHM=10 pm and an increase in the FOCL length from 10 to 100 km (by a factor of 10), the standard deviation of the photon delay time also increases by a factor of 10 (from 4.16 to 41.6 ps). To achieve the best performance of the QKD system as a whole, it is advisable to use a laser with a minimum width of the radiation spectrum, for example, an EML-laser. However, EML-lasers are considered the most complex and expensive of all the considered types of lasers, so the use of EML-lasers significantly increases the cost of the entire QKD system.

Quantum key distribution; B92 protocol; energy characteristics; time characteristics; probabilistic characteristics.

Введение. Обеспечение безопасности данных при передаче между легитимными пользователями является актуальной задачей. Существует множество классических систем и способов шифрования и передачи информации, которые с развитием вычислительной техники не обеспечивают секретность передаваемых данных.

В отличие от классических систем, применение систем квантового распределения ключа (КРК) обеспечивает абсолютную безопасность данных и контроль несанкционированного доступа (НСД) злоумышленниками [1–14], что достигается благодаря принципу неопределённости Гейзенберга: попытка измерения в квантовой системе искажает её состояние, а полученная в результате такого измерения информация не полностью соответствует состоянию системы до начала измерений.

Анализируемая система КРК состоит из приёмопередающей и кодирующей станций, связанных посредством волоконно-оптической линии связи (ВОЛС). Используя методику расчёта параметров как системы КРК в целом, так и отдельных функциональных модулей [15], разработано программное обеспечение для анализа влияния отдельных характеристик функциональных элементов на энергетические, временные и вероятностные параметры системы.

Оценка влияния протяжённости направляющей линии коммуникации на параметры системы КРК. Проводим моделирование работы системы КРК по протоколу B92 для оценки вероятностных, временных и энергетических характеристик системы КРК при протяжённости ВОЛС в диапазоне 10...100 км для ряда значений ширины спектра лазера FWHM. Диапазон изменений значений параметра FWHM 10...100 пм характерен для электроабсорбционных модулированных лазеров (EML – Electroabsorptive Modulated Laser). Особенностью EML-лазеров является присутствие в спектре излучения одной ярко выраженной моды.

Центральная длина волны лазерного излучения равна $\lambda_s=1550$ нм, а среднее число регистрируемых фотонов за длительность квантового импульса на выходе однофотонного источника $\overline{N_{FA}}=0,2$. Длительность импульса лазера принята равной $\tau_s=300$ пс. При моделировании исключается нестабильность генерации лазерных импульсов $\Delta T_s=0$.

Погонное затухание одномодового оптического волокна SMF-28 ULL, используемого в ВОЛС, составляет $\alpha_{SMF}=0,18$ дБ/км, а эффективный показатель преломления в сердечнике $n_{SMF} = 1,4682$.

Погонное затухание одномодового оптического волокна Panda, сохраняющего состояние поляризации и применяемого внутри передающей и приёмной станций, составляет $\alpha_{PMF}=1$ дБ/км, а эффективный показатель преломления в сердцевине $n_{PMF}=1,468$.

При моделировании исключается нестабильность генерации импульсов стробирования $\Delta T_{strob}=0$.

Частота генерации импульсов темного тока в однофотонном лавинном фотодиоде (ОЛФД) $f_b=100$ Гц.

На рис. 1 представлена зависимость вероятности записи правильного бита в сырую квантовую ключевую последовательность от длины ВОЛС. По оси ОУ используется логарифмический масштаб.

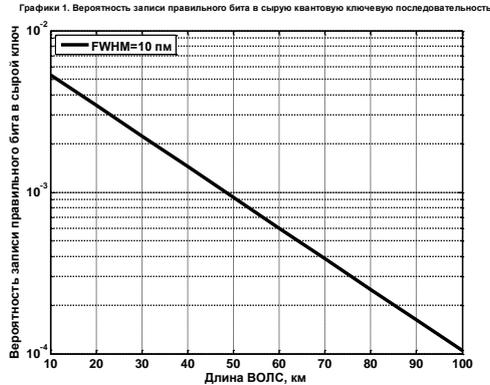


Рис. 1. Вероятность записи правильного бита в сырую квантовую ключевую последовательность при различных значениях длины ВОЛС

Необходимо отметить значительное снижение вероятности записи правильного бита при увеличении длины ВОЛС. При увеличении длины ВОЛС в 5 раз с 10 до 50 км вероятность записи правильного бита снижается в 5,7 раза, а с 10 до 100 км (в 10 раз) – уже в 48,5 раз. Причём, при каждом увеличении длины ВОЛС на 10 км вероятность записи правильного бита падает в 1,55 раза (на ~2 дБ).

Такое поведение вероятности следует из роста потерь оптического излучения непосредственно в одномодовом волокне на 1,8 дБ при изменении протяжённости ВОЛС на 10 км. Незначительный дополнительный рост потерь на 0,01 дБ при каждом приросте протяжённости ВОЛС на 10 км связан с потерями на сварных соединениях.

Если при длине ВОЛС в 10 км вероятность записи правильного бита в сырую квантовую ключевую последовательность составляет порядка 0,005, то при длине ВОЛС в 100 км не превышает 0,0001. Следовательно, только 1 бит из 10 000 сгенерированных будет правильно записан в сырую ключевую последовательность.

Как и следовало ожидать, ширина спектра лазера FWHM не влияет на вероятность записи правильного бита в сырую квантовую ключевую последовательность при оптимальном выборе длительности импульса стробирования в соответствии с формулой $\tau_{strob} = \tau_s + 2 \cdot \Delta\tau_{TF} + 2 \cdot \Delta T_{strob}$.

На рис. 2 представлена зависимость вероятности записи ошибочного бита в сырую квантовую ключевую последовательность от длины ВОЛС. По оси ОУ здесь также использован логарифмический масштаб.

При изменении протяжённости ВОЛС с 10 до 100 км при FWHM=80 пм вероятность записи ошибочного бита в сырую квантовую ключевую последовательность увеличивается в 4,6 раза (на 6,6 дБ) с $2,5 \cdot 10^{-8}$ до $11,7 \cdot 10^{-8}$. При увеличении длины ВОЛС с 10 до 20 км вероятность записи ошибочного бита увеличивается в 1,4 раза, при увеличении с 20 до 30 км – в 1,3 раза, а при увеличении с 90 до 100 км – всего в 1,10 раза. Напомним, что вероятность записи правильного бита изменяется в 1,55 раза при любом изменении длины ВОЛС на 10 км.

Сравнение с данными на рис. 1 показывает, что изменения вероятности записи правильного бита в сырую квантовую ключевую последовательность значительно более весомы, чем изменения вероятности записи ошибочного бита (50,9 раза против 3,3 раза при FWHM=80 пм и изменении протяжённости ВОЛС с

10 до 100 км). Это связано с тем, что с ростом протяжённости ВОЛС резко растёт вероятность отсутствия регистрации на приёмной станции фотонов или импульсов темнового тока (ИТТ).

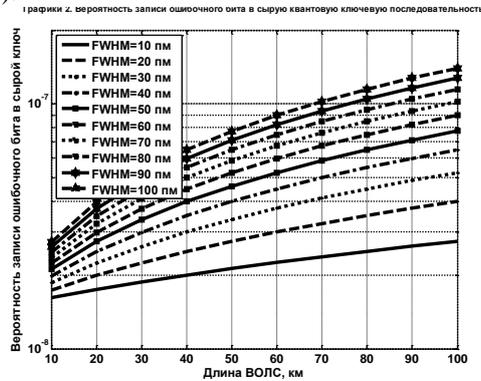


Рис. 2. Вероятность записи ошибочного бита в сырую квантовую ключевую последовательность при различных значениях длины ВОЛС

В процессе моделирования установлено, что влиянием поляризационной модовой дисперсии в одномодовом оптическом волокне можно пренебречь. Действительно, даже при минимальной ширине спектра лазера $\text{FWHM}=10$ нм и минимальной протяжённости ВОЛС в 10 км среднее квадратичное отклонение (СКО) времени задержки фотона после распространения через ВОЛС из-за хроматической дисперсии равно 4,16 пс, в то время как СКО из-за поляризационной модовой дисперсии – 0,13 пс. Различие превышает 32 раза (15 дБ). Причём различие растёт с расширением спектра лазера и превышает 320 раз при $\text{FWHM}=100$ нм (на верхней границе для электроабсорбционных модулированных лазеров).

На рис. 3 представлены результирующие зависимости среднее квадратичное отклонение задержки фотона после распространения через ВОЛС, которые совпадают с зависимостями СКО задержки фотона после распространения через ВОЛС из-за хроматической дисперсии.

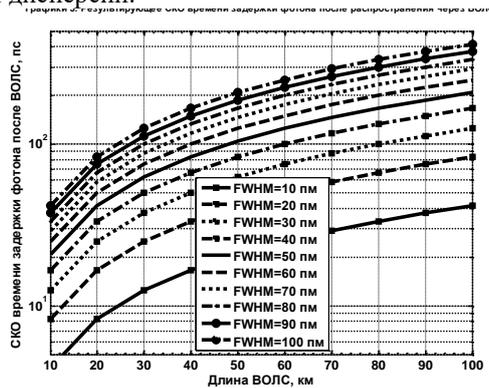


Рис. 3. Среднее квадратичное отклонение времени распространения фотона через ВОЛС

Среднее квадратичное отклонение времени задержки фотона прямо пропорционально длине ВОЛС и ширине спектра лазера. При увеличении длины ВОЛС с 10 до 100 км (в 10 раз) среднее квадратичное отклонение времени задержки фотона

также увеличивается в 10 раз (с 4,16 до 41,6 пс). Причём при каждом увеличении длины ВОЛС на 10 км при FWHM=10 пм СКО времени задержки фотона увеличивается на 4,16 пс.

Из графиков видно, что уже при протяжённости ВОЛС в 100 км среднеквадратичное отклонение времени задержки фотона превышает длительность импульса лазера $\tau_s=300$ пс при значениях FWHM более 75 пм.

Графики на рис. 4 отражают зависимость длительности импульса стробирования от длины ВОЛС. Учитывая то, что нестабильность генерации импульсов стробирования исключена при моделировании, длительность импульса стробирования зависит только от временного разброса времени задержки фотона после распространения через ВОЛС и длительности импульса лазера, которая фиксирована.

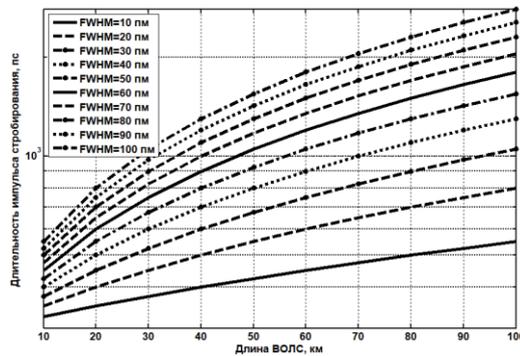


Рис. 4. Зависимости длительности импульса стробирования от длины линии коммуникации

В целом при увеличении длины ВОЛС с 10 до 100 км длительность импульса стробирования увеличивается в 1,69 раза при ширине спектра FWHM=10 пм, в 3,64 раза при FWHM=50 пм и уже в 5,09 раза при FWHM=100 пм. Различие при каждом увеличении длины ВОЛС на 10 км между предыдущим и текущим значением длительности импульса стробирования сокращается. Например, при FWHM=100 пм при увеличении длины ВОЛС с 10 до 20 км значение длительности импульса стробирования увеличивается в 1,45 раза, с 20 до 30 км – в 1,31 раза, а с 90 до 100 км – всего в 1,1 раза.

При каждом увеличении длины ВОЛС на 10 км длительность импульса стробирования возрастает. Причём если при ширине спектра FWHM=100 пм увеличение длины ВОЛС с 10 до 20 км требует увеличения длительности импульса стробирования на 45 %, то с 90 до 100 км – уже только на 10%.

Важно, что требуемая длительность импульса стробирования во много раз может превышать длительность импульса лазера. Действительно, при наименьшей ширине спектра FWHM=10 пм для группы электроабсорбционных модулированных ЕМЛ-лазеров изменение длины ВОЛС с 10 до 100 км потребует выбор длительностей импульса стробирования на 8...83 % больше длительности импульса лазера. Однако при худших параметрах лазера длительность импульса стробирования будет уже в десять раз превышать длительность импульса лазера. Естественно, что последнее негативно сказывается на параметрах системы КРК, увеличивая вероятность записи ошибочного бита.

На рис. 5 представлена зависимость максимальной требуемой частоты следования оптических импульсов от длины ВОЛС.

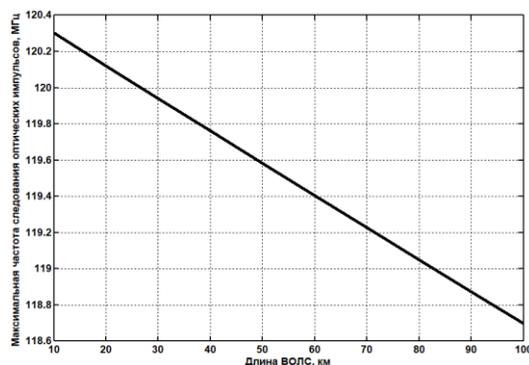


Рис. 5. Максимальная требуемая частота следования оптических импульсов при различных значениях длины ВОЛС

При каждом увеличении длины ВОЛС на 10 км максимальная частота следования оптических импульсов уменьшается на 170...180 кГц. Например, при длине ВОЛС 10 км максимальная частота следования оптических импульсов составляет 120,3 МГц, при длине ВОЛС 20 км – 120,12 МГц, при длине ВОЛС 100 км – 118,7 МГц. В целом при увеличении длины ВОЛС в 10 раз (с 10 до 100 км) максимальная требуемая частота следования оптических импульсов уменьшается на 1,6 МГц (120,3 МГц против 118,7 МГц).

Оценка влияния типов лазеров на параметры системы КРК. В системах КРК применяются различные типы лазеров. Выбор последнего зависит от необходимых характеристик реализуемой системы, например ширина спектра излучения, скорость передачи данных, стоимость и другие. Приведённые выше результаты характерны для электроабсорбционных модулированных лазеров (EML – Electroabsorptive Modulated Laser), отличительной особенностью которых является присутствие в спектре излучения одной ярко выраженной моды.

Специфика применения различных типов лазеров в системах КРК рассмотрена в [16]. Показано, что кроме EML-лазеров в системах КРК используют ещё три типа лазеров:

- ◆ простейшие излучатели Фабри-Перо, используемые в одномодовых и многомодовых передатчиках;
- ◆ поверхностно-излучающие лазеры с вертикальным резонатором (VCSEL-лазеры), получившие распространение в многомодовых системах на коротких расстояниях и отличающиеся высокой температурной стабильностью, малой потребляемой мощностью и минимальной стоимостью производства передатчиков на их основе;
- ◆ лазеры с распределённой обратной связью (DFB-лазеры), обеспечивающие высокую мощность излучения при сохранении узкой ширины спектра излучения. Также лазеры DFB обладают высокой температурной стабильностью и возможностью подстройки длины волны.

Характеристики указанных типов лазеров сведены в табл. 1 [17–19].

Таблица 1

Характеристики различных типов лазеров

Параметр лазера	FP	VCSEL	DFB	EML
Длина волны излучения, нм	1310/1550 850/1300	850	850 1550	1550
Скорость передачи данных, не более, Гбит/с	1	25	1 до 150 км 10 до 40 км	100
Ширина спектра излучения, нм	3...5	0,5...1	0,1...0,5	0,01...0,08
Типовая длительность оптического импульса, пс	300		300	
Стоимость, USD	25...500	5...100	100...3000	

Проведён вычислительный эксперимент для сравнительного анализа вероятностных, временных и энергетических характеристик рассмотренных типов лазеров при изменении протяжённости направляющей ВОЛС в диапазоне 10...100 км для крайних (минимального и максимального) значений ширины спектра излучения лазеров FWHM (для FP-лазера – 3 и 5 нм, для VCSEL-лазера – 0,5 и 1 нм, для DFB-лазера – 0,1 и 0,5 нм, для EML-лазера – 0,01 и 0,08 нм).

Использование лазеров с большой шириной спектра излучения значительно сказывается на разбросе времени распространения фотонов в ВОЛС. Так, при ширине спектра излучения лазера 5 нм при использовании FP-лазера СКО времени распространения фотона достигает 20,8 нс при протяжённости ВОЛС 100 км, что в 62,5 раза больше СКО времени распространения при использовании EML-лазера с шириной спектра излучения 80 нм. Различие результирующего СКО времени распространения фотона в ВОЛС при максимальном и минимальном значениях ширины спектра излучения прямо пропорционально изменению ширины спектра излучения и составляет 8 (332,80 пс против 41,60 пс); 5 (2079,98 пс против 416,00 пс); 2 (4159,96 пс против 2079,98 пс) и 1,67 раза (20799,80 пс против 12479,88 пс) при использовании EML-лазера, DFB-лазера, VCSEL-лазера и FP-лазера соответственно.

Расчёты показывают, что результирующее среднеквадратичное отклонение времени распространения фотона в ВОЛС определяется только хроматической дисперсией. Влиянием поляризационной модовой дисперсии в одномодовом оптическом волокне можно пренебрегать при протяжённости ВОЛС менее 100 км. Это справедливо и для EML-лазеров, где могут генерироваться импульсы существенно меньшей длительности.

На рис. 6 представлена зависимость требуемой длительности импульса стробирования однофотонных фотодетекторов от длины ВОЛС. Поскольку длительность импульсов применяемых в системах КРК лазеров не превышает 300 пс, то из графиков видно значительное превышение их значений длительностями импульсов стробирования. Так в диапазоне возможных значений ширины спектра излучения для FP-лазера (см. табл. 1) длительность импульса стробирования составляет 75,2...125,1 нс при протяжённости ВОЛС в 100 км. Длительности импульсов стробирования для лазеров VCSEL, DFB и EML лежат в диапазонах 12,8...25,3 нс, 2,8...12,8 нс и 0,55...2,3 нс соответственно.

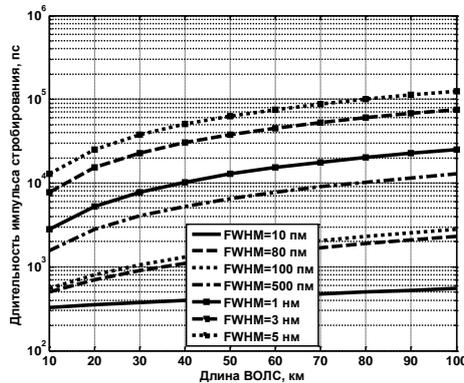


Рис. 6. Зависимость длительности импульса стробирования от длины ВОЛС

С учётом данных табл. 1 различия в длительностях импульсов лазера и стробирования превышают 250,6...417 раз для FP-лазера, 42,6...84,2 для VCSEL-лазера, 9,32...42,6 для DFB-лазера, и 1,83...7,66 раз для EML-лазера при протяжённости ВОЛС 100 км.

Многokратное превышение длительности импульса лазера длительностью импульса стробирования увеличивает вероятность записи ошибочного бита.

На рис. 7 представлена зависимость вероятности записи ошибочного бита в сырую квантовую ключевую последовательность от длины ВОЛС при различных значениях ширины спектра излучения, соответствующих различным типам лазеров. По оси OY использован логарифмический масштаб.

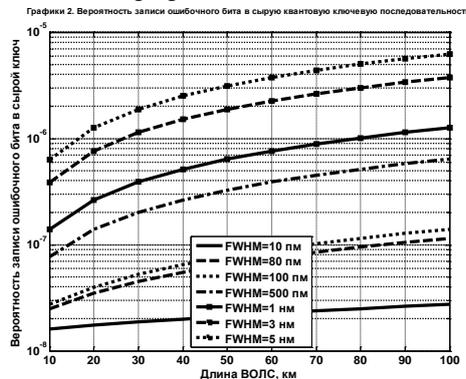


Рис. 7. Вероятность записи ошибочного бита в сырую квантовую ключевую последовательность при различных значениях длины ВОЛС

При протяжённости ВОЛС 100 км максимальное значение вероятности записи ошибочного бита в сырой ключ при использовании FP-лазера в 54 раза больше значения вероятности при использовании EML-лазера ($62,54 \cdot 10^{-7}$ при ширине спектра излучения 5 нм против $1,15 \cdot 10^{-7}$ при ширине спектра излучения 80 пм). При использовании VCSEL-лазера и DFB-лазера различие максимальных вероятностей значительно сокращается и составляет 11 и 5,6 раза соответственно, по сравнению со значением вероятности записи ошибочного бита при использовании EML-лазера ($1,15 \cdot 10^{-7}$ при ширине спектра излучения 80 пм). Следовательно, для достижения минимальной вероятности записи ошибочного бита целесообразно использование EML-лазера с шириной спектра излучения 10...80 пм.

Анализируя рассмотренные зависимости, отметим целесообразность использования лазера с минимальной шириной спектра излучения, например, EML-лазера, для достижения наилучших характеристик системы КРК в целом. Однако EML-лазеры считаются самыми сложными и дорогостоящими из всех рассмотренных типов лазеров [20], поэтому использование EML-лазеров значительно повышает стоимость всей системы КРК.

Влияние параметров фотоприёмного модуля на эффективность квантового распределения ключа. На эффективность квантового распределения ключа значительное влияние оказывают следующие параметры однофотонного приёмного модуля: частота поступления шумовых импульсов ОЛФД, время для восстановления работоспособности ОЛФД после регистрации фотона.

Для сравнительного анализа влияния параметров фотоприёмного модуля на вероятностные, временные и энергетические характеристики системы проведён вычислительный эксперимент. В качестве источника излучения выбран DFB-лазер с шириной спектра FWHM в 500 пм (наихудший случай с позиций FWHM) и длительностью импульса 300 пс. Нестабильность генерации импульсов стробирования исключена. В ВОЛС использовано одномодовое оптическое волокно SMF-28 ULL, а внутри передающей и приёмной станций – одномодовое оптическое волокно Panda. Центральная длина волны излучения равна 1550 нм, а среднее число фотонов за длительность квантового импульса на выходе однофотонного источника 0,2.

Проведённый ранее анализ показывает, что влияние частоты поступления шумовых импульсов ОЛФД f_b сказывается лишь на вероятность записи ошибочного бита в ключевую последовательность. Причём в силу того, что за длительность строб-импульса среднее число ИТТ значительно уступает среднему числу принимаемых фотонов влиянием на вероятность записи правильного бита можно пренебречь. Действительно, как следует из табл. 2, изменение частоты генерации шумовых импульсов ОЛФД с 1 до 100 Гц приводит к изменению вероятности записи правильного бита всего на 0,002 % при протяжённости ВОЛС 10 км, на 0,03 % при 50 км и на 0,67 % при 100 км.

Таблица 2

Умноженная на 100 000 вероятность записи правильного бита в сырую ключевую последовательность

Длина ВОЛС, км	Частота генерации шумовых импульсов ОЛФД, Гц						
	1	2	5	10	20	50	100
10	534,09	534,09	534,09	534,09	534,09	534,10	534,10
20	345,17	345,17	345,17	345,17	345,17	345,17	345,18
50	93,02	93,02	93,02	93,02	93,03	93,04	93,05
100	10,44	10,44	10,44	10,45	10,45	10,47	10,57

Это указывает на возможность при проектировании систем КРК на малошумящих ОЛФД не учитывать вклад шумовых импульсов ОЛФД и рассчитывать вероятность записи правильного бита по приближённой формуле

$$P_{bit} = \frac{(P_{bit}\{0|0\} + P_{bit}\{1|1\})}{2} \approx [1 - \exp(-\bar{n}_0)] \cdot [1 - \exp(-\bar{n}_1)] \approx \bar{n}_0 \cdot \bar{n}_1.$$

Здесь \bar{n}_0 и \bar{n}_1 представляют средние числа регистрируемых событий за длительность строб-импульса в приёмной станции при условии передачи соответственно логического нуля «0» и логической единицы «1».

На рис. 8 представлены графики зависимостей вероятности записи ошибочного бита в сырую квантовую ключевую последовательность от длины ВОЛС при семи различных частотах генерации шумовых импульсов ОЛФД. По оси ОУ использован логарифмический масштаб.

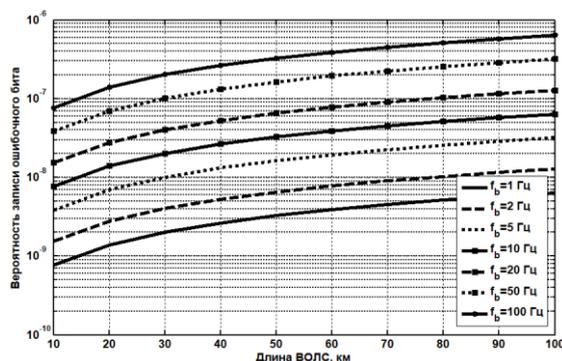


Рис. 8. Вероятность записи ошибочного бита в сырую квантовую ключевую последовательность

Числовой материал указывает на прямую пропорциональную зависимость вероятности записи ошибочного бита от частоты генерации шумовых импульсов ОЛФД.

Характеристики применяемых ОЛФД отличны от характеристик идеального однофотонного фотодетектора. Во-первых, ОЛФД регистрирует только первый фотон за время анализа. Во-вторых, в случае приёма фотона ОЛФД потребует время для восстановления рабочего состояния. Основные параметры ОЛФД, применяемых в системах КРК с рабочей длиной волны оптического излучения в диапазоне 900...1700 нм, сведены в табл. 3 [15].

Таблица 3

Параметры фотоприёмных модулей на основе ОЛФД

Наименование изделия	Частота появления ИТТ, не более, Гц	Типовое (максимальное) время восстановления рабочего состояния, нс
id201	100	0,1
id210	40	100 000
id220	6 000	25 000
id230	50	100 000

На рис. 9 представлены графики зависимостей вероятности записи ошибочного бита в сырую квантовую ключевую последовательность от длины ВОЛС при использовании конкретных фотоприёмных модулей.

Необходимо отметить, при увеличении частоты появления ИТТ в 60 раз (с 100 до 6000 Гц) вероятность записи ошибочного бита также увеличивается в 60 раз (например, при длине ВОЛС 100 км – 6,39 против 383,3), что указывает на прямую пропорциональную зависимость вероятности записи ошибочного бита от частоты появления ИТТ. Оптимальным выбором с точки зрения соотношения частоты появления ИТТ и типового времени восстановления рабочего состояния фотоприёмного

модуля является модуль id201. Вероятность записи ошибочного бита при использовании модуля id201 в 2...2,5 раза больше при длине ВОЛС 100 км ($3,20 \cdot 10^{-7}$ и $2,56 \cdot 10^{-7}$ против $6,40 \cdot 10^{-7}$), по сравнению с модулями id230 и id210 соответственно.

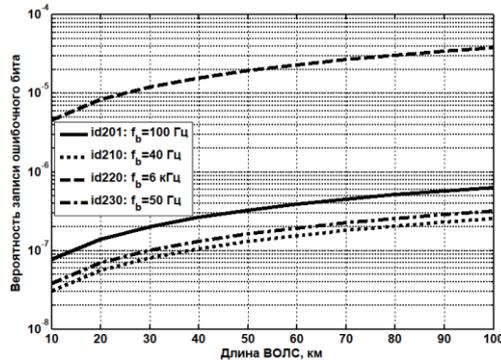


Рис. 9. Вероятность записи ошибочного бита в сырую квантовую ключевую последовательность

Отметим, что время восстановления рабочего состояния фотоприёмного модуля не сказывается на вероятностных характеристиках системы. С другой стороны, время восстановления работоспособности ОЛФД ограничивает максимально-допустимое значение частоты следования оптических импульсов, а следовательно, и предельную скорость формирования ключа. Это подтверждают графики на рис. 10, где представлены зависимости максимальной требуемой частоты следования оптических импульсов от длины ВОЛС при использовании 4-х модулей на основе ОЛФД, в которых время восстановления рабочего состояния различается в 1 000 000 раз (от 0,1 нс до 100 мкс). На рис. 10 графики для модулей id210 и id230 совпадают, что является следствием равенства $\tau_{SAPD} = 100$ мкс.

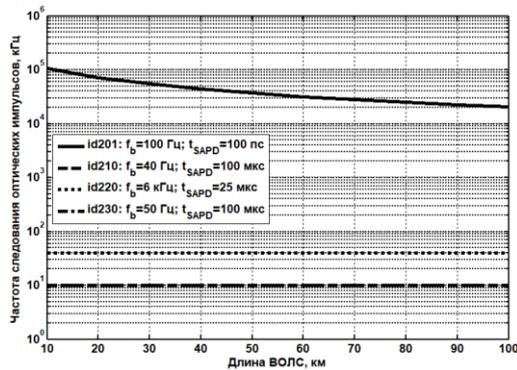


Рис. 10. Максимальная частота следования оптических импульсов

Следствием разного времени рабочего состояния ОЛФД является различие на 3...4 порядка частот следования оптических импульсов (предельных скоростей формирования ключа). При переходе с модуля id201 с $\tau_{SAPD} = 100$ пс на модуль id210 с $\tau_{SAPD} = 100$ мкс максимально-допустимая частота следования оптических импульсов уменьшается в 10 452 раз при протяжённости ВОЛС 10 км, в 7 176 раз – при протяжённости ВОЛС 20 км, и в 2 046 раз – при протяжённости ВОЛС 100 км.

В связи со значительным временем восстановления рабочего состояния модулей id210, id220 и id230, максимальная частота следования оптических импульсов постоянна при изменении протяжённости ВОЛС. Лишь для модуля id201, где время восстановления рабочего состояния 100 пс меньше исходных длительностей оптических импульсов (а тем более длительностей импульсов стробирования), при увеличении длины ВОЛС в 10 раз (с 10 до 100 км) максимальная частота следования оптических импульсов уменьшается в 5,1 раза.

Исследования показывают, что большое время восстановления рабочего состояния ОЛФД может значительно снизить скорость и увеличить время формирования ключа. За счёт этого злоумышленник получает большее время для атак на систему КРК. Следовательно, при выборе фотоприёмного модуля необходимо учитывать как частоту генерации ИТТ, так и время восстановления его рабочего состояния.

Выводы. Проанализировано влияние параметров функциональных элементов на энергетические, временные и вероятностные характеристики системы КРК на основе протокола B92.

Установлено, при увеличении длины ВОЛС значительно снижается вероятность записи правильного бита. Так, при увеличении длины ВОЛС в 10 раз (с 10 до 100 км) вероятность записи правильного бита снижается в 50,9 раза (при ширине спектра FWHM=80 пм). С другой стороны, изменение вероятности записи ошибочного бита менее значительно и составляет 3,3 раза при увеличении длины ВОЛС с 10 до 100 км. Это обусловлено резким ростом вероятности отсутствия регистрации на приёмной станции фотонов и/или ИТТ при увеличении протяжённости ВОЛС.

Отметим, среднеквадратичное отклонение времени задержки фотона прямо пропорционально длине ВОЛС и ширине спектра, причём при увеличении длины ВОЛС в 10 раз среднеквадратичное отклонение времени задержки фотона также увеличивается в 10 раз.

Анализируя рассмотренные зависимости, отметим целесообразность использования лазера с минимальной шириной спектра излучения, например, EML-лазера, для достижения наилучших характеристик системы КРК в целом. Однако EML-лазеры считаются самыми сложными и дорогостоящими из всех рассмотренных типов лазеров, поэтому использование EML-лазеров значительно повышает стоимость всей системы КРК.

Полученный числовой материал указывает, что оптимальным выбором с точки зрения соотношения частоты появления ИТТ и типового времени восстановления рабочего состояния фотоприёмного модуля является модуль id201.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Scarani V.* Quantum Physics: A First Encounter: Interference, Entanglement, and Reality. Translated by Rachael Thew. – Oxford: University Press, Mar 2006. – 125 p.
2. Физика квантовой информации: Квантовая криптография. Квантовая телепортация. Квантовые вычисления / под ред. Д. Боумейстера, А. Экерта, А. Цайлингера: пер. с англ. С.П. Кулика, Е.А. Шапиро. – М.: Постмаркет, 2002. – 376 с.
3. *Gisin N., Ribordy G., Tittel W., Zbinden H.* Quantum cryptography // *Reviews of Modern Physics.* – 2002. – Vol. 74, No. 1. – P. 145-195.
4. Ростелеком объявил о внедрении квантовой криптографии на своих сетях. – URL: <https://tass.ru/ekonomika/5685597> (дата обращения 17.10.2018).
5. *Румянцев К.Е.* Системы квантового распределения ключа: монография. – Таганрог: Изд-во ТТИ ЮФУ, 2011. – 264 с.
6. *Bennett C., Brassard G.* Quantum cryptography: Public key distribution and coin tossing // *Proceedings of IEEE international conference on computers, systems and signal processing.* Bangalore, India. – New York: Institute of Electrical and Electronics Engineers. – 1984. – P. 175-179.

7. *Shor P.W., Preskill J.* Simple proof of security of the BB84 quantum key distribution protocol // *Physical Review Letters*. – 2000. – Vol. 85. – P. 441-444.
8. *Mironov Y.K., Rumyantsev K.E.* Single-Photon Algorithm for Synchronizing the System of Quantum Key Distribution with Polling Sections of a Fiber-Optic Line // *Futuristic Trends in Networks and Computing Technologies*. – 2020. – P. 87-97. – DOI: https://doi.org/10.1007/978-981-15-4451-4_8.
9. *Румянцев К.Е.* Синхронизация в системе квантового распределения ключа с автоматической компенсацией поляризационных искажений // *Телекоммуникации*. – 2017. – № 2. – С. 32-40.
10. *Румянцев К.Е., Плёткин А.П.* Безопасность режима синхронизации системы квантового распределения ключей // *Известия ЮФУ. Технические науки*. – 2015. – № 5 (166). – С. 135-153.
11. *Курочкин В.Л. и др.* Экспериментальные исследования в области квантовой криптографии // *Фотоника*. – 2012. – Т. 5. – С. 54-66.
12. *Румянцев К.Е., Плёткин А.П.* Синхронизация системы квантового распределения ключа при использовании фотонных импульсов для повышения защищённости // *Известия ЮФУ. Технические науки*. – 2014. – № 8. – С. 81-96.
13. *Rumyantsev K.E., Pljonkin A.P.* Preliminary Stage Synchronization Algorithm of Autocompensation Quantum Key Distribution System with an Unauthorized Access Security // *International Conference on Electronics, Information, and Communications (ICEIC)*. 2016. Vietnam, Danang. – P. 1-4. – DOI: 10.1109/ELINFOCOM.2016.7562955. WOS:000389518100035. IDS: BG5KP.
14. *Rumyantsev K., Rudinsky E.* Parameters of the two-stage synchronization algorithm for the quantum key distribution system // *Proceedings of the 10th International Conference on Security of Information and Networks (SIN'17)*. – 2017. – P. 140-147. – DOI: 10.1145/3136825.3136888.
15. *Румянцев К.Е., Шакир Н.Н.* Проектирование системы квантового распределения ключа с интерферометрами Маха-Цендера: учеб. пособие. – Ростов-на-Дону; Таганрог: Изд-во ЮФУ, 2020. – 108 с.
16. *Румянцев К.Е.* Квантовые технологии в телекоммуникационных системах: учебник. – Ростов-на-Дону; Таганрог: Изд-во ЮФУ, 2021. – 346 с.
17. Лазерные диоды Фабри-Перо. – URL: http://www.electroncom.ru/product/nanoplus/fp_diodes.php (дата обращения: 10.01.2022).
18. Поверхностно-излучающие лазеры с вертикальным резонатором. – URL: <http://msd.com.ua/optoelektronika/poverxnostno-izluchayushhielazery-s-vertikalnym-rezonatorom-vcsel/> (дата обращения: 10.01.2022).
19. Лазеры с внешним резонатором. – URL: http://fmnauka.narod.ru/lazery_s_vneshnim-rezonatorom.pdf (дата обращения: 10.01.2022).
20. Основные параметры и сертификация оптических SFP модулей. – URL: <https://deps.ua/knowegable-base-ru/articles/1961-osnovnye-parametry-i-sertifikatsiia-opticheskikh-sfp-modulei.html> (дата обращения: 10.01.2022).

REFERENCES

1. *Scarani V.* Quantum Physics: A First Encounter: Interference, Entanglement, and Reality. Translated by Rachael Thew. Oxford: University Press, Mar 2006, 125 p.
2. *Fizika kvantovoy informatsii: Kvantovaya kriptografiya. Kvantovaya teleportatsiya. Kvantovye vychisleniya [Physics of Quantum Information: Quantum Cryptography. Quantum teleportation. Quantum computing]*, ed. by D. Boumeystera, A. Ekerta, A. TSaylingera, transl. from eng. by S.P. Kulik, E.A. Shapiro. Moscow: Postmarket, 2002, 376 p.
3. *Gisin N., Ribordy G., Tittel W., Zbinden H.* Quantum cryptography, *Reviews of Modern Physics*, 2002, Vol. 74, No. 1, pp. 145-195.
4. Rostelekom ob'yavil o vnedrenii kvantovoy kriptografii na svoikh setyakh [Rostelecom announced the introduction of quantum cryptography on its networks]. Available at: <https://tass.ru/ekonomika/5685597> (accessed 17 October 2018).
5. *Rumyantsev K.E.* Sistemy kvantovogo raspredeleniya klyucha [Systems of quantum key distribution: monograph]. Taganrog: Izd-voTTI YuFU, 2011, 264 p.
6. *Bennett C., Brassard G.* Quantum cryptography: Public key distribution and coin tossing. *Proceedings of IEEE international conference on computers, systems and signal processing. Bangalore, India*. New York: Institute of Electrical and Electronics Engineers, 1984, pp. 175-179.

7. Shor P.W., Preskill J. Simple proof of security of the BB84 quantum key distribution protocol. *Physical Review Letters*, 2000, Vol. 85, pp. 441-444.
8. Mironov Y.K., Rumyantsev K.E. Single-Photon Algorithm for Synchronizing the System of Quantum Key Distribution with Polling Sections of a Fiber-Optic Line. *Futuristic Trends in Networks and Computing Technologies*, 2020, pp. 87-97. DOI: https://doi.org/10.1007/978-981-15-4451-4_8.
9. Rumyantsev K.E. Sinhronizatsiya v sisteme kvantovogo raspredeleniya klyucha s avtomaticheskoy kompensatsiey polarizatsionnykh iskazheniy [Synchronization in a quantum key distribution system with automatic compensation of polarization distortions]. *Telekommunikatsii* [Telecommunications], 2017, No. 2, pp. 32-40.
10. Rumyantsev K.E., Plenkin A.P. Bezopasnost' rezhima sinhronizatsii sistemy kvantovogo raspredeleniya klyuchey [Security of the synchronization mode of a system of quantum key distribution]. *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2015, No. 5 (166), pp. 135-153.
11. Kurochkin V.L. i dr. Eksperimental'nye issledovaniya v oblasti kvantovoy kriptografii [Experimental research in the field of quantum cryptography], *Fotonika* [Photonics], 2012, Vol. 5, pp. 54-66.
12. Rumyantsev K.E., Plenkin A.P. Sinhronizatsiya sistemy kvantovogo raspredeleniya klyucha pri ispol'zovanii fotonnykh impul'sov dlya povysheniya zashchishchyonnosti [Synchronization of the system of quantum key distribution when using photon pulses to increase security]. *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2014, No. 8, pp. 81-96.
13. Rumyantsev K.E., Plenkin A.P. Preliminary Stage Synchronization Algorithm of Autocompensation Quantum Key Distribution System with an Unauthorized Access Security. *International Conference on Electronics, Information, and Communications (ICEIC)*, 2016. Vietnam, Danang. pp. 1-4. DOI: 10.1109/ELINFOCOM.2016.7562955. WOS:000389518100035. IDS: BG5KP.
14. Rumyantsev K., Rudinsky E. Parameters of the two-stage synchronization algorithm for the quantum key distribution system. *Proceedings of the 10th International Conference on Security of Information and Networks (SIN'17)*, 2017, pp. 140-147. DOI: 10.1145/3136825.3136888.
15. Rumyantsev K.E., Shakir H.H. Proektirovanie sistemy kvantovogo raspredeleniya klyucha s interferometrami Makha-Tsendera: ucheb. posobie [Designing a quantum key distribution system with Mach-Zehnder interferometers: a tutorial]. Rostov-on-Don; Taganrog: Izd-vo YuFU, 2020, 108 p.
16. Rumyantsev K.E. Kvantovye tekhnologii v telekommunikatsionnykh sistemakh: uchebnik [Quantum technologies in telecommunication systems: textbook]. Rostov-on-Don; Taganrog: Izd-vo YuFU, 2021, 346 p.
17. Lazernye diody Fabri-Pero [Fabry-Perot laser diodes]. Available at: http://www.electroncom.ru/product/nanoplus/fp_diodes.php (accessed 10 January 2022).
18. Poverkhnostno-izluchayushchie lazery s vertikal'nym rezonatorom [Surface-emitting lasers with a vertical cavity]. Available at: <http://msd.com.ua/optoelektronika/poverkhnostno-izluchayushhielazery-s-vertikalnym-rezonatorom-vcse/> (accessed 10 January 2022).
19. Lazery s vneshnim rezonatorom [External cavity lasers]. Available at: http://fmnauka.narod.ru/lazery_s_vneshnim_rezonatorom.pdf (accessed 10 January 2022).
20. Osnovnye parametry i sertifikatsiya opticheskikh SFP module [Main parameters and certification of optical SFP modules]. Available at: <https://deps.ua/knowegable-base-ru/articles/1961-osnovnye-parametry-i-sertifikatsiya-opticheskikh-sfp-modulei.html> (accessed 10 January 2022).

Статью рекомендовала к опубликованию к.т.н. К.Б. Дахкильгова.

Румянцев Константин Евгеньевич – Южный федеральный университет; e-mail: rke2004@mail.ru; г. Таганрог, Россия; тел.: 89281827209; кафедра информационной безопасности телекоммуникационных систем; зав. кафедрой; д.т.н.; профессор.

Миронова Полина Демьяновна – e-mail: linenkopdem@gmail.com; тел.: 89081924053; кафедра информационной безопасности телекоммуникационных систем; аспирант.

Шакир Хайдер Хуссейн – e-mail: hyder.almansoor@yahoo.com; кафедра информационной безопасности телекоммуникационных систем; аспирант.

Rumyantsev Konstantin Evgenievich – Southern Federal University; e-mail: rke2004@mail.ru; Taganrog, Russia; phone: +79281827209; the department of information security of telecommunication systems; head of department; dr. of eng. sc.; professor.

Mironova Polina Demyanovna – e-mail: linenkopdem@gmail.com; phone: +79081924053; the department of information security of telecommunication systems; graduate student.

Shakir Hayder Hussein – e-mail: hyder.almansoor@yahoo.com; the department of information security of telecommunication systems; graduate student.

УДК 550.343.3+550.34.016

DOI 10.18522/2311-3103-2022-4-77-94

А.С. Черепанцев

ЗАКОНОМЕРНОСТИ ПЕРЕХОДНОГО РЕЖИМА В ДИССИПАТИВНОЙ КЛЕТОЧНОЙ МОДЕЛИ ЗЕМЛЕТРЯСЕНИЙ

Целью данной работы был анализ механизмов роста кластеров сбросов, приводящего на решетке конечных размеров к состоянию, близкому к критическому, со степенным распределением по размерам кластеров, подобных наблюдаемым в сейсмическом процессе. В то же время вопрос о применимости модели для описания процессов в реальной геофизической среде остается открытым. Анализ связи элементов в одномерной модели OFC с открытыми граничными условиями позволяет оценить изменчивость поступающей энергии к элементам решетки расположенными на разном расстоянии от границ. Построенная расчетная модель позволяет оценить размер граничных областей высокой изменчивости средней поступающей энергии при различных значениях параметра связи α . Показано, что с ростом α граница область неоднородности расширяется. Показано что существуют два различных режима синхронного образования системы сбросов, имитирующих землетрясение. Оба механизма определяются захватом соседнего элемента и последующей синхронизацией их сбросов. Этот процесс формирует устойчивый сброс большого размера. Наличие пограничных областей с высоким градиентом скорости вводимой энергии определяет основной механизм образования кластеров элементов решетки и демонстрирующей синхронный сброс накопленной энергии. Такая синхронизация достигается за счет высокой взаимной изменчивости энергии на каждом шаге итерации. Вторым важным механизмом роста кластеров характерен для формирующихся кластеров, размер которых превышает размер приграничной области высокой неоднородности притока энергии. По мере роста размера кластера область захвата соседних элементов, не входящих в кластер, расширяется. Соответственно вероятность того, что энергия соседнего элемента находится в зоне захвата, увеличивается. Расчеты показывают, что среднее время достижения заданного размера кластера на решетке при разных размерностях пространства d и при разных параметрах связи α подтверждает наличие двух временных интервалов с разным механизмом образования кластеров. В таком случае, рост больших кластеров носит степенной характер с показателем степени, определяемым размерностью пространства d .

Модель Олами–Федера–Кристенсена; самоорганизованная критичность; степенное распределение в критических системах.

A.S. Cherepantsev

THE TRANSIENT REGIME PATTERNS IN THE DISSIPATIVE CELL MODEL OF EARTHQUAKES

The purpose of this work was to analyze the mechanisms of the growth of drop clusters, leading on a finite-size lattice to a state close to a critical one with a power-law size distribution of clusters similar to that observed in a seismic process. At the same time, the question of applicability of this model to the description of processes in a real geophysical medium remains. Analysis of the elements coupling in the one-dimensional OFC model with open boundary conditions allows

an estimation of the variability of the incoming energy to the lattice elements located at different distances from the boundaries. The constructed computational model makes it possible to estimate the size of the boundary areas of high average incoming energy variability at different values of the coupling parameter a . It is shown that, as a grows, the boundary region of inhomogeneity expands. It is shown that there are two different modes of synchronous drop formation, simulating an earthquake. Both mechanisms are determined by the capture of a neighboring element and the subsequent synchronization of the drops. This process forms a stable drop of a larger size. The presence of boundary regions with a high gradient of the input energy rate is the main mechanism for the formation of clusters of lattice elements, demonstrating the simultaneous drop of the accumulated energy. Such a synchronization is achieved due to the high mutual variability of energy at each iteration step. The second important mechanism of cluster growth is typical for the formed clusters that exceed the size of the near-boundary region of high inhomogeneity of the energy inflow. As the cluster size grows, the capture area of neighboring elements that are not included in the cluster expands. Accordingly, the probability that the energy of the neighboring element is in the capture area increases. The calculations show that the mean time of reaching the given size of the cluster on the lattice at different spatial dimensions d and at different coupling parameters confirms the presence of two time intervals with a different mechanism of cluster formation. In this case, the growth of large clusters has a power-law character, with an exponent determined by the dimension d .

Olami–Feder–Christensen model, self-organized criticality, power-law distribution in critical systems.

Введение. К настоящему времени проведено большое количество исследований, посвященных анализу возникновения самоорганизованного критического состояния и его параметров в клеточной модели Олами–Федера–Кристенсена (OFC). Замечательным свойством этой модели является то, что она повторяет статистические закономерности реального сейсмического процесса [1–3]. При определенных условиях механистическая модель Burridge–Knopoff (BK) [4] сводится к клеточной модели. Модель BK представляет собой динамическую систему упруго связанных блоков, расположенных на поверхности с трением и подверженных действию постоянной внешней силы. Модель BK – простейшая физико-механическая модель сложной динамической системы литосферных плит в земной коре. Прежде чем рассматривать механизмы синхронизации сбросов соседних элементов решетки в модели OFC со степенным распределением по размерам, кратко рассмотрим современные представления о закономерностях динамики реальной геофизической среды и некоторые физические и математические модели, разработанные для описания данного явления.

Принятая концепция тектоники плит позволяет рассматривать земную кору как открытую динамическую систему. Движение тектонических плит и их сцепление приводит к постоянному процессу роста упругой энергии. Энергия выделяется при достижении критических значений и сопровождается землетрясением с перераспределением упругой энергии с соседними участками и частичной диссипации.

Современные модели землетрясений основаны на концепции Рейда для подготовки для и возникновения землетрясения. Согласно этой концепции, разломы – это ослабленные участки земной коры. Накопленная упругая энергия периодически сбрасывается вдоль них. Землетрясение возникает при наличии препятствия смещению по разлому. Это определяет постепенное увеличение действующих здесь напряжений до критических значений. В районе, окружающем будущий источник, падение внутренней механической энергии резко замедляется, а энергия упругих деформаций накапливается [5]. После того, как напряжения достигают предела прочности, горная порода разрушается с образованием сдвиговых разрывов, что и является очагом землетрясения.

Другой механизм землетрясений связан с процессами динамического разрушения первоначально однородная порода. В этом случае рассматривается хрупкое разрушение породы и предполагается что для распространения трещины требуется дополнительная энергия разрушения материала.

С точки зрения наблюдаемого землетрясения эти два подхода математически эквивалентны, но различаются механизмом процесса разрыва и его параметрами.

Для землетрясения, как распространяющейся сдвиговой трещины, энергетический баланс [6] без учета изменения гравитационной энергии можно записать как

$$\Delta E_e = E_s + E_f + E_d ,$$

где ΔE_e – изменение упругой энергии при землетрясении, E_s – сейсмическая энергия землетрясения, E_f – работа, совершаемая силами трения на разломе, E_d – поверхностная энергия связи породы по разлому.

Соотношение энергетического баланса указывает на диссипативную природу динамической системы и физических процессов (правая часть уравнения), участвующих в этом. Под диссипацией в данном случае понимаются потери в системе упругих напряжений, вызванные не только потерями тепла на трение, но и также за счет энергии сейсмических волн. При этом потери, обусловленные сейсмической энергией, составляют лишь единицы процентов от общих потерь упругой энергии [7].

Фундаментальным свойством организации сейсмичности Земли является свойство масштабной инвариантности в энергетической и пространственной областях. Концепция масштабной инвариантности первоначально разработана для систем, близких к критической точке [8,9] со степенной функцией распределения. Распределения сейсмических событий в энергетической области определяется законом Гуттенберга-Рихтера:

$$N(E_s) \propto E_s^{-\gamma} ,$$

где N – число землетрясений, имеющих энергию, не меньшую E_s , $\gamma \approx 0.5$ – константа для тектонических землетрясений в различных регионах Земли.

Закон Гуттенберга-Рихтера и фрактальное распределение сейсмических событий, показывают степенное распределение размеров землетрясений с показателями γ, d , достаточно стабильными во времени и пространстве. Это предполагает сохранение масштабной инвариантности при сейсмическом процессе. При этом, принципиальное различие между сейсмичностью и стандартными критическими явлениями состоит в том, что для систем вблизи критической точки, например, фазовых переходов второго рода, масштабная инвариантность достигается только путем настройки параметра (температуры) на определенное значение. Ситуация, когда системы без настройки параметров спонтанно переходят в масштабно-инвариантное состояние, обычно определяется как состояние самоорганизованной критичности (SOC).

Модель, определяющая выбранные закономерности степенных распределений, включает переход от модели поведения отдельного сейсмического события к модели, включающей систему землетрясений и их взаимодействие. Основываясь на том факте, что землетрясение является формой диссипации накопленной потенциальной энергии в геологической среде, модель, описывающая автомодельные свойства сейсмичности, должна определять основной процесс-накопление напряжения и сброс напряжений в геологической среде. Это означает, что параметрами модельной динамической системы должны быть силовые характеристики, такие как тензор напряжений, кулоновское напряжение и сила взаимодействия блоков.

Лабораторные эксперименты показали, что прерывистое скольжение является общей чертой геологических материалов. Оно рассматривается как основной механизм землетрясения. Модель прерывистого скольжения легко получить в лабораторном эксперименте, когда отдельный блок располагается на шероховатой поверхности с трением и приводится в движение упругой пружиной с постоянной упругостью K , свободный конец которой движется с постоянной скоростью V_D . Динамика блока определяется законом трения. В [4] данный процесс смоделирован как одномерная цепочка блоков, соединенная между собой упругими пружинами и подверженная локальному трению и постоянно растущему линейному воздействию. Если сила, действующая на любой блок, превышает порог, она становится нестабильной и происходит смещение. В результате силы, действующие на другие блоки, могут измениться, так что другие блоки могут стать нестабильными. Это может привести к сходу лавин блоков, которые интерпретируются как землетрясения разной силы. ВК показали, что статистическое распределение сбросов потенциальной энергии $P(E_p)$ в данной экспериментальной модели согласуется со степенным законом распределения землетрясений: $P(E_p) \propto E^{-\eta}$.

Для уменьшения вычислительных затрат в расчете блочной модели обычно используются модели клеточных автоматов [1, 10, 11]. Математически эти модели включают карты сбросов вместо дифференциальных уравнений. Физически это соответствует мгновенно происходящим событиям скольжения без учета инерционных эффектов.

При учете пространственной неоднородности и сегментации разломов необходимо учитывать множество взаимодействующих блоков или сегментов разломов. Это приводит к задаче обработки пространственно-временного поля напряжений. Как правило, и прочность материала также зависит от рассматриваемого размера исследуемого объекта. Такая сложная модель может быть обработана с помощью методологии статистической физики, подобной модели Изинга или перколяционным моделям [12–14].

Особый интерес представляет класс моделей [1], демонстрирующих состояние самоорганизованной критичности (SOC), которые берут начало в простой модели клеточного автомата Bak, Tang, Wiesenfeld (BTW) [15]. В этом случае система постоянно приближается к критической точке с почти безмасштабными характеристиками. Следовательно, каждое небольшое событие с некоторой вероятностью может перерасти в сильное землетрясение (BTW) [15]. В этом случае система всегда приближается критической точке с практически безмасштабными характеристиками. Следовательно, каждое малое событие с некоторой вероятностью может перерасти в сильное землетрясение.

Диссипативная OFC-модель клеточных автоматов, представленная в [1], эквивалентна двумерной модели блоков с упругим взаимодействием в квазистатическом приближении. Она сочетает в себе простота построения и нетривиальный характер поведения. В отличие от исходного консервативных моделей происхождения SOC, модель OFC является диссипативной. Показатель степени закона спада в распределении амплитуд разломов, полученный в модели, близок к наблюдаемому значению в сейсмичности.

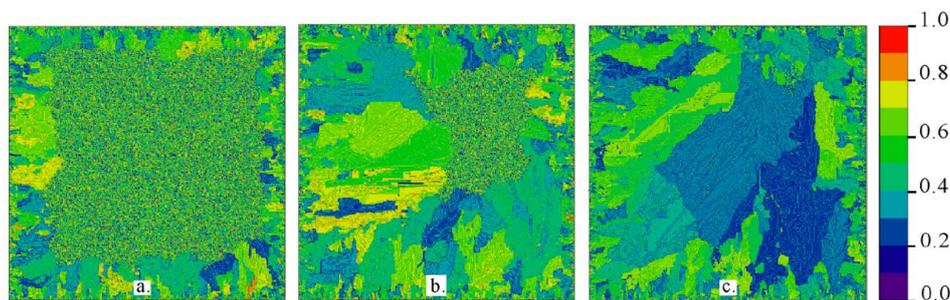


Рис. 1. Распределение энергии в 2d OFC модели после T сбросов в каждом элементе. $L \times L = 500 \times 500$, $\alpha = 0,125$, $E_c = 1$. (a) $T = 4 \cdot 10^3$. (б) $T = 2 \cdot 10^4$. (в) $T = 6 \cdot 10^4$

Как отмечено в [16], анализ распределения лавин сбросов по размеру требует предварительного расчета начального переходного режима выхода на стационарное состояние. В этом случае, время для расчета начального режима, измеряемого параметром, связанным с числом шагов приращения энергии в ячейках решетки, оказывается чрезвычайно большим. Оно зависит от размер решетки L и параметр диссипации α [17]. В [16] также указано, что многие распределения размеров лавин сбросов, представленных за последнее десятилетие, фактически были получены в переходном состоянии системы и не являются конечными. Таким образом, исследование параметров переходного режима является необходимой стадией анализа процесса формирования устойчивого состояния в модели OFC.

Одно из первых исследований переходного режима представлено в [18]. Авторы установили что образование начальных лавин происходит в граничных областях решетки. Неоднородный характер энерго вклада в граничных элементах определяет начальное появление кластеров сбросов в этих областях из-за эффекта синхронизации. Впоследствии этот эффект распространяется по всей решетке. В этом случае тип граничных условий задает тип амплитудного распределения сбросов. На рис. 1 показаны этапы переходного режима до устойчивого состояния, формируемого в двумерной модели OFC с открытыми граничными условиями.

Здесь L – размер решетки, α – параметр связи, E_c – предельная энергия сброса. В [18] синхронизированные элементы названы самоорганизующимися кластерами. Было показано, что элементы внутри кластеров имеют примерно одинаковую энергию. Кроме того, они стабильны в течение много циклов [1]. Элементы в одном кластере ведут себя одинаково при сбросе и сбрасываются в одной и той же последовательности. Кластеры, формируемые при разных значениях α , различаются по размеру. Размер кластера растет с увеличением α . При этом, размер кластеров растет по мере удаления от границы решетки [3].

Первоначально, модельные расчеты [18] показали, что время переходного режима установления увеличивается со временем как

$$t \propto L^{\eta(\alpha)}.$$

Однако, позднее было показано, что показатель степени $\eta \approx 1,3$ и не зависит от α [19].

Основная часть опубликованных исследований переходного режима сосредоточена на анализе расчетных характеристики при изменении параметров системы α , L и, как следствие, построение количественных соотношений. В то же время выводы часто противоречивы. Поэтому получение теоретических соотношений путем оценки величины граничной области неоднородности энергии при различных α и L является актуальной темой.

Имеющиеся данные указывают на различный характер устойчивого состояния системы в случае одно- и двумерной решетки элементов (рис. 2). Это ставит вопросы о характеристиках переходного режима в системах с различной размерностью d и выяснении механизма образования кластеров в данных моделях.

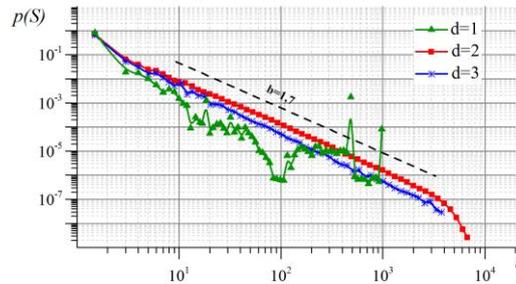


Рис. 2. Распределение размеров сбросов при размерностях решетки $d=1,2,3$

Описание модели OFC. Рассмотрим кубическую решетку размерности d . Число элементов решетки равно $N = L^d$. Присвоим каждому i -му элементу некоторый динамический параметр E_i . В простейшем случае E_i внутренняя энергия, запасенная в i -м элементе. Будем считать, что в единицу времени все клетки занимают такое же дополнительное значение приращения энергии ΔE

$$E_i \rightarrow E_i + \Delta E b, \quad i=1, \dots, N. \quad (1)$$

При отсутствии влияния соседних элементов, такое увеличение энергии происходит пока $E_i < E_c$. Здесь E_c – некоторое пороговое максимальное значение внутренней энергии. Когда достигнута предельная энергия, элемент сбрасывает накопленную энергию, часть α которой передается соседним клеткам, где $0 \leq \alpha \leq 1/2d$. В случае $d = 2$

$$E_{i,j} \geq E_{\max} \rightarrow \begin{cases} E_{i,j} \rightarrow 0 \\ E_{i\pm 1,j} \rightarrow E_{i\pm 1,j} + \alpha E_{i,j} \\ E_{i,j\pm 1} \rightarrow E_{i,j\pm 1} + \alpha E_{i,j} \end{cases} \quad (2)$$

При сбросе энергии (2) часть внутренней энергии системы теряется. Система будет диссипативной при условии

$$(1 - 2d\alpha)E_i > 0. \quad (3)$$

Условие консервативности системы определяется как $\alpha = 1/2d$. Дальнейшая эволюция элемента определяется (1) до тех пор, пока она снова не достигнет критического состояния (2). В зависимости от состояния системы, когда один из элементов достигает критического значения, это может мало сказаться на состоянии всей системы, а может вызвать лавинный процесс сброса энергии и переход всей системы в новое энергетическое состояние.

Граничные условия являются важной частью определения модели.

Открытые граничные условия определяются тем, что всем элементам, независимо от их позиции на решетке, получают постоянное приращение αE_i . Это определяет условие сброса, например, для левой границы $2d$ решетки

$$E_{0,j} \geq E_c \rightarrow \begin{cases} E_{0,j} \rightarrow 0 \\ E_{0,j\pm 1} \rightarrow E_{0,j\pm 1} + k\alpha E_{0,j} \\ E_{1,j} \rightarrow E_{1,j} + k\alpha E_{0,j} \end{cases}$$

В этом случае общие потери энергии системы определяются не только потерями энергии при сбросе ($\alpha < 1/4$), но и оттоком энергии через границы.

При выборе открытых граничных условий система после конечного числа итераций переходит в устойчивое состояние, характеризующееся степенным законом распределения сбросов по размерам [1].

Периодические краевые условия формулируются для систем, в которых отсутствует характеристика размер системы L . В $2d$ -модели с такими условиями решетка сворачивается в тор и условия на границе решетки размером $L \times L$ следующие:

$$\begin{aligned} E_{1,j} \geq E_c &\rightarrow E_{L,j} \rightarrow E_{L,j} + \alpha E_{1,j}; \\ E_{L,j} \geq E_c &\rightarrow E_{1,j} \rightarrow E_{1,j} + \alpha E_{L,j}; \\ E_{i,1} \geq E_c &\rightarrow E_{i,L} \rightarrow E_{i,L} + \alpha E_{i,1}; \\ E_{i,L} \geq E_c &\rightarrow E_{i,1} \rightarrow E_{i,1} + \alpha E_{i,L}. \end{aligned}$$

Расчеты показали, что для модели с периодическими условиями при значениях $\alpha < 1/4$ самоорганизованное состояние и, соответственно, степенное распределение величин сбросов не происходит. Сброс отдельного элемента почти никогда не сопровождается нарушением стабильности соседних ячеек и не происходит лавинообразного увеличения пространственного масштаба кластеров [18].

На рис. 2 показаны устойчивые распределения размеров сброса S для модели OFC с открытыми условиями на границе. Полученные соотношения показывают, что в одномерном случае устойчивое степенное распределение отсутствует.

Распределение числа сбросов на решетке. Для одномерной OFC модели с открытыми граничными условиями оценим распределение подводимой энергии по элементам вдоль решетки.

Определим энергию i -го элемента как z_i^k ($i = 1, \dots, N$) после k приращений энергии во всех элементах решетки. В этом случае k – параметр условного времени эволюции системы, с условной единицей $\Delta T = E_c / \Delta E$.

Предполагая, что превышение энергии в элементах $\max(z_i / E_c) \ll 1$ z_i^k можно приближенно определить как:

$$\begin{cases} z_i^k = z_i^0 + \Delta E \cdot k - n_i^k E_c + \alpha E_c (n_{i-1}^k + n_{i+1}^k), & i = 2, \dots, N-1 \\ z_1^k = z_1^0 + \Delta E \cdot k - n_1^k E_c + \alpha E_c n_2^k \\ z_N^k = z_N^0 + \Delta E \cdot k - n_N^k E_c + \alpha E_c n_{N-1}^k \end{cases} \quad (4)$$

Здесь ΔE – приращение энергии на одном шаге итерации, n_i^k – число сбросов, произошедшем в i -ом элементе за k итераций, E_c – пороговое значение энергии, α – параметр связи соседних элементов.

Выражение (4) можно переписать в виде

$$\begin{cases} \left(n_{i+1}^k - 2n_i^k + n_{i-1}^k \right) - \frac{1-2\alpha}{\alpha} n_i^k = \frac{1}{\alpha E_c} (z_i^k - z_i^0 - \Delta E \cdot k), & i = 2, \dots, N-1 \\ \left(n_2^k - n_1^k \right) + \frac{\alpha-1}{\alpha} n_1^k = -\frac{\Delta E \cdot k}{\alpha E_c}, \\ \left(n_N^k - n_{N-1}^k \right) - \frac{\alpha-1}{\alpha} n_N^k = \frac{\Delta E \cdot k}{\alpha E_c}. \end{cases} \quad (5)$$

Для оценки распределения сбросов на решетке, перейдем от разностного представления к краевой задаче для непрерывной переменной $n(x)$, где x – координата элемента на одномерной решетке:

$$\begin{cases} n'' - \frac{1-2\alpha}{\alpha} n = B, \\ n'(0) - \frac{1-\alpha}{\alpha} n(0) = B, \\ n'(L) + \frac{1-\alpha}{\alpha} n(L) = -B. \end{cases} \quad (6)$$

Здесь $B = -\frac{\Delta E \cdot k}{\alpha \cdot E_c}$. Такое приближение справедливо при числе итераций $k \gg 1$:

$$\Delta F \cdot k \gg |z_i^k - z_i^0|.$$

В случае пространства решетки $d > 1$ уравнение (6) сводится к уравнению эллиптического типа (уравнению Гельмгольца) со смешанными граничными условиями. Аналитическое решение (9) представляет функцию:

$$n(x) = \frac{1}{1-2\alpha} \cdot \frac{\Delta E \cdot k}{E_c} \cdot \left(1 - \frac{\operatorname{ch}\left(\sqrt{\frac{1-2\alpha}{\alpha}}\left(x - \frac{L}{2}\right)\right)}{\sqrt{\frac{1-2\alpha}{\alpha}} \cdot \operatorname{sh}\left(\frac{L}{2}\sqrt{\frac{1-2\alpha}{\alpha}}\right) + \frac{1-\alpha}{\alpha} \operatorname{ch}\left(\frac{L}{2}\sqrt{\frac{1-2\alpha}{\alpha}}\right)} \right). \quad (7)$$

На рис. 3 показано распределение числа сбросов вдоль сетки $n(x, \alpha)$ в соответствии с (7). Также на рис. 3 приведено распределение (сплошная линия), полученное прямым расчетом модели OFC.

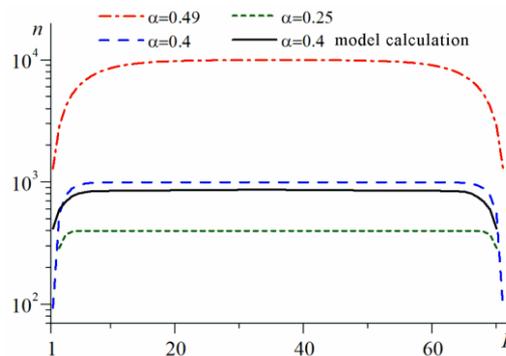


Рис. 3. Распределение числа сбросов в соответствии с (7) для различных значений параметра α . $L = 70$, $\Delta T = 200$ ($T = k \cdot \Delta E / E_c$)

Близость экспериментальной кривой и теоретической указывают на адекватность предложенной схемы оценки (7). Рост числа сбросов с увеличением параметра связи α связан с уменьшением диссипации энергии при увеличении α : $\Delta E = E_c (1 - 2\alpha)$. Это приводит к увеличению частоты сбросов как основного механизма компенсации поступающей в систему энергии в стационарном режиме.

Полученное небольшое расхождение теоретической и рассчитанной в модели кривых при одном и том же параметре $\alpha=0.4$ указывает на приближенный характер теоретической оценки, обусловленный сделанным допущением равенства сбрасываемой энергии величине E_c . В реальном процессе формирования лавины путем последовательности сбросов в соседних элементах, происходит рост величины превышающей энергии относительно порогового значения E_c и соответственно большая величина диссипации энергии при сбросе. Это приводит к меньшему числу необходимых сбросов.

Условно данный механизм показан на рис. 4 для сформированного кластера с близкими значениями энергии в элементах. Здесь и в дальнейшем под кластером понимается совокупность соседних элементов решетки, одновременно сбрасываемых при лавинообразном процессе и имеющих близкие значения энергии. Для удобства представления показан случай, когда все элементы достигли порогового значения и сбросы начинаются слева направо.

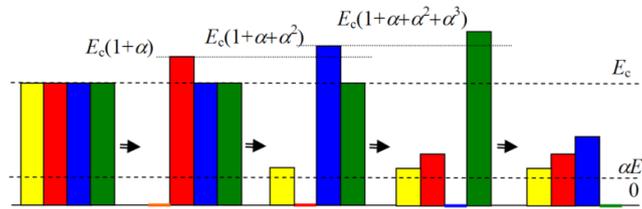


Рис. 4. Схема распространения сбросов в модели OFC

На рис. 4 показан механизм распространения лавины сбросов при котором с ростом номера сброса увеличивается энергия диссипации. Следовательно, величина сброшенной удельной энергии, приходящаяся на один элемент $\Delta \varepsilon$, зависит от размера лавины. То есть, лавина размером S является более эффективным средством сброса энергии ΔE , чем набор лавин суммарно того же размера:

$$S = \sum_{i=1}^m s_i \Rightarrow \Delta E_S > \sum_{i=1}^m \Delta E_{s_i} .$$

Условие $z_i \geq E_c \rightarrow \begin{cases} z_i \rightarrow 0 \\ z_{z \pm 1} \rightarrow z_{i \pm 1} + \alpha z_i \end{cases}$ в формулировке эволюции OFC модели носит принципиальный характер. При задании условия сброса в виде

$z_i \geq F_c \rightarrow \begin{cases} z_i \rightarrow 0 \\ z_{z \pm 1} \rightarrow z_{i \pm 1} + F_c \end{cases}$ в диссипативной модели степенное распределение

размеров лавин отсутствует и модель не обладает свойством SOC. Состояние SOC в этом случае достигается в пределе $\alpha \rightarrow 0.5$, когда модель переходит в модель BTW.

Также следует отметить, что в модели OFC при значениях α , приближающихся к критической величине $\alpha=0.5$ возможно возникновение повторных сбросов в ходе развития лавин большого размера. Этот эффект возникает, когда в сброшенный элемент поступает приращение энергии от соседних сброшенных элементов, превышающее пороговое значение.

Устойчивость сформированных кластеров сбросов. Расчет OFC модели показывает, что система стремится к статистически устойчивому соотношению набора кластеров элементов по размеру с близкими значениями параметра z_i . Оценим устойчивость подобного кластера исходя из механизма связи между элементами.

Пусть кластер состоит из N элементов: $z_i, i=1, \dots, N$. Условием формирования лавины сбросов, является

$$E_c(1-\alpha) \leq z_i \leq E_c. \quad (8)$$

Передними верхними индексами определим состояния элемента в ходе сброса: 0 – начальное значение энергии до сброса, 1 – перед сбросом, 2 – после сброса.

Тогда для i -ого элемента величина энергии после сбросов ($i-1$) элементов с младшими индексами:

$$\begin{aligned} {}^1z_i &= {}^0z_i + \alpha \cdot {}^1z_{i-1} = {}^0z_i + \alpha \left({}^0z_{i-1} + \alpha \cdot {}^1z_{i-2} \right) = {}^0z_i + \alpha \cdot {}^0z_{i-1} + \alpha^2 \left({}^0z_{i-2} + \alpha \cdot {}^1z_{i-3} \right) = \\ &= {}^0z_i + \alpha \cdot {}^0z_{i-1} + \alpha^2 \cdot {}^0z_{i-2} + \dots + \alpha^{i-1} \cdot {}^0z_1 = \sum_{k=0}^{i-1} \alpha^k \cdot {}^0z_{i-k} \end{aligned} \quad (9)$$

Аналогично:

$${}^2z_i = \alpha \cdot {}^1z_{i+1} = \alpha \left({}^0z_{i+1} + \alpha \cdot {}^1z_i \right) = {}^0z_{i+1} + \alpha^2 \sum_{k=0}^{i-1} \alpha^k \cdot {}^0z_{i-k}. \quad (10)$$

Из (9) следует, что приращения энергии в элементе после сброса кластера

$$\Delta z_i = {}^1z_i - {}^0z_i = \sum_{k=1}^{i-1} \alpha^k \cdot {}^0z_{i-k}$$

при лавинном процессе растет по мере развития лавины. Исходя из условия (8), для приращения энергии в больших кластерах $N \gg 1, i \gg 1$ справедливы оценки:

$$\alpha E_c \leq \Delta z_i \leq \frac{\alpha}{1-\alpha} E_c, \quad (11)$$

$$\alpha E_c \leq {}^2z_i \leq \frac{\alpha}{1-\alpha} E_c. \quad (12)$$

Из (11), (12) следует, что невозможно возникновение повторного сброса в узле кластера для диссипативной $1d$ модели ($0 < \alpha < 0.5$) при расчете сбросов в одном направлении. При этом в случае расчета сбросов одновременно с двух направлений - от элемента 1 к элементу 2 и т.д. и от элемента N к элементу $N-1$ такая возможность существует.

Для анализа устойчивости кластера, рассмотрим, как изменяется разность энергий в двух произвольных элементах после сбросов. Из (12) следует:

$$\begin{aligned} 0 &\leq \left| {}^0z_i - {}^0z_j \right| \leq E_c(1-\alpha), \\ 0 &\leq \left| {}^2z_i - {}^2z_j \right| \leq \frac{\alpha}{1-\alpha} E_c - \alpha E_c = \alpha^2 E_c. \end{aligned}$$

Для значений $0 < \alpha < 0.5$ всегда справедливо: $\alpha^2 < 1-\alpha$. Это означает, что диапазон разброса энергий в элементах кластера после сброса уменьшился по сравнению с диапазоном до сброса. Следовательно, значения энергии внутри кластера при последовательных лавинных сбросах стягиваются к одной величине и кластер демонстрирует устойчивость.

В заключении отметим, что после лавинного сброса энергии в кластере, наибольшей энергией обладают последние сбрасываемые элементы. В ходе последующего роста энергии именно они достигнут первыми порогового значения E_c .

Поэтому процесс лавинного сброса будет происходить в обратном направлении по сравнению с рассмотренным выше. Это также приводит к выравниванию энергий в элементах кластера.

Механизм формирования кластера. Рост размера лавины обусловлен захватом граничными элементами кластера соседних элементов. При этом происходит выравнивание значения энергии захваченного элемента с значениями энергии в элементах кластера с дальнейшей синхронизацией сбросов.

Условием синхронного сброса ($N+1$) элемента, граничащего со сброшенным элементом N кластера, является: ${}^0z_{N+1} + \alpha \cdot {}^1z_N \geq E_c$.

И с учетом (8):

$$F_c > {}^0z_{N+1} \geq F_c - \alpha \cdot \sum_{k=0}^{N-1} \alpha^k \cdot {}^0z_{N-k}. \quad (13)$$

Определим порог захвата элемента E_Z как величину минимальной энергии состояния элемента ${}^0z_{N+1}$, достаточной для захвата соседним кластером.

Для больших кластеров ($N \gg 1$) порог захвата z_{N+1} находится в диапазоне:

$$E_Z \geq E_c - \alpha \cdot \sum_{k=0}^{N-1} \alpha^k \cdot {}^0z_{N-k} \geq E_c \left(1 - \frac{\alpha}{1-\alpha}\right) = E_c \frac{1-2\alpha}{1-\alpha},$$

$$E_Z \leq E_c (1-\alpha).$$

На рис. 5 представлен диапазон изменения порога захвата при различных параметрах связи α :

$$E_c (1-\alpha) \leq E_Z \leq E_c \frac{(1-2\alpha)}{(1-\alpha)}. \quad (14)$$

Полученные зависимости определяют увеличение скорости роста кластеров с ростом параметра α .

Из (14) следует, что для больших кластеров при значениях параметра связи α , близких к консервативному пределу $\alpha=0.5$, вероятность захвата может приближаться к 1 при любых значениях ${}^0z_{N+1}$ ($0 \leq {}^0z_{N+1} < F_c$).

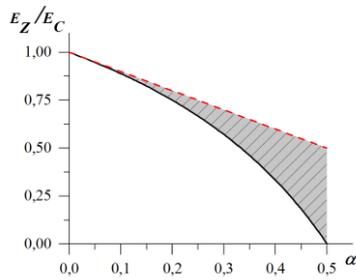


Рис. 5. Диапазон уровня захвата ячейки кластером (заштрихованная область)

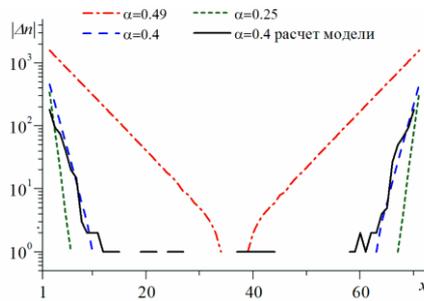


Рис. 6. Скорость изменения числа сбросов исходя из (4) и в расчетной модели при различных значениях параметра α . $L=70$, $T=200$ ($T = k \cdot \Delta E/E_c$).

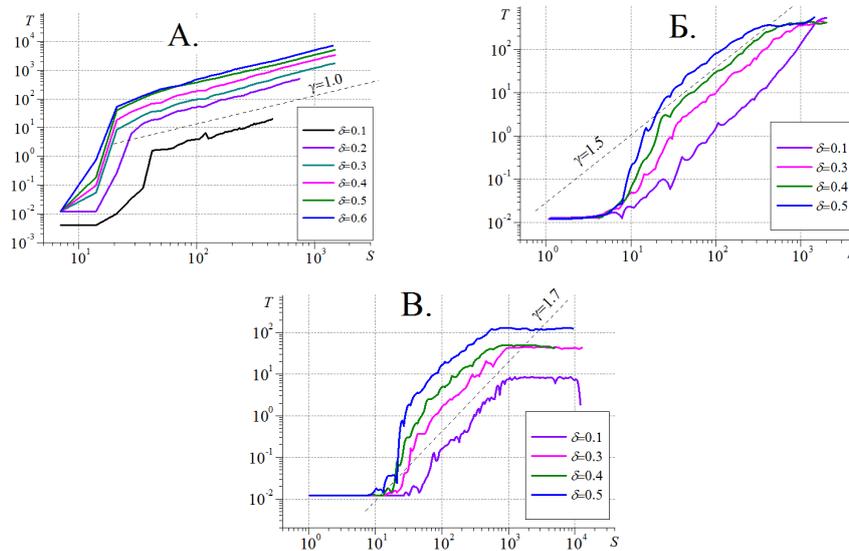


Рис. 7 Зависимость времени возникновения кластера от размера при различных параметрах диссипации δ . А.-1d, Б.-2d, В-3d

При малых N и α захват определяется вероятностью нахождения z_{N+1} в диапазоне (13). В случае если ${}^0z_{N+1}$ меньше порога захвата, то сброса данного узла и его захвата не происходит. При данном механизме роста кластера и равенстве приращений энергий ΔE на каждом шаге итераций при следующем лавинном сбросе, узел кластера ${}^0z_{N+1}$ остается вне области захвата.

Рассмотрим другой механизм захвата. Как было выше показано, приращение энергии и соответственно число сбросов при открытом типе граничных условий изменяется вдоль решетки (рис. 3). Уменьшение скорости изменения приращения энергии от граничных элементов к центральным имеет экспоненциальный характер. На рис. 6 представлены зависимости изменения числа сбросов для узлов одномерной сетки.

Значительные различия скорости приращения энергии для соседних элементов характерны для приграничных областей. Эта закономерность повышает вероятность возникновения в соседних элементах приграничных областей на следующих шагах итераций относительных значений энергий, достаточных для синхронизации сбросов, формирования кластера и его дальнейшего роста. В качестве иллюстрации, на рис. 1,а показана начальная стадия эволюции модели OFC в $2d$ пространстве с формированием кластеров в приграничных областях решетки.

Следует отметить, что длина граничной области с данным механизмом захвата слабо зависит от размера сетки L . Рассмотрим соотношение (7). Учтем, что для

$L \geq 30$ и $0.1 \leq \alpha \leq 0.45$ справедливо: $\frac{L}{2} \sqrt{\frac{1-2\alpha}{\alpha}} \gg 1$. Тогда:

$$\text{sh}\left(\frac{L}{2} \sqrt{\frac{1-2\alpha}{\alpha}}\right) \approx \text{ch}\left(\frac{L}{2} \sqrt{\frac{1-2\alpha}{\alpha}}\right).$$

И с учетом разложения гиперболического косинуса разности:

$$n(x) \approx \frac{T}{1-2\alpha} \cdot \left(1 - \frac{\operatorname{ch}\left(\sqrt{\frac{1-2\alpha}{\alpha}}x\right) - \operatorname{sh}\left(\sqrt{\frac{1-2\alpha}{\alpha}}x\right)}{\sqrt{\frac{1-2\alpha}{\alpha}} + \frac{1-\alpha}{\alpha}} \right) = \frac{T}{1-2\alpha} \cdot \left(1 - \frac{e^{-\sqrt{\frac{1-2\alpha}{\alpha}}x}}{\sqrt{\frac{1-2\alpha}{\alpha}} + \frac{1-\alpha}{\alpha}} \right), \quad (15)$$

где $T = k \cdot \Delta E / E_c$.

В (15) отсутствует зависимость от размера сетки L . При этом длина граничной области зависит только от параметра связи α , увеличиваясь с ростом α .

Переходного режима к устойчивому состоянию в OFC модели на сетках различной размерности. Для анализа поведения OFC модели при формировании картины устойчивого степенного распределения рассмотрим развитие кластеров лавин на решетках различной размерности. Имеющиеся данные [1, 2] указывают, что $2d$ и $3d$ модели демонстрируют степенную функцию плотности распределения размеров сбросов с показателем, близким к 2. При этом в $1d$ модели предельное степенное распределение отсутствует [20]. Предельное поведение в этом случае формируется либо как синхронный сброс всех узлов сетки либо как наличие нескольких сегментов сбрасываемых в различное время.

На рис. 7 представлены зависимости времени достижения сброса от его размера для размерности сетки $1d, 2d, 3d$. Представленные на рис. 7,А, кривые роста размера сброса со временем указывают на существование начального участка роста с резким ростом размеров сбросов и дальнейший рост $T \propto S^\gamma$ с показателем степени $\gamma \approx 1$. Такой линейный рост получен для $1d$ модели с различными значениями параметра диссипации $\delta = 1 - 2\alpha$.

Величина параметра связи α определяет скорость линейного роста размера сброса во времени $s'(\alpha)$. Исходя из результатов предыдущего раздела, вероятность захвата узла (11) при $N \gg 1$ имеет зависимость $s'(\alpha) \sim \frac{\alpha}{1-\alpha}$. На рис. 8 показаны данная

зависимость и расчетные значения модели $\frac{\Delta s}{\Delta T}(\alpha)$. Удовлетворительное согласие теоретической и модельной кривых может служить подтверждением адекватности предложенного механизма развития размера сброса в OFC модели.

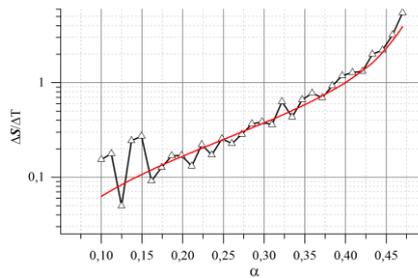


Рис. 8. Зависимость скорости роста размера сброса от параметра связи α .

Как следует из зависимостей на рис. 6, 7 диапазон области начальных амплитуд с резким ростом размера сбросов коррелирует с диапазоном граничных областей с большим значением изменчивости поступающей энергии в элементы. Это указывает на то, что основным механизмом роста сбросов является захват соседних узлов в сбросе, обусловленный высокой скоростью изменения взаимных энергий соседних ячеек.

В отличие от $1d$ модели, в случае $2d$ и $3d$ пространственной решетки узлов, рост времени достижения кластера заданного размера имеет степенную зависимость с $\gamma > 1$ (рис. 7,Б,В). При этом показатель γ слабо зависит от параметра связи $\alpha = \frac{1-\delta}{2d}$. Это указывает на иной механизм формирования кластеров. Будем предполагать, что подобно $1d$ модели, вероятность захвата граничным узлом кластера соседнего узла вне кластера определяется величиной (11), зависящей от α .

С ростом α вероятность $P(\alpha)$, а следовательно и скорость роста кластера также увеличивается. В $1d$ модели граничных ячеек две. Скорость роста кластера является постоянной величиной при фиксированном α :

$$\frac{dT}{dS_1} = c(\alpha).$$

В случае же моделей больших размерностей число граничных узлов кластера зависит от размера кластера и размерности пространства. Рассмотрим простейший механизм роста кластера исходя из условия, что один граничный узел кластера может захватить один внешний узел с вероятностью $P(\alpha)$. Число граничных узлов и размер кластера можно определить как

$$N_b = \lambda \cdot L^{d-1},$$

$$S_d = \chi \cdot L^d.$$

где λ, χ – параметры формы кластера, L – линейный размер кластера. Тогда:

$$\frac{dT}{dS_d} = N_b \cdot \frac{dT}{dS_1} = c(\alpha) \cdot \lambda L^{d-1} = c(\alpha) \cdot \lambda \left(\frac{S_d}{\chi} \right)^{\frac{d-1}{d}}. \quad (16)$$

Проинтегрировав (16):

$$T = \frac{d}{2d-1} \frac{c(\alpha) \cdot \lambda}{\chi^{\frac{d-1}{d}}} S_d^{2-\frac{1}{d}}. \quad (17)$$

Соотношение (17) дает линейную связь $T \sim S^1$ при $d=1$, $T \sim S^{\frac{3}{2}}$ при $d=2$ и $T \sim S^{\frac{5}{3}}$ при $d=3$. Как показано на рис. 7 степенные показатели модельных зависимостей оказываются близкими к зависимости (17).

Как следует из полученных соотношений, в переходном режиме рост пространственной корреляции и, соответственно, время появления кластера заданного размера имеет степенной характер (рис. 7) и не зависит от значения α при заданной размерности пространства d . Это подтверждает вывод, сделанный в [19]. Результат о наличии такой зависимости, указанный в [3], может иметь место, когда $T(S)$ получено в стационарном предельном состоянии. В этом случае зависимость показателя степени $b(\alpha)$ в распределениях сбросов по размеру определяет разную частоту появления сбросов заданного размера и соответственно меняет показатель степени в $T(S)$.

В качестве иллюстрации чувствительности показателя степенной зависимости $T(S)$, на рис. 9 показаны расчетные кривые потока сейсмических событий, соответствующие разным режимам. Фоновый режим соответствует стационарному состоянию и переходной режим афтершоковой последовательности после сильного землетрясения магнитудой $m = 5,9$.

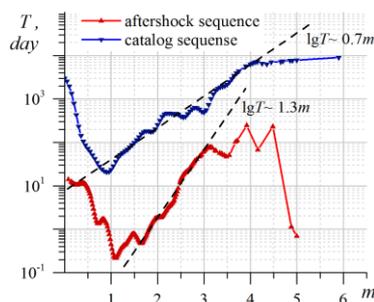


Рис. 9. Зависимость времени возникновения землетрясения от величины магнитуды по данным каталога USGS NCSN (1980-2010) в эпицентральной области Паркфилдского землетрясения 2004 года

Начало степенного роста кривой, соответствует величине представительности каталога $m_{\min} \approx 1$ и характеризует качество каталога. (Надежная регистрация землетрясений с магнитудой $m \geq m_{\min}$) Приведенные оценки показателей $\eta \approx 0,7$ и $\eta \approx 1,3$ для фоновой и афтершоковой сейсмичности соответствуют значению показателя степени b в степенной зависимости распределения магнитуд землетрясений по приведенным данным [21].

Заключение. Целью данной работы был анализ механизмов роста кластеров сбросов, приводящего на решетке конечных размеров к состоянию, близкому к критическому, со степенным распределением по размерам кластеров, подобных наблюдаемым в сейсмическом процессе. В то же время вопрос о применимости модели для описания процессов в реальной геофизической среде остается открытым. Решетка элементов модели представляет собой однородную систему, а организация геологической среды имеет блочную иерархическую структуру, разделенную разломами и разномасштабными неоднородностями. При этом наблюдаемая сейсмичность приурочена к межплитовым пограничным областям. И в этом смысле, она может отражать фрактальную структуру распределения неоднородностей земной коры а не результат нелинейного взаимодействия отдельных блоков, как это представлено в модели. Землетрясение может свидетельствовать о формировании новой неоднородности в свойствах среды. Модель после сбросов не меняет свойств в области сброса. Более того, поведение модели чрезвычайно чувствительно даже к малым шумовым возмущениям пороговой энергии E_c , или параметру взаимодействия с соседними ячейками α [22]. Даже небольшие возмущения в этом случае приводят к разрушению степенного распределения.

В то же время можно выделить фундаментальное свойство модели, отражающее реальный процесс. Нелинейное взаимодействие отдельных блоковых структур литосферы проявляется себя в накоплении упругой энергии до критического уровня, определяемого прочностными характеристиками с последующим сбросом с перераспределением напряжений в соседние области и возможное достижение при этом ими критического значения прочности. Имеющиеся данные лабораторных экспериментов по разрушению образцов горных пород [22] показывают, что образовавшаяся макротрещина представляет собой последовательность микротрещин твердого тела с чередованием сдвиговых и нормальных трещин. В этом смысле диссипативная OFC модель представляет собой базовую модель для построения более сложных моделей сейсмичности в реальной геологической среде. В основе таких моделей лежит принцип нелинейного взаимодействия отдельных структурных элементов.

Основные результаты работы сформулированы ниже.

1. Анализ связи элементов в одномерной модели OFC с открытыми граничными условиями позволяет оценить изменчивость поступающей энергии к элементам решетки расположенными на разном расстоянии от границ. Построенная расчетная модель позволяет оценить размер граничных областей высокой изменчивости средней поступающей энергии при различных значениях параметра связи α . Показано, что с ростом α граница область неоднородности расширяется.

2. Наличие пограничных областей с высоким градиентом скорости вводимой энергии определяет основной механизм образования кластеров элементов решетки и демонстрирующий синхронный сброс накопленной энергии. Такая синхронизация достигается за счет высокой взаимной изменчивости энергии на каждом шаге итерации. Этот процесс увеличивает вероятность одновременного сброса нескольких соседних элементов. Это происходит, когда энергия соседнего элемента достигает порога сброса из-за сброса соседнего элемента. Показано, что этот механизм кластерного образования характерен для начальной стадии организации системы и протекает в пограничных областях.

3. Второй важный механизм роста кластеров характерен для формирующихся кластеров, размер которых превышает размер приграничной области высокой неоднородности притока энергии. По мере роста размера кластера область захвата соседних элементов, не входящих в кластер, расширяется. Соответственно вероятность того, что энергия соседнего элемента находится в зоне захвата, увеличивается. Это означает, что после сброса соседних элементов кластера этот элемент также будет сброшен и станет новым граничным элементом кластера. Показано, что с увеличением параметра α область захвата расширяется. Это приводит к увеличению скорости роста кластера и, соответственно, более быстрое достижение предельного состояния SOC. В пределе $\alpha \rightarrow 1/2$ область захвата стремится ко всему диапазону возможных энергетических состояний элемента $z_i \in [0, E_c)$.

4. Исследование показывает устойчивость кластера в процессе многократных сбросов. В соответствии с оценками, во время сброса кластера среднее значение разброса энергий элементов кластера уменьшается. При этом повышается однородность энергии внутри кластера. Это можно рассматривать как показатель стабильности состояния СОК.

5. Расчеты показывают, что среднее время достижения заданного размера кластера на решетка при разных размерностях пространства d и при разных параметрах связи α подтверждает наличие двух временных интервалов с разным механизмом образования кластеров. В таком случае, рост больших кластеров носит степенной характер с показателем степени, определяемым размерностью пространства d :

$$N(t) = N_0(\alpha) \cdot t^{\eta(d)}.$$

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Olami Z., Feder H.S., and Christensen K. Self-Organized Criticality in a Continuous, Nonconservative Cellular Automaton Modeling Earthquakes // Phys. Rev. Lett. – 1992. – Vol. 68, No. 8. – P. 1244-1247.
2. Lise S. and Paczuski M. Scaling in a Nonconservative Earthquake Model of Self-Organized Criticality // Phys. Rev. E. – 2001. – Vol. 64, No. 4, 046111. – 5 p.
3. Drossel B. Complex Scaling Behavior of Nonconserved Self-Organized Critical Systems // Phys. Rev. Lett. – 2002. – Vol. 89, No. 23, 238701. – 4 p.
4. Burridge R. and Knopoff L. Model and Theoretical Seismicity // Bull. Seism. Soc. Am. – 1967. – Vol. 57, No. 3. – P. 341-371.

5. *Bullen K.E.* On Strain Energy in the Earth's Upper Mantle // *Trans. Am. Geophys. Union.* – 1953. – Vol. 34, No. 1. – P. 107-109.
6. *Kostrov B.V. and Das S.* Principles of Earthquake Source Mechanics. – New York: Cambridge Univ.Press, 1989.
7. *McGarr A.*, On Relating Apparent Stress to the Stress Causing Earthquake Fault Sleep // *J. Geophys.Res. Solid Earth.* – 1999. – Vol. 104, No. B2. – P. 3003-3011.
8. *Fisher M.E.* The Theory of Equilibrium Critical Phenomena // *Rep. Progr. Phys.* – 1967. – Vol. 30, No. 2. – P. 615-730.
9. *Kosobokov V.G. and Mazhkenov S.A.* On Similarity in the Spatial Distribution of Seismicity, Computational Seismology and Geodynamics: Vol. 1. D.K. Chowdhury, N.N. Biswas, A.T. Hsui, et al. (Eds.). – Washington: Am. Geophys. Union, 1994. – P. 6-15.
10. *Lomnitz-Adler J.* Automaton Models of Seismic Fracture: Constraints Imposed by the Magnitude-Frequency Relation // *J. Geophys. Res.* – 1999. – Vol. 98, No. B10. – P. 17745-17756.
11. *Черепанцев А.С.* Эффект частотной фильтрации в оценке параметров динамической системы // *Известия вузов. Прикладная нелинейная динамика.* – 2012. – Т. 20, № 6. – С. 48-55.
12. *Bramwell S.T., Christensen K., Fortin J.-Y., Holdsworth P.C., Jensen H.J., Lise S., Lopez J.M., Nicodemi M., Pinton J.-F., and Sellitto M.* Universal Fluctuations in Correlated Systems // *Phys. Rev. Lett.* – 2000. – Vol. 84, No. 17. – P. 3744-3747.
13. *Bak P.*, How Nature Works: The Science of Self-Organised Criticality. – New York: Springer, 1996.
14. *Binney J., Dowrick N., Fisher A., and Newman M.* The Theory of Critical Phenomena: An Introduction to the Renormalization Group. – Oxford: Clarendon, 1992.
15. *Bak P. and Tang C.* Earthquakes As a Phenomenon of Self-Organised Criticality // *J. Geophys. Res.* – 1989. – Vol. 94, No. B11. – P. 15635-156637.
16. *Wissel F. and Drossel B.* Transient and Stationary Behavior of the Olami – Feder –Christensen Model // *Phys. Rev. E.* – 2006. – Vol. 74, No. 6, 066109. – 9 p.
17. *Grassberger P.* Efficient Large-Scale Simulations of a Uniformly Driven System // *Phys. Rev. E.* – 1994. – Vol. 49, No. 3. – P. 2436-2444.
18. *Middleton A.A. and Tang Ch.* Self-Organized Criticality in Nonconserved Systems // *Phys. Rev. Lett.* – 1995. – Vol. 74, No. 5. – P. 742-745.
19. *Lise S.*, Self-Organisation to Criticality in a System without Conservation Law // *J. Phys. A.* – 2002. – Vol. 35, No. 22. – P. 4641-4649.
20. *Wissel F. and Drossel B.* The Olami – Feder –Christensen Earthquake Model in One Dimension // *New J. Phys.* – 2005. – Vol. 7, No. 1, Art. 5. – 19 p.
21. *Черепанцев А.С.* Временные вариации параметров динамических систем геодформационных процессов // *Физика Земли.* – 2018. – № 4S. – С. 20-38.
22. *Ставрогин А.Н., Протосеня А.Г.* Механика деформирования и разрушения горных пород, – М.: Недра, 1992. – 224 с.

REFERENCES

1. *Olami Z., Feder H.S., and Christensen K.* Self-Organized Criticality in a Continuous, Nonconservative Cellular Automaton Modeling Earthquakes, *Phys. Rev. Lett.*, 1992, Vol. 68, No. 8, pp. 1244-1247.
2. *Lise S. and Paczusi M.* Scaling in a Nonconservative Earthquake Model of Self-Organized Criticality, *Phys. Rev. E.*, 2001, Vol. 64, No. 4, 046111, 5 p.
3. *Drossel B.* Complex Scaling Behavior of Nonconserved Self-Organized Critical Systems, *Phys. Rev. Lett.*, 2002, Vol. 89, No. 23, 238701, 4 p.
4. *Burridge R. and Knopoff L.* Model and Theoretical Seismicity, *Bull. Seism. Soc. Am.*, 1967, Vol. 57, No. 3, pp. 341-371.
5. *Bullen K.E.* On Strain Energy in the Earth's Upper Mantle, *Trans. Am. Geophys. Union.*, 1953. – Vol. 34, No. 1, pp. 107-109.
6. *Kostrov B.V. and Das S.* Principles of Earthquake Source Mechanics. New York: Cambridge Univ.Press, 1989.
7. *McGarr A.*, On Relating Apparent Stress to the Stress Causing Earthquake Fault Sleep, *J. Geophys.Res. Solid Earth.*, 1999, Vol. 104, No. B2, pp. 3003-3011.

8. *Fisher M.E.* The Theory of Equilibrium Critical Phenomena, *Rep. Progr. Phys.*, 1967, Vol. 30, No. 2, pp. 615-730.
9. *Kosobokov V.G. and Mazhkenov S.A.* On Similarity in the Spatial Distribution of Seismicity, *Computational Seismology and Geodynamics: Vol. 1.* D.K. Chowdhury, N.N. Biswas, A.T. Hsui, et al. (Eds.). Washington: Am. Geophys. Union, 1994, pp. 6-15.
10. *Lomnitz-Adler J.* Automaton Models of Seismic Fracture: Constraints Imposed by the Magnitude-Frequency Relation, *J. Geophys. Res.*, 1999, Vol. 98, No. B10, pp. 17745-17756.
11. *Cherepantsev A.S.* Effekt chastotnoy fil'tratsii v otsenke parametrov dinamicheskoy sistemy [The effect of frequency filtering in the evaluation of dynamic system parameters], *Izvestiya vuzov. Prikladnaya nelineynaya dinamika [Izvestiya VUZ. Applied Nonlinear Dynamics]*, 2012, Vol. 20, No. 6, pp. 48-55.
12. *Bramwell S.T., Christensen K., Fortin J.-Y., Holdsworth P.C., Jensen H.J., Lise S., Lopez J.M., Nicodemi M., Pinton J.-F., and Sellitto M.* Universal Fluctuations in Correlated Systems, *Phys. Rev. Lett.*, 2000, Vol. 84, No. 17, pp. 3744-3747.
13. *Bak P.*, How Nature Works: The Science of Self-Organised Criticality. New York: Springer, 1996.
14. *Binney J., Dowrick N., Fisher A., and Newman M.* The Theory of Critical Phenomena: An Introduction to the Renormalization Group. Oxford: Clarendon, 1992.
15. *Bak P. and Tang C.* Earthquakes As a Phenomenon of Self-Organised Criticality, *J. Geophys. Res.*, 1989 Vol. 94, No. B11, pp. 15635-156637.
16. *Wissel F. and Drossel B.* Transient and Stationary Behavior of the Olami – Feder –Christensen Model, *Phys. Rev. E.*, 2006, Vol. 74, No. 6, 066109, 9 p.
17. *Grassberger P.* Efficient Large-Scale Simulations of a Uniformly Driven System, *Phys. Rev. E*, 1994, Vol. 49, No. 3, pp. 2436-2444.
18. *Middleton A.A. and Tang Ch.* Self-Organized Criticality in Nonconserved Systems, *Phys. Rev. Lett.*, 1995, Vol. 74, No. 5, pp. 742-745.
19. *Lise S.*, Self-Organisation to Criticality in a System without Conservation Law, *J. Phys. A*, 2002, Vol. 35, No. 22, pp. 4641-4649.
20. *Wissel F. and Drossel B.* The Olami – Feder –Christensen Earthquake Model in One Dimension, *New J. Phys.*, 2005, Vol. 7, No. 1, Art. 5, 19 p.
21. *Cherepantsev A.S.* Vremennyye variatsii parametrov dinamicheskikh sistem geodeformatsionnykh protsessov [Time variations of parameters of dynamic systems of geodeformation processes], *Fizika Zemli [Physics of the Earth]*, 2018, No. 4S, pp. 20-38.
22. *Stavrogin A.N., Protosenya A.G.* Mekhanika deformirovaniya i razrusheniya gornyykh porod [Mechanics of deformation and destruction of rocks]. Moscow: Nedra, 1992, 224 p.

Статью рекомендовал к опубликованию д.ф.-м.н., профессор Г.В. Куповых.

Черепанцев Александр Сергеевич – Южный федеральный университет; e-mail: acherepantsev@sfedu.ru; г. Таганрог, Россия; тел. 88634371663, к.ф.-м.н.; доцент.

Cherepantsev Alexandr Sergeevich – Southern Federal University; e-mail: acherepantsev@sfedu.ru; Taganrog, Russia; phone: +78634371663; cand. of phys. and math. sc.; associate professor.

В.А. Павский, К.В. Павский**ОЦЕНКА ОСУЩЕСТВИМОСТИ РЕШЕНИЯ ЗАДАЧ
НА ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ ПРИ ГРУППОВОМ
ОБСЛУЖИВАНИИ***

Рост производительности вычислительных систем (ВС) связан как с масштабируемостью, так и с развитием архитектуры вычислительных элементов системы. Кластерные ВС, которые являются масштабируемыми, составляют 93% суперкомпьютеров списка Top500 и относятся к высокопроизводительным. При этом по – прежнему остается проблема эффективного и полного использования всего имеющегося вычислительного ресурса суперкомпьютера и ВС для решения пользовательских задач. Отказы элементарных машин (узлов, вычислительных модулей) снижают технико-экономическую эффективность вычислительных систем и эффективность решения пользовательских задач. Поэтому при планировании процесса решения задач, уменьшение потерь времени на восстановление ВС от сбоев, отказов является важной задачей. Для количественной оценки потенциальных возможностей вычислительных систем используются показатели осуществимости решения задач. Эти показатели характеризуют качество работы систем с учетом надежности, временных характеристик и параметров обслуживания поступающих задач. В работе предлагается математическая модель функционирования вычислительной системы с накопителем при групповом обслуживании потока задач. Математическая модель использует методы теории массового обслуживания, основанных на теории вероятностей и системах дифференциальных уравнений. Следует заметить, что методика составления систем дифференциальных уравнений достаточно проста, если представлена соответствующая им граф-схема. Однако точное решение систем уравнений и, как правило, в элементарных функциях, не существует, либо формулы труднообозримы. Здесь решение получено в стационарном режиме функционирования системы массового обслуживания. Рассчитаны показатели, позволяющие оценить наполненность накопителя. Полученные аналитические решения просты, могут быть использованы для экспресс-анализа функционирования вычислительных систем.

Вычислительные системы; накопитель; поток задач; групповое обслуживание; показатели осуществимости решения задач.

V.A. Pavsky, K.V. Pavsky**ESTIMATION OF REALIZABILITY OF SOLVING TASKS ON COMPUTER
SYSTEMS IN GROUP MAINTENANCE**

The increase in the performance of computer systems (CS) is associated with both scalability and the development of the architecture of the computing elements of the system. Cluster CS, which are scalable, make up 93% of the Top500 supercomputers and are high-performance. At the same time, there is still the problem of efficient and complete use of all available computer resources of the supercomputer and CS for solving user tasks. Failures of elementary machines (nodes, computing modules) reduce the technical and economic efficiency of CS and the efficiency of solving user tasks. Therefore, when planning the process of solving problems, reducing the loss of time to restore CS from failures is an important problem. To quantify the potential capabilities of computer systems, indices of the realizability of solving tasks are used. These indices characterize the quality of the systems, taking into account reliability, time characteristics and service parameters of incoming tasks. The paper proposes a mathematical model of the functioning of a computer system with a buffer memory for group maintenance of a task flow. The mathematical model uses queuing theory methods based on probability theory and systems of differential equations. It should be noted that the method of composing systems of differential equations is simple

* Работа выполнена в рамках государственного задания ИФП СО РАН (ГЗ 0242-2021-0011).

enough if the corresponding graph scheme is presented. However, the exact solution of systems of equations and, as a rule, in elementary functions, does not exist, or formulas are difficult to see. Here the solution is obtained in the stationary mode of operation of the queuing system. The indices allowing to estimate the fullness of the buffer memory are calculated. The obtained analytical solutions are simple, can be used for express analysis of the functioning of computer systems.

Computer systems; buffer memory; task flow; group maintenance; indices of realizability of solving tasks.

Введение. Рост производительности вычислительных систем связан как с развитием архитектуры вычислительных элементов, так и с масштабируемостью [1, 2]. Развитие высокопроизводительных систем наглядно демонстрирует список TOP 500 [3], например, кластерные ВС составляют 93% суперкомпьютеров этого списка. Для таких систем сохраняется крайне сложная проблема эффективного использования всего имеющегося вычислительного ресурса суперкомпьютера для решения пользовательских задач. Данная проблема еще более усложняется тем обстоятельством, что при современном уровне надежности элементной базы время между отказами ЭМ (элементарные машины, узлы, вычислительные модули) вычислительной системы может измеряться часами или даже минутами [4, 5].

Отказы ЭМ снижают технико-экономическую эффективность ВС и эффективность решения задач [2, 6]. Поэтому актуальным является организация функционирования ВС и анализ надежности их потенциальных возможностей [7–14].

Для оценки потенциальных возможностей вычислительных систем используют показатели осуществимости решения задач. Эти показатели характеризуют качество работы систем с учетом надежности ВС и параметров поступающих задач [2]. В работе предлагается математическая модель для расчета показателей осуществимости решения задач, в режиме обслуживания потока задач на ВС с накопителем.

Постановка задачи. В накопитель вычислительной системы поступают задачи, из которых формируются пакеты фиксированного размера с последующей их обработкой в ВС. Требуется оценить эффективность работы накопителя (рис. 1). Считается, что накопитель может находиться в двух состояниях, либо переполненное, если число задач в нем больше критического уровня, либо нормальное – число задач меньше значения заданного уровня.

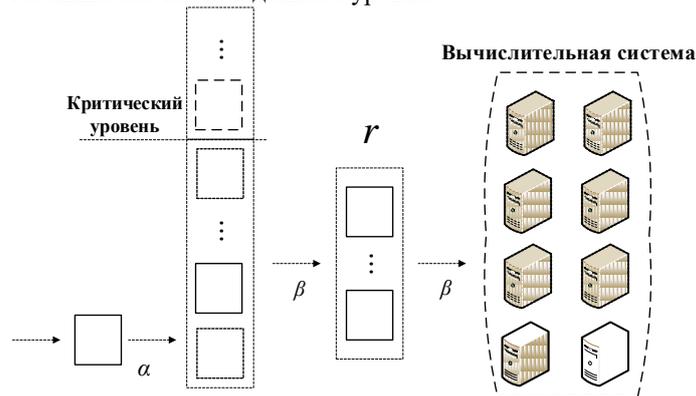


Рис. 1. Модель системы с накопителем

Математическая модель выполнена в рамках теории массового обслуживания [15–19], где под потоком требований понимаем: в первом случае – поток поступающих задач на обслуживание с известной интенсивностью, в другом случае –

поток отказов элементарных машин (ЭМ) с последующим, также известным, восстановлением. Предполагается, что случайные потоки простейшие. Параметр α – интенсивность поступающих задач, а β – интенсивность решения задач, зависящая от числа исправных ЭМ в масштабируемых ВС [1, 14].

1. Математическая модель. На систему массового обслуживания (СМО) поступает пуассоновский поток требований интенсивностью α . Требования обслуживаются группами; время обслуживания группы из r требований подчинено экспоненциальному распределению с параметром β , $\alpha < \beta$ [17]. В каждый момент времени $t \in [0, \infty)$, СМО находится в одном из множества C_k несовместных состояний (см. рис.2), k число требований в системе. Если система находится в состоянии C_k , то $l = \begin{cases} r, & \text{если } k \geq r, \\ k, & \text{если } 0 \leq k < r \end{cases}$ требований покидают систему с интенсивностью β и система переходит в состояние C_{k-l} . Если в систему приходит требование, то система переходит из состояния C_k в состояние C_{k+1} . На рис. 2 представлена граф-схема состояний СМО.

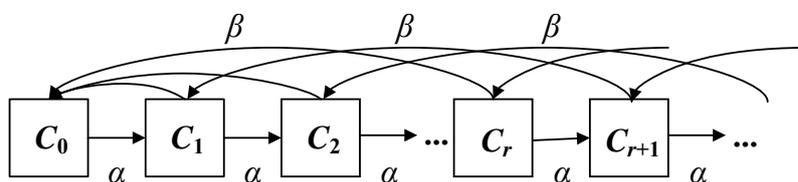


Рис. 2. Размеченный граф состояний, описывающий процесс с групповым восстановлением

Граф-схема формализуется системой дифференциальных уравнений как функций времени [15–17] для неизвестных вероятностей $P_k(t)$, того, что СМО находится в состоянии C_k в момент времени t . Имеем, систему

$$\begin{cases} \frac{dP_0(t)}{dt} = -\alpha P_0(t) + \beta(P_1(t) + P_2(t) + \dots + P_r(t)), \\ \frac{dP_k(t)}{dt} = -(\alpha + \beta)P_k(t) + \beta P_{k+r}(t) + \alpha P_{k-1}(t), \quad k \geq 1 \end{cases} \quad (1)$$

с условием нормировки $\sum_{k=1}^{\infty} p_k(t) = 1$, $t \in [0, \infty)$. Рассмотрим стационарный режим функционирования СМО, где $p_k = \lim_{t \rightarrow \infty} P_k(t)$, тогда из (1) получаем

$$\begin{cases} \alpha p_0 = \beta(p_1 + p_2 + \dots + p_r), \\ (\alpha + \beta)p_k = \beta p_{k+r} + \alpha p_{k-1}, \quad k \geq 1. \end{cases} \quad (2)$$

Для решения алгебраической системы уравнений (2) используем метод производящих функций [20], которые при $|z| \leq 1$ обычно имеют вид

$$H(z) = \sum_{k=0}^{\infty} p_k z^k. \quad (3)$$

Каждое уравнение k системы (2) умножаем на z^k , где $k=0,1,..$ соответствует номеру уравнения, и суммируем. После стандартных преобразований, получаем алгебраическое уравнение

$$(\alpha + \beta)(H(z) - p_0) = \frac{\beta}{z^r} (H(z) - \sum_{k=0}^r p_k z^k) + \alpha z H(z),$$

из которого, предварительно, приведя к общему знаменателю z^r , находим

$$H(z) = (\beta \sum_{k=0}^r p_k z^k - (\alpha + \beta)p_0 z^r) / (\alpha(z - (\alpha + \beta))z^r + \beta). \quad (4)$$

Из системы (2) и определения функции $H(z)$ (3) следует, что

$$-(\alpha + \beta)p_0 z^r = -\beta z^r \sum_{k=0}^r p_k z^k.$$

Подставляя правую часть в уравнение (4) и объединяя суммы, получим

$$H(z) = \frac{\beta \sum_{k=0}^r p_k (z^k - z^k)}{(\alpha z - (\alpha + \beta))z^r + \beta}.$$

Знаменатель дроби - многочлен степени $r+1$. Один корень $z=1$, другой $z=0$, а оставшиеся $r-1$ корней, по теореме Руше, имеют один $|z_0|>1$, а остальные $|z|<1$. Числитель и знаменатель дроби у $H(z)$ должны быть пропорциональны, поскольку производящая функция является многочленом, отсюда

$$\frac{A\beta \sum_{k=0}^r p_k (z^k - z^k)}{z - 1} = \frac{(\alpha z - (\alpha + \beta))z^r + \beta}{(z - 1)(z - z_0)}$$

где A – коэффициент пропорциональности.

Перепишем последнее выражение в виде

$$\frac{K\beta \sum_{k=0}^r p_k (z^k - z^k)}{1 - z} = \frac{(\alpha z - (\alpha + \beta))z^r + \beta}{(1 - z)(1 - z/z_0)},$$

где $K = Az_0$, отсюда $H(z) = \frac{1}{K(1 - z/z_0)}$. При $z=1$, $H(1) = 1$, следовательно, $K = 1/(1 - 1/z_0)$. Таким образом

$$H(z) = \frac{1 - 1/z_0}{1 - z/z_0}. \quad (5)$$

Из (3) имеем, что $H(0) = p_0$, тогда из (5)

$$p_0 = 1 - 1/z_0. \quad (6)$$

Последовательно дифференцируя производящую функцию (5) и полагая $z=0$, находим

$$p_k = (1 - 1/z_0)(1/z_0)^k, \quad k = 0, 1, 2, \dots, z_0 > 1.$$

Следовательно,

$$p_k = p_0(1 - p_0)^k, \quad k = 0, 1, 2, \dots \quad (7)$$

Воспользуемся первым уравнением из (2).

$$\alpha p_0 = \beta(p_1 + p_2 + \dots + p_r).$$

Рассмотрим два крайних случая этого уравнения при $r=1$ и $r \rightarrow \infty$,

$$\begin{aligned} \beta p_1 &\leq \alpha p_0 \leq \beta \sum_{k=1}^{\infty} p_k, \\ \beta p_0(1 - p_0) &\leq \alpha p_0 \leq \beta(1 - p_0). \end{aligned} \quad (8)$$

Из левой части неравенства (8) (сравнение со случаем при $r=1$)

$$\beta(1 - p_0) \leq \alpha,$$

Следовательно,

$$\frac{\beta - \alpha}{\beta} \leq p_0.$$

Из правой части неравенства (8) (т.е. имеем сравнение со случаем $r \rightarrow \infty$) следует, что

$$p_0 \leq \frac{\beta}{\alpha + \beta}.$$

Тогда, можем сформулировать следующее утверждение.

Утверждение. Для математической модели с групповым обслуживанием, представленной системой уравнений (2), оценкой для вероятности p_0 является двойное неравенство

$$\frac{\beta - \alpha}{\beta} \leq p_0 \leq \frac{\beta}{\alpha + \beta}. \quad (9)$$

Используя (2), (7) и (9) получаем, что
при $r=1$

$$p_0 = \frac{\beta - \alpha}{\beta}; \quad (10)$$

при $r=2$

$$p_0 = \frac{1}{2} \left(3 - \sqrt{1 + 4 \frac{\alpha}{\beta}} \right); \quad (11)$$

при $r \rightarrow \infty$

$$p_0 = \frac{\beta}{\alpha + \beta}. \quad (12)$$

2. Расчет моментов случайных величин. Производящая функция (3), на основании (6), может быть записана в следующем виде

$$H(z) = \frac{p_0}{1 - z(1 - p_0)} \quad (13)$$

На основании свойства производящих функций [20] для первого и второго центрального момента для случайной величины X – числа требований в системе

$$M(X) = H'(1),$$

$$D(X) = H''(1) + H'(1) - H'^2(1),$$

получаем, что производная n -го порядка для (13) равна

$$H(z)^{(n)} \Big|_{z=1} = \frac{n! p_0 (1 - p_0)^n}{(1 - z(1 - p_0))^{n+1}} \Big|_{z=1} = \frac{n! (1 - p_0)^n}{p_0^n}.$$

Откуда математическое ожидание и дисперсия равны

$$M(X) = \frac{p_0(1 - p_0)}{(1 - z(1 - p_0))^2} \Big|_{z=1} = \frac{1 - p_0}{p_0}, \quad (14)$$

$$D(X) = \frac{2(1 - p_0)^2}{p_0^2} + \frac{1 - p_0}{p_0} - \left(\frac{1 - p_0}{p_0} \right)^2 = \frac{1 - p_0}{p_0^2}. \quad (15)$$

Используя (9)-(10), легко найти значения $M(X)$ и $D(X)$ (см (14) и (15)), при $r=1, 2$ и $r \rightarrow \infty$.

Пример. Пусть интенсивности входящего потока задач $\alpha=15$ 1/ч, исходящего потока $\beta=20$ 1/ч, тогда на основании (10)-(12) и (14), (15):

при $r=1$ получаем, что средняя наполняемость накопителя равна $M(X)=3$ с соответствующей дисперсией $D(X)=12$;

при $r=2$ средняя наполняемость накопителя равна $M(X)=1$ с дисперсией $D(X)=2$;

при $r \rightarrow \infty$ средняя наполняемость накопителя равна $M(X)=3/4$ с дисперсией $D(X)=21/16$.

3. Функция распределения времени перехода накопителя из переполненного состояния в нормальное. Пусть η – случайная величина, отражающая время перехода из переполненного состояния накопителя в нормальное состояние, тогда функция $F(t, r)$ нахождения накопителя в переполненном состоянии в течение времени $t \in [0, \infty)$ при обработке требований группами, размером r , запишется в виде

$$F(t, r) = P\{\eta \geq t\}.$$

Пусть s – число задач в накопителе, начиная с которого считаем, что накопитель переполнен. Тогда, если система функционирует достаточно долго, то $P\{\eta=0\} = p_{отк}$ где $p_{отк} = 1 - \sum_{k=0}^{s-1} p_k$ постоянна для любого $t \in [0, \infty)$.

Пусть $a = 1 - p_0$, тогда на основании (7) $p_0 = (1 - a)a^k$, $p_{отк} = a^s$.

Далее,

$$P\{\eta \geq t\} = \sum_{k=s}^m p_k P_k\{\eta \geq t\},$$

где $P_k\{\eta \geq t\}$ – вероятность того, что на обработку k требований будет затрачено времени не менее t , а m – допустимое количество требований, ожидающих обработку в накопителе (в дальнейшем это число будет использовано при оценке погрешности $F(t, r)$).

Принимаем, что в стационарном режиме поток обслуживания простейший, т.е. имеем распределение Пуассона.

Вывод функции $F(t, r)$. Итак, при $k \geq s$ накопитель находится в переполненном состоянии, а обработка требований осуществляется группами по r требований, с интенсивностью β .

Имеем

$$P_k\{\eta \geq t\} = \sum_{j=0}^{\lfloor (k-s)/r \rfloor} \frac{(\beta t)^j}{j!} \exp(-\beta t).$$

где $\lfloor x \rfloor$ целая часть числа x .

$$P\{\eta \geq t\} = \sum_{k=s}^m (1-a)a^k \sum_{j=0}^{\lfloor (k-s)/r \rfloor} \frac{(\beta t)^j}{j!} \exp(-\beta t).$$

Допускаем $m \rightarrow \infty$, тогда

$$\tilde{F}(t, r) = \sum_{k=s}^{\infty} (1-a)a^k \sum_{j=0}^{\lfloor (k-s)/r \rfloor} \frac{(\beta t)^j}{j!} \exp(-\beta t),$$

Учитывая, что при $k \geq s$ (по условию $s > 0$) имеем $p_k = p_{отк} p_{k-s}$, откуда

$$\tilde{F}(t, r) = \exp(-\beta t) a^s \sum_{k=0}^{\infty} (1-a)a^k \sum_{j=0}^{\lfloor k/r \rfloor} \frac{(\beta t)^j}{j!}$$

Изменяем порядок суммирования и замечаем, что $\sum_{k=0}^{\infty} (1-a)a^k = 1$, тогда

$$\tilde{F}(t, r) = \exp(-\beta t) a^s \sum_{j=0}^{\infty} \frac{(\beta t)^j}{j!} (a^r)^j.$$

$$\tilde{F}(t, r) = a^s \exp(-\beta t) \exp(a^r \beta t),$$

$$F(t, r) \approx \tilde{F}(t, r) = p_{отк} \exp(-(1-a^r)\beta t). \quad (16)$$

Приведем погрешность $\Delta(t, m)$ функции $F(t, r) \approx \tilde{F}(t, r)$, определяемой формулой (16), и соответственно, с принятым допущением, что $m \rightarrow \infty$. Имеем

$$\Delta(t, m) = a^{m-r+1} \sum_{k=r}^{\infty} (1-a)a^k \sum_{j=0}^{\lfloor k/r \rfloor - 1} \frac{(\beta t)^j}{j!} \exp(-\beta t),$$

$$\Delta(t, m) = \sum_{k=m+1}^{\infty} (1-a)a^k \sum_{j=0}^{k-r} \frac{(\beta t)^j}{j!} \exp(-\beta t).$$

После преобразований, аналогичных выводу функции (16), при $a = 1 - p_0$, получаем

$$\Delta(t, m) = (1-p_0)^{m-r+1} F(t, r). \quad (17)$$

Погрешность $\Delta(t, m)$ определяет точность расчета функции $F(t, r) \approx \tilde{F}(t, r)$

Тогда из (15) функция распределения времени перехода накопителя из переполненного состояния в нормальное запишется в виде:

$$G(t, r) = 1 - \frac{F(t, r)}{p_{отк}} = 1 - \exp(-(1 - (1-p_0)^r)\beta t). \quad (18)$$

Полученные формулы определяют вероятность (16) нахождения накопителя в переполненном состоянии и функцию (18) – распределения времени перехода накопителя из переполненного состояния в нормальное при групповой обработке задач с погрешностью (17).

На рис. 3 представлен пример для функции распределения $G(t, r)$ (см. (18)) при $r=1, 2$ и $r \rightarrow \infty$. Из графика видно, как влияет размер группы на обработку задач.

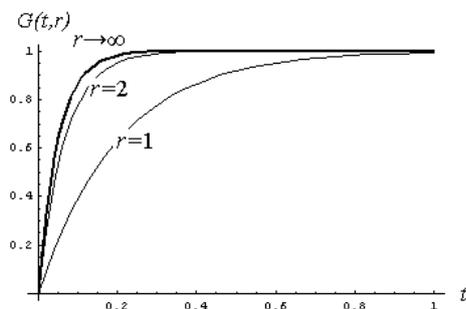


Рис. 3. Расчет функции $G(t, r)$ для группы $r=1, 2$ и $r \rightarrow \infty$ при $\alpha=15$ 1/4, $\beta=20$ 1/4

Заключение. Представлена математическая модель вычислительной системы с накопителем при групповом обслуживании потока задач. Получена вероятность нахождения накопителя в переполненном состоянии и функция распределения времени перехода накопителя из переполненного состояния в нормальное (штатное), при групповой обработке задач. Предложены формулы для расчета показателей, позволяющих оценить наполненность накопителя в среднем – математическое ожидание, дисперсия. Все формулы обладают наглядностью и простотой, удобны в экспресс анализе функционирования вычислительных систем. Для точных расчетов полученных показателей, как функций времени, требуется решить предложенную систему дифференциальных уравнений, решение которой доступно численными методами

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Dongarra J.J., A.J. van der Steen. High-performance computing systems: Status and outlook // Acta Numerica. – 2012. – P. 1-96.
2. Хорошевский В.Г. Архитектура вычислительных систем. – М.: МГТУ им. Баумана, 2008. – 520 с.
3. TOP500 Supercomputers Official Site. TOP500 Lists 2021. – <http://www.top500.org>.
4. Gupta S., Patel T., Engelmann C. and Tiwari D. Failures in large scale systems: long-term measurement, analysis, and implications SC '17: Proc. of the International Conference for High Performance Computing, Networking, Storage and Analysis (Denver, Colorado). – 2017. – Vol. 44.
5. Schroeder B. and Gibson Garth. A 2006 large-scale study of failures in high-performance computing systems // Proceedings of the International Conference on Dependable Systems and Networks (DSN2006) (Philadelphia, PA, USA). – P. 10.
6. Korobkin V., Melnik E., Klimenko A. Fault-tolerant architecture for the hazardous object information control systems // Application of information and communication technologies - AICT2015 (IEEE catalog number CFPI556H-PRT): conference proceedings (Rostov-on-Don, Russia 14-16 October 2015). – Rostov-on-Don: SFedU, 2015. – P. 274-276.
7. Хорошевский В.Г. Модели анализа и организации функционирования большемасштабных распределенных вычислительных систем // Электронное моделирование. – Киев, 2003. – Т. 25, № 6.
8. Xie M., Dai Y.S. and Poh K.L. Computing system reliability: models and analysis. – New York: Kluwer academic publishers, 2004.
9. Blischke W.R. and Murthy D.N.P. Reliability. – New York: Wiley, 2000
10. Hoyland A., Rausand M. System reliability theory. – New York: Wiley, 1994.
11. Kuo W., Zuo M.J. Optimal reliability modeling: principles and applications. – New York: Wiley, 2003.

12. *Mor Harchol-Balter*. Performance Modeling and Design of Computer Systems: Queuing Theory in Action. – Cambridge University Press, 2013.
13. *Чечельницкий А.А., Кучеренко О.В.* Стационарные характеристики параллельно функционирующих систем обслуживания с двумерным входным потоком // Сб. научных статей. – Минск, 2009. Вып. 2. – С. 262-268.
14. *Pavsky V.A., Pavsky K.V.* Stochastic models and calculations of distributed computer systems reserve size // Proc. of 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), Vladivostok, Russia, 2019. – P. 1-5.
15. *Назаров А.А., Терпугов А.Ф.* Теория массового обслуживания. – Томск: Изд-во НТЛ, 2010. – 228 с.
16. *Саати Т.Л.* Элементы теории массового обслуживания и ее приложения. – 3-е изд. – М.: Книжный дом «ЛИБРОКОМ», 2010. – 520 с.
17. *Клейнрок Л.* Теория массового обслуживания. – М.: Машиностроение, 1979. – 432 с.
18. *Боровков А.А.* Вероятностные процессы в теории массового обслуживания. – М.: Наука, 1972. – 368 с.
19. *Вентцель Е.С.* Теория случайных процессов и ее инженерные приложения. – М.: Наука, 1991. – 384с.
20. *Феллер В.* Введение в теорию вероятностей и ее приложения: в 2-х т. – Т. 1. – М.: ЛИБРОКОМ, 2010. – 528 с.

REFERENCES

1. *Dongarra J.J., A.J. van der Steen*. High-performance computing systems: Status and outlook, *Acta Numerica*, 2012, pp. 1-96.
2. *Khoroshevskiy V.G.* Arkhitektura vychislitel'nykh system [Architecture of computing systems]. Moscow: MGTU im. Bauman, 2008, 520 p.
3. TOP500 Supercomputers Official Site. TOP500 Lists 2021. Available at: <http://www.top500.org>.
4. *Gupta S., Patel T., Engelmann C. and Tiwari D.* Failures in large scale systems: long-term measurement, analysis, and implications *SC '17: Proc. of the International Conference for High Performance Computing, Networking, Storage and Analysis (Denver, Colorado), 2017*, Vol. 44.
5. *Schroeder B. and Gibson Garth*. A 2006 large-scale study of failures in high-performance computing systems, *Proceedings of the International Conference on Dependable Systems and Networks (DSN2006) (Philadelphia, PA, USA)*, pp. 10.
6. *Korobkin V., Melnik E., Klimenko A.* Fault-tolerant architecture for the hazardous object information control systems, *Application of information and communication technologies - AICT2015 (IEEE catalog number CFP1556H-PRT): conference proceedings (Rostov-on-Don, Russia 14-16 October 2015)*. Rostov-on-Don: SFedU, 2015, pp. 274-276.
7. *Khoroshevskiy V.G.* Modeli analiza i organizatsii funktsionirovaniya bol'shemasshtabnykh raspredelennykh vychislitel'nykh sistem [Models of analysis and organization of functioning of large-scale distributed computing systems], *Elektronnoe modelirovanie* [Electronic modeling]. Kiev, 2003, Vol. 25, No. 6.
8. *Xie M., Dai Y.S. and Poh K.L.* Computing system reliability: models and analysis. New York: Kluwer academic publishers, 2004.
9. *Blischke W.R. and Murthy D.N.P.* Reliability. New York: Wiley, 2000
10. *Hoyland A., Rausand M.* System reliability theory. New York: Wiley, 1994.
11. *Kuo W., Zuo M.J.* Optimal reliability modeling: principles and applications. New York: Wiley, 2003.
12. *Mor Harchol-Balter*. Performance Modeling and Design of Computer Systems: Queuing Theory in Action. Cambridge University Press, 2013.
13. *Chechel'nitskiy A.A., Kucherenko O.V.* Statsionarnye kharakteristiki parallel'no funktsioniruyushchikh sistem obsluzhivaniya s dvumernym vkhodnym potokom [Stationary characteristics of parallel functioning service systems with a two-dimensional input stream], *Sb. nauchnykh statey* [Collection of scientific articles]. Minsk, 2009. Issue 2, pp. 262-268.
14. *Pavsky V.A., Pavsky K.V.* Stochastic models and calculations of distributed computer systems reserve size, *Proc. of 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), Vladivostok, Russia, 2019*, pp. 1-5.

15. *Nazarov A.A., Terpugov A.F.* Teoriya massovogo obsluzhivaniya [Queuing theory]. Tomsk: Izd-vo NTL, 2010, 228 p.
16. *Saati T.L.* Elementy teorii massovogo obsluzhivaniya i ee prilozheniya [Elements of the theory of queuing and its applications]. 3 ed. Moscow: Knizhnyy dom «LIBROKOM», 2010, 520 p.
17. *Kleyurok L.* Teoriya massovogo obsluzhivaniya [Theory of queuing]. Moscow: Mashinostroenie, 1979, 432 p.
18. *Borovkov A.A.* Veroyatnostnye protsessy v teorii massovogo obsluzhivaniya [Probabilistic processes in the theory of queuing]. Moscow: Nauka, 1972, 368 p.
19. *Venttsel' E.S.* Teoriya sluchaynykh protsessov i ee inzhenernye prilozheniya [Theory of random processes and its engineering applications]. Moscow: Nauka, 1991, 384 p.
20. *Feller V.* Vvedenie v teoriyu veroyatnostey i ee prilozheniya [Introduction to probability theory and its applications]: In 2 vol. Vol. 1. Moscow: LIBROKOM, 2010, 528 p.

Статью рекомендовал к опубликованию д.т.н., профессор Б.М. Глинский.

Павский Валерий Алексеевич – Кемеровский государственный университет (КемГУ); e-mail: pavva46@mail.ru; г. Кемерово, Россия; тел.: +73842396832; д.т.н.; профессор; профессор кафедры общей математики и информатики

Павский Кирилл Валерьевич – Институт физики полупроводников им. А.В. Ржанова СО РАН (ИФП СО РАН); e-mail: pkv@isp.nsc.ru; г. Новосибирск, Россия; тел.: +7383332171, 3305626; д.т.н.; доцент; зав. лабораторией вычислительных систем; профессор кафедры вычислительных систем Сибирского государственного университета телекоммуникаций и информатики (СибГУТИ).

Pavsky Valery Alekseevich – Kemerovo State University; e-mail: pavva46@mail.ru; Kemerovo, Russia; phone: +73842396832; dr. of eng. sc.; professor; professor of department of general mathematics and informatics.

Pavsky Kirill Valerievich – Rzhanov Institute of Semiconductor Physics Siberian Branch of Russian Academy of Sciences; e-mail: pkv@isp.nsc.ru; Novosibirsk, Russia; phone: +7383332171, 3305626; dr. of eng. sc.; head of computer systems laboratory; professor of computer systems department, SibSUTIS.

УДК 004.056

DOI 10.18522/2311-3103-2022-4-103-112

И.Д. Русаловский, Л.К. Бабенко, О.Б. Макаревич

РАЗРАБОТКА МЕТОДОВ ГОМОМОРФНОГО ДЕЛЕНИЯ*

Рассматриваются проблемы гомоморфной криптографии. Гомоморфная криптография – одно из молодых направлений криптографии. Её отличительная особенность заключается в том, что можно обрабатывать зашифрованные данные без их предварительной расшифровки таким образом, что результат операций над зашифрованными данными эквивалентен после расшифровки результату операции над открытыми данными. Гомоморфное шифрование может эффективно применяться для реализации защищенных облачных вычислений. Для решения различных прикладных задач требуется поддержка всех математических операций, в том числе и операции деления, однако эта тема недостаточно проработана. Возможность выполнить операцию деления гомоморфно позволит расширить возможности прикладного применения гомоморфного шифрования и позволит выполнить гомоморфную реализацию многих алгоритмов. В работе рассматриваются существующие гомоморфные алгоритмы и возможность реализации операции деления в рамках этих алгоритмов. Также в работе предлагаются два метода гомоморфного деления. Первый метод основан на представлении шифротекстов в виде простых дробей и

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-37-90140.

выражении операции деления через операцию умножения. В рамках второго метода предлагается представление шифротекстов в виде массива гомоморфно зашифрованных бит, а все операции, в том числе и рассматриваемую в данной статье операцию деления, выражать через бинарные гомоморфные операции. Рассматриваются возможные подходы к реализации деления через бинарные операции и выбирается подход, наиболее подходящий для гомоморфной реализации. Выполняется анализ предложенных методов и указываются их преимущества и недостатки.

Гомоморфное шифрование; криптографическая защита; методы и алгоритмы; гомоморфное деление.

I.D. Rusalovsky, L.K. Babenko, O.B. Makarevich

DEVELOPMENT OF HOMOMORPHIC DIVISION METHODS

The article deals with the problems of homomorphic cryptography. Homomorphic cryptography is one of the young areas of cryptography. Its distinguishing feature is that it is possible to process encrypted data without decrypting it first, so that the result of operations on encrypted data is equivalent to the result of operations on open data after decryption. Homomorphic encryption can be effectively used to implement secure cloud computing. To solve various applied problems, support for all mathematical operations, including the division operation, is required, but this topic has not been sufficiently developed. The ability to perform the division operation homomorphically will expand the application possibilities of homomorphic encryption and will allow performing a homomorphic implementation of many algorithms. The paper considers the existing homomorphic algorithms and the possibility of implementing the division operation within the framework of these algorithms. The paper also proposes two methods of homomorphic division. The first method is based on the representation of ciphertexts as simple fractions and the expression of the division operation through the multiplication operation. As part of the second method, it is proposed to represent ciphertexts as an array of homomorphically encrypted bits, and all operations, including the division operation considered in this article, are implemented through binary homomorphic operations. Possible approaches to the implementation of division through binary operations are considered and an approach is chosen that is most suitable for a homomorphic implementation. The proposed methods are analyzed and their advantages and disadvantages are indicated.

Homomorphic encryption; cryptographic protection; methods and algorithms; homomorphic division.

Введение. В современном мире информационные технологии активно применяются во всех сферах жизни. В результате процесса информатизации вырос объем информации, возросли информационные потоки. В условиях бурного роста информационных технологий как никогда ранее стала актуальна проблема обеспечения информационной безопасности, в частности проблема обеспечения конфиденциальности информации. Актуальность необходимости обеспечения конфиденциальности информации обострилась с появлением и широким распространением облачных технологий. Классические криптографические средства обеспечивают необходимый уровень защиты данных при их передаче от клиента на облачный сервис по незащищенному каналу связи. Однако после передачи сервис получает неограниченный доступ к данным клиента для обработки. В этом кроется потенциальная уязвимость, так как сервис может быть оказаться недобросовестным, либо может быть подменен или скомпрометирован. Для решения этой проблемы может быть применена гомоморфная криптография [1–7]. Гомоморфное шифрование – это особый вид шифрования, позволяющий выполнять операции над зашифрованными данными и получать зашифрованный результат, соответствующий результату выполнения операции над открытыми данными. Эта особенность позволяет эффективно применять данный вид шифрования для решения любых задач

обработки данных без раскрытия самих данных, что актуально, например, для реализации защищенных облачных вычислений и защищенного поиска информации. Данные в этом случае шифруются на стороне клиента, а передаются и обрабатываются в зашифрованном виде. Ключ расшифровки при этом никуда не передается, следовательно, при соблюдении криптографической стойкости алгоритма выполнить расшифровку может только клиент.

Анализ актуальности. Для решения прикладных задач необходима поддержка как можно большего числа гомоморфных операций над целыми числами – сложение, разность, умножение и деление. Одной из часто решаемых прикладных задач является решение систем линейных алгебраических уравнений (СЛАУ). Решение СЛАУ требуется во многих задачах и алгоритмах, например, используется для нахождения некоторых неизвестных коэффициентов на основе ряда экспериментов. При решении подобной задачи для каждого эксперимента строится линейное алгебраическое уравнение, а неизвестные коэффициенты вычисляются на основе решения СЛАУ порядка n , где n – число поставленных экспериментов. Одним из наиболее популярных методов решения СЛАУ является алгоритм Гаусса. Для выполнения его гомоморфной реализации необходима поддержка следующих гомоморфных операций: сложение, разность, умножение, деление, сравнение (для исключения нулевых элементов с главной диагонали). Также операция гомоморфного деления необходима для гомоморфной реализации других алгоритмов.

На данный момент существует большое количество алгоритмов полностью гомоморфного шифрования, основанных на различных принципах [8–16]. Для ряда из них выполнены практические реализации, которые находятся в общем доступе, есть и коммерческие продукты. Однако ни в одной из найденных во время анализа реализаций не было поддержки операции деления. Таким образом, актуальным является разработка метода или алгоритма, позволяющего расширить существующие гомоморфные алгоритмы функциональностью деления.

Операция деления является обратной к операции умножения. Следовательно, чтобы реализовать гомоморфное деление гомоморфный алгоритм шифрования должен проявлять мультипликативные свойства. Рассмотрим несколько алгоритмов гомоморфного шифрования и проанализируем возможность выполнения операции деления на их базе, а также предложим методы реализации гомоморфного деления.

Алгоритмы в кольце вычетов. Одним из вариантов построения гомоморфных алгоритмов является отображение в кольцо вычетов. Примером такого алгоритма является RSA. Алгоритм RSA проявляет мультипликативный гомоморфизм, а в кольце вычетов можно найти обратный элемент. Следовательно, возможно реализовать операцию деления как умножение на обратное. Однако, операция деления во множестве действительных чисел \mathbb{R} и в кольце вычетов Z_n не во всех случаях эквивалентна. Результатом выполнения операции деления во множестве целых чисел будут частное и остаток, и частное будет округлено до целого числа и точность результата деления будет снижена. Однако для решения некоторых задач было бы достаточно деления с низкой точностью, в результате которого остаток бы полностью отбрасывался.

Рассмотрим кольцо Z_5 в качестве примера. Обратный элемент можно найти, воспользовавшись следствием из формулы Эйлера (1):

$$m^{-1} \bmod p \equiv m^{p-2} \bmod p. \quad (1)$$

Продемонстрируем несколько численных примеров. Пусть $m_1 = 4$, $m_2 = 2$, тогда: $\frac{m_1}{m_2} \bmod 5 \equiv \frac{4}{2} \bmod 5 \equiv 4 * 2^{-1} \bmod 5 \equiv 4 * 3 \bmod 5 \equiv 2 \bmod 5$

В результате получаем, что $4 / 2 = 2$, что верно. Однако, в рассматриваемой схеме операции выполняются в кольце, в то время как для прикладного использования нас интересуют целочисленные операции во множестве целых чисел. В качестве наглядного примера разделим 2 на 4:

$$\frac{m_1}{m_2} \bmod 5 \equiv \frac{2}{4} \bmod 5 \equiv 2 * 4^{-1} \bmod 5 \equiv 2 * 4 \bmod 5 \equiv 3 \bmod 5.$$

Как видно из примера, $2 / 4 = 3$ в кольце Z_5 , в то время как мы ожидали получить 0 или 1, в зависимости от выбранной стратегии округления результата. Следовательно, данное решение не подходит для реализации гомоморфного деления.

Алгоритмы на основе полиномов. Еще один вариант реализации гомоморфного алгоритма шифрования – соотнесение открытому тексту некоторого полинома. К примеру, Ф. Буртыка предложил алгоритм шифрования на основе матричных полиномов (полиномов, каждый коэффициент которых представлен матрицей) [17]. Также в выпускной квалификационной работе Яковлева [18] предлагается алгоритм шифрования посредством преобразования целого числа полиному с целочисленными коэффициентами. Между двумя полиномами можно выполнить операцию деления, результатом которой будут частное и остаток от деления. Как было указано в предыдущем примере, для решения некоторых задач может хватить точности деления, при которой остаток полностью отбрасывается. Однако в случае с шифротекстами на основе полиномов возникает ряд проблем.

Величина открытого текста никак не связана с порядком полинома, следовательно большему числу может соответствовать меньший полином и наоборот, а делимое полностью будет остатком.

Результатом деления будут частное и остаток, каждый из которых представлен полиномом. Если расшифровать их, разделить расшифрованный остаток на расшифрованный делитель, то мы получим корректный результат. Но частное и остаток в зашифрованном виде не соответствуют частному и остатку в расшифрованном виде. Таким образом, остаток от деления в зашифрованном виде может содержать большую часть частного, что делает операцию отбрасывания остатка некорректной, но не отбросить остаток мы не можем, так как в рамках алгоритма могут обрабатываться только полиномы.

Рассмотрим численный пример на основе алгоритма Яковлева. Пусть даны целые числа $m_1 = 4$, $m_2 = 1$, $p = 4$, $q = 2$, $x_0 = p / q = 2$ – секретный ключ. Выполним шифрование:

$$f_1(x) = 5x + 2; f_1(x_0) = 12$$

$$g_1(x) = 22 * f_1(x) - 22 * 12 + m_1 = 20x + 8 - 48 + 4 = 20x - 36$$

$$f_2(x) = 3x - 5; f_2(x_0) = 1$$

$$g_2(x) = 22 * f_2(x) - 22 * 1 + m_2 = 12x - 20 - 4 + 1 = 12x - 23$$

В результате деления получим $20x - 36$ на $12x - 23$ получим $g_3(x) = 1$, остаток $g_4(x) = 8x - 13$. После расшифрования получим:

$$D(g_3(x)) = g_3(x_0) = 1$$

$$D(g_4(x)) = g_4(x_0) = 8 * 2 - 13 = 3$$

Таким образом в результате деления получаем $1+3=4$, однако на остаток пришлось 3, из-за чего мы не можем отбросить остаток от деления, а следовательно, продолжать вычисления без перезашифрования результата.

Метод гомоморфного деления на основе представления шифротекстов в виде простых дробей. Данный метод [19] позволяет реализовать гомоморфное деление на основе любого полностью гомоморфного алгоритма шифрования над целыми числами. Метод хорошо подходит для программной реализации и на его

основе может быть построен программный комплект, позволяющий реализовать математический аппарат для вычислений в рамках гомоморфной криптографии. С помощью предлагаемого метода можно шифровать не только целые, но и рациональные числа Q . В случае, если шифруется рациональное число, первый пункт можно пропустить.

Пусть дан некоторый полностью гомоморфный алгоритм шифрования над целыми, для которого определены $E(m)$ – алгоритм шифрования, $D(c)$ – алгоритм расшифрования, обратный к $E(m)$, \otimes, \oplus – операторы гомоморфного умножения и сложения над зашифрованными данными соответственно, где m – открытый текст, c – шифротекст. Тогда схема шифрования целого числа m с поддержкой операции деления может быть построена следующим образом:

1. Представляем шифруемый открытый текст в виде рационального числа: $m = \frac{m_1}{m_2}$, если число m – целое, то $m_2 = 1$.

2. Шифруем делимое и делитель полученного рационального числа: $a_1 = E(m_1)$, $b_2 = E(m_2)$.

3. Шифротекст в предлагаемой схеме шифрования будет представлен в виде пары шифротекстов: $C = (a_1; b_2)$.

Алгоритм расшифрования:

1. Расшифруем гомоморфно зашифрованные делимое и делитель, в виде которых представлен шифротекст: $m_1 = D(a_1)$; $m_2 = D(b_2)$

2. Результат расшифрования будет равен рациональному числу $m = \frac{m_1}{m_2}$. Полученный результат можно при необходимости округлить до целого.

Реализация математических операций:

1. Сложение. $C_1 + C_2 = (a_1 \otimes b_2 \oplus a_2 \otimes b_1; b_1 \otimes b_2)$.

2. Умножение. $C_1 * C_2 = (a_1 \otimes a_2; b_1 \otimes b_2)$.

3. Деление. $C_1 / C_2 = (a_1 \otimes b_2; b_1 \otimes a_2)$.

К преимуществам данного подхода можно отнести возможность реализации операции гомоморфного деления на основе любого полностью гомоморфного алгоритма шифрования над целыми числами. Сама операция деления выполняется только над конечным результатом единой операции на шифрующей стороне (на стороне клиента). Поэтому при гомоморфных вычислениях отсутствует округление промежуточных результатов, поэтому данный метод обеспечивает высокую точность вычислений. Конечно, предлагаемый подход имеет негативные стороны – увеличение объема шифротекста и сложности вычислений. Размер шифротекста увеличивается вдвое, так как вместо одного гомоморфно зашифрованного числа мы храним пару. Сложность вычислений тоже возрастает – сложность операции умножения возрастает в два раза, сложность операции деления эквивалентна сложности операции умножения, а сложность операции сложения увеличивается приблизительно в четыре раза. Однако возможность выполнения операции деления сильно расширяет возможности прикладного применения гомоморфного шифрования, что нивелирует снижение скорости вычислений.

Метод гомоморфного деления на основе битовых операций. Гомоморфное деление возможно реализовать через битовые операции на основе некоторого полностью гомоморфного алгоритма шифрования над битами. Для этого рассмотрим возможные подходы к выполнению операции деления над числами, представленными в двоичном коде. Как правило операция деления строится на последовательных операциях разности и сдвигов и возможны два варианта:

- ◆ деление со сдвигом делителя вправо;
- ◆ деление со сдвигом делимого влево.

Рассмотрим каждый из подходов и проанализируем, какой из них лучше подходит для гомоморфной реализации.

Алгоритм деления со сдвигом делителя вправо:

1. Проверка делителя на ноль. Это необходимо для корректной работы алгоритма, иначе выполнить шаг 2 будет невозможно. В случае, если делитель равен нулю, алгоритм завершается ошибкой.

2. Нормализация делителя. Для выполнения деления необходимо, чтобы старший разряд делителя был значащим, то есть содержал единицу. Для достижения этого мы предварительно сдвигаем делитель влево необходимое число раз, а количество сдвигов k запоминаем. Шаги 3-5 мы будем повторять $k+1$ число раз.

3. Вычитаем делитель из делимого и получаем текущий остаток. На основе комбинации знаков делимого, делителя и текущего остатка формируем очередной разряд частного. Также на основе комбинации знаков либо восстанавливается предыдущий остаток, либо используется текущий.

4. Сдвиг делителя на один разряд вправо

5. Уменьшаем значение счетчика итераций на 1 и проверяем его значение. Если он стал равен нулю, то завершаем процесс деления, иначе возвращаемся к шагу 3.

Плюсом данного подхода является оптимизация процесса деления за счет нормализации делителя и уменьшения числа итераций. Но это является огромным недостатком при построении гомоморфной реализации данного алгоритма, так как итерации выполняются внутри внешней системы, а информация о количестве сдвигов зашифрована гомоморфно. Из-за этого мы не можем получить информацию о числе шагов и придется выполнять максимально возможное число шагов гомоморфно, что значительно усложнит реализацию алгоритма. Поэтому рассмотрим вариант деления со сдвигом делимого.

Алгоритм деления со сдвигом делимого влево:

1. Формируем начальный остаток

2. Сдвигаем содержимое делимого влево, вдвигая выдвигаемый старший разряд в остаток.

3. Вычитаем делитель из остатка. Фактически, это вычитание из старших разрядов делимого, которые мы постепенно перемещаем на место остатка. Как и в первом варианте алгоритма, в зависимости от знака результата формируются биты частного и при необходимости выполняется восстановление остатка.

4. Повторяем шаги 2-3, пока не будут сформированы $n+1$ битов частного (где n – разрядность частного)

5. Прибавляем к частному единицу для округления результата.

Данный подход лучше подходит для гомоморфной реализации, так как не требует нормализации и выполняется всегда за фиксированное число шагов. Также алгоритм может обработать ноль в качестве делителя, однако в этом случае в частном будет некоторое ошибочное значение. Поэтому проверка на ноль и генерация ошибки необходима в любом случае. На основе данного подхода предложим метод гомоморфного деления на основе гомоморфных бинарных операций.

Пусть дан некоторый полностью гомоморфный алгоритм шифрования над битами, для которого определены $E(m)$ – алгоритм шифрования, $D(c)$ – алгоритм расшифрования, обратный к $E(m)$, \otimes, \oplus – гомоморфные логические операции, определённые в данном алгоритме шифрования, где m – открытый текст, c – шифротекст.

В общем виде алгоритм шифрования можно представить следующим образом:

1. Целое число m представляем побитно: $m = m_1m_2\dots m_n$.

2. Каждый бит шифруется отдельно: $C = E(m_1)E(m_2)\dots E(m_n)$.

Алгоритм битового деления будет аналогичен алгоритму машинного деления, с той лишь разницей, что обрабатываются гомоморфно зашифрованные биты, из-за чего будут некоторые особенности реализации. Также может возникнуть выразить недостающие логические операции через те, которые поддерживаются гомоморфной схемой шифрования. Как было рассмотрено в статье [20], операцию гомоморфного сравнения возможно реализовать на основе любого полностью гомоморфного алгоритма шифрования над битами, поэтому реализация алгоритма деления возможна. Операции суммы, разности и умножения в предлагаемой криптосистеме будут рассмотрены в другой статье.

Для облегчения выполнения операции разности в данной криптосистеме числа представляются в дополнительном коде. Алгоритм гомоморфного деления целых чисел, представленных в виде массива гомоморфно зашифрованных битов, можно представить следующим образом:

1. Нормализация делимого и делителя. В случае, если делитель и делимое имеют разную разрядность, нормализуем их разрядность к одной величине – наибольшей разрядности среди них. В шифротекст с меньшей разрядностью для нормализации необходимо продублировать самый левый разряд (знаковый) необходимое число раз.

2. Определяем разрядность частного. Минимально допустимое значение модуля делителя равно 1, поэтому числитель по модулю не может быть больше делимого. Следовательно, разрядность частного будет равна разрядности делимого.

3. Формируем начальный остаток, заполняя его знаковым битом делителя.

4. Сдвигаем содержимое делимого влево, игнорируя знаковый бит. Выдвинутый бит вдвигаем справа в промежуточный остаток.

5. Прибавляем модуль делителя к остатку, взяв его с противоположным знаком. Фактически, это вычитание из старших разрядов делимого, которые мы постепенно перемещаем на место остатка. Если знак полученного остатка совпадает со знаком делителя, то значение следующего бита частного устанавливаем равным 1, иначе 0. Если знак полученного остатка совпадает со знаком делимого, то используем этот остаток на следующей итерации, иначе выполняем операцию «восстановления» остатка – используем предыдущее значение остатка.

6. Если сформированы $n+1$ битов частного, где n – разрядность частного, то переходим к следующему шагу. Иначе возвращаемся к шагу 4.

7. Прибавляем к частному единицу округления. В случае, если мы получаем отрицательный результат, начальное заполнение частного будет в обратном коде. Поэтому прибавление единицы автоматически округляет результат и переводит его в дополнительный код, в случае отрицательного результата.

8. Отбрасываем $(n+1)$ -й бит частного. Полученное частное является результатом деления.

Криптосистема на основе битовых операций обладает рядом преимуществ. Самое главное заключается в том, что в рамках данной криптосистемы можно реализовать практически любой алгоритм гомоморфно, так как поддерживаются все арифметические и логические операции. Однако сложность операций высока ввиду использования гомоморфного шифрования над битами, и размеры шифротекстов сильно увеличиваются. Также важно учитывать, что гомоморфное деление в рамках данной криптосистемы будет выполняться с округлением.

Заключение. В рамках данной статьи рассмотрена проблема гомоморфного деления целых чисел, предложены и описаны методы реализации гомоморфного деления. Практическая ценность работы состоит в решении одной из проблем гомоморфного шифрования – реализация гомоморфного деления, что позволяет расширить область практического применения гомоморфного шифрования. Метод

гомоморфного деления может использоваться в ряде алгоритмов, например, для нахождения среднего арифметического нескольких чисел или решения СЛАУ методом Гаусса.

Предложенные методы обладают как плюсами, так и минусами, поэтому выбор метода должен опираться на общую постановку задачи, которая должна быть решена с применением гомоморфной криптографии. В дальнейшем планируется рассмотреть возможности повышения эффективности предложенных методов, а также рассмотреть возможность обработки чисел с фиксированной или плавающей запятой.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Батура Т.В., Мурзин Ф.А., Семич Д.Ф. Облачные технологии: основные понятия, задачи и тенденции развития // Программные продукты и системы. – 2014. – № 3. – С. 64-72.
2. Жиров А.О., Жирова О.В., Кренделев С.Ф. Безопасные облачные вычисления с помощью гомоморфной криптографии // Безопасность информационных технологий. – 2013. – Т. 1. – С. 6-12.
3. Беккер М.Я., Гатчин Ю.А., Кармановский Н.С., Терентьев А.О., Федоров Д.Ю. Информационная безопасность при облачных вычислениях: проблемы и перспективы // Научно-технический вестник информационных технологий, механики и оптики. – 2011. – С. 97-102.
4. Денисов Д.В. Перспективы развития облачных вычислений // Прикладная информатика. – 2009. – № 5 (23). – С. 52-58.
5. Ковалев Д. Информационная безопасность облачных вычислений // T-Comm. – 2011. – № S1. – С. 14-16.
6. Трубей А.И. Гомоморфное шифрование: безопасность облачных вычислений и другие приложения (обзор) // Информатика. – 2015. – № 1 (45). – С. 90-101.
7. Бабенко Л.К., Буртыка Ф.Б., Макаревич О.Б., Трепачева А.В. Защищенные вычисления и гомоморфное шифрование. // III Национальный суперкомпьютерный форум (25-27 ноября 2014, г. Переславль-Залесский). – ИПС им. А.К. Айламазяна РАН, 2014.
8. Gentry C. A Fully homomorphic encryption using ideal lattices // Symposium on the Theory of Computing (STOC). – Bethesda, USA, 2009. – P. 169-178.
9. Gentry C., Sahai A., Waters B. Homomorphic Encryption from Learning With Errors: Conceptually-Simpler, Asymptotically-Faster, Attribute-Based // Advances in cryptology – CRYPTO-2013, 33rd Annual Cryptology Conf. – Santa Barbara, CA, USA, 2013. – Part 1. – P. 73-93.
10. Parmar P.V. Survey of various homomorphic encryption algorithms and schemes // Intern. J. of Computer Applications. – 2014. – Vol. 91, No. 8. – P. 26-32.
11. Jain N., Pal S.K., Upadhyay D.K. Implementation and analysis of homomorphic encryption schemes // Intern. J. on Cryptography and Information Security (IJCIS). – 2012. – Vol. 2, No. 2. – P. 27-44.
12. Smart N., Vercauteren F. Fully Homomorphic Encryption with Relatively Small Key and Ciphertext Sizes // Public Key Cryptography – PKC 2010: 13th International Conference on Practice and Theory in Public Key Cryptography, Paris, France, May 26-28, 2010, Proceedings / P.Q. Nguyen, D. Pointcheval. – Berlin, Heidelberg, New York, NY, London [etc.]: Springer Science+Business Media, 2010. – P. 420-443.
13. Smart N., Vercauteren F. Fully homomorphic SIMD operations // Des. Codes Cryptogr. – Springer US, Springer Science+Business Media, 2014. – Vol. 71, Iss. 1. – P. 57-81.
14. Gentry C., Halevi S. Implementing Gentry's Fully-Homomorphic Encryption Scheme // Advances in Cryptology – EUROCRYPT 2011: 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011, Proceedings / K.G. Paterson. – Springer Science+Business Media, 2011. – P. 129-148.
15. Dijk M. v., Gentry C., Halevi S., Vaikuntanathan V. Fully Homomorphic Encryption over the Integers // Advances in Cryptology – EUROCRYPT 2010: 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, French Riviera, May 30 - June 3, 2010. Proceedings / H. Gilbert. – Berlin: Springer Berlin Heidelberg, 2010. – P. 24-43.
16. Coron J., Mandal A., Naccache D., Tibouchi M. Fully Homomorphic Encryption over the Integers with Shorter Public Keys // Advances in Cryptology – CRYPTO 2011: 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011, Proceedings / P. Rogaway. – Springer Science+Business Media, 2011. – P. 487-504.

17. Буртыка Ф.Б. Пакетное симметричное полностью гомоморфное шифрование на основе матричных полиномов // Тр. Института системного программирования РАН. – 2014. – Т. 26. № 5. – С. 99-116.
18. Яковлев М.О. Защищенный калькулятор. Разработка клиентского компонента // Выпускная квалификационная работа бакалавра. – URL: http://www.nsu.ru/xmlui/bitstream/handle/nsu/471/Text_YakovlevMO.pdf (дата обращения 20.09.2022).
19. Бабенко Л.К., Русаловский И.Д. Метод реализации гомоморфного деления // Известия ЮФУ. Технические науки. – 2020. – № 4 (214). – С. 212-221.
20. Бабенко Л.К., Русаловский И.Д. Масштабирование цифровых изображений с применением гомоморфного шифрования // Вопросы кибербезопасности. – 2021. – № 3(43). – С. 2-10.

REFERENCES

1. Batura T., F.A. Murzin, D.F. Semich Oblachnye tekhnologii: osnovnye ponyatiya, zadachi i tendentsii razvitiya [Cloud technologies: basic concepts, tasks and development trends], *Programmnye produkty i sistemy* [Software products and systems], 2014, No. 3, pp. 64-72.
2. Zhiron A.O., Zhirova O.V., Krendelev S.F. Bezopasnye oblachnye vychisleniya s pomoshch'yu gomomorfnoy kriptografii [Secure Cloud Computing with Homomorphic Cryptography], *Bezopasnost' informacionnykh tekhnologiy* [Information technology security], 2013, No. 1, pp. 6-12.
3. Bekker M., Gatchin Yu., Karmanovskiy N., Terent'ev A., Fedorov D. Informatsionnaya bezopasnost' pri oblachnykh vychisleniyakh: problemy i perspektivy [Information security in cloud computing: problems and prospects], *Nauchno-tekhnicheskiiy vestnik informatsionnykh tekhnologii, mekhaniki i optiki* [Scientific and technical bulletin of information technologies, mechanics and optics], 2011, pp. 97-102.
4. Denisov D. Perspektivy razvitiya oblachnykh vychisleniy [Prospects for the development of cloud computing], *Prikladnaya informatika* [Applied Informatics], 2009, No. 5 (23), pp. 52-58.
5. Kovalev D. Informatsionnaya bezopasnost' oblachnykh vychisleniy [Information security of cloud computing], *T-Comm*, 2011, No. S1, pp. 14-16.
6. Trubey A. Gomomorfnoe shifrovaniye: bezopasnost' oblachnykh vychisleniy i drugie prilozheniya (obzor) [Homomorphic Encryption: Cloud Computing Security and Other Applications (Review)], *Informatika* [Informatics], 2015, No. 1 (45), pp. 90-101.
7. Babenko L., Burtyka Ph., Makarevich O., Trepacheva A. Zawiennyye vychisleniya i gomomorfnoe shifrovaniye [Secure computing and homomorphic encryption], *III Natsional'nyy superkomp'yuternyy forum (25-27 noyabrya 2014, g. Pereslavl'-Zalesskiy)* [III National Supercomputer Forum (November 25-27, 2014, Pereslavl'-Zalessky)]. IPS im. A.K. Aylamazyan RAN, 2014.
8. Gentry C. A Fully homomorphic encryption using ideal lattices, *Symposium on the Theory of Computing (STOC)*. Bethesda, USA, 2009, pp. 169-178.
9. Gentry C., Sahai A., Waters B. Homomorphic Encryption from Learning With Errors: Conceptually-Simpler, Asymptotically-Faster, Attribute-Based // *Advances in cryptology – CRYPTO-2013, 33rd Annual Cryptology Conf.* Santa Barbara, CA, USA, 2013. Part 1, pp. 73-93.
10. Parmar P.V. Survey of various homomorphic encryption algorithms and schemes, *Intern. J. of Computer Applications*, 2014, Vol. 91, No. 8, pp. 26-32.
11. Jain N., Pal S.K., Upadhyay D.K. Implementation and analysis of homomorphic encryption schemes, *Intern. J. on Cryptography and Information Security (IJCIS)*, 2012, Vol. 2, No. 2, pp. 27-44.
12. Smart N., Vercauteren F. Fully Homomorphic Encryption with Relatively Small Key and Ciphertext Sizes, *Public Key Cryptography – PKC 2010: 13th International Conference on Practice and Theory in Public Key Cryptography, Paris, France, May 26-28, 2010, Proceedings* / P.Q. Nguyen, D. Pointcheval. Berlin, Heidelberg, New York, NY, London [etc.]: Springer Science+Business Media, 2010, pp. 420-443.
13. Smart N., Vercauteren F. Fully homomorphic SIMD operations, *Des. Codes Cryptogr.* Springer US, Springer Science+Business Media, 2014, Vol. 71, Iss. 1, pp. 57-81.
14. Gentry C., Halevi S. Implementing Gentry's Fully-Homomorphic Encryption Scheme, *Advances in Cryptology – EUROCRYPT 2011: 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011, Proceedings*, K.G. Paterson. Springer Science+Business Media, 2011, pp. 129-148.

15. Dijk M. v., Gentry C., Halevi S., Vaikuntanathan V. Fully Homomorphic Encryption over the Integers, *Advances in Cryptology – EUROCRYPT 2010: 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, French Riviera, May 30 - June 3, 2010. Proceedings*, H. Gilbert. – Berlin: Springer Berlin Heidelberg, 2010, pp. 24-43.
16. Coron J., Mandal A., Naccache D., Tibouchi M. Fully Homomorphic Encryption over the Integers with Shorter Public Keys, *Advances in Cryptology – CRYPTO 2011: 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011, Proceedings*, P. Rogaway. Springer Science+Business Media, 2011, pp. 487-504.
17. Burtyka F.B. Paketnoe simmetrichnoe polnost'yu gomomorfnoe shifrovaniye na osnove matrichnykh polinomov [Batch symmetric fully homomorphic encryption based on matrix polynomials], *Tr. Instituta sistemnogo programmirovaniya RAN [Proceedings of the Institute for System Programming RAS]*, 2014, Vol. 26. No. 5, pp. 99-116.
18. Yakovlev M.O. Zashchishchenny kal'kulyator. Razrabotka klientskogo komponenta [Secure calculator. Development of the client component], *Vypusknaya kvalifikatsionnaya rabota bakalavra [Bachelor's final qualification work]*. Available at: http://www.nsu.ru/xmlui/bitstream/handle/nsu/471/Text_YakovlevMO.pdf (accessed 20 September 2022).
19. Babenko L.K., Rusalovskiy I.D. Metod realizatsii gomomorfno go deleniya [Homomorphic division implementation method], *Izvestiya YuFU. Tekhnicheskie nauki [Izvestiya SFedU. Engineering Sciences]*, 2020, No. 4 (214), pp. 212-221.
20. Babenko L.K., Rusalovskiy I.D. Masshtabirovaniye tsifrovyykh izobrazheniy s primeneniem gomomorfno go shifrovaniya [Digital Image Scaling Using Homomorphic Encryption], *Voprosy kiberbezopasnosti [Cybersecurity issues]*, 2021, No. 3 (43), pp. 2-10.

Статью рекомендовал к опубликованию д.т.н., профессор И.А. Калмыков.

Русаловский Илья Дмитриевич – Южный федеральный университет; e-mail: ilya.rusalovskiy@mail.ru; г. Таганрог, Россия; тел.: +79885526701; кафедра безопасности информационных технологий; аспирант.

Бабенко Людмила Климентьевна – e-mail: blk@tsure.ru; тел.: +79054530191; кафедра безопасности информационных технологий; кафедра безопасности информационных технологий; д.т.н.; профессор.

Макаревич Олег Борисович – e-mail: mak@sfedu.ru; тел.: +78634312018; кафедра безопасности информационных технологий; д.т.н.; профессор.

Rusalovsky Ilya Dmitrievich – Southern Federal University; e-mail: ilya.rusalovskiy@mail.ru; Taganrog, Russia; phone: +79885526701; the department of information technologies security; postgraduate student.

Babenko Lyudmila Kliment'evna – e-mail: blk@tsure.ru; phone: +79054530191; the department of information technologies security; dr. of eng. sc.; professor.

Makarevich Oleg Borisovich – e-mail: mak@sfedu.ru; phone: +78634312018; the department of information technologies security; dr. of eng. sc.; professor.

Раздел II. Алгоритмы обработки информации

УДК 004.896

DOI 10.18522/2311-3103-2022-4-113-121

О.Б. Лебедев, А.А. Жиглатый

БИОИНСПИРИРОВАННЫЙ АЛГОРИТМ РЕШЕНИЯ ИНВАРИАНТНЫХ ГРАФОВЫХ ЗАДАЧ*

Предлагается биоинспирированный метод решения набора инвариантных комбинаторно-логических задач на графах: формирования паросочетания графа, выделения внутренне-устойчивого множества вершин, выделения клики графа. Описывается модифицированная парадигма муравьиной колонии использующая, в отличие от канонического метода, механизмы формирования решений на модели пространства поиска в виде звездного графа. Задача формирования в графе внутренне-устойчивого множества вершин может быть сформулирована, как задача разбиения. На начальном этапе на всех ребрах звездного графа H откладывается одинаковое (небольшое) количество феромона ζ/m , где $m=|E|$. Процесс поиска решений итерационный. Каждая итерация l включает три этапа. Агенты обладают памятью. На каждом шаге t в памяти агента a_k имеется количество феромона $\phi_j(t)$, отложенного на каждом ребре графа H . На первом этапе каждый агент a_k популяции конструктивным алгоритмом находит решение U^r_{ok} , рассчитывает оценку решения $\xi_k(U^r_{ok})$ и значение степени пригодности полученного агентом решения ϕ_k (количество феромона, соответствующее оценке). На втором этапе, после полного формирования всеми агентами решений на текущей итерации, феромон ω_j , накопленный в j -ой ячейке в буферном массиве $KЭП_b$, добавляется в каждую j -ю ячейку основного массива $Q_2=\{q_j|j=1,2,\dots,m\}$ коллективной эволюционной памяти $KЭП_e$. На третьем этапе происходит общее испарение феромона на множестве ребер E звездного графа H . Временная сложность алгоритма, полученная экспериментальным путем, совпадает с теоретическими исследованиями и для рассмотренных тестовых задач составляет $O(n^2)$.

Паросочетание; внутренне-устойчивое множество; клика; адаптация; модификация; муравьиная колония; модели пространства поиска; звездный граф.

O.B. Lebedev, A.A. Zhiglatiy

BIOINSPIRED ALGORITHM FOR SOLVING INVARIANT GRAPH PROBLEMS

A bioinspired method for solving a set of invariant combinatorial-logical problems on graphs is proposed: the formation of a graph matching, the selection of an internally stable set of vertices, and the selection of a graph clique. A modified paradigm of the ant colony is described, which uses, in contrast to the canonical method, the mechanisms for generating solutions on the search space model in the form of a star graph. The problem of forming an internally stable set of vertices in a graph can be formulated as a partitioning problem. At the initial stage, the same (small) amount of pheromone ζ/m , where $m=|E|$, is deposited on all edges of the star graph H . The process of finding solutions is iterative. Each iteration l includes three stages. Agents have memory. At each step t , the memory of the agent a_k contains the amount of pheromone $\phi_j(t)$ deposited on each edge of the graph H . At the first stage, each agent a_k of the population uses a constructive algorithm to find the solution U^r_{ok} , calculates the estimate of the solution $\xi_k(U^r_{ok})$ and the value of the degree of suitability of the solution obtained by the agent ϕ_k (the amount of phero-

* Работа выполнена при финансовой поддержке гранта РФФИ № 20-07-00260 А.

mone corresponding to the estimate). At the second stage, after the complete formation of solutions by all agents at the current iteration, the pheromone ω_j accumulated in the j -th cell in the CEP_σ buffer array is added to each j -th cell of the main array $Q_2=\{q_j|j=1,2,\dots,m\}$ of the CEP_0 collective evolutionary memory. At the third stage, the general evaporation of the pheromone occurs on the set of edges E of the star graph H . The time complexity of the algorithm, obtained experimentally, coincides with theoretical studies and for the considered test problems is $O(n^2)$.

Matching; internally stable set; clique; adaptation; modification; ant colony; search space models; star graph.

Введение. Задача определения паросочетаний – одна из фундаментальных задач в области математической оптимизации. Подмножество ребер $M \subseteq U$ графа $G=(V,U)$ называется паросочетанием в графе G , если каждой вершине $v_i \in V$ инцидентно не более одного ребра из M . **Максимальное паросочетание** – это такое паросочетание M в графе G , которое не содержится ни в каком другом паросочетании этого графа, то есть к нему невозможно добавить ни одно ребро, которое бы являлось несмежным ко всем ребрам паросочетания [1]. Наибольшее паросочетание (или максимальное по размеру паросочетание) [2] – это такое паросочетание, которое содержит максимальное количество рёбер. **Совершенным паросочетанием** называется паросочетание, в котором участвуют все вершины графа.

Алгоритмы построения паросочетания применяются в задачах о назначении, планировании и размещении при проектировании топологии СБИС, покрытия, транспортных задачах, при проектировании инженерных сетей, коммуникаций, построения систем поддержки принятия решений в неопределённых условиях и т.п. Задачи такого типа относятся к переборным задачам с экспоненциальной временной сложностью. В этой связи разрабатывают различные эвристики для построения алгоритмов с полиномиальной временной сложностью. Существуют алгоритмы определения паросочетаний в графе, основанные на использовании потоков в сетях [1, 2], имитационного моделирования [3], генетического поиска [4] и других эвристик, которые обеспечивают приемлемые результаты при решении задач малой и средней сложности. Часто эта процедура используется в итерационных структурах. Это предъявляет повышенные требования к качеству и времени решения задачи нахождения максимального паросочетания. Возникшие потребности в решении задач большой и очень большой размерности является побудительным мотивом исследований и разработок новых эффективных алгоритмов. Анализ литературы показывает, что наиболее успешными в этих условиях являются математические методы, в которых заложены принципы природных механизмов принятия решений [4–6]. К таким методам можно отнести, прежде всего, методы моделирования отжига [7], методы эволюционного моделирования и генетические алгоритмы [8], эволюционной адаптации [9], алгоритмы роевого интеллекта [10] и муравьиные алгоритмы (Ant Colony Optimization – ACO) [11, 12]. Идея муравьиного алгоритма – моделирование поведения муравьёв, связанного с их способностью быстро находить кратчайший путь от муравейника к источнику пищи. Основу поведения муравьиной колонии составляет самоорганизация, обеспечивающая достижения общих целей колонии на основе низкоуровневого взаимодействия, благодаря которому, в целом, колония представляет собой разумную многоагентную систему.

Для повышения эффективности, усиления сходимости алгоритма и способности выхода из локальных оптимумов, предложен подход к построению алгоритма паросочетания на основе модели адаптивного поведения муравьиной колонии, **отличающийся** от канонического тем, что в качестве модели пространства поиска решений используется звездный граф.

Предложенный подход обеспечивает более широкий обзор пространства решений и более высокую вероятность локализации глобального экстремума задачи.

В работе излагается методика решения задачи нахождения максимального паросочетания в графе, и родственных ей задач раскраски графа и выделения клик в графе, основанная на моделировании адаптивного поведения муравьиной колонии.

1. Постановка задачи нахождения паросочетания в графе. Пусть дан двудольный граф $G=(V,U)$ (рис. 1). $V=\{v_i/i=1,2,\dots,n\}$. $U=\{u_j/j=1,2,\dots,m\}$. На рис. 1 $n=10, m=11$.

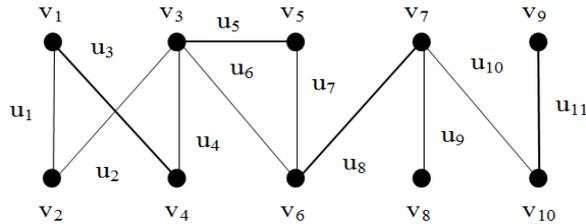


Рис. 1. Пример графа

Паросочетание P такого графа определяется как множество попарно несмежных рёбер, то есть рёбер, не имеющих общих вершин [1]. Например: для графа G (рис.1) паросочетание $P=\{u_3, u_5, u_8, u_{11}\}$. Будем говорить, что вершина графа G принадлежит паросочетанию, понимая под этим, что она инцидентна некоторому ребру паросочетания.

Построим граф $G_r=(U^r,V^r)$ – реберный для графа G . Вершины U^r графа G_r – соответствуют рёбрам U графа G . Пара вершин u_i^r, u_j^r в графе G_r связаны ребром в том и только в том случае, если в графе G пара рёбер (u_i, u_j) смежны, т.е. инцидентны одной вершине [1].

Для примера, представленного на рис. 1, реберный граф G_r графа G имеет вид, представленный на рис. 2.

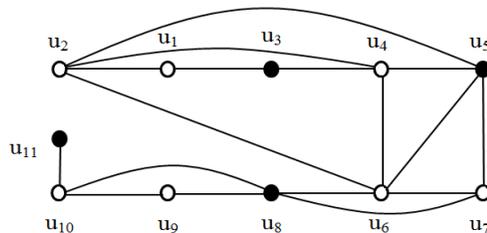


Рис. 2. Реберный граф G_r графа G

Множество $U^r_0 \subset U^r$ вершин графа $G_r=(U^r,V^r)$ называется внутренне устойчивым, если любые две вершины $u_i^r \in U^r_0$ и $u_j^r \in U^r_0$ не являются смежными [2]. Максимальное число вершин во внутренне устойчивом множестве графа U^r называется числом внутренней устойчивости и обозначается как $\alpha(G_r)$. Иногда число внутренней устойчивости называют также числом независимости графа G_r [3].

Подмножество вершин $U^r_0=\{u_3, u_5, u_8, u_{11}\}$ (см. рис. 2) графа $G_r=(U^r,V)$ является внутренне устойчивым, т.к. любые две вершины подмножества U^r_0 не смежны. Таким образом, паросочетанию в графе G соответствует внутренне-устойчивое подмножество реберного графа G_r графа G . Максимальному по мощности паросочетанию в графе G соответствует предельное внутренне-устойчивое подмножество (содержащее наибольшее число вершин) реберного графа G_r .

Рассмотрим задачу о *кликке*. Кликкой графа G называется максимальное по включению множество $V_0 \subset V$ вершин графа G , любые две из которых являются смежными. Другими словами, клика графа есть подмножество его вершин, такое, что между каждой парой вершин этого подмножества существует ребро.

Максимальная клика – это клика, которая не может быть расширена путём включения дополнительных смежных вершин, то есть нет клики большего размера, включающей все вершины данной клики. Наибольшая клика – это клика максимального размера для данного графа [4].

Дополнение графа (обратный граф) – граф G_0 , имеющий то же множество вершин, что и заданный граф G , но в котором две несовпадающие вершины смежны тогда и только тогда, когда они не смежны в G [2].

Нетрудно видеть, что при переходе от графа G к его дополнению G_0 каждая клика в G переходит в независимое множество в G_0 . Отсюда следует, что задача выделения клики в графе G сводится к задаче выделения независимого множества вершин в графе G_0 .

Таким образом, в основе процедур построения максимального паросочетания и выделения в графе клик лежит одна общая процедура формирования в графе $G(V, U)$ внутренне-устойчивого множества вершин. Задачу нахождения максимальных независимых множеств иногда называют «упаковкой вершин».

2. Механизмы выделения в графе независимого подмножества вершин на основе моделирования адаптивного поведения муравьиной колонии. Пусть дан граф $G(V, U)$, где V – множество вершин, $|V| = n$, U – множество ребер, $|U| = m$. Построим граф $G_r = (U^r, V^r)$ – реберный для графа G . Множество вершин $U^r = \{u^r_j | j = 1, 2, \dots, m\}$ графа G_r соответствует множеству ребер U графа G . Множество ребер $V^r = \{v^r_i | i = 1, 2, \dots, n\}$ графа G_r соответствует множеству вершин V графа G .

Сформулируем задачу формирования в графе $G_r = (U^r, V^r)$ внутренне – устойчивого множества вершин $U^r_0 \subset U^r$ как задачу разбиения.

Необходимо разбить множество U^r графа G_r на два непустых и непересекающихся подмножества U^r_0 и U^r_2 , таких, что любые две вершины $u^r_i \in U^r_0$ и $u^r_j \in U^r_0$ не являются смежными, $U^r_0 \cup U^r_2 = U^r$, $U^r_0 \cap U^r_2 = \emptyset$. Пусть $|U^r_0| = n_0$, $|U^r_2| = n_2$, $n_0 + n_2 = m$. Критерий оптимизации – число вершин $F = n_0$ в подмножестве U^r_0 . Цель оптимизации – максимизация критерия F .

Для поиска решения задачи формируется звездный граф поиска решений $H(U^r \cup S_0, E)$, определяемый корневой вершиной S_0 , который состоит из всех ребер E , связывающих корневую вершину S_0 с множеством всех вершин U^r (рис. 3). Корневая вершина S_0 связывается со всеми вершинами множества U^r .

В процессе решения участвует коллектив агентов. Каждый из агентов a_k формирует свое решение – свое подмножество $U^r_{0k} \subset U^r$ вершин звездного графа H , такое, что любые две вершины $u^r_i \in U^r_{0k}$ и $u^r_j \in U^r_{0k}$ в графе $G_r = (U^r, V^r)$ не являются смежными. k – номер агента.

Обозначим множество ребер звездного графа H , связывающих вершины сформированного внутри устойчивого подмножества U^r_{0k} с корневой вершиной S_0 как E_{0k} , $E_{0k} \subset E$.

Моделирование поведения агентов в процессе формирования каждым из них своего подмножества U^r_{0k} связано с распределением и учетом количества феромона на множестве E_{0k} ребер графа H .

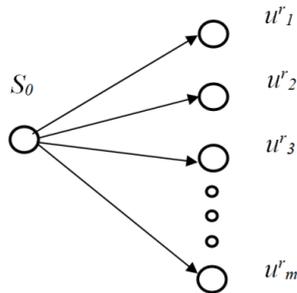


Рис. 3. Звездный граф H

На начальном этапе на всех ребрах звездного графа H откладывается одинаковое (небольшое) количество феромона ξ/m , где $m=|E|$.

Процесс поиска решений итерационный. Каждая итерация l включает три этапа. На первом этапе каждый агент a_k популяции конструктивным алгоритмом находит решение U^r_{ok} , рассчитывает оценку решения $\xi_k(U^r_{ok})$ и значение степени пригодности полученного агентом решения ϕ_k (количество феромона, соответствующее оценке).

Первый этап осуществляется следующим образом.

Агенты обладают памятью. На каждом шаге t в памяти агента a_k имеется:

- ◆ количество феромона $\phi_j(t)$, отложенного на каждом ребре графа H ;
- ◆ список вершин, уже включенных в формируемое множество – $U^r_{ok}(t-1)$ и список оставшихся (свободных) вершин $U^r_{ck}(t-1)$. $U^r_{ck}(t-1)=U^r \setminus U^r_{ok}(t-1)$, $U^r_{ok}(t-1) \cup U^r_{ck}(t-1)=U^r$.

Формирование решения (множества U^r_{ok}) осуществляется последовательно (пошагово). На каждом шаге t агент a_k применяет вероятностное правило выбора следующей вершины u^r_j для включения ее формируемое множество $U^r_{ok}(t)$.

На первом шаге в формируемое множество $U^r_{ok}(t)$, где $t=1$, включается одна их вершин $u^r_j \in U^r$ графа H , смежная корневой вершине S_0 .

На конечном шаге $t=n_0$ агентом a_k будет сформировано множество $U^r_{ok}(n_0)=U^r_{ok}$, $|U^r_{ok}(n_0)|=n_0$ такое, что любые две вершины $u^r_i \in U^r_{ok}$ и $u^r_j \in U^r_{ok}$ в графе $G_r=(U^r, V^r)$ не являются смежными. Шаг $t=n_0$ является конечным, если после его выполнения ни одна из вершин множества $U^r_{ck}(t)$ не может быть включена в $U^r_{ok}(t)$.

На шаге t формируется множество $U^r_{nk}(t) \subset U^r_{ck}(t-1)$ вершин, не смежных вершинам множества $U^r_{ok}(t-1)$ (рис. 4).

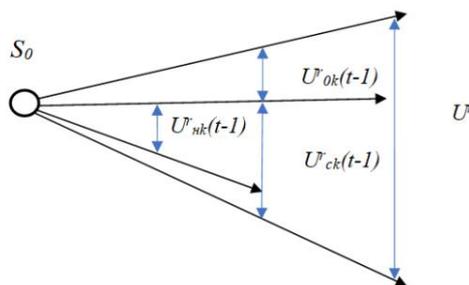


Рис. 4. Формирование множества $U^r_{nk}(t-1)$ вершин, не смежных с вершинами множества $U^r_{ok}(t-1)$

Вероятность включения в формируемое агентом a_k на базе звездного графа H множество $U^r_{ok}(t)$ вершины $u^r_j \in U^r_{nk}(t)$, не имеющей в графе $G_r=(U^r, V^r)$ связей с вершинами множества $U^r_{ok}(t-1)$, пропорциональна суммарному количеству феромона на ребре, связывающему вершину $u^r_j \in U^r_{nk}(t)$, $U^r_{nk}(t) \subset U^r$ с корневой вершиной S_0 в графе H .

Формируется множество ребер $E_{ok}(t-1) \subset E$ звездного графа H , связывающих вершины сформированного подмножества $U^r_{nk}(t-1)$ с корневой вершиной S_0 .

Агент просматривает все ребра $(S_0, u^r_j) \in E_{ok}(t-1)$, $u^r_j \in U^r_{nk}(t-1)$. Для каждого ребра $e_j \in E_{ok}(t-1)$ рассчитываются два параметра:

$\phi_j(t)$ – суммарное количество феромона на ребре e_j (в массиве коллективной эволюционной памяти КЭП_o);

$s_j(t)$ – число связей на графе $G_r=(U^r, V^r)$ между u^r_j и $U^r_{ck}(t-1)$.

По формуле (1) – при мультипликативной свертке, либо по формуле (2) – при аддитивной свертке определяется потенциальная степень предпочтения F_{ik} включения u^r_j

$$F_{jk} = (\phi_j(t)) / (s_j(t) + 1)^\beta \quad (1)$$

$$F_{jk} = (\alpha \phi_j(t)) + (\beta / (s_{jk} + 1)), \quad (2)$$

где α, β – управляющие параметры, которые подбираются экспериментально.

Вероятность $P_{ik}(t)$ включения вершины $u^r_j \in U^r_{nk}(t)$ в формируемое $U^r_{ok}(t)$, определяется следующим соотношением:

$$P_{ik}(t) = F_{ik} / \sum_i F_{ik} . \quad (3)$$

Агент с вероятностью $P_{ik}(t)$ выбирает одну из вершин, которая включается в подмножество $U^r_{ok}(t)$ и исключается из подмножества $U^r_{ck}(t-1)$.

Количество φ_k откладываемого агентом a_k феромона на ребрах $E_{ok} \subset E$ звездного графа H , связывающих вершины $u^r_j \in U^r_{ok}$ сформированного внутри устойчивого подмножества U^r_{ok} с вершиной S_0 , пропорционально числу вершин n_0 сформированного подмножества U^r_{ok} : $\varphi_k = \mu n_0$, где μ – коэффициент пропорциональности. Чем больше n_0 , тем больше феромона φ_k будет отложено агентом a_k на каждом ребре $(S_0, u^r_j) \in E_{ok}$ звездного графа H , построенного на вершинах множества U^r_{ok} . Следовательно, с большей вероятностью агенты будут выбирать такие вершины при синтезе собственного внутренне-устойчивого подмножества.

В работе используется циклический (ant-cycle) метод муравьиных систем. В этом случае рассчитанное агентом количество феромона φ_k предварительно откладывается в буферном массиве $Q_1 = \{q_j | j=1, 2, \dots, m\}$ коллективной эволюционной памяти (КЭП_o) в ячейках q_j , соответствующих ребрам множества $E_{ok} \subset E$.

На втором этапе, после полного формирования всеми агентами решений на текущей итерации, феромон ω_j , накопленный в j -ой ячейке в буферном массиве КЭП_o, добавляется в каждую j -ю ячейку основного массива $Q_2 = \{q_j | j=1, 2, \dots, m\}$ коллективной эволюционной памяти КЭП_o.

$$\phi_j(t) = \phi_j(t) + \omega_j. \quad (4)$$

Для избежания преждевременной сходимости используется отрицательная обратная связь в виде испарения феромона. После отложения феромона в основной массив коллективной эволюционной памяти КЭП_o, на третьем этапе происходит общее испарение феромона на множестве ребер E звездного графа H в соответствии с формулой (5).

$$\phi_j(t) = \phi_j(t) \cdot (1 - \rho), \quad (5)$$

где $\phi_j(t)$ – уровень феромона на ребре $e_j \in E_{ok}$, ρ – коэффициент обновления.

После выполнения всех действий на итерации находится агент с лучшим решением, которое запоминается. Далее осуществляется переход на следующую итерацию. Временная сложность этого алгоритма зависит от времени жизни колонии l (число итераций), количества вершин графа n и числа муравьев m , и определяется как $O(l * n^2 * m)$.

После формирования в реберном графе $G_r=(U^r, V)$ графа $G(X, U)$ внутренне – устойчивого множества вершин $U^r_0 \subset U^r$ осуществляется переход от графа G_r к исходному графу G для построения паросочетания или выделения в графе клики.

Для преодоления локального барьера используются подходы, основанные на сочетании различных видов эволюции. В частности, эффективность показала комбинированная архитектура бионического поиска, заключающаяся в последовательной работе муравьиного и эволюционного алгоритма нахождения максимального паросочетания, рассмотренного в работе [11–14].

3. Экспериментальные исследования. Алгоритм нахождения паросочетания в графе реализован в виде программы ПАР.

Исследование программы ПАР проводилось на примерах с известным оптимумом $F_{\text{опт}}$ по методологии ВЕКУ [13–20].

Контрольные примеры для исследования содержали до 1000 вершин.

Для оценки качества используется безразмерная величина $\xi = F_{\text{опт}}/F$, где F – оценка решения полученного, полученного с помощью разработанного алгоритма. Зависимости показателя качества ξ от числа итераций и размера популяции представлены на рис. 4 и 5.

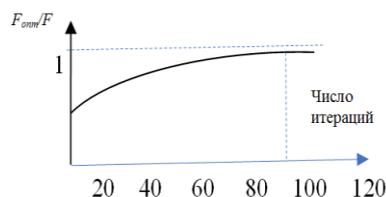


Рис. 5. Зависимость показателя ξ от числа итераций

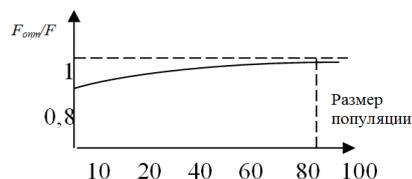


Рис. 6. Зависимости показателя ξ от размера популяции

Временная сложность алгоритма $O(n^2)$. Сравнение с известными алгоритмами показало, что при меньшем времени работы новый алгоритм дает более качественные решения.

Заключение. В работе рассматривается поисковый популяционный алгоритм. Разработана модифицированная метаэвристика по аналогии с моделями адаптивного поведения МА. Для поиска решения задачи в отличие от канонической парадигмы муравьиной колонии используется звездный граф поиска решений.

Основными достоинствами разработанного алгоритма являются:

- ◆ Низкая трудоемкость: временная сложность алгоритма $O(n^2)$.
- ◆ Реализация алгоритма отличается быстродействием и точностью вычислений.

Полученные в работе результаты можно использовать при решении задач, таких как: задачи о выделении внутренне-устойчивых множеств в графе, максимальной клике, проблема определения паросочетаний, раскраске графов и составление расписаний, задаче о назначениях.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Андерсон Д.* Дискретная математика и комбинаторика. – М.: Вильямс, 2003.
2. *Кормен К., Лейзерсон Ч., Ривест Р.* Алгоритмы построения и анализ. – М.: МЦМНО, 2000.
3. *Свами М., Тхуласираман К.* Графы, сети и алгоритмы. – М.: Мир, 1984.
4. *Лебедев Б.К., Лебедев О.Б.* Гибридный биоинспирированный алгоритм размещения базовых стандартных библиотечных элементов при проектировании топологии полужаказной СБИС // Известия ЮФУ. Технические науки. – 2019. – № 3. – С. 97-110.
5. *Карпенко А.П.* Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой: учеб. пособие. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2016. – 448 с.
6. *Лебедев Б.К., Лебедев О.Б.* Покрытие на основе методов роевого интеллекта // Проблемы разработки перспективных микро- и нанoeлектронных систем: Сб. трудов / под общ. ред. академика РАН А.Л. Стемпковского, 2016. – С. 187-193.
7. *Лебедев О.Б.* Распределение ресурсов на основе гибридных моделей роевого интеллекта // Научно-технический вестник информационных технологий, механики и оптики. – 2017. – № 6. – С. 1063-1073.
8. *Lebedev O.B.* Hybrid bioinspired algorithm of 1.5 dimensional bin-packing // Proceedings of the Third International Scientific Conference «Intelligent Information Technologies for Industry (ITI'18)». – 2018. – P. 254-263.
9. *Лебедев О.Б.* Модели адаптивного поведения муравьиной колонии в задачах проектирования. – Таганрог: Изд-во ЮФУ, 2013. – 199 с.
10. *Лебедев Б.К., Лебедев О.Б.* Ко-эволюционный алгоритм разбиения // Проблемы разработки перспективных микро- и нанoeлектронных систем / под ред. А.Л. Стемпковского, 2020. – № 3. – С. 79-86.
11. *Лебедев Б.К., Лебедев О.Б.* Муравьиные алгоритмы разбиения, использующие представления задачи, отличные от канонического // Вестник Ростовского государственного университета путей сообщения. – 2016. – № 3 (63). – С. 42-47.
12. *Лебедев Б.К., Лебедев О.Б., Жиглатый А.А.* Размещение элементов СБИС на основе моделей роевого интеллекта // Проблемы разработки перспективных микро- и нанoeлектронных систем: Сб. трудов / под общ. ред. академика РАН А.Л. Стемпковского. – М.: ИППМ РАН, 2020. – С. 118-126.
13. *Cong J., Romesis M., Xie M.* Optimality, Scalability and Stability Study of Partitioning and Placement Algorithms // Proc. of the International Symposium on Physical Design. – Monterey, CA. – P. 88-94.
14. *Sherwani N.A.* Algorithms for VLSI Physical Design Automation. – Third Edition, Kluwer Academic Publisher. – USA, 2013. – 572 p.
15. *Mourelle M.* Swarm intelligent systems. – Berlin: Heidelberg: Springer Verlag, 2006. – 217 p.
16. *Норенков И.П.* Основы автоматизированного проектирования: учебник. – М.: Изд-во МГТУ имени Н.Э. Баумана, 2006. – 336 с.
17. MCNC Electronic and Information Technologies (Online). Available: www.mcnc.org.
18. hMetis [Online]. Available: <http://www-users.cs.umn.edu/karypis/memis/hmetis300>. HB Floorplan Benchmarks [Online]. Available: <http://cadlab.cs.ucla.edu/cpmo/HBsuite.html>.
19. IBM-PLACE 2.0 benchmark suits. – <http://er.cs.ucla.edu/benchmarks/-ibm-place2/bookshelf/ibm-place2-all-bookshelf-nopad.tar.gz>.
20. *Adya S.N.* ISPD02 IBM-MS Mixed-size Placement Benchmarks. – <http://vlsicad.eecs.umich.edu/BK/ISPD02bench/>.

REFERENCES

1. *Anderson D.* Diskretnaya matematika i kombinatorika [Discrete mathematics and combinatorics]. Moscow: Vil'yams, 2003.
2. *Kormen K., Leyzerson Ch., Rivest R.* Algoritmy postroenie i analiz [Algorithms construction and analysis]. Moscow: MTSMNO, 2000.
3. *Svami M., Tkulasiraman K.* Grafy, seti i algoritmy [Graphs, networks and algorithms]. Moscow: Mir, 1984.
4. *Lebedev B.K., Lebedev O.B.* Gibridny bioinspirirovanny algoritm razmeshcheniya bazovykh standartnykh biblioteknykh elementov pri proektirovanii topologii poluzakaznoy SBIS [Hybrid bioinspired algorithm for placing basic standard library elements in the design of the topology of a semi-custom VLSI], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2019, No. 3, pp. 97-110.

5. *Karpenko A.P.* Sovremennyye algoritmy poiskovoy optimizatsii. Algoritmy, vdokhnovlennyye prirodoy: ucheb. posobie [Modern search engine optimization algorithms. Algorithms inspired by nature]. Moscow: Izd-vo MGTU im. N.E. Baumana, 2016, 448 p.
6. *Lebedev B.K., Lebedev O.B.* Pokrytie na osnove metodov roevogo intellekta [Coverage based on swarm intelligence methods // *Problems of development of promising micro- and nanoelectronic systems*], *Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem: Sb. trudov* [Collection of works], under the total. ed. academician of the RAS A.L. Stempkovskogo, 2016, pp. 187-193.
7. *Lebedev O.B.* Raspredelenie resursov na osnove gibridnykh modeley roevogo intellekta [Resource distribution based on hybrid models of swarm intelligence], *Nauchno-tehnicheskii vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and technical bulletin of information technologies, mechanics and optics], 2017, No. 6, pp. 1063-1073.
8. *Lebedev O.B.* Hybrid bioinspired algorithm of 1.5 dimensional bin-packing, *Proceedings of the Third International Scientific Conference «Intelligent Information Technologies for Industry (IITI'18)»*, 2018, pp. 254-263.
9. *Lebedev O.B.* Modeli adaptivnogo povedeniya murav'inoy kolonii v zadachakh proektirovaniya [Models of adaptive behavior of an ant colony in design problems]. Taganrog: Izd-vo YuFU, 2013, 199 p.
10. *Lebedev B.K., Lebedev O.B.* Ko-evolyutsionnyy algoritm razbieniya [Co-evolutionary partitioning algorithm], *Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem* [Problems of development of promising micro- and nanoelectronic systems], ed. by A.L. Stempkovskogo, 2020, No. 3, pp. 79-86.
11. *Lebedev B.K., Lebedev O.B.* Murav'inye algoritmy razbieniya, ispol'zuyushchie predstavlenie zadachi, otlichnye ot kanonicheskogo [Ant partitioning algorithms using problem representation other than the canonical one], *Vestnik Rostovskogo gosudarstvennogo universiteta putey soobshcheniya* [Bulletin of the Rostov State University of Communications], 2016, No. 3 (63), pp. 42-47.
12. *Lebedev B.K., Lebedev O.B., Zhiglaty A.A.* Razmeshchenie elementov SBIS na osnove modeley roevogo intellekta [Placement of VLSI elements based on swarm intelligence models], *Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem: Sb. trudov* [Problems of development of promising micro- and nanoelectronic systems. Collection of works], under the total. ed. academician of the RAS A.L. Stempkovskogo. Moscow: IPPM RAN, 2020, pp. 118-126.
13. *Cong J., Romesis M., Xie M.* Optimality, Scalability and Stability Study of Partitioning and Placement Algorithms, *Proc. of the International Symposium on Physical Design*. Monterey, CA, pp. 88-94.
14. *Sherwani N.A.* Algorithms for VLSI Physical Design Automation. Third Edition, Kluwer Academic Publisher. USA, 2013, 572 p.
15. *Mourelle M.* Swarm intelligent systems. Berlin: Heidelberg: Springer Verlag, 2006, 217 p.
16. *Norenkov I.P.* Osnovy avtomatizirovannogo proektirovaniya: uchebnik [Fundamentals of computer-aided design: textbook]. Modvov: Izd-vo MGTU imeni N.E. Baumana, 2006, 336 p.
17. MCNC Electronic and Information Technologies (Online). Available: www.mcnc.org.
18. hMetis [Online]. Available: <http://www-users.cs.umn.edu/karypis/memis/hmet300>. HB Floorplan Benchmarks [Online]. Available: <http://cadlab.cs.ucla.edu/cpmo/HBSuite.html>.
19. IBM-PLACE 2.0 benchmark suits. Available at: <http://er.cs.ucla.edu/benchmarks/-ibm-place2/bookshelf/ibm-place2-all-bookshelf-nopad.tar.gz>.
20. Adya, S.N. ISPD02 IBM-MS Mixed-size Placement Benchmarks. Available at: <http://vlsicad.eecs.umich.edu/BK/ISPD02bench/>.

Статью рекомендовал к опубликованию д.т.н., профессор А.Г. Коробейников.

Лебедев Олег Борисович – Южный федеральный университет; e-mail: lebedev.ob@mail.ru; г. Таганрог, Россия; тел.: 89085135512; кафедра систем автоматизированного проектирования; доцент.

Жиглатый Артемий Александрович – e-mail: artemiy.zhiglaty@gmail.com; тел.: 89185916819; кафедра систем автоматизированного проектирования; аспирант.

Lebedev Oleg Borisovich – Southern Federal University; e-mail: lebedev.ob@mail.ru; Taganrog, Russia; phone: +79085135512; the department of computer aided design; associate professor

Zhiglatiy Artemy Alexandrovich – e-mail: artemiy.zhiglaty@gmail.com; phone: +79185916819; the department of computer aided design; graduate student.

В.В. Бова, Ю.А. Кравченко, С.И. Родзин

МЕТОДЫ И АЛГОРИТМЫ КЛАСТЕРИЗАЦИИ ТЕКСТОВЫХ ДАННЫХ (ОБЗОР)

Рассматривается одна из важных задач искусственного интеллекта – машинная обработка естественного языка. Решение данной задачи на основе кластерного анализа позволяет выявлять, формализовывать и интегрировать большие объемы лингвистической экспертной информации в условиях информационной неопределенности и слабой структурированности исходных текстовых ресурсов, полученных из различных предметных областей. Кластерный анализ является мощным средством разведочного анализа текстовых данных, позволяющий провести объективную классификацию любых объектов, которые охарактеризованы рядом признаков и имеют скрытые закономерности. Проведен обзор и анализ современных модифицированных алгоритмов агломеративной кластеризации CURE, ROCK, CHAMELEON, неиерархической кластеризации PAM, CLARA и алгоритма аффинного преобразования, используемых на различных этапах кластеризации текстовых данных, эффективность которых проверяется экспериментальными исследованиями. В работе обоснованы требования к выбору наиболее эффективного метода кластеризации для решения задачи повышения эффективности интеллектуальной обработки лингвистической экспертной информации. Также в работе рассмотрены способы визуализации результатов кластеризации для интерпретации кластерной структуры и зависимостей на множестве элементов текстовых данных и графические средства их представления в виде дендограмм, диаграмм рассеивания, диаграмм сходства VOS и карт интенсивности. Для сравнения качества работы алгоритмов использовались внутренние и внешние метрики эффективности: «V-мера», «Adjusted Rand index», «Силуэт». На основании проведенных экспериментов выявлено, что необходимо использовать гибридный подход, в котором для первоначального выбора числа кластеров и распределения их центров использовать иерархический подход, основанный на последовательном объединении и максимизации близости данных ограниченной выборки, когда нет возможности выдвинуть гипотезу о начальном количестве кластеров. Далее подключать алгоритмы итерационной кластеризации, обеспечивающие высокую устойчивость по отношению к шумовым признакам и наличию выбросов. За счет гибридизации повышается эффективность работы алгоритмов кластеризации. Результаты исследований показали, что для повышения вычислительной эффективности и преодоления чувствительности при инициализации параметров алгоритмов кластеризации для оптимизации параметров модели обучения и поиска глобального оптимального решения необходимо использовать метаэвристические подходы.

Кластерный анализ текстовых данных; агломеративная кластеризация; метрики качества; неиерархическая кластеризация; метод аффинного преобразования; дендограммы; диаграммы рассеивания.

V.V. Bova, Y.A. Kravchenko, S.I. Rodzin

METHODS AND ALGORITHMS FOR TEXT DATA CLUSTERING (REVIEW)

The article deals with one of the important tasks of artificial intelligence – machine processing of natural language. The solution of this problem based on cluster analysis makes it possible to identify, formalize and integrate large amounts of linguistic expert information under conditions of information uncertainty and weak structure of the original text resources obtained from various subject areas. Cluster analysis is a powerful tool for exploratory analysis of text data, which allows for an objective classification of any objects that are characterized by a number of features and have hidden patterns. A review and analysis of modern modified algorithms for agglomerative clustering CURE, ROCK, CHAMELEON, non-hierarchical clustering PAM, CLARA and the affine transformation algorithm used at various stages of text data clustering, the effectiveness of which is verified by experimental studies, is carried out. The paper substantiates the requirements for choosing the most efficient clustering method for solving the problem of increas-

ing the efficiency of intellectual processing of linguistic expert information. Also, the paper considers methods for visualizing clustering results for interpreting the cluster structure and dependencies on a set of text data elements and graphical means of their presentation in the form of dendograms, scatterplots, VOS similarity diagrams, and intensity maps. To compare the quality of the algorithms, internal and external performance metrics were used: "V-measure", "Adjusted Rand index", "Silhouette". Based on the experiments, it was found that it is necessary to use a hybrid approach, in which, for the initial selection of the number of clusters and the distribution of their centers, use a hierarchical approach based on sequential combining and averaging the characteristics of the closest data of a limited sample, when it is not possible to put forward a hypothesis about the initial number of clusters. Next, connect iterative clustering algorithms that provide high stability with respect to noise features and the presence of outliers. Hybridization increases the efficiency of clustering algorithms. The research results showed that in order to increase the computational efficiency and overcome the sensitivity when initializing the parameters of clustering algorithms, it is necessary to use metaheuristic approaches to optimize the parameters of the learning model and search for a global optimal solution.

Text data cluster analysis; agglomerative clustering; quality metrics; non-hierarchical clustering; affine transformation method; dendograms; scatterplots.

Введение. Одним из актуальных направлений в области машинной обработки естественного языка (natural language processing (NLP)) является кластерный анализ слабоструктурированных естественно-языковых (ЕЯ) текстовых данных, позволяющий проводить первоначальную обработку информации для ее структурирования, выделения семантических признаков, обобщения и сортировки.

Важной задачей кластерного анализа текстовых данных является поиск скрытых закономерностей в больших объемах информации. При ее решении возникает проблема извлечения и отбора признаков существенных для анализа лингвистической информации [1–3]. Необходимость выбора большого количества семантических признаков многомерных текстовых данных приводит к увеличению вычислительной и временной сложности алгоритма кластеризации. Но при этом исключение из пространства признаков шумовых объектов, может приводить к потере значимой для идентификации неявных закономерностей информации [4–6]. Определение связи шумовых признаков с целевой переменной в решаемой задаче обработки ЕЯ текстовых данных является фундаментальной проблемой алгоритмов кластеризации. Поэтому при выборе алгоритма необходимо основываться на максимальном сохранении информационных признаков, используемых для кластеризации.

Анализ научных исследований в области кластерного анализа текстовых данных [2–8] показывает, что в настоящее время присутствует существенное многообразие алгоритмов кластеризации, отличных по своей природе и происхождению, каждым из них можно получить различные разбиения исходного множества текстовых данных. По этой причине актуальной проблемой остается выбор и применение эффективного метода кластеризации для решения задач NLP. Для решения данных проблем можно сформулировать ряд требований, которым должен удовлетворять метод кластеризации текстовых данных: обеспечение высокой размерности пространства признаков данных; масштабируемость при работе с большим объемом текстовых данных; обеспечение смешанного типа измерений при выборе метода сходства (вычисления семантической близости) и метода объединения (агрегации связей на основе поиска скрытых закономерностей); выявление семантически однородных кластеров, содержащих объекты близкие по совокупности свойств и признаков; распределение объектов по всей совокупности взаимоисключающих кластеров.

В данной работе проводится сравнительный анализ современных модификаций известных алгоритмов кластеризации текстовых данных, наиболее подходящих для решения задачи повышения эффективности интеллектуальной обработки лингвистической экспертной информации.

1. Особенности выбора и классификация методов кластерного анализа текстовой информации. Построение моделей кластерного анализа данных, представленных на ЕЯ основывается на семантическом подходе. Семантическая кластеризация является многоэтапной процедурой, на каждом этапе которой должна решаться отдельная задача выбора наиболее эффективного алгоритма извлечения признаков и уменьшения размерности данных [3, 9–11]. Последовательность ключевых этапов кластерного анализа представлена на рис. 1. На первом этапе предварительной обработки текста происходит преобразование текстовых документов в структурированные данные на основе извлечения признаков, которые представлены в онтологии определенной предметной области, удаление избыточной информации и уменьшение размерности пространства признаков [4, 10, 11]. Для предварительной обработки применяются методы фильтрации – удаления специфических символов и общеупотребительных слов, токенизации – разбиения текста на лингвистические тематические единицы, стемминг, удаление стоп-слов, сокращение низкочастотных терминов [1, 9].

На втором этапе реализуются задачи выбора признаков, их скрытое семантическое индексирование и преобразование в векторную модель данных, которые можно подавать на вход алгоритмам машинного обучения. Подробный обзор методик векторизации текстовых документов приведен в работах [12–14].



Рис. 1. Этапы кластерного анализа

Векторное представление модели текстовых данных характеризуется высокой размерностью и разреженностью семантических признаков, что требует применения улучшенных алгоритмов кластеризации для обработки данных большой размерности [4, 13].

В последнее время ведутся активные разработки новых масштабируемых алгоритмов кластеризации, способных обрабатывать текстовые данные большой размерности [3, 14–19]. Разработанные авторами модели кластерного анализа представляют собой гибридное решение на основе известных алгоритмов и эффективно решают задачи машинного перевода, распознавания речи и генерации реалистичных текстов. В работах [8, 9] представлена современная классификация модифицированных алгоритмов кластеризации. На рис. 2 приведена классификация основных типов репрезентативных алгоритмов кластеризации текстовых данных на основе иерархических и итерационных (неиерархических) методов.

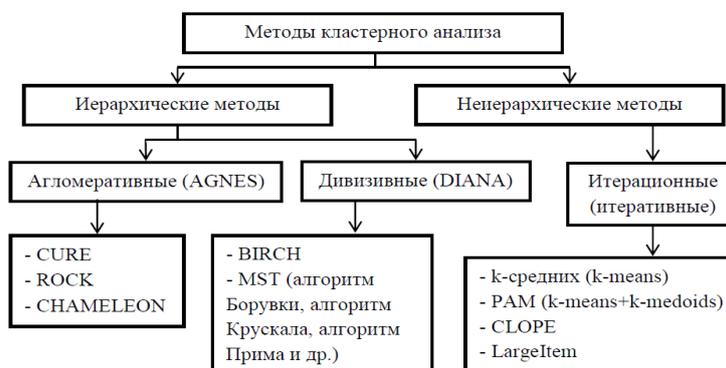


Рис. 2. Классификация методов кластерного анализа

Выбор определенного метода зависит от набора данных, к которому он применяется, требований к точности и скорости работы алгоритма, критериям оптимальности параметров кластеризации при решении задач NLP [19]. Также уточнения и обоснования требуют следующие вопросы:

- ◆- выбор числа кластеров и метода оптимизации параметров их первоначального формирования;
- ◆- применение на различных этапах кластеризации эффективных методов измерения сходства и вычисления семантической близости связанных текстовых данных в многомерном пространстве их информационных признаков;
- ◆- выбор метрик точности и качества кластеризации текстовых данных большой размерности, позволяющих определить семантическую однородность полученных кластеров;
- ◆- выбор способа визуализации результатов кластеризации с отражением семантических зависимостей на множестве элементов текстовых данных для графического представления связанных между собой элементов кластерной структуры, их анализа и уточнения параметров метода кластеризации.

2. Методы иерархической агломеративной кластеризации. Алгоритмы иерархической кластеризации рекурсивно находят систему вложенных разбиений анализируемой выборки данных на непересекающиеся кластеры, применяя агломеративные (объединяющие) или дивизионные (разделительные) стратегии [17]. В агломеративных алгоритмах каждый объект из множества наблюдений рассматривается как независимый кластер. На каждом шаге алгоритма вычисляется матрица расстояний на основе критериев сходства признаков и последовательно объединяются объекты, обладающие непустым набором общих семантических свойств или закономерностей, пока все объекты не будут составлять один кластер.

Агломеративные методы различаются используемыми мерами расстояния для вычисления семантической близости и алгоритмами объединения, использующие различные критерии связи [8, 19].

1. Одиночная связь (Single link). Условие объединения кластеров C_1 и C_2 по принципу ближайшего соседа представлено как:

$$D(C_1, C_2) = \min_{x \in C_1, y \in C_2} dist(x, y), \quad (1)$$

где $dist(x, y)$ – мера семантической близости, используемая при создании матрицы расстояний, $X = (x_0, x_1, \dots, x_n)$, $Y = (y_0, y_1, \dots, y_n)$ – множество n векторов признаков объектов кластеров C_1 и C_2 соответственно.

2. Полная связь (Complete link). Два кластера объединяются на основе вычисления расстояния между их самыми дальними объектами по формуле:

$$D(C_1, C_2) = \max_{x \in C_1, y \in C_2} \text{dist}(x, y). \quad (2)$$

3. Метод попарного среднего (Average link). Расстояние между двумя кластерами рассчитывается как среднее арифметическое между всеми возможными парами объектов из каждого кластера согласно:

$$D(C_1, C_2) = \frac{1}{|C_1||C_2|} \sum_{x \in C_1, y \in C_2} \text{dist}(x, y). \quad (3)$$

4. Центроидная связь (Centroids link). В этом методе объединяются два кластера с минимальным расстоянием до центроида, межкластерное расстояние между центрами тяжести определяется как:

$$D(C_1, C_2) = \min_{x \in C_1, y \in C_2} \text{dist} \left(\left(\frac{1}{|C_1|} \sum_{x \in C_1} \vec{x} \right), \left(\frac{1}{|C_2|} \sum_{y \in C_2} \vec{y} \right) \right). \quad (4)$$

5. Связь Уорда (Ward's method). В данном методе критерием выбора пары кластеров для слияния является оптимальное значение целевой функции (ЦФ) связи, определяющей расстояние между двумя кластерами, которая вычисляется как сумма квадратов отклонений (ESS) от среднего вектора (центроида) после объединения кластеров. Способы вычисления расстояния между кластерами задаются формулой Ланса – Уильямса [7], а расстояния между объектами можно задавать любой метрикой семантической близости. Метод Уорда основан на объединении не максимально близких кластеров, а тех, слияние которых дает минимальный прирост внутрикластерной дисперсии.

Преимуществами иерархических методов является наглядность представления агломератов в древообразном представлении наблюдений, называемом дендрограммой, возможность получать сбалансированные кластеры произвольных форм и идентифицировать выбросы в наборе текстовых данных. К сложностям можно отнести: ограничение объема набора данных; выбор меры близости; отсутствие глобальной целевой функции, каждый шаг иерархической кластеризации является локальным решением; вычислительная и временная сложность значительно ограничивают способность алгоритмов обрабатывать текстовые данные большой размерности; систему полных разбиений, которая может являться излишней в контексте решаемой задачи NPL, поэтому в алгоритме необходимо задавать пороговое значение расстояния для усечения иерархии или максимальное количество объектов в кластере.

Рассмотрим модифицированные алгоритмы Cure, Chameleon, Rock, в которых методы иерархической кластеризации интегрированы с другими методами, преодолевающими ограничения агломеративных алгоритмов.

Алгоритм CURE (Clustering Using REpresentatives) выполняет агломеративную кластеризацию с использованием репрезентативных объектов - представителей окрестности семантических признаков текстовых данных каждого кластера [17]. Для работы с данными большой размерности CURE использует комбинацию случайной выборки набора данных и сегментирования распределенных объектов-представителей кластера [20]. Выбор представителей, принадлежащих кластеру начинается с наиболее удаленных, с последующим сжатием расстояния между ними и центром кластера на некоторую долю α . Объединение кластеров происходит на основе вычисления расстояния между наиболее близкими представителями двух кластеров. Если $\alpha = 0$, то алгоритм объединения приближается к методам кластеризации по принципу ближайшего соседа, в случае $\alpha = 1$, к центроидным методам. Такой подход позволяет ослабить влияние выбросов, фиксировать форму и размер кластера.

Исходными данными алгоритма являются: m объектов размерности n , k – количество результирующих кластеров, c – количество объектов-представителей кластера, α – параметр сдвига объектов к центроиду кластера, γ – доля объектов

выборки на наборе данных. Вычислительное ядро алгоритма представляет собой итерационное построение заданного числа кластеров, вычисление их центроидов и назначение объектов-представителей для каждого кластера [20].

Шаг 1. Построение дерева кластеров на основе случайной выборки репрезентативных объектов на исходном наборе данных $[y, m]$. Каждый объект принимается за кластер, центроид которого равен самому объекту. Далее выполняются следующие итерации, пока количество кластеров не станет равно k :

- 1) для каждого i -го кластера вычисляется центроид μ_i по формуле:
$$\mu_i = \frac{1}{K_i} \sum_{m_j \in K_i} m_j$$
, где m_j – множество объектов-представителей, находящихся в K_i -м кластере;
- 2) выполняется поиск пары кластеров, расстояние между центроидами которых минимально и определяется по двум ближайшим объектам-представителям согласно формуле (1).

Шаг 2. Пары, имеющие наибольшее сходство объединяются и для образованного кластера определяются представительные объекты:

- 1) выбирается произвольный объект-представитель кластера;
- 2) пока количество представителей меньше s выбирается представителем объект, расстояние от которого, до уже выбранных является максимальным согласно формуле (2);
- 3) для каждого представителя выполняется сдвиг к центроиду соответствующего кластера в α раз.

Шаг 3. Пересчет наиболее близкого кластера для некоторого поднабора объединенных кластеров по формуле (4). Если полученные на этом шаге кластеры отвечают ЦФ, то завершить кластеризацию, иначе – перейти к шагу 3.

Для ускорения кластеризации, на шаге 2 CURE производит разбиение полученных объектов-представителей. Наборы n объектов сегментируются на p равных частей, и на них выполняется иерархическая кластеризация, пока количество агломератов каждой части не станет равно n/pq , где $q > 0$.

Алгоритм CURE менее чувствителен к выбросам и позволяет получать семантически однородные кластеры сложной формы и различных размеров, хорошо масштабируем для больших данных без ущерба для качества кластеризации. Общая временная сложность алгоритма равна $O(n^2 \log(n))$, где n – количество объектов кластеризации, пространственная сложность этого алгоритма за счет использования динамических структур данных в виде дерева кластеров равна $O(n)$.

К недостаткам данного метода можно отнести то, что при объединении кластеров не учитывается их внутреннее сходство и необходимость задания параметров порога плотности и количества кластеров, от которых во многом зависят результаты анализа.

Алгоритм ROCK (RObust Clustering using linKs) – алгоритм агломеративной кластеризации, который использует стратегию поиска связей в категориальном наборе данных по принципу взаимного соседства [21]. В ROCK алгоритме сходство кластеров основано на количестве объектов из разных кластеров, имеющих общих соседей. Особенностью алгоритма является введение функции связи для поиска скрытых закономерностей в окрестностях векторного пространства признаков текстовых данных. Соседями являются объекты x_i и y_j , для которых значение функции сходства $sim(x_i, y_j) \geq \theta$ превышает заданный порог θ . Предполагается, что 1 соответствует абсолютной близости, и 0 – наоборот. Выбор функции sim и значения θ зависят от входных данных и особенностей реализации алгоритма. Значение функции связи $link(x_i, y_j)$ между x_i и y_j соответствует количеству общих соседей в окрестности поиска. Чем больше значение связи, тем больше вероятность, что объекты имеют одинаковую окрестность признаков и принадлежат одному кластеру. Кластеризация с использованием ROCK является более глобальной и не сводится только к локальному поиску объектов [17].

Задача кластеризации сводится к максимизации суммы связей объектов, принадлежащих одному кластеру, и минимизации ее для объектов из разных кластеров. ЦФ разбиения на k кластеров имеет следующий вид [21]:

$$F_i = \sum_{i=1}^k n_i^{f(\theta)_i} \times \sum_{x_i, y_j \in C_i} \frac{\text{link}(x_i, y_j)}{n_i^{1+2f(\theta)_i}}, \quad (5)$$

где C_i – кластер i размера n , ожидаемая сумма связей – $n_i^{1+2f(\theta)_i}$, ожидаемая сумма соседей – $n_i^{f(\theta)_i}$, для функции $f(\theta)$ выбирается взвешенное представление $\frac{1+\theta}{1-\theta}$, что предупреждает отнесение объектов с низкими значениями связи к одному кластеру.

Для слияния кластеров вводится ЦФ качества (полезности), определяющая насколько выгодно объединять полученные кластеры:

$$Q(C_i, C_j) = \frac{\text{link}(C_i, C_j)}{(n_i + n_j)^{1+2f(\theta)} - n_i^{1+2f(\theta)} - n_j^{1+2f(\theta)}}. \quad (6)$$

Значение связи между кластерами представлено суммой попарных связей объектов, принадлежащих разным кластерам и определяется как:

$$\text{link}(C_i, C_j) = \sum_{x_i \in C_i, y_j \in C_j} \text{link}(x_i, y_j). \quad (7)$$

Алгоритм состоит из двух основных этапов. На входе имеется n объектов и k – количество кластеров для разбиений. Для каждого n задана функция схожести sim . На первом этапе каждый объект представляется отдельным кластером, для него вычисляется значение link и значение ЦФ для объединения с другим кластером по формуле (5). Для полученных решений создается глобальное дерево, которое содержит все подмножества кластеров, связь между которыми не нулевая.

На втором этапе выполняется цикл, на каждом шаге которого объединяются два подмножества кластеров с максимальным значением функции полезности согласно формуле (6), после чего вносятся соответствующие изменения в дерево кластеров. После объединения производится пересчет значений связи и качества для новообразованного кластера. Алгоритм завершает работу в двух случаях: когда получено k кластеров или, когда все связи между оставшимися кластерами равны 0.

К преимуществам алгоритма ROCK можно отнести высокую эффективность кластеризации и масштабируемость, эффективность работы с выбросами, ROCK подходит для данных произвольной формы и большой размерности. Недостатки: результат кластеризации чувствителен к параметрам алгоритма. Временная сложность алгоритма равна $O(n^2 \log(n))$, где n – количество объектов.

Алгоритм CHAMELEON (hierarchical clustering using dynamic modeling) – гибридный алгоритм агломеративной кластеризации на основе графоориентированных и классических иерархических методов, использующий динамическую модель определения сходства между парами кластеров [22]. Существующие алгоритмы применяют статическую модель кластеризации и не используют информацию о характере отдельных кластеров по мере их слияния. Так алгоритм CURE не учитывает информацию о совокупной взаимосвязанности элементов в двух кластерах, а алгоритм ROCK – информацию о близости двух кластеров, определяемую сходством ближайших элементов в них. Ключевой особенностью Chameleon является то, что при вычислении сходства между кластерами используются их относительная взаимосвязь и относительная близость, а объединение наиболее похожей пары кластеров происходит с учетом нормализации их абсолютных значений. Chameleon позволяет выявлять стабильные кластеры в наборе текстовых данных, используя трехэтапный алгоритм [7, 22].

Первый этап заключается в построении графа $G=(X,E)$ по матрице сходства объектов (по множеству объектов $x_i, x_j \in E$ и мере их близости $dist(x_i, x_j)$) по принципу k ближайших соседей $dist(x_i, x_j) \in sorted(\{r | r = dist(x_i, x_k), k = 1, \dots, n\})[1..k]$, причем ребра – взвешенное значение близости объекта с одним из k -ближайших соседей другого объекта $weight(e) = dist(x_i, x_j)$. Вершина графа, соответствующая некому объекту, соединяется с k вершинами, расстояние до которых минимально.

На втором этапе Chameleon использует алгоритм разделения графа для получения набора относительно малых подкластеров. Количество кластеров, на которые будет разбит граф, заранее неизвестно, и зависит от свойств объекта и выбранной метрики. Для разбиения графа создается множество $C=G$ и выполняется цикл со следующими операциями на каждой его итерации:

- ◆ выделить C_{gr} такой, что $|C_{gr}| = \max_{1 \leq j \leq |C|} |C_j|$;
- ◆ если $|C_{gr}|/|X| > min$, то разделить C_{gr} на 2 кластера C_1 и C_2 , удалить C_{gr} из C , и добавить C_1 и C_2 в C ;
- ◆ иначе процесс дальнейшего разбиения прекращается, переход на 3-й этап.

На третьем этапе выполняется агломеративная кластеризация для формирования результирующего множества подлинных кластеров путем многократного объединения полученных ранее подкластеров. Формируется начальный набор кластеров $C_{mek} = C$ и на каждой итерации слияния кластеров выполняется следующий цикл.

1. Для каждой пары кластеров $C_i, C_j \in C_{mek}$ вычисляются показатели: абсолютной взаимной связности $EC(C_i, C_j)$, рассчитывается как сумма весов ребер, соединяющих вершину 1-го кластера C_i с вершинами второго кластера C_j по формуле: $EC(C_i, C_j) = \sum_{e \in conn(C_i, C_j)} weight(e)$, где $conn(C_i, C_j) = \{(v, w) | v \in C_i, w \in C_j\}$ и внутренней связности кластера $EC(C_i)$ определяется как сумма весов ребер, разбивающих C_i на 2 равных подграфа: $EC(C_i) = \sum_{e \in mincut(C_i)} weight(e)$, где $mincut(C_i)$ – минимальный разрез C_i . Аналогичным образом вычисляются $EC(C_j)$.

2. Определяется относительная взаимная связность пары кластеров $RI(C_i, C_j)$ по формуле $RI(C_i, C_j) = 2 * EC(C_i, C_j) / (EC(C_i) + EC(C_j))$. Нормализация с учетом $EC(C_i)$ используется для исключения тенденции к преимущественному объединению крупных кластеров, у которых значение взаимной связности достаточно велико, вследствие их размерности.

3. Далее вычисляется относительное взаимное сходство $RC_{C_i, C_j} = \frac{S_{EC(C_i, C_j)}}{S_{EC(C_i)} + S_{EC(C_j)}}$, где $S_{EC(C_i, C_j)} = EC(C_i, C_j) / conn(C_i, C_j)$ – абсолютное взаимное сходство пары кластеров и $S_{EC(C_i)} = EC(C_i) / mincut(C_i)$ – среднее сходство между связанными объектами C_i .

4. Последовательное объединение пары кластеров происходит на основе вычисления ЦФ $F = \max RI(C_i, C_j) * RCa(C_i, C_j)$, где a – заданный заранее параметр. Процесс слияния останавливается, когда количество кластеров, удовлетворяющих ЦФ менее одного.

Преимуществами Chameleon является высокая точность и эффективность результатов кластеризации. К недостаткам метода можно отнести то, что временная сложность резко возрастает с увеличением сложности графа. Общая временная сложность алгоритма равна $O(n^3)$, где n – количество объектов кластеризации.

Методы неиерархической (итерационной) кластеризации. Существующие методы неиерархической кластеризации отличаются подходами к разделению набора текстовых данных на определенное количество непересекающихся кластеров [8]. Первый заключается в идентификации неявной кластерной структуры, на основе «сгущения объектов» для определения границ кластеров с высокой локаль-

ной плотностью в многомерном пространстве исходных данных, а области низкой плотности соотносятся с шумовыми признаками. Кластер определяется локальным максимумом оценочной функции плотности. Этот подход реализован в плотностных алгоритмах DBSCAN, OPTICS, DBCLASD, позволяющих с высокой точностью идентифицировать кластеры произвольной формы и размера для данных большой размерности, искаженных шумом [7, 23].

Второй подход основан на разделении векторного пространства текстовых данных на оболочки смысловых кластеров, с минимальными ограничениями на признаки отбора, во избежание исключения скрытых закономерностей, поведение которых имеет характер шума [8]. Алгоритмы, поддерживающие данную идею кластеризации путем итеративных вычислений выполняют случайную инициализацию центроидов и определяют ближайшие вектора данных для каждого из них. После чего, выполняется итеративное смещение центров кластеров и пересчет ближайших векторов данных, расчет суммарного квадратичного отклонения данных от центров полученных кластеров и выбор наилучшего локального решения, для которого ЦФ будет минимальной. К таким алгоритмам относятся k-means и его модификации K-modes, K-medoids: PAM (Partitioning Around Medoids), CLARA (Clustering Large Applications), CLARANS (Randomized CLARA), отличающихся применением различных стратегий инициализации центров кластеров и метрик несходства для вычисления расстояния от каждого наблюдения до центров их кластеров [6–9].

Неиерархические методы проявляют более высокую устойчивость по отношению к шумам и выбросам, выбору метрики схожести, включению незначимых переменных в набор, участвующий в кластеризации, но слишком чувствительны к инициализации параметров центроидов, требуют задания начальных условий (количество образуемых кластеров, порог завершения кластеризации и др.), задача нахождения глобального оптимального решения для данных алгоритмов является NP-полной.

Алгоритм кластеризации K-medoids: PAM (Partitioning Around Medoids).

Алгоритм k-medoids является модификацией k-means и предназначен для решения задач выделения кластеров в случаях, когда проводится кластеризация объектов без использования свойств линейного пространства [5]. В отличие от k-means, центром кластера может быть не любой объект признакового пространства (центроид), а только, принадлежащий кластеризуемой выборке представитель – медоид, среднее несоответствие (сумма различий) которого по отношению ко всем другим объектам в кластере минимальна [8]. Принадлежность объекта x_c к соответствующему кластеру C определяется индексом наиболее близкого к нему медоида x_m .

$$M(C) = \arg \min_{x_m \in C} \sum_{x_c \in C} \text{dist}(x_c, x_m). \quad (8)$$

K-medoid использует произвольные меры сходства для вычисления близости dist и минимизирует абсолютную ошибку или «полное отклонение» TD суммы попарных различий между объектами в кластере C_i и медоидами m_i как:

$$TD = \min \sum_{i=1}^k \sum_{x_c \in C_i} \text{dist}(x_c, m_i). \quad (9)$$

В общем случае задачу полного разбиения алгоритмом k-medoids, применяющего стратегию полного перебора решить точно NP-сложно. Для решения задачи нахождения лучшего разбиения с использованием медоидов предназначена PAM-реализация алгоритма k-medoids [14]. PAM использует стратегию «жадного» поиска, преобразует каждый шаг алгоритма из детерминированных вычислений в задачу статистической оценки и уменьшает размерность выборки текстовых данных n до $O(n \log n)$.

Входными данными алгоритма являются множество объектов $X=\{x_1, \dots, x_n\}$ и их индексов $z = \{1, \dots, n\} \subset Z$ с заданной функцией – метрикой расчета расстояний между объектами, $d_{ij} = \text{dist}(x_i, x_j)$, либо матрица расстояний $D \in R^{n \times n}$, характеризующая дистанции между каждой парой объектов $D=(d_{ij}), i \in z, j \in z$, а также число кластеров $k \leq n$, на которое необходимо разбить множество объектов. На выходе алгоритм выдает множество индексов объектов, принятых в качестве медоидов $m = \{m_1, \dots, m_k\}$. PAM состоит из двух последовательных алгоритмов: BUILD для выбора начальной кластеризации и SWAP для улучшения кластеризации в сторону локального оптимума [16].

BUILD. Построение начального набора k медоидов путем их последовательного выбора. Первый из них выбирается как объект, сумма различий которого до всех других объектов минимальна согласно (8). Далее осуществляется последовательный выбор следующего медоида из оставшихся объектов по формуле (9) до тех пор, пока $m = m \cup \{ \underset{i \in z \setminus m}{\text{argmax}} \sum_{j \in z \setminus (m \cup i)} \max(D_j - d_{ij}, 0) \}$, D_j – индекс первого ближайшего j -медоида. Выбранный набор медоидов характеризует конфигурацию кластеров, для которой может быть вычислена ее ЦФ стоимости.

SWAP. На каждой итерации производится перебор всех комбинаций пар медоид-немедоид и для них рассчитывается ЦФ стоимости смены текущей конфигурации как:

$$T_m(i, h) = \sum_{j \in o \setminus \{m \cup h\}} C_{jih}, \quad (10)$$

где C_{jih} вклад объекта j в перестановку медоида i и немедоида h . Данная операция производится $k(n-k)$ раз для каждой итерации. Из всех комбинаций выбирается та замена, стоимость которой максимальна. Критерий останова PAM $T_m(i, h) < 0$.

PAM обладает высокой точностью результатов кластеризации, устойчив к шумовым признакам и выбросам. Временная сложность алгоритма $O(n^2)$. Ключевой проблемой PAM является большое время выполнения на данных большой размерности и неэффективный поиск новых кандидатов на места медоидов.

Алгоритмы CLARA (Clustering Large Applications) и CLARANS (Randomized CLARA). Рассмотрим модификации PAM алгоритма CLARA и CLARANS, которые позволяют генерировать оптимальный набор медоидов из выборки текстовых данных на основе рандомизированной инициализации (рис. 3), что обеспечивает ускорение в $O(n)$ раз на фазе «SWAP» и дают точные результаты кластеризации [18].

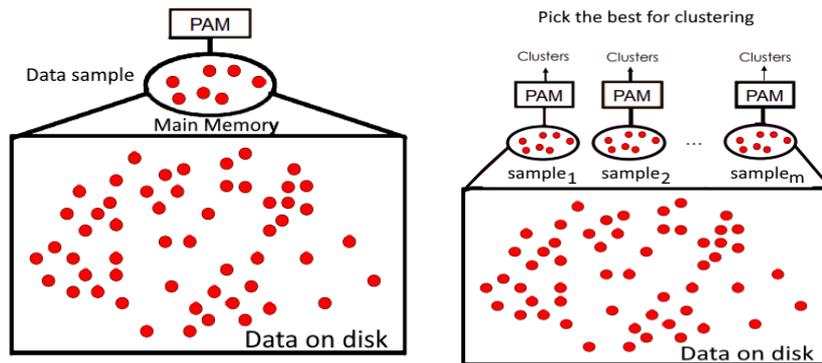


Рис. 3. Стратегия поиска медоидов алгоритмами CLARA и CLARANS

Вместо поиска медоидов для всего набора данных алгоритм CLARA случайным образом выбирает из исходного набора подмножество объектов данных фиксированного размера. Далее для создания оптимального набора медоидов выборки применяется алгоритм PAM, производится вычисление k репрезентативных объектов (медоидов), назначение каждого наблюдения из всего набора данных ближайшему медоиду, вычисление среднего (или суммы) различий наблюдений с их ближайшим медоидом и поиск центра кластера с наименьшей функцией стоимости из лучших k медоидов, кластеризованных для каждой выборки. CLARA повторяет процессы выборки и кластеризации заранее заданное количество раз. Окончательные результаты кластеризации соответствуют набору медоидов с минимальной стоимостью. Однако алгоритм может давать неправильные результаты кластеризации, если один или несколько выбранных медоидов отличаются от фактических лучших медоидов [8].

Алгоритм улучшает масштабируемость PAM, сокращает время вычислений при работе с большим набором данных, но качество кластеризации зависит от используемого метода формирования выборки, а на эффективность влияет ее размер и заданное количество результирующих кластеров. Временная сложность алгоритма равна $O(n^2 \log(n))$.

Алгоритм CLARANS комбинирует методы PAM и CLARA, выполняя рандомизированный поиск только в подмножестве набора данных и формулирует задачу кластеризации как случайный поиск в графе [14]. Из исходного графа $G = (X, E)$, представляющего собой разбиение множества данных на число кластеров выбирается подмножество вершин S и кластеризуется подграф. Каждая вершина представлена набором признаков объектов данных, каждому из которых может быть назначен k медоид. Вершины подграфа являются соседями (т. е. соединенными дугой), если их наборы отличаются только одним объектом. Далее поиск оптимального набора медоидов осуществляется по алгоритму PAM [23].

Суть эвристики CLARANS заключается в сокращении множества для выбора нового медоида из всего графа до множества замен медоид-немедоид в каждом кластеризованном подграфе G_m . Это позволяет искать оптимальную замену не по всем вершинам графа, а только по одному кластеру, перебирая s соседних вершин внутри него [7]. Если $T_m(i, h)$ выбранного соседа согласно формуле (10) выше, чем у текущей вершины CLARANS переходит к этому соседу и продолжает процесс случайного выбора и сравнения. Поиск решения останавливается в том узле, где достигается локальный минимум. CLARANS минимизирует ту же целевую функцию, что и CLARA, путем изменения текущих медоидов таким образом, что сумма различий внутри кластера сводится к минимуму. На вычисление новых медоидов может влиять шум, поэтому выбор медоидов в CLARANS диктуется расположением преобладающей доли объектов внутри кластера и, следовательно, он менее чувствителен к наличию выбросов [8].

Алгоритм CLARANS устраняет недостатки как алгоритмов K-medoids, так и алгоритмов CLARA, он поддерживает баланс между вычислительными затратами и влиянием выборки данных на формирование семантически однородных кластеров. Временная сложность алгоритма равна $O(n^2)$.

4. Алгоритм кластеризации на основе аффинного преобразования. Алгоритм кластеризации Affinity propagation (AP), основанный на распространении сходства, является еще одним представителем алгоритмов неиерархической кластеризации. Основная идея: AP получает на вход матрицу схожести между элементами набора данных и возвращает набор меток, присвоенных этим элементам. Изначально AP рассматривает все объекты данных как потенциальные центры кластеров – «образцы», а сходство объекта данных из входного набора с «образцом» определяет

вероятность того, что этот объект будет соотнесен с центром кластера. Алгоритм AP не требует предварительного задания числа кластеров, и как k -medoids, находит «экземпляры» на входном наборе, которые представляют начальное представление центра кластера. Алгоритм использует жадную стратегию, которая максимизирует значение глобальной функции сходства на каждой итерации [24].

Пусть $X = (x_1, x_2, \dots, x_n)$ кластеризуемый набор объектов данных без ограничений их внутренней структуры, $f = \text{sim}(x_i, x_j)$ – функция, количественно определяющая сходство между любыми двумя объектами по правилу: если $\text{sim}(x_i, x_j) > \text{sim}(x_i, x_k)$, то x_j больше похож на x_i , чем на x_k . Мера близости sim может вычисляться различными методами. Выполнение алгоритма сводится к итеративному выполнению двух этапов [8, 24].

1 этап – предварительная обработка каждого измерения и формирование матриц подобия: R – матрицы «ответственности» с элементами $r(i, k)$ – значениями близости между парами, которые представляют пригодность экземпляра x_i быть образцом x_i , по сравнению с другими кандидатами для x_i и A – матрицы «доступности» с элементами $a(i, k)$ значения которых представляют насколько «уместно» для x_i взять x_k как его экземпляр, принимая во внимание значения предпочтений других кандидатов данного «образца». Формализуем представления $r(i, k)$ и $a(i, k)$ как:

$$\begin{aligned} r(i, k) &\leftarrow s(i, k) - \max [a(i, k') + \text{sim}(i, k') \quad \forall k' \neq k] \text{ и} \\ a(i, k) &\leftarrow \min [0, r(k, k) + \sum_{i' \neq k, i' \notin \{i, k\}} r(i', k)]. \end{aligned} \quad (11)$$

2 этап – обновление матриц R и A . Начальное значение этих матриц для $r(i, k)$ и $a(i, k)$ установлены в ноль, их расчет и обновление происходит на каждой итерации, и повторяется до сходимости. Как обсуждалось в работе [22], чтобы избежать числовых колебаний при обновлении значений, в итерационный процесс вводится коэффициент демпфирования λ :

$$r_{t+1}(i, k) = \lambda * r_t(i, k) + (1 - \lambda) * r_{t+1}(i, k) \text{ и } a_{t+1}(i, k) = \lambda * a_t(i, k) + (1 - \lambda) * a_{t+1}(i, k), \quad (12)$$

где t – время итерации. Итерации выполняются до тех пор, пока либо границы кластера останутся неизменными на протяжении ряда итераций, либо не будет достигнуто заданное число итераций. Из окончательных матриц извлекаются такие экземпляры диагоналей, чья «ответственность + доступность» положительна.

Главный недостаток метода AP – его сложность. Алгоритм имеет временную сложность порядка $O(n^2T)$, где n – количество «образцов» и T – количество итераций до сходимости. Преимущества: простая и понятная идея алгоритма, нечувствительная к выбросам и количеству кластеров, не требующая предварительной настройки параметров.

5. Методы визуализации результатов кластеризации. Визуальный анализ многомерных текстовых данных это частный случай решения задачи нелинейного понижения размерности векторного пространства признаков. Методы визуализации позволяют осуществлять точный их перенос на плоскость или в трёхмерное пространство так, чтобы графическое представление наглядно отражало связанные между собой элементы кластерной структуры [25]. Из-за отсутствия однозначности в ее решении и предъявления различных требований к характеристикам получаемых визуализаций – остается актуальным вопрос о выборе метода визуального отображения многомерных данных (каждый из которых имеет свою целевую направленность, сильные и слабые стороны). Также очевидно, что совместное взаимодополняющее применение различных методов кластеризации и визуализации результатов позволяет провести более глубокий и многосторонний анализ исследуемых данных.

Дендограмма – это визуализатор представления результатов иерархической кластеризации в виде древовидной структуры, построенной по матрице мер близости между кластерами. Дендограмма является наиболее эффективной формой визуального представления и анализа структуры взаимосвязей между объектами при агломеративной кластеризации, а также наглядно демонстрирует в графическом виде последовательность их объединения или разделения [17]. Количество уровней дендограммы соответствует числу этапов слияния кластеров. На каждом этапе построения дендограммы вычисляются расстояния между кластерами и производится пересчет расстояния между образованными кластерами $D(C_i, C_j) = \text{dist}(x, y)$. В зависимости от специфики решаемой задачи для вычисления расстояния $D(C_i, C_j)$ между кластерами используются различные функции связи, рассмотренные ранее в пункте 2.

Дендограмма позволяет представлять зависимости между множеством объектов с любым числом заданных характеристик на двумерном графике, где по одной из осей откладываются все объекты, а по другой D_t – расстояние между кластерами, выбранными на шаге t для объединения (рис. 4). Для того, чтобы любой кластер мог быть представлен в виде непрерывного отрезка на оси объектов и ребра не пересекались (исключение самопересечений), на D_t накладывается ограничение монотонности.

Функция расстояния D_t является монотонной, если на каждом следующем шаге расстояние между кластерами не уменьшается: $D_2 \leq D_3 \leq \dots \leq D_m$ и, для коэффициентов в формуле Ланса-Уильямса верна теорема Миллигана [7]. Следующим ограничением являются свойства растяжения $D_t \geq \text{dist}(\mu_x, \mu_y), \forall t$ и сжатия $D_t \leq \text{dist}(\mu_x, \mu_y), \forall t$, которые способствуют четкому отделению кластеров.

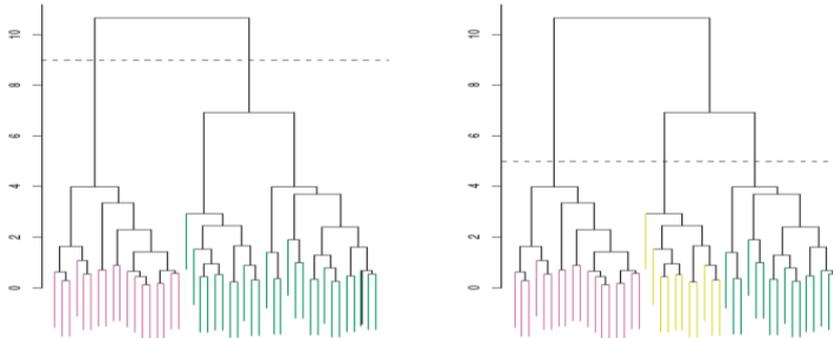


Рис. 4. Дендограмма с отсечением кластеров

Для определения числа кластеров находится интервал максимальной длины $|Dt+1-Dt|$. В качестве итоговых результатов выдаются кластеры, полученные на шаге t . При этом число кластеров равно $m-t+1$.

Для оценки производительности неиерархической кластеризации на наборах текстовых данных, их сравнительного анализа и интерпретации результатов используются средства непараметрической визуализации, поддерживающие следующие методы: t-SNE (t-distributed Stochastic Neighbor Embedding), Sammon, Isomap и локально линейное встраивание (LLE) [14, 25]. Данные алгоритмы предоставляют различные методы оценки внутренней связности и анализа скрытых закономерностей на множестве элементов кластерной структуры. Основными способами непараметрической визуализации на основе данных методов являются диаграммы рассеивания (разброса), диаграммы сходства и диаграммы интенсивности (тепловые карты).

Диаграммы рассеивания – способ интерактивной визуализации кластеризации многомерных текстовых данных путем проецирования их в низкоразмерное двух- или трехмерное пространство. На диаграмме каждому многомерному объекту соответствует точка, координаты которой равны значениям двух или более параметров этого наблюдения [13]. Если предполагается, что один из параметров зависит от другого, то сходные объекты моделируются близлежащими точками, а разнородные объекты моделируются удаленными точками с высокой вероятностью. Диаграмма рассеивания позволяет отражать несколько зависимостей: параметры одной переменной X откладываются по горизонтальной оси, а по вертикальной оси – значения нескольких переменных Y . Для каждой переменной Y используется разный цвет и вид точек (рис. 5, 6).

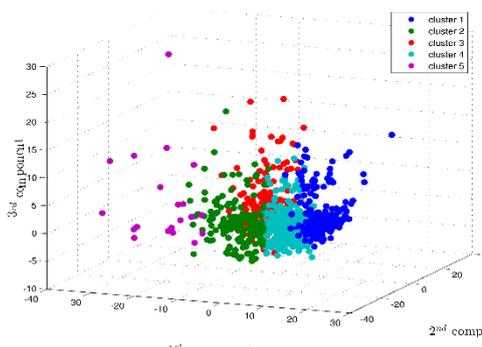


Рис. 5. Трехмерное представление диаграммы

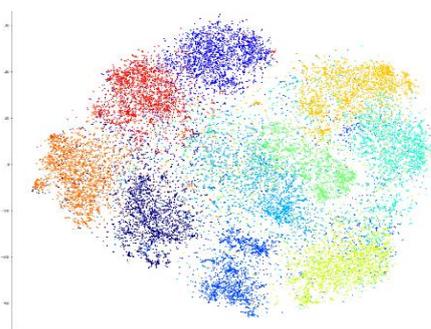


Рис. 6. Двухмерное представление диаграммы

Исследование диаграмм рассеивания позволяет определять формы зависимостей в связанных между собой данных кластерной структуры, проводить их анализ и уточнение параметров метода кластеризации. Другое важное преимущество диаграмм рассеивания состоит в том, что они позволяют находить «выбросы» (нетипичные данные), которые искусственным образом увеличивают или уменьшают («смещают») коэффициент корреляции.

Диаграмма сходства VOS (Visualization of Similarities) – это еще один инструмент исследования качества моделей кластеризации, основанный на визуализации зависимостей между многомерными данными [26]. Цель VOS – обеспечить низкоразмерную визуализацию, в которой объекты расположены таким образом, чтобы расстояние между любой парой объектов максимально точно отражало их сходство. Объекты распределяются по кластерам путем максимизации функции качества. Графическое представление результатов кластеризации текстовых данных показано на рис. 7, а визуальный анализ связности терминов для сформированных кластеров 2, 7, 8, 10 на рис. 8. На диаграммах VOS размер объекта характеризует общую мощность связей («total link strength»), а ширина линий – мощность связи («link strength») между двумя терминами.

Анализ диаграммы помогает исследовать следующие аспекты данных: различные уровни корреляции между точками данных необходимы для понимания взаимосвязи внутри данных; для заданных данных можно провести линию наилучшего соответствия и использовать ее для дальнейшего прогнозирования новых значений данных; точки данных, лежащие за пределами данного набора, могут быть легко идентифицированы, чтобы найти выбросы; группировку точек данных на диаграмме сходства можно определить как разные кластеры в данных.

близости данных [27]. Для оценки кластеризации будем использовать в качестве внешних мер – скорректированный индекс Ранда (Adjusted Rand Index - ARI) и V-меру, а внутренних – коэффициент «силуэта».

V-мера является средним гармоническим значением между двумя другими метриками: гомогенностью и полнотой [19]. Однородность характеризует степень соответствия тому, что каждый кластер содержит только объекты одного класса. Полнота показывает, в какой мере все объекты данного класса собраны в одном кластере. $V = 2 * Homogeneity * Completeness / (Homogeneity + Completeness)$. Величина V-меры изменяется в диапазоне от 0 до 1. V-мера комбинирует в себе гомогенность и полноту таким образом, чтобы максимизация итоговой метрики не приводила к тривиальным решениям.

ARI – это метрика качества, построенная на основе меры близости, вычисляющей семантическое сходство между двумя кластерами. ARI измеряет степень соответствия между двумя распределениями данных, которые находятся как в одном кластере, так и в разных кластерах, с поправкой на случайность [12].

$$ARI = \frac{a+b}{C_2^{n_{samples}}}, \quad (13)$$

где a – количество объектов, которые фактически относятся к одному и тому же кластеру и предсказаны для одного кластера; b – количество объектов, которые и верны, и предсказаны в разных кластерах; в знаменателе количество комбинаций любых двух объектов, попадающих в один кластер, которые могут быть сформированы в наборе данных. Диапазон значений ARI составляет [0,1]. Чем больше значение, тем более согласован результат кластеризации с реальной ситуацией.

Коэффициент «силуэта» – метрика кластеризации, не требующая разметки, измеряет схожесть объектов своего кластера по сравнению с объектами других кластеров. Коэффициент определяется для каждого объекта выборки, а метрика определяется как средний коэффициент «силуэта» для всех объектов выборки [6].

$$Sil(C) = \frac{1}{N} \sum_{c_k \in C} \sum_{x_i \in c_k} \frac{z(x_i, c_k) - q(x_i, c_k)}{\max\{q(x_i, c_k), z(x_i, c_k)\}}, \quad (14)$$

где $q(x_i, c_k)$ – компактность распределения, определяется как среднее расстояние близости $x_i \in c_k$ до других объектов кластера c_k , $z(x_i, c_k)$ – отделимость, рассчитывается как среднее расстояние близости от $x_i \in c_k$ до объектов другого кластера $c_l, l \neq k$. Коэффициент принимает значения $-1 \leq Sil(C) \leq 1$ и максимизируется, когда кластеры «кучные» и хорошо отделены друг от друга.

Данные, используемые для экспериментального исследования, были получены из текстового корпуса базы данных научных публикаций Hindawi, находящейся в открытом доступе. В качестве предметной области для проверки результатов работы алгоритмов кластеризации была сформирована выборка из 11628 оригинальных исследовательских статей с термином «Knowledge management» в названиях, аннотациях и терминах индексирования (ключевых словах). Для фильтрации текстов был разработан парсер текстов, реализующий возможность выполнять операцию нахождения основы слова с помощью алгоритмов стемминга и лемматизации. Составление векторов текстов в виде моделей «документ-термин с TF-IDF», «документ - ассоциативно-семантическая группа с TF-IDF» выполнялось алгоритмом Sem_group из библиотеки scikit-learn [13].

Для сравнения эффективности методов агломеративной и неиерархической кластеризации были выбраны алгоритмы: «k-medoids:PAM», «Affinity Propagation», «ROCK» и «CHAMELEON». Так как алгоритм «k-medoids:PAM» определяет необходимость задания количества кластеров, для него исследуемая выборка текстовых данных была размечена в полуавтоматическом режиме и со-

ставила 80 кластеров различных смысловых тематик. Алгоритмы «Affinity Propagation», «ROCK» и «CHAMELEON» без предварительной параметризации определили количество кластеров в процессе эксперимента. На следующем этапе проведены исследования по сравнению качества разбиения данными алгоритмами кластерного анализа на сформированной выборке текстовых данных, для которой был предварительно размечен словарь терминов на 1000 элементов и выявлено 59 наиболее часто встречающихся терминов. Результаты сравнения алгоритмов кластеризации по внешним и внутренним метрикам качества и времени работы представлены в табл. 1.

Таблица 1

Сравнение результатов кластеризации текстовых данных различными алгоритмами

Алгоритм кластеризации	Алгоритм векторизации	Время, с	V-мера	ARI	Коэфф. «силуэта»
k-medoids:PAM	tf_idf	397	0,721	0,357	0,544
k-medoids:PAM	sem_group	457	0,753	0,360	0,549
CHAMELEON	tf_idf	1378	0,714	0,329	0,752
CHAMELEON	sem_group	1056	0,741	0,336	0,856
ROCK	tf_idf	986	0,613	0,128	0,349
ROCK	sem_group	875	0,661	0,135	0,351
Affinity Propagation	tf_idf	197	0,762	0,208	0,434
Affinity Propagation	sem_group	236	0,794	0,334	0,536

Из представленных в табл. 1 результатов можно сделать вывод о том, что все алгоритмы показывают достаточно близкую точность работы. По временным затратам алгоритм «Affinity Propagation» показал лучший результат, он на 200 секунд быстрее «k-medoids:PAM». По метрике «V-мера», характеризующей семантическую однородность распределения статей в кластеры алгоритмы кластеризации «k-medoids:PAM» со значением 0,753 и «Affinity Propagation» со значением 0,762 представили качественное разбиение. Лучший результат измерения согласованности между результатами кластеризации и данными реальных категорий получен алгоритмом кластеризации «k-medoids:PAM» со значением «Adjusted Rand Index» – 0,360. Статистически значимый показатель локальной плотности по метрике «силуэт» показал агломеративный алгоритм CHAMELEON – коэффициент 0,856, что говорит о высоком уровне формирования семантически плотных кластеров и хорошем их разделении.

Алгоритм «k-medoids:PAM» производит кластеризацию текстовой выборки с выделением обобщенных тематических направлений. Алгоритмом «Affinity Propagation» определяются подтемы, так, что узко специфические тематики находятся в каждом кластере. Алгоритм «CHAMELEON» по результатам кластеризации близок к «k-medoids:PAM», и для всех алгоритмов векторизации точно распределяет темы статей по кластерам. Алгоритм «ROCK» показывает слабые результаты кластеризации, статьи по одному тематическому направлению оказываются в разных кластерах, а во многих кластерах определились статьи по нескольким различным тематикам. Это может быть объяснено тем, что в аннотациях слишком кратко изложена сущность текстового документа или используются специализированные термины.

Так же по результатам работы всех алгоритмов видно, что качество кластеризации при использовании модели представления текста «документ – ассоциативно-семантическая группа с TF-IDF» значительно улучшается.

По результатам проведенной кластеризации для визуального анализа и оценки семантической связности элементов кластерной структуры из словаря были выбраны 835 ключевых терминов с частотой упоминания, превышающей или равной пяти. Для снижения полученного пространства признаков использовался алгоритм t-SNE, который позволяет выполнять визуализацию элементов многомерного пространства путём снижения размерности до двух [25]. Так как построенные вектора текстов имеют большую размерность, длина их векторов зависит от длины набора терминов, которые выбраны в качестве основных признаков. Диаграмма сходства, построенная на основе t-SNE для выбранных терминов представлена на рис. 11. Графическое представление получено с использованием VOSviewer программного инструмента для построения и интерактивной визуализации результатов кластерного анализа текстовых данных [26].

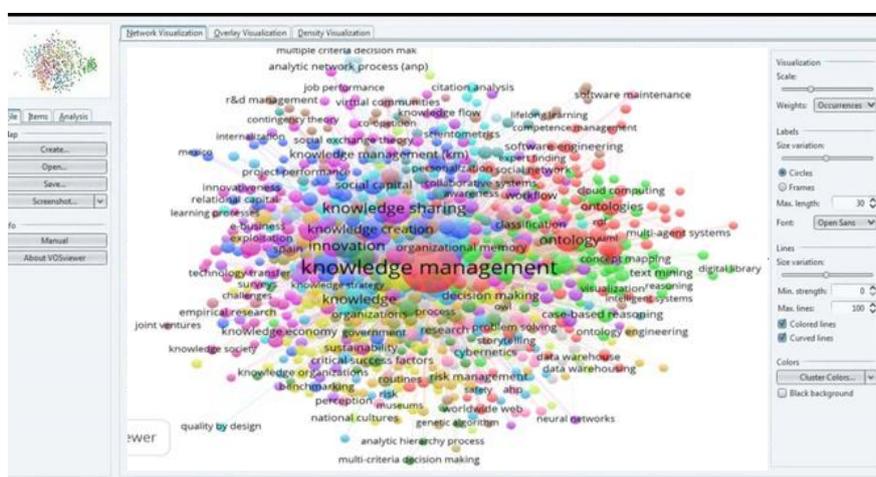


Рис. 11. Диаграмма VOS

Интерпретация кластеров основана на встречающихся в них ключевых словах, которые тесно связаны внутри одних и тех же кластеров, но слабо связаны между разными кластерами. На основе визуального анализа частоты встречаемости терминов определена мощность (числовое значение близости данных) семантической связи между элементами кластеров и плотность каждого кластера по количеству связанных элементов, что позволило выделить шумовые признаки и выбросы.

Заключение. В работе рассмотрена одна из важных задач искусственного интеллекта - кластерный анализ слабоструктурированных текстовых данных. В статье проведен обзор и анализ модифицированных алгоритмов агломеративной кластеризации CURE, ROCK, CHAMELEON, неиерархической кластеризации k-medoids:PAM, CLARA, CLARANS и алгоритма аффинного преобразования. Выявлены их достоинства и недостатки, а также рассмотрены особенности их реализации. Рассмотрены методы интерактивной визуализации результатов кластеризации, оценки внутренней связности и анализа скрытых закономерностей на множестве элементов кластерной структуры: дендограммы, диаграммы рассеивания, диаграммы сходства VOS и диаграммы интенсивности. Данные способы графического представления позволяют работать с многомерными наборами данных для снижения размерности векторного пространства признаков. Проведены эксперимен-

тальные исследования рассмотренных алгоритмов кластеризации текстовых данных. Сравнительный анализ алгоритмов показал, что наилучшее качество при наименьшем времени работы показывают неиерархические алгоритмы – «k-medoids:PAM» и «Affinity Propagation» с использованием модели векторизации «документ – лексико-семантическая группа с TF-IDF». Проведенный обзор и анализ показал, что для повышения эффективности кластеризации текстовых данных необходимо использовать гибридные подходы и метаэвристические алгоритмы.

Исследование выполнено за счет гранта Российского научного фонда № 22-21-00316, <https://rscf.ru/project/22-21-00316/> в Южном федеральном университете.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Junkai Yi, Yacong Zhang, Xianghui Zhao, Jing Wan* A Novel Text Clustering Approach Using Deep-Learning Vocabulary Network // *Mathematical Problems in Engineering*. – 2017. – Vol. 2017. – 13 p.
2. *Yujia Sun, Jan Platoš*. High-Dimensional Text Clustering by Dimensionality Reduction and Improved Density Peak // *Wireless Communications and Mobile Computing*. – 2020. – Vol. 2020. Article ID 8881112. – 16 p.
3. *Бова В.В., Кулиев Э.В., Щеглов С.Н.* Метод семантической кластеризации распределенных ресурсов знаний с динамическими компонентами на основе контентной фильтрации // *Информатика, вычислительная техника и инженерное образование*. – 2019. – № 1 (34).
4. *Bova V., Kureichik V., Leshchanov D.* The model of semantic similarity estimation for the problems of big data search and structuring // *11th IEEE International Conference AICT 2017*. – P. 27-32.
5. *Zhang W., Tang X., Yoshida T.* TESC: an approach to TExt classification using Semi-supervised Clustering // *Knowledge-Based Systems*. – 2015. – Vol. 75. – P. 152-160.
6. *Wei T., Lu Y., Chang H., Zhou Q., Bao X.* A semantic approach for text clustering using WordNet and lexical chains // *Expert Systems with Applications*. – 2015. – Vol. 42, No. 4. – P. 2264-2275.
7. *Xu D., Tian Y.* A Comprehensive Survey of Clustering Algorithms // *Ann. Data. Sci.* – 2015. – No. 2. – P. 165-193.
8. *Sabha Firdaus, Md. Ashraf Uddin* A Survey on Clustering Algorithms and Complexity Analysis // *International Journal of Computer Science Issues*. – 2015. – Vol. 12. Issue 2. – P. 62-85.
9. *Sara Saad Soliman, Maged F. El-Sayed, Yasser F. Hassan* Semantic Clustering of Search Engine Results // *The Scientific World Journal*. – 2015. – Vol. 2015. Article ID 931258. – 9 p.
10. *Kravchenko Y.A., Rodzin S.I., Kuliev E.V., Bova V.V.* Simulation of the semantic network of knowledge representation in intelligent assistant systems based on ontological approach // *Communications in Computer and Information Science* this link is disabled. – 2021. – 1396 CCIS. – P. 241-252.
11. *Kravchenko Y., Bova V.* Assessment of ontological structures semantic similarity based on a modified cuckoo search algorithm // *IOP Conference Series: Materials Science and Engineering*. – 2020. – No. 12018.
12. *Отрадных К.К., Раев В.К.* Экспериментальное исследование эффективности методик векторизации текстовых документов и алгоритмов их кластеризации // *Вестник РГРТУ*. – 2018. – № 64. – С. 73-84.
13. *Zhou S., Xu X., Liu Y., Chang R., Xiao Y.* Text similarity measurement of semantic cognition based on word vector distance decentralization with clustering analysis // *IEEE Access*. – 2019. – Vol. 7. – P. 107247-107258.
14. *Krömer P., Platoš J.* Cluster analysis of data with reduced dimensionality: an empirical study // *Intelligent Systems for Computer Modelling*. – Springer. Cham, 2016. – P. 121-132.
15. *Бова В.В., Щеглов С.Н., Лещанов Д.В.* Модифицированный алгоритм EM-кластеризации для задач интегрированной обработки больших данных // *Известия ЮФУ. Технические науки*. – 2018. – № 4 (165). – С. 197-211.

16. *Jingdong Yan, Wuwei Liu.* An Ensemble Clustering Approach (Consensus Clustering) for High-Dimensional Data // *Security and Communication Networks.* – 2022. – Vol. 2022. Article ID 5629710. – 9 p.
17. *Olson C.F.* Parallel algorithms for hierarchical clustering // *Pattern Analysis & Machine Intelligence IEEE Transactions on.* – 2016. – Vol. 12, No. 11. – P. 1088-1092.
18. *Xueli X.U., Xuejing Z.* Application of sparse spectral clustering algorithm in high-dimensional data // *Journal of University of Science and Technology of China.* – 2017. – Vol. 47. – P. 311-319.
19. *Пархоменко П.А., Григорьев А.А., Астраханцев Н.А.* Обзор и экспериментальное сравнение методов кластеризации текстов // *Тр. ИСП РАН.* – 2017. – Т. 29. – Вып. 2. – С. 161-200.
20. *Guha S., Rastogi R. Shim K.* CURE: an efficient clustering algorithm for large databases // *ACM SIGMOD. Rec.* – 2017. – Vol. 27. – P. 73-84.
21. *Guha S., Rastogi R. Shim K.* ROCK: a robust clustering algorithm for categorical attributes // *Proceedings of the 15th international conference on data engineering.* – 2016. – P. 512-521.
22. *Karypis G. Han E. Kumar V.* Chameleon: hierarchical clustering using dynamic modeling // *ACM SIGMOD Rec.* – Aug. 1999. – Vol. 32. – P. 68-75.
23. *Huu Hiep Nguyen.* Clustering Categorical Data Using Community Detection Techniques // *Computational Intelligence and Neuroscience.* – 2017. – Vol. 2017. Article ID 8986360. – 11 p.
24. *Yancheng He, Qingcai Chen, Xiaolong Wang, Ruifeng Xu.* An adaptive affinity propagation document clustering // *The 7th International Conference on Informatics and Systems.* – 2010.
25. *Laurens van der Maaten, Geoffrey Hinton* Visualizing data using t-SNE // *Journal of Machine Learning Research.* – 2008. – No. 9. – P. 2579-2605.
26. *Van Eck N.J., Waltman L.* Software survey: VOSviewer, a computer program for bibliometric mapping // *Scientometrics.* – 2015. – Vol. 84, No. 2. – P. 523-538.
27. *Бова В.В., Лещанов Д.В.* Метод оценки эффективности семантической кластеризации гипертекстовых динамических структур на основе DOM-фильтра // *Тр. конгресса IS&IT.* – 2018. – Т. 2. – С. 59-70.

REFERENCES

1. *Junkai Yi, Yacong Zhang, Xianghui Zhao, Jing Wan* A Novel Text Clustering Approach Using Deep-Learning Vocabulary Network, *Mathematical Problems in Engineering*, 2017, Vol. 2017, 13 p.
2. *Yujia Sun, Jan Platoš.* High-Dimensional Text Clustering by Dimensionality Reduction and Improved Density Peak, *Wireless Communications and Mobile Computing*, 2020, Vol. 2020. Article ID 8881112, 16 p.
3. *Bova V.V., Kuliev E.V., Shcheglov S.N.* Metod semanticheskoy klasterizatsii raspredelennykh resursov znaniy s dinamicheskimi komponentami na osnove kontentnoy fil'tratsii [The method of semantic clustering of distributed knowledge resources with dynamic components based on content filtering], *Informatika, vychislitel'naya tekhnika i inzhenernoe obrazovanie* [Informatics, computer engineering and engineering education], 2019, No. 1 (34).
4. *Bova V., Kureichik V., Leshchanov D.* The model of semantic similarity estimation for the problems of big data search and structuring, *11th IEEE International Conference AICT 2017*, pp. 27-32.
5. *Zhang W., Tang X., Yoshida T.* TESC: an approach to TExt classification using Semi-supervised Clustering, *Knowledge-Based Systems*, 2015, Vol. 75, pp. 152-160.
6. *Wei T., Lu Y., Chang H., Zhou Q., Bao X.* A semantic approach for text clustering using WordNet and lexical chains, *Expert Systems with Applications*, 2015, Vol. 42, No. 4, pp. 2264-2275.
7. *Xu D., Tian Y.* A Comprehensive Survey of Clustering Algorithms, *Ann. Data. Sci.*, 2015, No. 2, pp. 165-193.
8. *Sabha Firdaus, Md. Ashraf Uddin* A Survey on Clustering Algorithms and Complexity Analysis, *International Journal of Computer Science Issues*, 2015, Vol. 12. Issue 2, pp. 62-85.
9. *Sara Saad Soliman, Maged F. El-Sayed, Yasser F. Hassan* Semantic Clustering of Search Engine Results, *The Scientific World Journal*, 2015, Vol. 2015. Article ID 931258, 9 p.
10. *Kravchenko Y.A., Rodzin S.I., Kuliev E.V., Bova V.V.* Simulation of the semantic network of knowledge representation in intelligent assistant systems based on ontological approach, *Communications in Computer and Information Science this link is disabled*, 2021, 1396 CCIS, pp. 241-252.

11. Kravchenko Y., Bova V. Assessment of ontological structures semantic similarity based on a modified cuckoo search algorithm, *IOP Conference Series: Materials Science and Engineering*, 2020, No. 12018.
12. Otradnov K.K., Raev V.K. Eksperimental'noe issledovanie effektivnosti metodik vektorizatsii tekstovykh dokumentov i algoritmov ikh klasterizatsii [Experimental study of the effectiveness of methods of vectorization of text documents and algorithms for their clustering], *Vestnik RGRU* [Vestnik of RSREU], 2018, No. 64, pp. 73-84.
13. Zhou S., Xu X., Liu Y., Chang R., Xiao Y. Text similarity measurement of semantic cognition based on word vector distance decentralization with clustering analysis, *IEEE Access*, 2019, Vol. 7, pp. 107247-107258.
14. Krömer P., Platoš J. Cluster analysis of data with reduced dimensionality: an empirical study, *Intelligent Systems for Computer Modelling*. Springer. Cham, 2016, pp. 121-132.
15. Bova V.V., Shcheglov S.N., Leshchanov D.V. Modifitsirovannyi algoritm EM-klasterizatsii dlya zadach integrirovannoy obrabotki bol'shikh dannykh [Modified EM clustering algorithm for integrated big data processing tasks], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2018, No. 4 (165), pp. 197-211.
16. Jingdong Yan, Wuwei Liu. An Ensemble Clustering Approach (Consensus Clustering) for High-Dimensional Data, *Security and Communication Networks*, 2022, Vol. 2022. Article ID 5629710, 9 p.
17. Olson C.F. Parallel algorithms for hierarchical clustering, *Pattern Analysis & Machine Intelligence IEEE Transactions on*, 2016, Vol. 12, No. 11, pp. 1088-1092.
18. Xueli X.U., Xuejing Z. Application of sparse spectral clustering algorithm in high-dimensional data, *Journal of University of Science and Technology of China*, 2017, Vol. 47, pp. 311-319.
19. Parkhomenko P.A., Grigor'ev A.A., Astrakhantsev N.A. Obzor i eksperimental'noe sravnenie metodov klasterizatsii tekstov [Review and experimental comparison of text clustering methods], *Tr. ISP RAN* [Proceedings of ISP RAS], 2017, Vol. 29, Issue 2, pp. 161-200.
20. Guha S., Rastogi R. Shim K. CURE: an efficient clustering algorithm for large databases, *ACM SIGMOD Rec.*, 2017, Vol. 27, pp. 73-84.
21. Guha S., Rastogi R. Shim K. ROCK: a robust clustering algorithm for categorical attributes, *Proceedings of the 15th international conference on data engineering*, 2016, pp. 512-521.
22. Karypis G. Han E. Kumar V. Chameleon: hierarchical clustering using dynamic modeling, *ACM SIGMOD Rec.*, Aug. 1999, Vol. 32, pp. 68-75.
23. Huu Hiep Nguyen. Clustering Categorical Data Using Community Detection Techniques, *Computational Intelligence and Neuroscience*, 2017, Vol. 2017. Article ID 8986360, 11 p.
24. Yancheng He, Qingcai Chen, Xiaolong Wang, Rui Feng Xu. An adaptive affinity propagation document clustering, *The 7th International Conference on Informatics and Systems*, 2010.
25. Laurens van der Maaten, Geoffrey Hinton Visualizing data using t-SNE, *Journal of Machine Learning Research*, 2008, No. 9, pp. 2579-2605.
26. Van Eck N.J., Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping, *Scientometrics*, 2015, Vol. 84, No. 2, pp. 523-538.
27. Bova V.V., Leshchanov D.V. Metod otsenki effektivnosti semanticheskoy klasterizatsii gipertekstovykh dinamicheskikh struktur na osnove DOM-fil'tra [A method for evaluating the effectiveness of semantic clustering of hypertext dynamic structures based on a DOM filter], *Tr. kongressa IS&IT* [Proceedings of the Congress IS&IT], 2018, Vol. 2, pp. 59-70.

Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Гатчин.

Бова Виктория Викторовна – Южный федеральный университет; e-mail: vvbova@sfedu.ru; г. Таганрог, Россия; тел.: 88634371651; кафедра систем автоматизированного проектирования; доцент.

Кравченко Юрий Алексеевич – e-mail: yakravchenko@sfedu.ru; кафедра систем автоматизированного проектирования; доцент.

Родзин Сергей Иванович – e-mail: srodzin@sfedu.ru; тел.: 88634371673; кафедра МОП ЭВМ; профессор.

Bova Victoria Victorovna – Southern Federal University; e-mail: vvbova@sfnedu.ru; Taganrog, Russia; phone: +78634371651; the department of computer aided design; associate professor.

Kravchenko Yury Alekseevich – e-mail: yakravchenko@sfnedu.ru; the department of computer aided design; associate professor.

Rodzin Sergey Ivanovich – e-mail: srodzin@sfnedu.ru; phone: +78634371673; the department of software engineering; professor.

УДК 004.896

DOI 10.18522/2311-3103-2022-4-143-157

Б.К. Лебедев, О.Б. Лебедев, Е.О. Лебедева

ЭВОЛЮЦИОННЫЙ ПОПУЛЯЦИОННЫЙ МЕТОД РЕШЕНИЯ ТРАНСПОРТНОЙ ЗАДАЧИ*

Рассматривается эволюционный популяционный метод решения транспортной задачи на основе метаэвристики кристаллизации россыпи альтернатив. Исследуется закрытая (или сбалансированная) модель транспортной задачи: сумма груза у поставщиков равно общей сумме потребностей в пунктах назначения. Цель оптимизации – минимизация стоимости (достижение минимума затрат на перевозку) или расстояний и критерий времени (затрачивается минимум времени на перевозку). В основу метаэвристики кристаллизации россыпи альтернатив положена стратегия, основанная на запоминании и повторении прошлых успехов. Стратегия делает упор на «коллективную память», под которой подразумевается любой вид информации, которая отражает прошлую историю развития и хранится независимо от индивидуумов. В качестве кода решения транспортной задачи рассматривается упорядоченная последовательность D_k маршрутов. Объектами являются маршруты, альтернативами – множество позиций P в списке, где n_p – число позиций в списке D_k . Множество объектов D_k соответствует множеству всех маршрутов. Множество альтернативных состояний P объекта соответствует множеству альтернативных вариантов размещения объекта списке D_k . Работа популяционного эволюционного алгоритма кристаллизации россыпи альтернатив опирается на коллективную эволюционную память, называемую россыпью альтернатив. Под россыпью альтернатив решения в работе называется структура данных, используемая в качестве коллективной эволюционной памяти, несущая информацию о решении, включающую сведения о реализованных альтернативах агентов в данном решении и о полезности решения. Разработан конструктивный алгоритм формирования опорного плана путем декодирования списка D_k . На каждом шаге t решается задача выбора очередного в последовательности D_k маршрута и определения количества груза, перевозимого из пункта отправления A_i в пункт назначения B_j по этому маршруту. Разработанный алгоритм является популяционным, реализующим стратегию случайного направленного поиска. Каждый агент является кодом некоторого решения транспортной задачи. На первом этапе каждой итерации l конструктивным алгоритмом на базе интегральной россыпи альтернатив формируется n_k кодов решений D_k . Формирование каждого кода решения D_k выполняется последовательно по шагам путем последовательного выбора объекта и позиции. Для построенного кода решения D_k рассчитывается оценка решения ξ_k и оценка полезности δ_k . Формируется индивидуальная россыпь альтернатив R_k и переход к построению следующего кода решения. На втором этапе итерации производится суммирование интегральной россыпи альтернатив, сформированной на предыдущих итерациях от l до $(l-1)$, со всеми индивидуальными россыпями альтернатив, сформированных на итерации l . На третьем этапе итерации l производится снижение всех интегральных оценок полезности $r_{\alpha\beta}^$ интегральной россыпи альтернатив $R^*(l)$ на величину δ^* . Алгоритм решения транспортной задачи был реализован на языке C++ в среде Windows. Сравнение значений критерия, на тестовых примерах, с*

* Работа выполнена при финансовой поддержке гранта РФФИ № 20-07-00260 А.

известным оптимумом показало, что у 90% примеров полученное решение было оптимальным, у 2% примеров решения были на 5% хуже, а у 8% примеров решения отличались менее, чем на 2%. Временная сложность алгоритма, полученная экспериментальным путем, лежит в пределах $O(n^2)$.

Транспортная задача; метаэвристика; кристаллизация россыпи альтернатив; оптимизация; популяционный алгоритм; коллективная память; агент; направленный поиск.

В.К. Lebedev, O.B. Lebedev, E.O. Lebedeva

EVOLUTIONARY POPULATION METHOD FOR SOLVING THE TRANSPORT PROBLEM

The paper considers an evolutionary population method for solving a transport problem based on the metaheuristics of crystallization of a placer of alternatives. We study a closed (or balanced) model of the transport problem: the amount of cargo from suppliers is equal to the total amount of needs at destinations. The goal of optimization is to minimize the cost (achieving a minimum of transportation costs) or distances and the criterion of time (a minimum of time is spent on transportation). The metaheuristics of the crystallization of a placer of alternatives is based on a strategy based on remembering and repeating past successes. The strategy emphasizes «collective memory», which refers to any kind of information that reflects the past history of development and is stored independently of individuals. An ordered sequence D_k of routes is considered as a code for solving the transport problem. The objects are routes, the alternatives are the set of positions P in the list, where n_p is the number of positions in the list D_k . The set of objects D_k corresponds to the set of all routes. The set of alternative states P of the object corresponds to the set of alternative options for placing the object in the list D_k . The operation of the population evolutionary algorithm for the crystallization of a placer of alternatives is based on a collective evolutionary memory called a placer of alternatives. A scattering of solution alternatives is a data structure used as a collective evolutionary memory that carries information about the solution, including information about the realized alternatives of agents in this solution and about the usefulness of the solution. A constructive algorithm for the formation of a reference plan by decoding the list D_k has been developed. At each step t , the problem of choosing the next route in the sequence D_k and determining the amount of cargo transported from the point of departure A_i to the point of destination B_j along this route is solved. The developed algorithm is population-based, implementing the strategy of random directed search. Each agent is a code for some solution of the transport problem. At the first stage of each iteration l , a constructive algorithm based on the integral placer of alternatives generates n_k decision codes D_k . The formation of each decision code D_k is performed sequentially in steps by sequentially selecting an object and position. For the constructed solution code D_k , the solution estimate ζ_k and the utility estimate δ_k are calculated. An individual scattering of alternatives R_k is formed and a transition to the construction of the next solution code is formed. At the second stage of the iteration, the integral placer of alternatives formed at previous iterations from l to $(l-1)$ is summed with all individual placers of alternatives formed at iteration l . At the third stage of iteration l , all integral utility estimates r_{ab}^* of the integral placer of alternatives $R^*(l)$ are reduced by δ^* . The algorithm for solving the transport problem was implemented in C++ in the Windows environment. Comparison of the values of the criterion, on test examples, with a known optimum showed that in 90% of the examples the solution obtained was optimal, in 2% of the examples the solutions were 5% worse, and in 8% of the examples the solutions differed by less than 2%. The time complexity of the algorithm, obtained experimentally, lies within $O(n^2)$.

Transport problem; metaheuristics; crystallization of a placer of alternatives; optimization; population algorithm; collective memory; agent; directed search.

Введение. Транспортная задача (классическая) – задача об оптимальном плане перевозок однородного продукта из однородных пунктов наличия в однородные пункты потребления на однородных транспортных средствах (предопределённом количестве) со статическими данными и линейном подходе (это основные условия задачи) [1, 2].

Распространенность в приложениях задач транспортного типа оправдывает неослабевающее внимание к ним. Транспортная задача является по теории сложности вычислений *NP*-трудной и входит в класс сложности *NP*. Транспортные задачи относятся к задачам линейного программирования и могут решаться симплекс-методом.

Когда суммарный объём предложений (грузов, имеющих в пунктах отправления) не равен общему объёму спроса на товары (грузы), запрашиваемые пунктами потребления, транспортная задача называется несбалансированной.

Для того чтобы решить транспортную задачу используется множество методов [4, 5]. В зависимости от особенностей задачи используются метод потенциалов, дифференциальных рент (для поиска оптимального плана), метод северо-западного угла, метод наименьшего элемента (для поиска опорного плана) и т.д. Одним из подходов к получению качественного решения задачи предполагает итерационное улучшение плана перевозок. Метод содержит три последовательных этапа [4, 5]:

1. Формирование опорного плана.
2. Проверка опорного плана на оптимальность.
3. Переход к новому опорному плану, если предыдущий не оптимален.

На первом этапе решается задача нахождения опорного плана. Опорным решением транспортной задачи называется любое допустимое решение, для которого векторы условий, соответствующие положительным координатам, линейно независимы.

После построения опорного плана перевозок нужно применить один из алгоритмов его улучшения, приближения к оптимальному результату.

Альтернативным подходом к решению этих задач является применение методов, инспирированными искусственными или естественными системами, способными за приемлемое время найти оптимум в пространстве поиска решений большой размерности [6–8].

Среди них особенно активно развиваются поведенческие методы роевого интеллекта (Swarm Intelligence) [6, 9]. Базовыми идеями поведенческих методов являются: децентрализованность, взаимодействие агентов, коллективная адаптация, простота поведения агентов. Являясь по своей сути итерационными, алгоритмы на основе моделирования естественных процессов отличаются от обычных итерационных процедур «слепого поиска». В основе работы этих алгоритмов лежат метаэвристики поиска в пространстве состояний (решений). Основными свойствами алгоритмов на основе методов метаэвристической оптимизации являются:

- ◆ стохастичность, то есть случайность поиска, за счёт чего исключается возможность закливания в локальном оптимуме;
- ◆ мультиагентность;
- ◆ скорость нахождения оптимального решения выше, чем у традиционных методов [9].

В работе наряду с метаэвристиками, на которых построены роевые алгоритмы [10–12], используется метаэвристика, имеющая тенденцию к использованию альтернатив (вариантов компонентов) из наилучших найденных решений [13, 14]. В процессе эволюционной коллективной адаптации методами дискриминантного анализа формируются оценки приспособленности альтернатив. Приспособленность альтернатив рассматривается как вероятность ее использования в формируемом решении. Совокупность данных об альтернативах и их оценках составляет россыпь альтернатив [15,16]. Дискриминантный анализ альтернатив в процессе эволюционной коллективной адаптации назван по аналогии с процессами вычленения объектов (формирования кристаллов) кристаллизацией. Другими словами, в процессе эволюционной коллективной адаптации производится вычленение из множества вариантов наиболее приспособленных альтернатив. Отсюда название метода оптимизации – метод кристаллизации россыпи альтернатив (КРА, Crystalli-

zation of alternatives field (CAF)) [13]. В работе рассматривается эволюционный популяционный метод решения транспортной задачи на основе метаэвристики кристаллизации россыпи альтернатив [13, 14].

1. Математическая постановка транспортной задачи. Общая постановка транспортной задачи заключается в определении оптимального плана перевозок некоторого однородного груза из пунктов отправления $A = \{A_i / i = 1, 2, \dots, m\}$ в пункты назначения $B = \{B_j / j = 1, 2, \dots, n\}$. Критерием оптимальности является минимальная стоимость перевозки или минимальное время доставки груза.

Рассмотрим транспортную задачу, где в качестве критерия оптимальности взята минимальная стоимость перевозок всего груза [2, 4].

Задается список $D = \{d_{ij} / i = 1, \dots, m, j = 1, \dots, n\}$ маршрутов между множеством пунктов отправления A и множеством пунктов назначения B . Обозначим через c_{ij} тарифы перевозки единицы груза из пункта отправления i в пункт назначения j . Обозначим через a_i запасы груза i -м пункте отправления, а через b_j потребности груза j -м пункте назначения, а через x_{ij} количество единиц груза перевозимого из пункта отправления i в пункт назначения j .

Тогда математическая модель транспортной задачи состоит в определении минимального значения функции [1, 2]

$$F = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij}, \quad (1)$$

при следующих условиях:

$$\sum_{i=1}^m x_{ij} = b_j \quad (j = 1, \dots, n), \quad (2)$$

$$\sum_{j=1}^n x_{ij} = a_i \quad (i = 1, \dots, m), \quad (3)$$

$$x_{ij} \geq 0 \quad (i = 1, \dots, m, j = 1, \dots, n). \quad (4)$$

Поскольку удовлетворяются условия (1)–(4), то обеспечивается доставка необходимого количества груза в каждый из пунктов назначения, вывоз груза из всех пунктов отправления, а также исключаются обратные перевозки.

В работе исследуется закрытая (или сбалансированная) модель транспортной задачи: сумма груза у поставщиков равно общей сумме потребностей в пунктах назначения:

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j \quad (5)$$

В задаче требуется найти такой план перевозок, т.е. такие $x_{ij} \geq 0$, чтобы суммарная стоимость перевозок была минимальной, и выполнялись все ограничения.

Исходные данные транспортной задачи представлены в виде табл. 1.

Таблица 1

Исходные данные транспортной задачи

Пункты отправления	Пункты назначения				Запасы
	B_1	B_2	B_3	B_4	
A_1	c_{11} x_{11}	c_{12} x_{12}	c_{13} x_{13}	c_{14} x_{14}	a_1
A_2	c_{21} x_{21}	c_{22} x_{22}	c_{23} x_{23}	c_{24} x_{24}	a_2
A_3	c_{31} x_{31}	c_{32} x_{32}	c_{33} x_{33}	c_{34} x_{34}	a_3
Потребности	b_1	b_2	b_3	b_4	

Содержательная постановка задачи.

Однородный продукт, сосредоточенный в m пунктах отправления в количествах a_1, a_2, \dots, a_m единиц соответственно, необходимо доставить в каждый из n пунктов назначения в количествах b_1, b_2, \dots, b_n единиц соответственно.

Стоимость (расстояние) перевозки единицы продукта из i -го пункта отправления в j -й пункт назначения равна c_{ij} (стоимость доставки) и известна для каждого маршрута d_{ij} . Пусть x_{ij} – количество продукта, перевозимого из i -го пункта отправления в j -й пункт назначения.

Задача заключается в определении таких величин x_{ij} для всех маршрутов, при которых суммарная стоимость или расстояние F перевозок были бы минимальными.

Обозначения:

m – количество пунктов отправления (поставщиков);

i – номер поставщика;

n – количество пунктов назначения (потребителей);

j – номер потребителя;

a_i – исходный объем однородного груза i -го поставщика (запасы);

b_j – исходный объем однородного груза, требуемого j -ому потребителю (спрос);

c_{ij} – стоимость доставки единицы груза i -го поставщика j -ому потребителю;

x_{ij} – количество груза, доставляемое от i -го поставщика к j -му потребителю;

D – набор маршрутов между множеством пунктов отправления A и множеством пунктов назначения B ;

d_{ij} – маршрут между пунктом отправления A_i и пунктом назначения B_j ;

C – общие затраты на перевозки.

2. Разработка алгоритма решения транспортной задачи методом кристаллизации россыпи альтернатив. В основу метаэвристики кристаллизация россыпи альтернатив (КРА) положена стратегия, основанная на запоминании и повторении прошлых успехов. Стратегия делает упор на «коллективную память», под которой подразумевается любой вид информации, которая отражает прошлую историю развития и хранится независимо от индивидуумов.

В работе рассматривается эволюционный популяционный метод решения транспортной задачи на основе метаэвристики кристаллизации россыпи альтернатив.

В работе используется кодированное представление опорного плана. В качестве кода решения Q_k рассматривается упорядоченная последовательность маршрутов – $D_k = \langle d_{kij} | i=1, \dots, m, j=1, \dots, n \rangle$. Возможное число маршрутов $R=mn$. Существенную роль в общем процессе нахождения решения играет декодер, осуществляющий трансформацию кода от списка к решению. Важной характеристикой декодера является его способность получить оптимальное решение по заранее известному оптимальному коду (приоритетному списку), т.е. способность правильно декодировать.

Формирование опорного плана путем декодирования списка D_k осуществляется путем последовательного формирования частичных планов перевозок маршрутами и их интеграции в общий план.

На каждом шаге t решается задача выбора очередного в последовательности D_k маршрута d_{kij} и определения количества $x_{ij}(t)$ груза, перевозимого из пункта отправления A_i в пункт назначения B_j по этому маршруту.

Введем табл. 2 для отражения текущих значений остаточных запасов $a_i(t)$ в пунктах отправления, текущих потребностей $b_j(t)$ в пунктах назначения и количества перевезенного груза $x_{ij}(t)$ после выполнения очередного шага t .

Таблица 2

Остаточные запасы

t	$a_i(t)$	$b_j(t)$	$x_{ij}(t)$
0	$a_i(0)=a_i$	$b_j(0)=0$	0
1	$a_i(1)$	$b_j(1)$	$x_{ij}(1)$
2
...
mn

Отметим, в процессе поиска плана параметры $a_i(t)$ уменьшаются, а параметры $b_j(t)$ растут.

Маршрут груза d_{ij} считается реализованным, если хотя бы один из параметров $a_i(t)=0$ (запас) или $b_j(t)$ (потребность)= b_j .

На шаге t выбирается t -й по списку неиспользованный маршрут $d_{ij}=(A_i, B_j)$ с параметрами $(a_i(t), b_j(t), x_{ij}(t))$.

Возможны 3 случая:

$$a_i(t) = b_j(t);$$

$$a_i(t) > b_j(t);$$

$$a_i(t) < b_j(t).$$

Рассмотрим тариф c_{ij} .

а) Если $a_i(t)=b_j(t)$, то $x_{ij}(t+1)=x_{ij}(t)+a_i(t)$, $a_i(t+1)=a_i(t)-a_i(t)=0$, $b_j(t+1)=b_j(t)-b_j(t)=0$. В этом случае остаточные запасы $a_i(t)$ пункта A_i полностью исчерпаны. Потребности пункта B_j полностью удовлетворены, поэтому все маршруты, у которых пунктом отправления является пункт A_i , а пунктом назначения является пункт B_j , исключаются из дальнейшего рассмотрения.

Если $a_i(t)>b_j(t)$, то $x_{ij}(t+1)=x_{ij}(t)+b_j(t)$, $a_i(t+1)=a_i(t)-b_j(t)$, $b_j(t+1)=b_j(t)-b_j(t)=0$. Потребности пункта B_j полностью удовлетворены, поэтому все маршруты, у которых пунктом назначения является пункт B_j , из дальнейшего рассмотрения исключаются.

Если $a_i(t)<b_j(t)$, то $x_{ij}=a_i(t)$. В этом случае остаточные запасы $a_i(t)$ пункта A_i полностью исчерпаны $a_i(t+1)=a_i(t)-a_i(t)=0$. Поэтому все тарифы и соответствующие им маршруты, у которых пунктом отправления является пункт A_i , из дальнейшего рассмотрения исключаются, а в пункте B_j потребность $b_j(t)$, уменьшится на $a_i(t)$, т.е. $b_j(t+1)=b_j(t)-a_i(t)$.

Рассмотрим метод декодирования на конкретном примере.

Пример 1. На три базы A_1, A_2, A_3 поступил очередной груз в количествах равных 50, 40, 30 единиц. Этот груз требуется перевезти в четыре пункта назначения B_1, B_2, B_3, B_4 в количествах 42, 33, 25, 20.

Решение. Запишем все данные в таблицу условий:

Таблица 3

Таблица условий

Пункты отправления	Пункты назначения				Запасы
	B_1	B_2	B_3	B_4	
A_1	$c_{11} - 4$	$c_{12} - 3$	$c_{13} - 2$	$c_{14} - 1$	$a_1 - 50$
A_2	$c_{21} - 3$	$c_{22} - 2$	$c_{23} - 4$	$c_{24} - 1$	$a_2 - 40$
A_3	$c_{31} - 2$	$c_{32} - 4$	$c_{33} - 1$	$c_{34} - 3$	$a_3 - 30$
Потребности	$b_1 - 42$	$b_2 - 33$	$b_3 - 25$	$b_4 - 20$	

Найти опорный план перевозок данной транспортной задачи, решение которой представлено в виде кода $D = \langle d_{14}, d_{24}, d_{34}, d_{33}, d_{23}, d_{13}, d_{22}, d_{12}, d_{32}, d_{21}, d_{31}, d_{11} \rangle$.

Число пунктов отправления $m=3$, а число пунктов назначения $n=4$.

Наличие груза у поставщиков равно: $\sum A_i = 50 + 40 + 30 = 120$.

Общая потребность в грузе в пунктах назначения равна: $\sum B_j = 42 + 33 + 25 + 20 = 120$.

$\sum A_i = \sum B_j$. Модель транспортной задачи является закрытой (сбалансированной). Следовательно, она разрешима. Отметим, что число маршрутов, используемых для перевозок в построенном плане может оказаться меньше числа маршрутов в коде.

На шаге $t=0$ формируются начальные значения параметров:

$$a_1(0) = 50, a_2(0) = 40, a_3(0) = 30.$$

$$b_1(0) = 42, b_2(0) = 33, b_3(0) = 25, b_4(0) = 20.$$

Для всех маршрутов $x_{ij}(0) = 0$.

Задается код решения:

$$D(0) = \langle d_{14}, d_{24}, d_{34}, d_{33}, d_{23}, d_{13}, d_{22}, d_{12}, d_{32}, d_{21}, d_{31}, d_{11} \rangle.$$

На шаге $t=1$ из списка $D(0)$ выбирается первый элемент – маршрут d_{14} : $A_1 \rightarrow B_4$, с тарифом $c_{14} = 1$:

$a_1(0) = 50, b_4(0) = 20$. Проверяем соотношение параметров $a_1(0)$ и $b_4(0)$:

$a_1(0) > b_4(0)$, ситуация соответствует пункту (а) и $a_1(0)$ полностью покрывает $b_4(0)$.

Отсюда $a_1(1) = a_1(0) - b_4(0) = 50 - 20 = 30, b_4(1) = 0. x_{14}(1) = b_4(0) = 20$.

Потребности пункта B_4 полностью удовлетворены, поэтому маршруты, у которых пунктом назначения является пункт B_4 , из дальнейшего рассмотрения исключаются, а в пункте A_1 запас $a_1(0)$ уменьшится на $b_4(0)$. Трансформированный код после первого шага примет вид:

$$D(1) = \langle d_{33}, d_{23}, d_{13}, d_{22}, d_{12}, d_{32}, d_{21}, d_{31}, d_{11} \rangle.$$

$$a_1(1) = 30, a_2(1) = 40, a_3(1) = 30.$$

$$b_1(1) = 42, b_2(1) = 33, b_3(1) = 25, b_4(1) = 0.$$

$$x_{14}(1) = 20.$$

На шаге $t=2$ в $D(1)$ выбирается маршрут d_{33} : $A_3 \rightarrow B_3$:

$a_3(1) = 30, b_3(1) = 25, a_3 > b_3$ ситуация соответствует пункту (а) и $a_3(1)$ полностью покрывает $b_3(1)$.

Отсюда в пункте A_3 запас $a_3(1)$ уменьшится на $b_3(1)$, т.е. $a_3(2) = a_3(1) - b_3(1) = 30 - 25 = 5, b_3(2) = 0. x_{33} = b_3(1) = 25$. Потребности пункта B_3 полностью удовлетворены, поэтому все тарифы и соответствующие им маршруты, у которых пунктом назначения является пункт B_3 , из дальнейшего рассмотрения исключаются

Трансформированный код после второго шага примет вид:

$$D(2) = \langle d_{22}, d_{12}, d_{32}, d_{21}, d_{31}, s_{11} \rangle.$$

$$a_1(2) = 30, a_2(2) = 40, a_3(2) = 5.$$

$$b_1(2) = 42, b_2(2) = 33, b_3(2) = 0, b_4(2) = 0.$$

$$x_{33}(2) = 25.$$

На шаге $t=3$ в $D(2)$ выбирается маршрут d_{22} : $A_2 \rightarrow B_2$.

$a_2(2) = 0, b_2(2) = 33. a_2 > b_2$ ситуация соответствует пункту (а) и $a_2(2)$ полностью покрывает b_2 .

Отсюда $a_2(3) = 40 - 33 = 7, b_2(3) = 0. x_{22}(3) = b_2(2) = 33$. Потребности пункта B_2 полностью удовлетворены, поэтому все тарифы и соответствующие им маршруты, у которых пунктом назначения является пункт B_2 , из дальнейшего рассмотрения исключаются, а в пункте A_2 запас $a_2(2)$ уменьшится на $b_2(2)$, т.е. $a_2(3) = a_2(2) - b_2(2) = 40 - 33 = 7$. Трансформированный код после третьего шага примет вид:

$$\begin{aligned}
 D(3) &= \langle d_{21}, d_{31}, d_{11} \rangle. \\
 a_1(3) &= 30, a_2(3) = 7, a_3(3) = 5. \\
 b_1(3) &= 42, b_2(3) = 0, b_3(3) = 0, b_4(3) = 0. \\
 x_{22}(3) &= 33.
 \end{aligned}$$

На 4, 5 и 6 шагах последовательно в пункт B_1 назначаются 5 единиц груза из пункта A_3 , 7 единиц груза из пункта A_2 и 30 единиц груза из пункта A_1 .

Знания параметров после шестого шага примут вид:

$$\begin{aligned}
 a_1(6) &= 0, a_2(6) = 0, a_3(6) = 0. \\
 b_1(6) &= 0, b_2(6) = 0, b_3(6) = 0, b_4(6) = 0. \\
 x_{31}(4) &= 5, x_{21}(5) = 7, x_{11}(6) = 30.
 \end{aligned}$$

Полученный план перевозок имеет вид:

$$x_{14} = 20, x_{33} = 25, x_{22} = 33, x_{31} = 5, x_{21} = 7, x_{11} = 30.$$

Суммарная стоимость перевозок P , соответствующая плану определяется как:

$$\begin{aligned}
 P &= c_{14} \cdot x_{14} + c_{33} \cdot x_{33} + c_{22} \cdot x_{22} + c_{31} \cdot x_{31} + c_{21} \cdot x_{21} + c_{11} \cdot x_{11}. \\
 P &= 1 \cdot 20 + 1 \cdot 25 + 2 \cdot 33 + 2 \cdot 5 + 3 \cdot 7 + 4 \cdot 30 = 262.
 \end{aligned}$$

Код, соответствующий построенному решению, имеет вид:

$$D(0) = \langle d_{14}, d_{33}, d_{22}, d_{31}, d_{21}, d_{11} \rangle.$$

Отметим, что для выполнения плана перевозок используется 6 маршрутов из 12.

3. Механизмы решения транспортной задачи на основе метаэвристики кристаллизации россыпи альтернатив. В методе кристаллизации россыпи альтернатив [14–16] каждое решение Q_k формируется (представляется) множеством объектов (маршрутов) $D_k = \langle d_{kij} | i=1, \dots, m, j=1, \dots, n \rangle$. Возможное число маршрутов $R = mn$. Каждому маршруту d_{kij} соответствует множество альтернативных состояний $P = \{p_\beta | \beta=1, 2, \dots, n\}$ где n – число состояний маршрута d_{kij} . Состоянию маршрута соответствует позиция в векторе D_k в которой размещен элемент (маршрут) d_{kij} . Каждый маршрут d_{kij} может находиться в одном из альтернативных состояний (позиции в векторе D_k). Решение Q_k определяется совокупностью альтернативных состояний множества маршрутов.

Применение любой метаэвристики для решения комбинаторной задачи заключается в представлении исходной формулировки задачи в виде компонент метаэвристики.

В качестве кода решения ТЗ Q_k рассматривается упорядоченная последовательность маршрутов – $D_k = \langle d_{kij} | i=1, \dots, m, j=1, \dots, n \rangle$. В разработанном алгоритме решения ТЗ объектами являются маршруты, альтернативами – множество позиций $P = \{p_\beta | \beta=1, 2, \dots, n\}$ в списке D_k , где n_p – число позиций в списке D_k в которые помещаются элементы d_{kij} . Другими словами множество объектов $D_k = \langle d_{kij} | i=1, \dots, m, j=1, \dots, n \rangle$, соответствует множеству всех маршрутов. Множество альтернативных состояний $P = \{p_\beta | \beta=1, 2, \dots, n\}$ объекта d_{kij} соответствует множеству альтернативных вариантов размещения объекта в списке D_k .

Каждому элементу d_{kij} списка D_k соответствует один единственный элемент p_β . Номер позиции p_β в которую помещен элемент d_{kij} определяется соотношением $p_\beta = F_k(d_{kij})$.

Работа популяционного эволюционного алгоритма КРА опирается на коллективную эволюционную память, называемую **россыпью альтернатив**.

Под россыпью альтернатив (РА) решения в работе называется структура данных, используемая в качестве коллективной эволюционной памяти (КЭП), несущая информацию о решении, включающую сведения о реализованных альтернативах агентов в данном решении и о полезности решения.

Установим биективное отображения $F_k = NP \rightarrow NP$ между множеством чисел натурального ряда NP от 1 до $r = mn$ и множеством NP всех пар (i, j) , таких, что $1 \leq i \leq m; 1 \leq j \leq n | NP| = mn$. Каждому элементу a множества NR соответствует один единственный элемент $(i, j)_l$ множества NP и наоборот. Номер элемента a определяется соотношением $a = F_k((i, j)_a)$ и наоборот. Тогда $d_{kij} = d_{ka}$.

Представим структуру данных для отображения одного решения Q_k , формируемого множеством объектов (маршрутов), в виде матрицы альтернатив $R_k = \|r_{k\alpha\beta}\|_{m \times n}$. α – индекс объекта, β – индекс позиции.

Строки матрицы соответствуют – объектам (маршрутам) $d_{k\alpha}$. Размерность строки (вектора $R_{k\alpha} = \{r_{k\alpha\beta} | \beta = 1, 2, \dots, n\}$) определяется числом n возможных альтернатив – состояний объекта (маршрута $d_{k\alpha}$) – (числом позиций, в которые назначаются объекты).

Столбцы матрицы $R_k = \|r_{k\alpha\beta}\|_{m \times n}$ соответствуют позициям. Размерность столбца (вектора $R_{k\beta} = \{d_{k\alpha} | \alpha = 1, 2, \dots, m\}$), определяется числом m объектов (маршрутов).

В каждой строке (векторе $R_{k\alpha}$) только один элемент $r_{k\alpha\beta}$, соответствующий позиции p_β , в которой находится объект $d_{k\alpha}$, имеет значение, отличное от нуля. Остальные элементы вектора $R_{k\alpha}$ имеют нулевые значения.

В каждом столбце (векторе $R_{k\beta}$) только один элемент $r_{k\alpha\beta}$, соответствующий объекту $d_{k\alpha}$, который расположен в позиции p_β , имеет значение, отличное от нуля. Остальные элементы вектора $R_{k\beta}$ имеют нулевые значения.

Каждый, отличный от нуля элемент $r_{k\alpha\beta}$ матрицы $R_k = \|r_{k\alpha\beta}\|_{m \times n}$, имеет значение, равное полезности δ_k решения Q_k , при котором объект $d_{k\alpha}$ (маршрут) назначен в позицию p_β , $\delta_k = f(\zeta_k)$, где ζ_k – оценка решения, а δ_k – оценка полезности этого решения.

Таким образом, в матрицу R_k заносится информация о состояниях, реализованных объектами (маршрутами) в решении Q_k , т.е. о позициях в которые они назначены (с не нулевыми значениями) и об оценке полезности – δ_k этих состояний.

Пример. Представим матрицу россыпи альтернатив R_k для рассмотренного выше примера. Пусть рассчитанная оценка полезности имеет значение $\delta_k = 6$ для кода:

$$D = \langle d_{14}, d_{24}, d_{34}, d_{33}, d_{23}, d_{13}, d_{22}, d_{12}, d_{32}, d_{21}, d_{31}, d_{11} \rangle.$$

$$F_k(D) = \langle d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}, d_{11}, d_{12} \rangle.$$

Тогда матрица россыпи альтернатив для одного решения имеет вид, представленный в табл. 4:

Таблица 4

Матрица россыпи альтернатив

P	1	2	3	4	5	6	7	8	9	10	11	12
d_1	6											
d_2		6										
d_3			6									
d_4				6								
d_5					6							
d_6						6						
d_7							6					
d_8								6				
d_9									6			
d_{10}										6		
d_{11}											6	
d_{12}												6

На базе сгенерированного популяцией агентов множества решений $Q = \{Q_k | k = 1, 2, \dots, n_k\}$ формируется множество индивидуальных россыпей альтернатив $R = \{R_k | k = 1, 2, \dots, n_k\}$.

Платформой для организации эволюционной процедуры поиска решений является интегральная россыпь альтернатив (ИРА) – R^* , которая формируется путем объединения всех россыпей альтернатив:

$$R^* = \|r^*_{\alpha\beta}\|_{m \times n}, \text{ где } r^*_{\alpha\beta} = \sum_k (r_{k\alpha\beta}), \text{ } k = \{1, n_k\}.$$

Фактически $r^*_{\alpha\beta}$ является суммарным значением полезностей решений, в которых агентами популяции маршрут d_α был назначен в позицию p_β . ИРА используется в качестве коллективной эволюционной памяти и служит базой для формирования новых решений.

Пример. Пусть агентами популяции построены 7 решений Q_1-Q_7 , на множестве маршрутов $D(0)=\langle d_{14}, d_{33}, d_{22}, d_{31}, d_{21}, d_{11} \rangle$, для которых рассчитаны оценки полезности – δ_k , приведенные в табл. 5.

Таблица 5

Решения с оценками полезности

P	p_1	p_2	p_3	p_4	p_5	p_6	
Q_1	d_{14}	d_{33}	d_{22}	d_{21}	d_{31}	d_{11}	$\delta_k = 6$
Q_2	d_{33}	d_{11}	d_{14}	d_{22}	d_{21}	d_{31}	$\delta_k = 4$
Q_3	d_{22}	d_{14}	d_{11}	d_{33}	d_{31}	d_{21}	$\delta_k = 5$
Q_4	d_{21}	d_{22}	d_{31}	d_{14}	d_{33}	d_{11}	$\delta_k = 7$
Q_5	d_{31}	d_{21}	d_{14}	d_{22}	d_{11}	d_{33}	$\delta_k = 9$
Q_6	d_{22}	d_{11}	d_{33}	d_{14}	d_{21}	d_{31}	$\delta_k = 8$
Q_7	d_{11}	d_{31}	d_{14}	d_{33}	d_{21}	d_{22}	$\delta_k = 10$

Переименуем индексы маршрутов $D(0)$ в соответствии с графиком соответствия F_k .

$$F_k(d_{14}, d_{33}, d_{22}, d_{31}, d_{21}, d_{11}) = (d_1, d_2, d_3, d_4, d_5, d_6), \text{ табл. 6.}$$

Таблица 6

Переименованные индексы маршрутов

P	p_1	p_2	p_3	p_4	p_5	p_6	
Q_1	d_1	d_2	d_3	d_4	d_5	d_6	$\delta_k = 6$
Q_2	d_2	d_6	d_1	d_3	d_4	d_5	$\delta_k = 4$
Q_3	d_3	d_1	d_6	d_2	d_5	d_4	$\delta_k = 5$
Q_4	d_4	d_3	d_5	d_1	d_2	d_6	$\delta_k = 7$
Q_5	d_5	d_4	d_1	d_3	d_6	d_2	$\delta_k = 9$
Q_6	d_3	d_6	d_2	d_1	d_4	d_5	$\delta_k = 8$
Q_7	d_6	d_5	d_1	d_2	d_4	d_3	$\delta_k = 10$

Для каждого решения Q_k , формируемого множеством объектов (маршрутов) формируются матрицы альтернатив $R_k = \|r_{k\alpha\beta}\|_{m \times n}$.

На рис. 1 в целях компактности приведена совмещенная матрица альтернатив, позволяющая просмотреть матрицу альтернатив для каждого решения.

	p_1	p_2	p_3	p_4	p_5	p_6
d_1	6	5	4+9+10	7+8		
d_2	4	6	8	5+10	7	9
d_3	5+8	7	6	4+9		10
d_4	7	9		6	4+8+10	5
d_5	9	10	7		6+5	4+8
d_6	10	4+8	5		9	6+7

Рис. 1. «Совмещенная» матрица альтернатив с отдельно показанными результатами всех решений

Интегральная россыпь альтернатив R^* имеет вид, представленный на рис. 2:

	p_1	p_2	p_3	p_4	p_5	p_6
d_1	6	5	23	15		
d_2	4	6	8	15	7	9
d_3	13	7	6	13		10
d_4	7	9		6	22	5
d_5	9	10	7		11	12
d_6	10	12	5		9	13

Рис. 2. Интегральная россыпь альтернатив R^*

Разработанный алгоритм является популяционным, реализующим стратегию случайного направленного поиска. Популяция – совокупность агентов, эволюционирующих в пространстве поиска. Каждый агент является кодом некоторого решения ТЗ. Процесс поиска решений итерационный. Алгоритм обладает памятью. На начальном этапе:

- ◆ Задается объем обучающей популяции решений – N_k .
- ◆ Генерируется начальное множество кодов решений $D = \{D_k | k=1, 2, \dots, N_k\}$ путем случайного выбора альтернатив p_β .
- ◆ Формирование начальной интегральной россыпи альтернатив $R^*(0)$: присвоение начального значения всем элементам матрицы $R^* = \|r^*_{\alpha\beta}\|_{m \times n}$, где $r^*_{\alpha\beta} = \sum_k (r_{k\alpha\beta})$, $k = \{1, n_k\}$.

Каждая итерация l начинается с приведения в текущего состоянии служебной памяти в исходное и выполнения трех этапов.

На первом этапе каждой итерации l конструктивным алгоритмом на базе интегральной россыпи альтернатив формируется n_k кодов решений D_k (популяция решений). Код решения D_k формируется (представляется) в виде упорядоченного множества объектов (маршрутов) $D_k = \langle d_{kij} | i=1, \dots, m, j=1, \dots, n \rangle$.

Каждый код решения D_k является биективным отображением $F_k = D \rightarrow P$ и формируется путем последовательного назначения элементов множества D в позиции множества P .

На первом этапе каждой итерации l конструктивным алгоритмом на базе интегральной россыпи альтернатив формируется n_k кодов решений D_k (популяция решений). На каждом шаге *процесс* назначения объекта в позицию включает две стадии. На первой стадии выбирается объект (d_{ka}), а на второй стадии – позиция p_β . При этом должно выполняться ограничение: каждому объекту d_{ka} множества D_k соответствует один единственный элемент множества P и наоборот.

В текущей памяти хранится множество объектов $D^*(t)$, не назначенных в позиции, и множество $P^*(t)$ свободных позиций.

На первой стадии для каждого $d_{ka} \in D^*(t)$ в соответствующей строке R^*_{ka} матрицы интегральной россыпи альтернатив определяется позиция $p_\beta^*(t) \in P^*(t)$ с максимальным значением суммарной полезности $\delta_k = r^*_{\alpha\beta}$ (где $r^*_{\alpha\beta} = \sum_k (r_{k\alpha\beta})$), которое обозначим как π_{ka} . π_{ka} характеризует перспективность выбора d_{ka} .

Среди множества не размещенных элементов $D^*(t)$ с вероятностью $P(d_{ka}) = \pi_{ka} / \sum_a (\pi_{ka})$, пропорциональной π_{ka} , выбирается элемент d_{ka} .

На второй стадии среди множества свободных позиций $P^*(t)$, соответствующих выбранному элементу d_{ka} , т.е. при выбранном a с вероятностью $P(p_\beta^*(t)) = r^*_{\alpha\beta} / \sum_\beta (r^*_{\alpha\beta})$, пропорциональной значению оценки полезности $\delta = r^*_{\alpha\beta}$ выбирается позиция $p_\beta^*(t)$. Объект d_{ka} помещается в позицию $p_\beta^*(t)$, затем объект d_{ka} удаляется из $D^*(t)$, а позиция $p_\beta^*(t)$ удаляется из множества $P^*(t)$ и переход к следующему шагу.

Для построенного кода решения D_k рассчитывается оценка решения ζ_k и оценка полезности δ_k . Формируется индивидуальная россыпь альтернатив R_k для Q_k и переход к построению следующего кода решения.

На втором этапе итерации производится суммирование интегральной россыпи альтернатив $R^*(l-1)$, сформированной на предыдущих итерациях от 1 до $(l-1)$, со всеми индивидуальными россыпями альтернатив R_k , сформированных на итерации l .

На третьем этапе итерации l производится снижение всех интегральных оценок полезности $r^*_{\text{ар}}$ интегральной россыпи альтернатив $R^*(l)$ на величину δ^* .

Фиксация и вывод лучшего решения.

4. Экспериментальные исследования. Алгоритм решения транспортной задачи был реализован на языке C++ в среде Windows. Экспериментальные исследования проводились на ЭВМ типа IBM PC/AT. Проведение экспериментов преследовало две цели: исследование эффективности и качества механизмов алгоритма кристаллизации россыпи альтернатив для решения транспортной задачи. Для этих целей была использована процедура синтеза контрольных примеров с известным оптимумом по аналогии с известным методом РЕКУ (Placement Examples with Known Upper bounds on wirelength) [17-21].

Сравнение значений критерия, на тестовых примерах, с известным оптимумом показало, что у 90% примеров полученное решение было оптимальным, у 2% примеров решения были на 5% хуже, а у 8% примеров решения были хуже не более, чем на 2%. Для проведения объективных экспериментов были использованы известные тестовые задачи, представленные в литературе и Интернет. Задачи, на которых был протестирован разработанный алгоритм, доступны в библиотеке OR-объектов (<http://www.ms.ic.ac.uk/info.html>).

Для составления достоверных выводов был проведен не один, а серия опытов-экспериментов. По сравнению с существующими алгоритмами достигнуто улучшение результатов на 2–3%.

Была произведена серия по измерению продолжительности работы алгоритма от количества маршрутов. Общая оценка временной сложности лежит в пределах $O(n^2)$ - $O(n^3)$. Временная сложность алгоритма (ВСА), полученная экспериментальным путем, практически совпадает с теоретическими исследованиями и для рассмотренных тестовых задач составляет (ВСА $\approx O(n^2)$).

На основе обработки экспериментальных исследований была построена средняя зависимость качества решений от числа итераций (рис. 3). Оценкой качества служит величина $F/F_{\text{опт}}$, где F – оценка полученного решения. Исследования показали, что число итераций, при которых алгоритм находил лучшее решение, лежит в пределах 130–150. Из графика видно, что в среднем на 140 итерации решение близко к оптимальному.

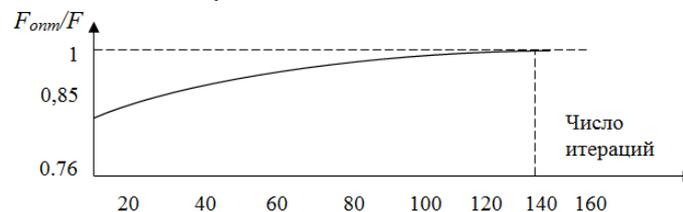


Рис. 3. Зависимость качества решений от числа итераций

Заключение. На основе сравнительного анализа существующих подходов и методов для решения транспортной задачи использованы мультиагентные методы интеллектуальной оптимизации.

Разработана новая модификация метаэвристики интеллектуальной оптимизации, учитывающая тенденцию к использованию альтернатив из наилучших найденных решений, базирующаяся на моделировании коллективного интеллекта. Совокупность данных об альтернативах и их оценках составляет россыпь альтернатив. В методе кристаллизации россыпи альтернатив каждое решение формируется (представляется) множеством объектов. Каждому объекту соответствует множество альтернативных состояний. Решение определяется совокупностью альтернативных состояний множества агентов.

Отличительной особенностью представленной метаэвристики является использование упорядоченной последовательности маршрутов в качестве кода решения ТЗ. Разработанный алгоритм построения опорного плана обладает линейной трудоемкостью и позволяет получать квазиоптимальные решения. Рассмотрены ключевые моменты анализа альтернатив в процессе эволюционной коллективной адаптации, названной по аналогии с процессами вычленения объектов (формирования кристаллов) кристаллизацией. Алгоритм на основе кристаллизации россыпи альтернатив был успешно применен для решения транспортной задачи и является эффективным способом поиска рациональных решений для задач оптимизации, допускающих интерпретацию в виде россыпи альтернатив.

Экспериментальные исследования показали, что алгоритмы на основе предлагаемого подхода могут давать лучшие результаты, чем при использовании классических методов для решения транспортных задач.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Корбут А.А., Финкельштейн Ю.Ю.* Дискретное программирование. – М.: Наука: Гл. ред. физ.-мат. лит., 1969.
2. *Вентцель Е.С.* Исследование операций: задачи, принципы, методология: учебник для вузов. – М.: Дрофа, 2004. – 208 с.
3. *Taha Х.А.* Введение в исследование операций. – 7-е изд. – М.: Изд. дом «Вильямс», 2005.
4. *Engelbrecht А.Р.* Fundamentals of Computational Swarm Intelligence. – John Wiley & Sons: Chichester, UK, 2005.
5. *Юдин Д.Б., Гольштейн Е.Г.* Задачи и методы линейного программирования: Задачи транспортного типа. – М.: Изд. 3, 2013. – 184 с.
6. *Рудик И.Д., Величко В.В.* Понятие, виды и методы решения транспортной задачи // Международный студенческий научный вестник. – 2017. – № 4-4.
7. *Dorigo M., Stützle T.* Ant Colony Optimization. – MIT Press: Cambridge, MA, 2004. – 154 p.
8. *Poli R.* Analysis of the publications on the applications of particle swarm optimization // Journal of Artificial Evolution and Applications: Article ID 685175. – 2008. – P. 10-21.
9. *Курейчик В.М., Лебедев Б.К., Лебедев О.Б.* Поисковая адаптация: Теория и практика. – М.: Физматлит, 2006. – 272 с.
10. *Лебедев О.Б.* Модели адаптивного поведения муравьиной колонии в задачах проектирования. – Таганрог. Изд-во ЮФУ, 2013. – 199 с.
11. *Карпенко А.П.* Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой: учеб. пособие. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2016. – 448 с.
12. *Гладков Л.А., Гладкова Н.В.* Решение динамических транспортных задач на основе гибридных интеллектуальных методов и моделей // Известия ЮФУ. Технические науки. – 2013. – № 7 (144). – С. 102-107.
13. *Курейчик В.М., Кажаров А.А.* Муравьиные алгоритмы для решения транспортных задач // Известия РАН. Теория и системы управления. – 2010. – № 1. – С. 32-45.
14. *Курейчик В.М., Емельянова Т.С.* Решение транспортных задач с использованием комбинированного генетического алгоритма // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008. – М.: Физматлит, 2008. – С. 158-164.
15. *Лебедев Б.К., Лебедев В.Б.* Оптимизация методом кристаллизации россыпи альтернатив // Известия ЮФУ. Технические науки. – 2013. – № 7. – С. 11-17.

16. *Лебедев Б.К., Лебедев В.Б.* Метод кристаллизации россыпи альтернатив // Сб. научных трудов VII Международной научно-практической конференции “Интегрированные модели и мягкие вычисления в искусственном интеллекте”. – М.: Изд-во Физматлит, 2013. – С. 903-912.
17. *Лебедев Б.К., Лебедев В.Б.* Глобальная трассировка методом кристаллизации россыпи альтернатив // Известия ЮФУ. Технические науки. – 2014. – № 7 (156). – С. 42-51.
18. *Лебедев Б.К., Лебедев В.Б.* Построение кратчайших связывающих соединений методом кристаллизации россыпи альтернатив // VI Всероссийская научно-техническая конференция «Проблемы разработки перспективных микро- и наноэлектронных систем – 2014»: Сб. трудов / под общ. ред. академика РАН А.Л. Стемпковского. – М.: ИПИМ РАН, 2014. – С. 177-183.
19. *Mourelle M.* Swarm intelligent systems. – Berlin: Heidelberg: Springer, 2006. – 217 p.
20. *Cong J., Romesis M., Xie M.* UCLA Optimality Study Project. – <http://cadlab.cs.ucla.edu/~pubbench>. 2004.
21. MCNC Electronic and Information Technologies (Online). Available: www.mcnc.org.

REFERENCES

1. *Korbut A.A., Finkel'shteyn Yu.Yu.* Diskretnoe programmirovaniye [Discrete programming]. Moscow: Nauka: Gl. red. fiz.-mat. lit., 1969.
2. *Ventsel' E.S.* Issledovanie operatsiy: zadachi, printsipy, metodologiya: uchebnik dlya vuzov [Operations research: tasks, principles, methodology: textbook for universities]. Moscow: Drofa, 2004, 208 p.
3. *Takha Kh.A.* Vvedenie v issledovanie operatsiy [Introduction to operations research]. 7th ed. Moscow: Izd. dom «Vil'yams», 2005.
4. *Engelbrecht A.P.* Fundamentals of Computational Swarm Intelligence. John Wiley & Sons: Chichester, UK, 2005.
5. *Yudin D.B., Gol'shteyn E.G.* Zadachi i metody lineynogo programmirovaniya: Zadachi transportnogo tipa [Problems and methods of linear programming: Problems of transport type]. Moscow: Izd. 3, 2013, 184 p.
6. *Rudik I.D., Velichko V.V.* Ponyatie, vidy i metody resheniya transportnoy zadachi [The concept, types and methods of solving the transport problem], *Mezhdunarodnyy studencheskiy nauchnyy vestnik* [International Student Scientific Bulletin], 2017, No. 4-4.
7. *Dorigo M., Stützle T.* Ant Colony Optimization. MIT Press: Cambridge, MA, 2004, 154 p.
8. *Poli R.* Analysis of the publications on the applications of particle swarm optimization, *Journal of Artificial Evolution and Applications: Article ID 685175*, 2008, pp. 10-21.
9. *Kureychik V.M., Lebedev B.K., Lebedev O.B.* Poiskovaya adaptatsiya: Teoriya i praktika [Search adaptation: Theory and practice]. Moscow: Fizmatlit, 2006, 272 p.
10. *Lebedev O.B.* Modeli adaptivnogo povedeniya murav'inoi kolonii v zadachakh proektirovaniya [Models of adaptive behavior of an ant colony in design problems]. Taganrog. Izd-vo YuFU, 2013, 199 p.
11. *Karpenko A.P.* Sovremennyye algoritmy poiskovoy optimizatsii. Algoritmy, vdokhnovlennyye prirodoy: ucheb. posobie [Modern search engine optimization algorithms. Algorithms inspired by nature]. Moscow: Izd-vo MGTU im. N.E. Baumana, 2016, 448 p.
12. *Gladkov L.A., Gladkova N.V.* Reshenie dinamicheskikh transportnykh zadach na osnove gibridnykh intellektual'nykh metodov i modeley [Solving dynamic transport problems based on hybrid intelligent methods and models], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2013, No. 7 (144), pp. 102-107.
13. *Kureychik V.M., Kazharov A.A.* Murav'inye algoritmy dlya resheniya transportnykh zadach [Ant algorithms for solving transport problems], *Izvestiya RAN. Teoriya i sistemy upravleniya* [Izvestiya RAS. Theory and control systems], 2010, No. 1, pp. 32-45.
14. *Kureychik V.M., Emel'yanova T.S.* Reshenie transportnykh zadach s ispol'zovaniem kombinirovannogo geneticheskogo algoritma [Solving transport problems using a combined genetic algorithm], *Odinnadsataya natsional'naya konferentsiya po iskusstvennomu intellektu s mezhdunarodnym uchastiem KII-2008* [Eleventh National Conference on Artificial Intelligence with International Participation CII-2008]. Moscow: Fizmatlit, 2008, pp. 158-164.

15. *Lebedev B.K., Lebedev V.B.* Optimizatsiya metodom kristallizatsii rossypi al'ternativ [Optimization by the method of crystallization of a placer of alternatives], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2013, No. 7, pp. 11-17.
16. *Lebedev B.K., Lebedev V.B.* Metod kristallizatsii rossypi al'ternativ [The method of crystallization of a placer of alternatives], *Sb. nauchnykh trudov VII Mezhdunarodnoy nauchno-prakticheskoy konferentsii «Integrirovannye modeli i myagkie vychisleniya v iskusstvennom intellekte»* [Collection of scientific papers of the VII International Scientific and Practical Conference “Integrated Models and Soft Computing in Artificial Intelligence”]. Moscow: Izd-vo Fizmatlit, 2013, pp. 903-912.
17. *Lebedev B.K., Lebedev V.B.* Global'naya trassirovka metodom kristallizatsii rossypi al'ternativ [Global tracing by the method of crystallization of a placer of alternatives], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2014, No. 7 (156), pp. 42-51.
18. *Lebedev B.K., Lebedev V.B.* Postroenie krachayshikh svyazyvayushchikh soedineniy metodom kristallizatsii rossypi al'ternativ [Construction of the shortest binding connections by the method of crystallization of a placer of alternatives], *VI Vserossiyskaya nauchno-tekhnicheskaya konferentsiya «Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem – 2014»: Sb. trudov* [VI All-Russian scientific and technical conference “Problems of developing promising micro- and nanoelectronic systems – 2014”: Collection of works], under the total. ed. academician of the RAS A.L. Stempkovskogo. Moscow: IPPM RAN, 2014, pp. 177-183.
19. *Mourelle M.* Swarm intelligent systems. Berlin: Heidelberg: Springer, 2006, 217 p.
20. *Cong J., Romesis M., Xie M.* UCLA Optimality Study Project. Available at: <http://cadlab.cs.ucla.edu/~pubbench>. 2004.
21. MCNC Electronic and Information Technologies (Online). Available at: www.mcnc.org.

Статью рекомендовал к опубликованию д.т.н., профессор А.Г. Коробейников.

Лебедев Борис Константинович – Южный федеральный университет; e-mail: lebedev.b.k@gmail.com; г. Таганрог, Россия; тел.: 89282897933; кафедра систем автоматизированного проектирования; профессор.

Лебедев Олег Борисович – e-mail: lebedev.ob@mail.ru; тел.: 89085135512; кафедра систем автоматизированного проектирования; доцент.

Лебедева Екатерина Олеговна – e-mail: lbedevakate@mail.ru; тел.: 89289591426; кафедра систем автоматизированного проектирования; аспирант.

Lebedev Boris Konstantinovich – Southern Federal University; e-mail: lebedev.b.k@gmail.com; Taganrog, Russia; phone: +79282897933; the department of computer aided design; professor.

Lebedev Oleg Borisovich – e-mail: lebedev.ob@mail.ru; phone: +79085135512; the department of computer aided design; associate professor.

Lebedeva Ekaterina Olegovna – e-mail: lbedevakate@mail.ru; phone: +79289591426; the department of computer aided design; graduate student.

УДК 004.428.4+336.76.066

DOI 10.18522/2311-3103-2022-4-157-169

Т.Н. Кондратьева, Е.Р. Мунтян, И.Ф. Развеева

РЕАЛИЗАЦИЯ ТОРГОВОГО СОВЕТНИКА ДЛЯ МУЛЬТИРЫНОЧНОЙ ПЛАТФОРМЫ METATRADER 5

В статье рассмотрен процесс создания гибкой торговой стратегии для алготрейдинга в специализированной среде разработки MQL5 IDE в мультирыночной платформе MetaTrader 5. Показаны преимущества и целесообразность использования платформ MetaTrader 5, MetaTrader 4 и соответствующих им торговых приложений Trade Assistant, Forex Trade Manager, Trade Time Manager, CAP Gold Albatross EA и Fast Copy. Проведен

сравнительный анализ имеющихся реализаций торговых советников, основанных на различных индикаторах, а также созданных с применением интеллектуальных технологий. В ранее реализованных торговых советниках для прогнозирования цен волатильности финансовых активов в основном использованы алгоритмы гибкого обучения, модели компенсаторной нечеткой логики, инструменты технического анализа, что влечет за собой высокие временные затраты в условиях высокой волатильности финансового рынка. Для решения данной проблемы авторами предлагается комплексный подход, основанный на применении инструментов технического анализа, встроенных в мультимедийную платформу MetaTrader 5 и алгоритма автоматизации торговой стратегии, что позволяет получить прогноз заданной точности по выбранному инструменту в режиме реального времени. В работе обоснована необходимость внедрения элементов автоматической торговли при анализе котировки финансовых инструментов и управлении торговым счетом во избежание механических, аналитических, организационных и психологических ошибок, совершаемых трейдерами. В ходе исследования поэтапно показан процесс создания, отладки, тестирования, оптимизации и исполнения реализуемого торгового советника. Разработан алгоритм автоматизации торговой стратегии и представлена его блок-схема. Определены исходные данные для алгоритма автоматизации торговой стратегии, а также описан математический аппарат вычисления показателей лимитных ордеров типа TakeProfit и StopLoss. Так как биржевая торговля связана со множеством рисков, то было проанализировано влияние различных значений лотов лимитных ордеров типа TakeProfit и StopLoss на возможную прибыль и ограничение на просадку (убыток). В результате советник корректно отработал в реальном режиме времени без участия человека в течение восьми недель по двум торговым стратегиям. Результаты тестирования разработанного программного обеспечения позволяют сделать следующие выводы: когда советник показывает высокую степень рекомендации, фактические финансовые активы показывают высокую эффективность.

Мультиязычная платформа; MetaTrader 5; Forex; алготрейдинг; торговая стратегия; индикатор.

T.N. Kondratieva, E.R. Muntyan, I.F. Razveeva

IMPLEMENTATION OF A TRADING ADVISOR FOR THE METATRADER 5 MULTI-MARKET PLATFORM

The article describes the process of creating a flexible trading strategy for algorithmic trading in a specialized development environment MQL5 IDE in the MetaTrader 5 multi-asset platform. The advantages and expediency of using the MetaTrader 5, MetaTrader 4 platforms and their respective trading applications Trade Assistant, Forex Trade Manager, Trade Time Manager, CAP Gold Albatross EA and Fast Copy are shown. A comparative analysis of the existing implementations of trading advisors based on various indicators, as well as those created using intelligent technologies, has been carried out. In the previously implemented trading advisors, for predicting the prices of the volatility of financial assets, flexible learning algorithms, compensatory fuzzy logic models, and technical analysis tools are mainly used, which entails high time costs, in conditions of high financial market volatility. To solve this problem, the authors propose an integrated approach based on the use of technical analysis tools built into the MetaTrader 5 multimedia platform and the trading strategy automation algorithm, which makes it possible to obtain a forecast of a given accuracy for the selected instrument in real time. The paper substantiates the need to introduce elements of automatic trading when analyzing the quotes of financial instruments and managing a trading account in order to avoid mechanical, analytical, organizational and psychological mistakes made by traders. The study shows step by step the process of creating, debugging, testing, optimizing and executing the implemented trading advisor. An algorithm for automating a trading strategy has been developed and its block diagram has been presented. The initial data for the trading strategy automation algorithm are determined, and the mathematical apparatus for calculating indicators of limit orders of the TakeProfit and StopLoss types is described. Since exchange trading is associated with many risks, we analyzed the impact of different values of lots of TakeProfit and StopLoss limit orders on possible profit and drawdown limit (loss). As a result, the EA worked correctly in real time without human intervention for eight

weeks using two trading strategies. The results of testing the developed software allow us to draw the following conclusions: when the EA shows a high degree of recommendation, the actual financial assets show high efficiency.

Multimarket platform; MetaTrader 5; Forex; algorithmic trading; trading strategy; indicator.

Введение. Платформа MetaTrader 5 – это инструмент трейдера, позволяющий проводить технический анализ и совершать торговые операции на Forex и фондовых биржах.

Платформа MetaTrader 5 предлагает инструменты технического и фундаментального анализа для комплексного разбора динамики цен [1–4]. Платформа позволяет одновременно изучать около 100 графиков котировок валют и акций. Большое количество тайм фреймов (торговых периодов) позволяет проводить исчерпывающий детальный анализ динамики цен в режиме 24/7. Более 80 технических индикаторов, осцилляторов и аналитических инструментов, включая графические объекты, позволяют детально анализировать динамику котировок [5–8].

Аналитические возможности терминала не ограничиваются встроенными инструментами. Благодаря специализированному языку программирования высокого уровня MetaQuotes Language 5 (MQL5) можно создать собственный индикатор со своими уникальными характеристиками. Все эти функции увеличивают аналитические возможности платформы и способность проводить более подробный и точный технический анализ котировок [9].

В данном исследовании рассмотрен один из наиболее эффективных способов работы на платформе – алгоритмическая торговля с использованием специализированных торговых роботов (экспертов). Роботы работают без участия трейдера – они могут анализировать котировки и совершать торговые операции по базовому алгоритму [10–12].

1. Анализ предметной области. Одной из первых задач, с решением которых сталкивается разработчик собственного торгового советника – это изучение, осмысление и анализ предметной области. Дело в том, что предметная область сильно влияет на все аспекты алгоритма: допустимые границы основных параметров, взаимодействие с пользователем, модель хранения и представления данных и т.д.

Чтобы избежать механические, аналитические, организационные и психологические ошибки на сегодняшний день создан ряд торговых советников, основанных на современных индикаторах. Так, в работе [13] рассмотрено создание эффективного торгового робота для торговли акциями ПАО «Сбербанк России», в основу которого заложен стохастический осциллятор, а также исследуются способы его применения и выявляются наиболее эффективные для получения максимальной прибыли. В работе [14] показано, как торговые роботы формируют значительную часть торгового оборота на мировых биржевых площадках, оказывая при этом немаловажное воздействие на всю биржевую инфраструктуру. Торговая стратегия, основанная на стохастическом осцилляторе (Stochastic Oscillator), при тестировании на исторических данных показала положительные результаты по всем основным показателям: прибыльность торговой системы, величина чистой и общей прибыли, максимальная и относительная просадка, общий убыток.

Говоря о необходимости автоматизации сферы трейдинга, исследователи отмечают, что уже сейчас 2/3 инвесторов готовы прислушаться к советам компьютера при вложении средств, как показано в работе [15]. Одной из ветвей развития автоматического трейдинга является внедрение методов искусственного интеллекта. Применение «умных» алгоритмов позволит вывести работу трейдера на новый качественный уровень. Торговые системы на основе нейронных сетей способны обрабатывать большие массивы исторических данных, при этом четко выявлять основные закономерности и внутренние логические связи в них [16].

Табл. 1 содержит торговые приложения для MetaTrader 4 (MT4), MetaTrader 5 (MT5) их стоимость при покупке и аренде, функционал и рейтинг. Мы не стали описывать все полезные функции торговых систем представленных в табл. 1, ограничились теми функциями, которые нам необходимы для создания собственного советника в терминале MetaTrader 5. Нужно отметить, что рейтинг MetaTrader 4 выше рейтинга MetaTrader 5, это объясняется тем, что MetaTrader 4 был выпущен в 2005 году и сразу же стал популярным и надёжным инструментом для трейдеров, а новая версия платформы MetaTrader 5 вышла в 2010 году дополненная новыми инструментами, в частности инструментами искусственного интеллекта и возможностью создания скрипта на Python. Стоимость торговых приложений для трейдинга может колебаться от одной сотни долларов до одной тысячи долларов и до несколько тысяч долларов, так например NorthEastWay MT5 стоимость 12483 долларов, год выпуска 2020, последнее обновление июль 2022 или Darwin Evolution MT5 стоимость от 1449 до 999 долларов (зависит от функционала), год выпуска 2022, последнее обновление февраль 2022.

Таблица 1

Торговые приложения

Торговые утилиты	Стоимость (USD)	Аренда (USD)		Функционал	РРейтинг
		3 месяца	1 год		
Trade Assistant MT4	100	30	50	Расчёт рисков по сделкам, простая установка нового ордера с помощью линий, 7 типов трейлинг – стопа	4,71
Trade Assistant MT5	100	30	50	Расчёт рисков по сделкам, простая установка нового ордера с помощью линий, 7 типов трейлинг – стопа и другой полезный функционал	4,63
Forex Trade Manager MT5	99	30	-	автоматический break even; trailing stop; частичное закрытие сделки; one-cancels-the-other order (OCO)	4,91
Trade Time Manager MT5	30	-	-	Советник открывает и закрывает новые ордера по времени. С его помощью можно реализовать торговые сценарии любой сложности, создав список задач	1,27
CAP Gold Albatross EA MT5	Free	-	-	Математический скальпер, который агрессивно торгует тиковыми данными, реализует простую и универсальную торговую стратегию	1,88
Fast Copy MT5	55	30	-	Копирование сделок между различными терминалами MT4 и MT5 по валютным парам	4,84

На сегодняшний день, компания MetaQuotes занимается разработкой новой версии платформы MetaTrader 6, которая будет являться объединённой версией платформ MetaTrader 4 и MetaTrader 5, поскольку и та и другая версия имеют свои преимущества и могут успешно дополнять друг друга.

В наших ранних работах для построения торговой стратегии мы использовали цены закрытия валютной пары [17], в данной работе предлагается использование только цен открытия, что обусловлено выбором финансового инструмента лимитных ордеров типа: TakeProfit и StopLoss.

Зарубежные ученые в своих работах по прогнозированию индекса волатильности финансового рынка в основном используют алгоритмы гибкого обучения [18], модели компенсаторной нечеткой логики [19, 20], инструменты технического анализа [21]. Недостатком применения аналитического подхода к решению задачи прогнозирования волатильности рынка является высокие временные затраты. Для решения данной проблемы авторами предлагается комплексный подход, основанный на применении инструментов технического анализа, встроенных в мультимедийную платформу MetaTrader 5 и алгоритма автоматизации торговой стратегии, что позволяет получить прогноз заданной точности по выбранному инструменту в режиме реального времени.

2. Постановка цели исследования. Целью работы является разработка программного модуля в виде торговой системы для тестирования стратегий и технических индикаторов на платформе Metatrader 5 на языке MetaQuotes Language 5 (MQL5) для оптимизации программ алготрейдинга с применением методов технического анализа.

3. Реализация алгоритма автоматизации торговой стратегии. При создании модуля на платформе Metatrader 5 первоочередной задачей является выбор библиотек. В составе скрипта интегрировано несколько различных торговых стратегий, которые применяются в зависимости от условий на рынке (выбора финансовых инструментов). Торговый советник производит анализ динамики курса валют, автоматически подбирает наиболее приемлемую стратегию.

Библиотека состоит из списков объектов, и предоставляет возможность выбора любого из элементов списка по любому критерию или свойству, запрошенному пользователем, поддерживаемому этим объектом.

Структура данных библиотеки – это набор разных объектов для описания свойств необходимых нам данных и коллекции данных – списки, хранящие соответствующие объекты. Для организации списков использован класс динамического массива указателей на экземпляре класса CObject и его наследников из коллекции данных стандартной библиотеки.

В каталоге библиотеки \MQL5\Include\DoEasy\Collections\ создан новый файл класса-коллекции тиковых данных с именем TickSeriesCollection.mqh и новый класс CBaseObj в файле BaseObj.mqh. Используемые методы: метод сравнения двух классов-событий объектов (Compare); метод TickTime; виртуальный метод SetTypeEvent; приватный метод IndexTickSeries.

Для создания объекта-коллекции использован метод, создающий список-коллекцию тиковых серий символов в виде массивов. Для связи созданной коллекции с «внешним миром» использован основной класс библиотеки CEngine, хранящийся по адресу\MQL5\Include\DoEasy\Engine.mqh.

Второй этап представляет собой настройку общих параметров скрипта. В проекте решается достаточно большой спектр вопросов, а именно: определение типа ордеров; открытие/закрытие позиций ордеров; удаление лимитных и стоповых ордеров.

Для открытия позиций использована функция *OrderSend*, которая позволяет учесть структуры торгового запроса *MqlTradeRequest* и результат (своего рода ответ на запрос) выполнения помещается в структуру *MqlTradeResult*, поле *redcode*, которой содержит код возврата торгового сервера (рис. 1).

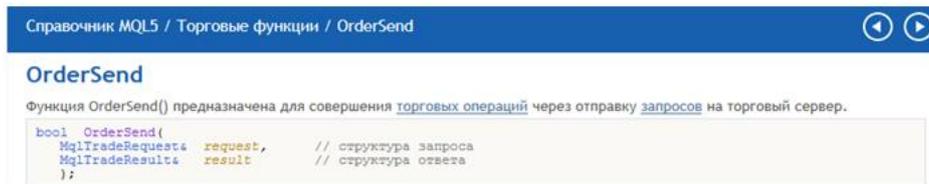


Рис. 1. Справочник MQL5. Функция OrderSend

При открытии ордера использованы *StopLoss* и *TakeProfit*, при этом известна текущая цена (работаем в режиме реального времени), используем *MqlTick*, имя новой переменной *cur_price* – текущая цена. Произведена нормализация данных *NormalizeDouble* и проверка условий по ордерам (рис. 2).

```

{
if (PositionGetInteger POSITION_TYPE = HJSITION_TYPE_BUY)
{
ask = PositionGetDouble (POSITION_PRICE_OPEN);
dsl = NormalizeDouble ( asc - slloss * _Point, _Digits);
dtp = NormalizeDouble ( asc + tprofit * _Point, _Digits);
}
if (PositionGetInteger POSITION_TYPE = HJSITION_TYPE_SELL)
{
bid = PositionGetDouble (POSITION_PRICE_OPEN);
dsl = NormalizeDouble ( bid + slloss * _Point, _Digits);
dtp = NormalizeDouble ( bid - tprofit * _Point, _Digits);
}
}

```

Рис. 2. Нормализация данных *NormalizeDouble*

Торговля осуществляется посредством отправки с помощью функции *OrderSend* «приказов» на открытие позиций, а также «приказов» на установку, модификацию и удаление отложенных ордеров. Каждый торговый «приказ» содержит указание на тип запрашиваемой торговой операции, согласно справочнику торговых операций *ORDER*.

В финальной части нашего продукта использованы коды возврата торгового сервера: 10008 – Ордер размещен (*TRADE_RETCODE_PLACED*); 10009 – Заявка выполнена (*TRADE_RETCODE_DONE*), как показано на рис. 3.

```

if (!OrderSend (mrequest,mresult))
{
Print ("Ошибка выполнения запроса открытия позиции на покупку" + IntegerToString(mresult.retcodes));
}
if (mresult.retcodes == 10008 || mresult.retcodes == 10009)
{
Print ("Ордер на покупку успешно открыт");
}
else
{
Print ("Запрос на установку ордера на покупку не выполнен + код ошибки: " + IntegerToString (GetLastError ()));
}

```

Рис. 3. Коды возврата торгового сервера

Далее создается стандартный шаблон (блок) для класса, который включает в себя конструктор (*order*). Данный конструктор позволяет открывать позиции, а выполнять инициализация всех предикторов для открытия позиций, включая текущий символ валютной пары. Таким, образом, в этом блоке заполняется структура торгового запроса всеми необходимыми параметрами.

Алгоритм автоматизации торговой стратегии (далее алгоритм) работает следующим образом: при открытии позиции *Buy*, позиция *Sell* закрывается. То есть, если инвертор выполнен, как: *trader.Buy(Lot)*, то открывается позиция на покупку, иначе открывается позиция на продажу (*trader.Sell(Lot)*). Далее, идёт проверка на

закрытие ордера в случае получения (достижения) прибыли: если $points \geq TP$, то поступает команда закрытия позиции (`CTrade::PositionClose`) и далее, $invertor=true$; если $points \leq -ST$, то так же поступает команда закрытия позиции и далее, $invertor=false$ (рис. 4).

```
double points;
if (!PositionSelect(_Symbol))
{
    if (Invertor)
        trader.Buy(Lot);
    else trader.Sell(Lot);
}
else
{
    if (PositionGetInteger(POSITION_TYPE) == POSITION_TYPE_BUY)
    {
        points = (SymbolInfoDouble(_Symbol, SYMBOL_BID) - PositionGetDouble(POSITION_PRICE_OPEN))/_Point;
        if (points >= TP)
        {
            trader.PositionClose(_Symbol);
            Invertor = true;
        }
        if (points <= -SL)
        {
            trader.PositionClose(_Symbol);
            Invertor = false;
        }
    }
}
```

Рис. 4. CTrade, trader.Buy(Lot)

Аналогичным образом поступим для ордера типа *Sell*. `Sell(const double volume,const string symbol=NULL,double price=0.0,const double sl=0.0,const double tp=0.0,const string comment="")`; осталось записать параметр объём (*Lot*). Алгоритм работает следующим образом: при открытии позиции *Sell*, позиция *Buy* закрывается. То есть, если *invertor* выполнен, как: `trader.Sell(Lot)`, то открывается позиция на продажу, иначе: `trader.Buy(Lot)`, открывается позиция на покупку.

Алгоритм работает и в обратном порядке так же. Для того, чтобы алгоритм заработал в обратную сторону необходимо переменной *invertor* присвоить значение *False*: `bool invertor=false` (это происходит сразу после объявления класса *CTrade*) и проверить условие: если $points \geq TP$, затем закрывается позиция: `CTrade::PositionClose` и $invertor=false$; если $points \leq -ST$, то так же поступает команда закрытия позиции `CTrade::PositionClose` и далее, $invertor=true$ (рис. 5).

```
}
if(PositionGetInteger(POSITION_TYPE) == POSITION_TYPE_SELL)
{
    points = (PositionGetDouble(POSITION_PRICE_OPEN) - SymbolInfoDouble(_Symbol, SYMBOL_ASK))/points;
    if(points >= TP)
    {
        trade.PositionClose(_Symbol);
        Invertor = false;
    }
    if(points <= -SL)
    {
        trade.PositionClose(_Symbol);
        Invertor = true;
    }
}
```

Рис. 5. CTrade, trader.Sell (Lot)

Реализуемый алгоритм автоматизации торговой стратегии можно представить в виде последовательности действий:

- 1) создание файла класса-коллекции тиковых данных;
- 2) подключение к нему файлы вновь созданных классов;
- 3) настройка общих параметров скрипта;
- 4) определение структуры торгового запроса и результата;
- 5) определение объёма ордера в лотах;
- 6) подача данных на вход (вектор из цен открытия по определённому периоду) их нормализация;

- 7) проверка условия по ордерам (тип торгового ордера);
 - 8) проверка отсутствия открытых позиций;
 - 9) открытие позиции на покупку/продажу;
 - 10) вычисление переменной k_i ;
 - 11) проверка условия на закрытие ордера;
 - 12) выводы результатов работы советника.
- На рис. 6 приведена блок-схема разработанного алгоритма.

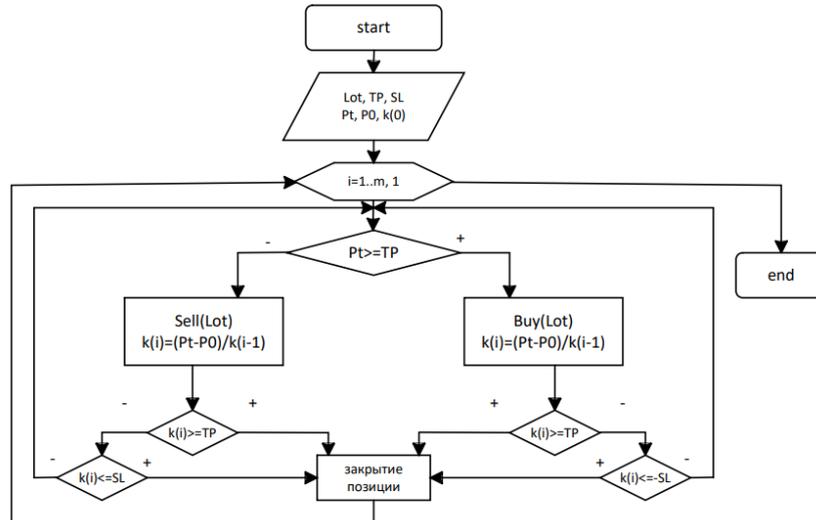


Рис. 6. Блок-схема алгоритма автоматизации торговой стратегии

В качестве исходных данных для алгоритма автоматизации торговой стратегии используются следующие параметры:

- Lot – размер лота, $Lot=1/10$;
 - $StopLoss$ – верхняя граница SL , $SL=a$, $a=10$;
 - $TakeProfit$ – нижняя граница TP , $TP=b$, $b=65$;
 - p_0 – цена открытия;
 - p_t – текущая цена;
 - $Buy(Lot)$ – открытие позиции на покупку;
 - $Sell(Lot)$ – открытие позиции на продажу;
 - k_i – количество тиков (пунктов) по открытому ордеру.
- Значение k_i можно вычислить по формуле:

$$k_i = \frac{p_t - p_0}{k_{i-1}},$$

где $i = \overline{1:m}$, m – число тиков по заданному периоду.

Процесс настройки входных параметров алгоритма отражен на рис. 7.

4. Тестирование торгового советника. Журнал ордеров представлен на рис. 8, можно наблюдать чередование ордеров: Buy; Sell, при этом красным цветом выделены позиции остановки (закрытия) ордера на покупку, а синим – на продажу.

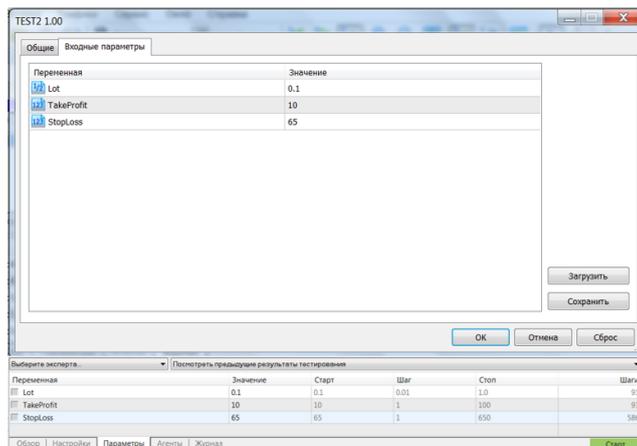


Рис. 7. Входные параметры тестирования Test2



Рис. 8. Журнал ордеров

Данная стратегия позволяет в режиме реального времени проводить работоспособность советника на исторических данных, прежде чем приступить к использованию её на котировках валютных пар.

Ордер *StopLoss/TakeProfit* позволяет решить задачу определения линии поддержки/сопротивления строя задачу на *infimum/supremum* и предназначен для минимизации/максимизации потери/прибыли (рис. 9). При достижении цены уровня поддержки/сопротивления ордер прекращает своё действие и фиксирует убытки/прибыль до экспирации ордера.

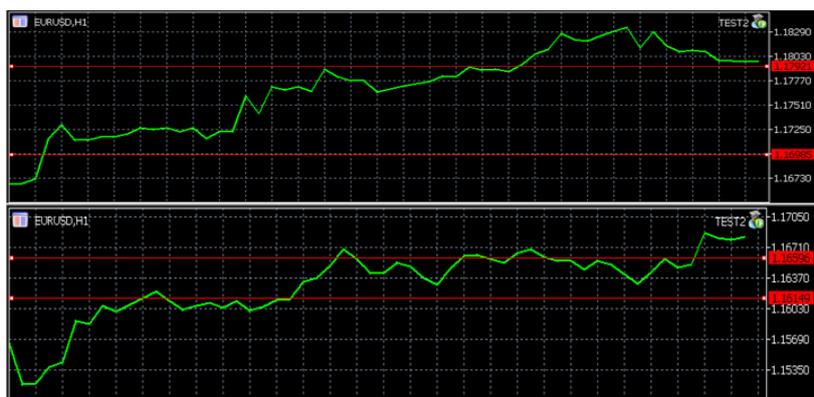


Рис. 9. Линии поддержки/сопротивления

На графике (рис. 9) торговые линии отражают уровни выставленного отложенного ордера (уровни *StopLoss* и *TakeProfit*). *StopLoss* соответствует значению 1,6985, *TakeProfit* – 1, 7921 на восьми недельных периодах (25.04.2022 – 20.06.2022) и *StopLoss* соответствует значению 1,6149, *TakeProfit* – 1,6596 (27.06.2022 – 22.08.2022).

Таблица 2 содержит два выбора *Choice1* и *Choice2* для двух наборов значений *StopLoss* (*SL*) и *TakeProfit* (*TP*).

Таблица 2

Влияние различных значений TP, SL на прибыль

Choice1	Buy TP 250	Sell TP 250	Buy SL 390	Sell SL 400
Choice2	Buy TP 610	Sell TP 610	Buy SL 790	Sell SL 980
Choice1 Profit	-76	235	107	236
Choice2 Profit	52	-34	231	187

В результате работы алгоритма получены следующие рекомендации советника: при размере лота 260 по ордеру типа *Sell TakeProfit* для *Choice1* прибыль составляет 235 пунктов; при размере лота 790 по ордеру типа *Buy StopLoss* для *Choice2* – 231 пункт; при размере лота 400 по ордеру типа *Sell StopLoss* для *Choice1* 236 пунктов. В случае реализации ордера типа *Buy TakeProfit* при *Choice1* убытки составляют 76 пунктов, размер лота 250, а по ордеру *Sell TakeProfit* при *Choice2* 34 пункта при размере лота 620. Разница в прибыли обусловлена принадлежностью стратегиям *Choice1* либо *Choice2*, либо, возможно, рыночные цены открытия позиций устарели в силу быстрого движения рынка. В любом случае, советник реализует остановку автоматически, тем самым предотвращает убытки, наращивая прибыль. Общая прибыль по ордеру типа *Buy* для *Choice1* составила 22730 пунктов; по ордеру типа *Sell* для *Choice1* – 145500 пунктов за период с 25.04.2022 по 20.06.2022. Общая прибыль по ордеру типа *Buy* для *Choice2* составила 214210 пунктов; по ордеру типа *Sell* для *Choice2* – 162180 пунктов за период с 27.06.2022 по 22.08.2022.

В целом, система отработала неоднократно в реальном режиме времени без участия человека в течение восьми недель достаточно корректно.

Заключение. Основным вкладом и новизной этой работы является реализация алгоритма автоматизации торговой стратегии для оптимизации прибыли/убытков по ордерам типа *TakeProfit* и *StopLoss*, построение собственного советника позволяющего оценить эффективность работы автоматической торговой системы на MQL для MetaTrader5.

В работе мы руководствовались принципом факторизации параметров по времени, исходя из этого, строилась дальнейшая стратегия. В наших ранних работах для построения торговой стратегии мы использовали цены закрытия валютной пары, в данной работе мы использовали только цены открытия, поскольку этого требует выбранный финансовый инструмент.

Зарубежные ученые в своих работах по прогнозированию индекса волатильности финансового рынка в основном используют алгоритмы гибкого обучения, модели компенсаторной нечеткой логики, инструменты технического анализа, мы же предсказываем возможную прибыль и ограничиваем просадку инвестора инструментами торговой системы с помощью торгового советника. Построенный нами торговый советник позволяет полностью автоматизировать торговлю в режиме 24/7; выполняет остановку торговли до объявления брокером второго уровня маржи или маржинколла; позволяет использовать фиксированное и свободное число лотов, как процент маржи; определяет ограничения по времени работы советника.

Перспективы исследований данного направления прослеживаются в решении большого спектра прикладных задач финансового рынка. Созданный советник может совершать аналитические и торговые операции по любым финансовым инструментам. Апробировав советника на одном инструменте, открывается возможность использовать его на других финансовых инструментах или на их комбинации.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Оноре Г.С.* Инструменты технического анализа с использованием Python и MetaTrader 5 // Образование и наука без границ: социально-гуманитарные науки. – 2022. – № 17. – С. 46-52.
2. *Кондратьева Т.Н., Развеева И.Ф.* Автоматическая торговая система в клиентском терминале MetaTrader 5 // Современные наукоемкие технологии. – 2018. – № 10. – С. 56–60.
3. *Кондратьева Т.Н.* Принцип работы трендовых индикаторов // Интернет-журнал Науковедение. 2013. – № 3. – URL: <https://naukovedenie.ru/PDF/63trgsu313.pdf> (дата обращения: 11.09.2022).
4. *Каспаринский Ф.О.* Принципы мультитрейдинга // Электронные библиотеки. – 2021. – Т. 24. № 5. – С. 808-869. – DOI 10.26907/1562-5419-2021-24-5-808-869.
5. *Еремин В.В.* Выбор технического индикатора для российского фондового рынка // Аллея науки. 2020. – Т. 1, № 11(50). – С. 97-105.
6. *Nazário R.T.F., Silva J.L., Sobreiro V.A., Kimura H.A.* literature review of technical analysis on financial asset markets // Q. Rev. Econ. Financ. – 2017. – Vol. 66. – P. 115-126.
7. *Scott G., Carr M, Cremonie M.* Technical Analysis: Modern Perspectives; CFA Institute Research Foundation: Charlottesville, VA, USA, 2016.
8. *Fernandez E., Navarro J., Solares E., Coello C.C.* A novel approach to select the best portfolio considering the preferences of the decision maker // Swarm Evol. Comput. – 2019. – Vol. 46. – P. 140-153.
9. *Saul J., Jun M.* Modeling Technical Analysis; EPSRC: Swindon, UK, 2018; Chapter 1. – P. 3.
10. *Irwin S.H., Park C.-H.* What do we know about the profitability of technical analysis? // J. Econ. Surv. – 2007. – No. 21. – P. 786-826.
11. Market Technicians Association. CMT Level I 2016: An Introduction to Technical Analysis; JohnWiley & Sons: Hoboken, NJ, USA, 2015.
12. *Северюгин Ю.В., Сошин Н.А.* Классификация автоматизированных биржевых торговых стратегий – торговых роботов // Актуальные проблемы и перспективы развития экономики: российский и зарубежный опыт. – 2022. – № 2 (40). – С. 7-11.
13. *Гасымова Л.А.* Создание торгового робота на основе стохастического осциллятора для торговли привилегированными акциями Сбербанка // Математическое и компьютерное моделирование в экономике, страховании и управлении рисками. – 2021. – № 6. – С. 47-52.
14. *Доловская Ю.А.* Создание торгового робота для автоматической торговли на фондовом рынке // Математическое и компьютерное моделирование в экономике, страховании и управлении рисками. – 2017. – № 2. – С. 46-49.
15. *Тимофеев А.Г., Лебединская О.Г.* Создание торговой программы-робота для алгоритмической торговли и биржевых операций // Экономика и управление: проблемы, решения. – 2018. – Т. 4, № 4. – С. 97-104.
16. *Белоконская Е.Г., Калягин И.И.* Развитие инструментальных средств алготрейдинга на основе применения нейронных сетей // Известия высших учебных заведений. Серия: Экономика, финансы и управление производством. – 2017. – № 1 (31). – С. 53-56.
17. *Кондратьева Т.Н.* Комплексный подход анализа современных индикаторов финансового рынка // Обозрение прикладной и промышленной математики. – 2015. – Т. 22, № 5. – С. 589-590.
18. *Daniali, S.M.; Barykin, S.E.; Kapustina, I.V.; Mohammadbeigi Khortabi, F.; Sergeev, S.M.; Kalinina, O.V.; Mikhaylov, A.; Veynberg, R.; Zasova, L.; Senjyu, T.* Predicting Volatility Index According to Technical Index and Economic Indicators on the Basis of Deep Learning Algorithm // Sustainability. – 2021. – No. 13. – P. 14011. – <https://doi.org/10.3390/su132414011>.
19. *Rodríguez-Cándido, N.P., Espin-Andrade R.A., Solares E., Pedrycz W.A.* Compensatory Fuzzy Logic Model in Technical Trading // Axioms. – 2021. – No. 10. – P. 36. – <https://doi.org/10.3390/axioms10010036>.

20. *Yanyun Yao, Shangzhen Cai & Huimin Wang*. Are technical indicators helpful to investors in china's stock market? A study based on some distribution forecast models and their combinations // *Economic Research-Ekonomika Istraživanja*. – 2022. – Vol. 35:1. – P. 2668-2692. – DOI: 10.1080/1331677X.2021.1974921.
21. *Han Yufeng, Liu Yang, Zhou Guofu and Zhu Yingzi*. Technical Analysis in the Stock Market: A Review (May 21, 2021). Available at SSRN: <https://ssrn.com/abstract=3850494> or <http://dx.doi.org/10.2139/ssrn.3850494>.

REFERENCES

1. *Honore G.S.* Instrumenty tekhnicheskogo analiza s ispol'zovaniem Python i MetaTrader 5 [Technical analysis tools using Python and MetaTrader 5], *Obrazovanie i nauka bez granic: social'no-gumanitarnye nauki* [Education and science without borders: social sciences and humanities], 2022, No 17, pp. 46-52.
2. *Kondratieva T.N., Rveveeva I.F.* Avtomaticheskaya torgovaya sistema v klientskom terminale MetaTrader 5 [Automated trading system in the MetaTrader 5 client terminal], *Sovremennye naukoemkie tekhnologii* [Modern high technologies], 2018, No 10. pp. 56-60.
3. *Kondratieva T.N.* Printsip raboty trendovykh indikatorov [The principle of operation of trend indicators], *Internet-zhurnal Naukovedenie* [Journal of Naukovedenie], 2013, No. 3. Available at: <https://naukovedenie.ru/PDF/63trgsu313.pdf> (accessed 09/11/2022).
4. *Kasparinsky F.O.* Printsipy mul'titreydinga [Principles of multitrading]. *Elektronnye biblioteki* [Electronic libraries], 2021, Vol. 24, No 5, pp. 808-869. DOI 10.26907/1562-5419-2021-24-5-808-869.
5. *Eremín V.V.* Vybor tekhnicheskogo indikatora dlya rossiyskogo fondovogo rynka [The choice of a technical indicator for the Russian stock market], *Alleya nauki* [Alley of Science], 2020, Vol. 1, No 11 (50), pp. 97-105.
6. *Nazário R.T.F., Silva J.L., Sobreiro V.A., Kimura H.A.* literature review of technical analysis on financial asset markets, *Q. Rev. Econ. Financ.*, 2017, Vol. 66, pp. 115-126.
7. *Scott G., Carr M, Cremonie M.* Technical Analysis: Modern Perspectives; CFA Institute Research Foundation: Charlottesville, VA, USA, 2016.
8. *Fernandez E., Navarro J., Solares E., Coello C.C.* A novel approach to select the best portfolio considering the preferences of the decision maker, *Swarm Evol. Comput.*, 2019, Vol. 46, pp. 140-153.
9. *Saul J., Jun M.* Modeling Technical Analysis; EPSRC: Swindon, UK, 2018; Chapter 1, pp. 3.
10. *Irwin S.H., Park C.-H.* What do we know about the profitability of technical analysis?, *J. Econ. Surv.*, 2007, No. 21, pp. 786-826.
11. Market Technicians Association. CMT Level I 2016: An Introduction to Technical Analysis; JohnWiley & Sons: Hoboken, NJ, USA, 2015.
12. *Sevryugin Yu.V., Soshin N.A.* Klassifikatsiya avtomatizirovannykh birzhevykh torgovykh strategiy – torgovykh robotov [Classification of automated exchange trading strategies – trading robots]. *Aktual'nye problemy i perspektivy razvitiya ekonomiki: rossiyskiy i zarubezhnyy opyt* [Actual problems and prospects for the development of the economy: Russian and foreign experience], 2022, No 2 (40), pp. 7-11.
13. *Gasimova L.A.* Sozdanie torgovogo robota na osnove stokhasticheskogo ostsillyatora dlya torgovli privilegirovannymi aktsiyami Sberbanka [Creation of a trading robot based on a stochastic oscillator for trading preferred shares of Sberbank], *Matematicheskoe i komp'yuternoe modelirovanie v ekonomike, strakhovanii i upravlenii riskami* [Mathematical and computer modeling in economics, insurance and risk management], 2021, No. 6, pp. 47-52.
14. *Dolgovskaya Yu.A.* Cozdanie torgovogo robota dlya avtomaticheskoy torgovli na fondovom rynke [Creation of a trading robot for automatic trading in the stock market], *Matematicheskoe i komp'yuternoe modelirovanie v ekonomike, strakhovanii i upravlenii riskami* [Mathematical and computer modeling in economics, insurance and risk management], 2017, No. 2, pp. 46-49.
15. *Timofeev A.G., Lebedinskaya O.G.* Sozdanie torgovoy programmy-robota dlya algoritmicheskoy torgovli i birzhevykh operatsiy [Creation of a trading robot program for algorithmic trading and exchange operations], *Ekonomika i upravlenie: problemy, resheniya* [Economics and management: problems, solutions]. 2018. Vol. 4. No 4. pp. 97-104.

16. *Belokonskaya E.G., Kalyagin I.I.* Razvitie instrumental'nykh sredstv algotreydinga na osnove primeneniya neyronnykh setey [Development of tools for algorithmic trading based on the use of neural networks]. *Izvestiya vysshikh uchebnykh zavedeniy. Seriya: Ekonomika, finansy i upravlenie proizvodstvom* [News of higher educational institutions. Series: Economics, finance and production management], 2017, No 1 (31), pp. 53-56.
17. *Kondratieva T.N.* Kompleksnyy podkhod analiza sovremennykh indikatorov finansovogo rynka [An integrated approach to the analysis of modern indicators of the financial market], *Obozrenie prikladnoy i promyshlennoy matematiki* [Review of Applied and Industrial Mathematics], 2015, Vol. 22, No 5, pp. 589-590.
18. *Daniali S.M., Barykin S.E., Kapustina I.V., Mohammadbeigi Khortabi F., Sergeev S.M., Kalinina O.V., Mikhaylov A., Veynberg R., Zasova L., Senjyu T.* Predicting Volatility Index According to Technical Index and Economic Indicators on the Basis of Deep Learning Algorithm, *Sustainability*, 2021, No. 13, pp. 14011. Available at: <https://doi.org/10.3390/su132414011>.
19. *Rodríguez-Cándido, N.P., Espin-Andrade R.A., Solares E., Pedrycz W.A.* Compensatory Fuzzy Logic Model in Technical Trading, *Axioms*, 2021, No. 10, pp. 36. Available at: <https://doi.org/10.3390/axioms10010036>.
20. *Yanyun Yao, Shangzhen Cai & Huimin Wang.* Are technical indicators helpful to investors in china's stock market? A study based on some distribution forecast models and their combinations, *Economic Research-Ekonomika Istraživanja*, 2022, Vol. 35:1, P. 2668-2692. DOI: 10.1080/1331677X.2021.1974921.
21. *Han Yufeng, Liu Yang, Zhou Guofu and Zhu Yingzi.* Technical Analysis in the Stock Market: A Review (May 21, 2021). Available at SSRN: <https://ssrn.com/abstract=3850494> or <http://dx.doi.org/10.2139/ssrn.3850494>.

Статью рекомендовал к опубликованию д.т.н. Э.В. Мельник.

Кондратьева Татьяна Николаевна – Донской государственный технический университет; e-mail: ktn618@yandex.ru; г. Ростов-на-Дону, Россия; тел.: 88632738628; кафедра математики и информатики; к.т.н.; доцент.

Развеева Ирина Фёдоровна – e-mail: razveevai@mail.ru; тел.: 88632738628; кафедра математики и информатики; старший преподаватель.

Мунтян Евгения Ростиславна – Южный федеральный университет; e-mail: ermuntyan@sfnu.ru; г. Таганрог, Россия; тел.: 88634371608; кафедра вычислительной техники; к.т.н.; доцент.

Kondratieva Tatyana Nikolaevna – Don State Technical University; e-mail: ktn618@yandex.ru; Rostov-on-Don, Russia; phone: +78632738628; the department of mathematics and informatics; associate professor.

Razveeva Irina Fedorovna – e-mail: razveevai@mail.ru; phone: +78632738628; the department of mathematics and informatics; senior lecturer.

Muntyan Evgenia Rostislavna – Southern Federal University; e-mail: ermuntyan@sfnu.ru; Taganrog, Russia; phone: +78634371608; the department of computer science; cand. of eng. sc.; associate professor.

И.И. Турулин, Аль-Карави Хуссейн Шукор Мукер

МЕТОДИКА И АЛГОРИТМ СИНТЕЗА УПРАВЛЯЕМЫХ ЦИФРОВЫХ ФИЛЬТРОВ ЧЕБЫШЕВА I РОДА НИЖНИХ ЧАСТОТ НА БАЗЕ МЕТОДА БИЛИНЕЙНОГО ПРЕОБРАЗОВАНИЯ

Приводятся методика синтеза управляемых цифровых рекурсивных фильтров нижних частот Чебышева I рода с бесконечной импульсной характеристикой. Амплитудно-частотная характеристика таких фильтров имеет пульсации в полосе пропускания и является максимально плоской в полосе заграждения. Под управляемостью понимается явная зависимость коэффициентов фильтра от частоты среза. Методика основана на билинейном преобразовании передаточной функции аналогового фильтра-прототипа нижних частот и частотном преобразовании амплитудно-частотных характеристик полученного цифрового фильтра. Основная идея методики состоит в том, что для аналогового фильтра-прототипа с частотой среза 1 рад/с параметры передаточной функции биквадратных или билинейных звеньев, имеющие размерность частоты, будут численно равны поправочным коэффициентам для аналогичных параметров управляемого фильтра с произвольной частотой среза. В качестве примера рассмотрен синтез цифрового фильтра Чебышева I рода V порядка. В данной статье передаточная функция фильтра произвольного порядка представляется в виде каскадного соединения звеньев II порядка, если фильтр чётного порядка. В случае нечётного порядка больше единицы добавляется одно каскадно включенное звено I порядка. Несмотря на относительную простоту частотного преобразования, при практическом использовании его для цифровых фильтров, синтезированных с помощью систем автоматизированного проектирования цифровых фильтров (или с помощью справочников, содержащих рассчитанные фильтры-прототипы нижних частот для различных аппроксимаций амплитудно-частотной характеристики идеального фильтра нижних частот) возникает ряд нетривиальных специфических моментов, затрудняющих инженерное использование такого способа синтеза управляемых цифровых фильтров. Поэтому кроме методики разработан пошаговый алгоритм, позволяющий синтезировать фильтр без знания этих моментов. Алгоритм реализован в среде Mathcad, в качестве примера рассчитан цифровой рекурсивный фильтр Чебышева I рода V порядка. В примере приводятся рассчитанные коэффициенты цифрового управляемого фильтра нижних частот, явно зависящие от частоты среза, амплитудно-частотные характеристики этого фильтра и его низкочастотного прототипа, преобразованного в фильтр с такой же частотой среза, амплитудно-частотные характеристики приведены в одних координатах. Благодаря хорошей формализации алгоритма последний пригоден для реализации систем автоматизированного проектирования управляемых цифровых фильтров нижних частот Чебышева I рода.

Фильтр; цифровой; управляемый; перестраиваемый; БИХ; бесконечная импульсная характеристика; синтез; методика; фильтр цифровой; цифровой управляемый фильтр; Чебышева I рода.

I.I. Turulin, Al-Karawi Hussein Shakor Mogheer

METHOD AND ALGORITHM FOR SYNTHESIS OF CONTROLLED CHEBYSHEV DIGITAL FILTERS OF THE FIRST KIND OF LOW FREQUENCIES BASED ON THE BILINEAR TRANSFORMATION METHOD

The article presents a technique for the synthesis of controlled digital recursive Chebyshev low-pass filters of the first kind with an infinite impulse response. The frequency response of such filters has ripples in the passband and is as flat as possible in the stopband. Controllability is understood as an explicit dependence of the filter coefficients on the cutoff frequency. The technique is based on the bilinear transformation of the transfer function of the analog low-pass filter prototype and the frequency transformation of the amplitude-frequency characteristics of the obtained digital filter. The main idea of

the technique is that for an analog prototype filter with a cutoff frequency of 1 rad/s, the parameters of the transfer function of biquadratic or bilinear links, which have the dimension of frequency, will be numerically equal to the correction factors for similar parameters of a controlled filter with an arbitrary cutoff frequency. As an example, the synthesis of a digital Chebyshev filter of the first kind of the fifth order is considered. In this article, the transfer function of an arbitrary order filter is represented as a cascade connection of Π order links if the filter is of an even order. In the case of an odd order greater than one, one cascaded link of the first order is added. Despite the relative simplicity of the frequency conversion, in its practical use for digital filters synthesized using computer-aided design of digital filters (or using reference books containing calculated prototype low-pass filters for various approximations of the frequency response of an ideal low-pass filter), a series arises non-trivial specific moments that complicate the engineering use of this method of synthesizing controlled digital filters. Therefore, in addition to the technique, a step-by-step algorithm has been developed that allows one to synthesize a filter without knowing these moments. The algorithm is implemented in the Mathcad environment; as an example, a digital recursive Chebyshev filter of the 1st kind of the 5th order is calculated. The example shows the calculated coefficients of a digital controlled low-pass filter, which explicitly depend on the cutoff frequency, the amplitude-frequency characteristics of this filter and its low-frequency prototype converted into a filter with the same cutoff frequency, the amplitude-frequency characteristics are given in the same coordinates. Due to the good formalization of the algorithm, the latter is suitable for the implementation of computer-aided design systems for controlled digital Chebyshev low-pass filters of the first kind.

Filter; digital; controlled; tunable; IIR; infinite impulse response; synthesis; technique; digital filter; digital controlled filter; Chebyshev type I.

Введение, актуальность. В современной технике, в том числе радиотехнике, инфокоммуникационных системах, применяются управляемые (перестраиваемые) цифровые фильтры [1–3]. Это, прежде всего, различные адаптивные системы, в которых реализуются адаптивное сжатие, адаптивная дискретизация, адаптивные фильтры [4–6], а также оптимальный прием сигналов на фоне помех, адаптивные антенные решетки [7, 8] и т.п.

В одних случаях возникает задача разработки перестраиваемых цифровых фильтров по возможности с малой вычислительной сложностью алгоритмов или минимальной вычислительной сложностью перестройки [9–11], которые могли бы быть реализованы на микроконтроллерах, универсальных микропроцессорах и/или программируемых логических интегральных схемах (ПЛИС). При этом требования к качеству амплитудно-частотной характеристики (АЧХ) (коэффициент прямоугольности, уровень пульсаций в полосах пропускания и заграждения) таких фильтров зачастую невысоки.

В другом случае, наоборот, на первом месте стоят коэффициент прямоугольности, уровень пульсаций в полосах пропускания и заграждения.

Иногда необходимо сочетание качества АЧХ и малой вычислительной сложности перестройки.

Примеры перестраиваемых фильтров приведены в [12–14].

Под управляемостью в данной статье понимается явная зависимость коэффициентов фильтра от частоты среза.

Несмотря на относительную простоту частотного преобразования, при практическом использовании его для цифровых фильтров, синтезированных с помощью систем автоматизированного проектирования (САПР) цифровых фильтров (либо с помощью данных справочной литературы), возникает ряд нетривиальных специфических моментов, затрудняющих инженерное использование такого способа синтеза управляемых цифровых фильтров.

Целью данной разработки является разработка методики и пошагового алгоритма синтеза управляемых цифровых рекурсивных фильтров нижних частот (ФНЧ) Чебышева I рода с бесконечной импульсной характеристикой (БИХ), позволяющая избавить разработчика от знания этих специфических моментов.

Основная идея предлагаемой методики синтеза управляемых цифровых фильтров. Методика основана на расчете нужного аналогового прототипа с частотой среза 1 рад/с, с помощью САПР аналоговых фильтров. Прототип также может быть взят из соответствующего справочника. Все дальнейшие расчеты вплоть до получения коэффициентов цифрового фильтра входят в математический аппарат методики. Напомним, что коэффициенты полностью определяют цифровой фильтр (не считая погрешностей квантования).

Далее с помощью билинейного преобразования аналоговый ФНЧ-прототип преобразуется в цифровой ФНЧ.

В этом случае для фильтра-прототипа с частотой среза 1 рад/с параметры передаточной функции биквадратных или билинейных звеньев, и имеющие размерность частоты, будут численно равны поправочным коэффициентам аналогичных параметров управляемого фильтра с произвольной частотой среза (в формулах для коэффициентов звеньев цифрового фильтра).

Управляемый цифровой фильтр может быть синтезирован путем частотного преобразования амплитудно-частотной характеристики неуправляемого цифрового фильтра-прототипа [15–17], однако в данной статье рассматривается методика, основанная на расчете аналогового прототипа с частотой среза 1 рад/с и синтезом цифрового фильтра методом билинейного преобразования.

Исходные данные к разработке методики. Передаточная функция аналогового фильтра или другой цепи (преобразование Лапласа) [18–20]

$$H(p) = \frac{c_m p^m + c_{m-1} p^{m-1} + \dots + c_1 p + c_0}{p^n + d_{n-1} p^{n-1} + \dots + d_1 p + d_0}, \quad (1)$$

где c_m и d_n – постоянные коэффициенты, те же, что и в дифференциальном уравнении, описывающем цепь или фильтр.

При $m \leq n$ (формула (1)), что допускает метод билинейного преобразования, возможны, в частности, два варианта представления $H(p)$: в виде произведения $H(p) = H_1(p)H_2(p)\dots H_K(p)$ либо суммы $H(p) = k + \sum_{i=1}^K H_i(p)$, где $K \leq n$; k – константа (как правило $k = 0$); $H_j(p)$ передаточная функция звена 2-го порядка (биквадратного) в общем случае [21–22]:

$$H_j(p) = \frac{c_{j2} p^2 + c_{j1} p + c_{j0}}{p^2 + d_{j1} p + d_{j0}} \quad (2)$$

или звена 1-го порядка (билинейного) в общем случае [21–22]

$$H_j(p) = \frac{c_{j1} p + c_{j0}}{p + d_{j0}}.$$

При этом j -е звено II порядка в формуле (1) имеет частоту j -го полюса ω_{0j} и добротность Q_j [21, 22]:

$$Q_j = \sqrt{d_{j0}}/d_{j1}, \quad \omega_{0j} = \sqrt{d_{j0}}. \quad (3)$$

Для фильтров Чебышева I рода предпочтительной является каскадная реализация, которой соответствует произведение звеньев $H(p) = H_1(p)H_2(p)\dots H_K(p)$, поскольку в этом случае звенья имеют только полюсы. Вследствие этого в статье используется именно такая реализация для упрощения выражений.

Поскольку импульсная характеристика аналогового ФНЧ-прототипа вещественная, $H(p)$ имеет комплексно-сопряженные полюсы (биквадратное звено) или вещественные полюсы (билинейное звено).

Выбор метода преобразования аналогового фильтра в цифровой. Метод билинейного преобразования [1–3] более прост, чем метод инвариантного преобразования импульсной характеристики, и, в отличие от последнего, как уже говорилось, допускает $m \leq n$, где m и n – соответственно степень числителя и знаменателя передаточной функции фильтра в целом или его звена. Кроме того, метод билинейного преобразования не имеет эффекта наложения АЧХ аналогового прототипа при переходе к цифровому фильтру. Вместо этого эффекта – деформация оси частот, которая легко учитывается в случае одной граничной частоты путем введения предвысказаний в значения граничных частот. Деформация оси частот является положительным моментом в том смысле, что ведет к росту крутизны АЧХ в переходной полосе и снижению уровня АЧХ в полосе заграждения.

Метод билинейного преобразования [1–3] состоит в замене переменной p в передаточной функции по Лапласу

$$p \rightarrow \frac{2}{T_d} \cdot \frac{1 - z^{-1}}{1 + z^{-1}} \quad (4)$$

и применяется для каждого аналогового звена II и/или I порядка при каскадной реализации ФНЧ-прототипа.

Для компенсации деформации оси частот перед расчетом фильтра граничные частоты среза ω_i пересчитываются по формуле [1–3]:

$$\omega_i^n = \frac{2}{T_d} \operatorname{tg} \frac{\omega_i T_d}{2}. \quad (5)$$

Рассмотрим аналоговые звенья I и II порядка.

Расчет поправочных коэффициентов. Под поправочными коэффициентами понимаются отношения частот нулей и полюсов аналогового прототипа к частоте среза этого прототипа. В случае частоты среза, равной 1 рад/с, поправочные коэффициенты численно равны частотам нулей (ω_n) и полюсов (ω_0) (формулы (8), (12)).

Например, расчет аналогового ФНЧ-прототипа Чебышева I рода V порядка с частотой среза 1 рад/с и неравномерностью в полосе пропускания 1 дБ дает

$$H(p) = 0.1228 \cdot \frac{1}{p^2 + 0.1789 p + 0.9883} \cdot \frac{1}{p^2 + 0.4684 p + 0.4293} \cdot \frac{1}{p + 0.2895}. \quad (7)$$

Для первой дроби (биквадратное звено) $Q_0 = 5.556$, $\omega_0 = 0.9941$; для второй дроби $Q_0 = 1.399$, $\omega_0 = 0.6552$; для третьей дроби (билинейное звено) $\omega_0 = 0.2895$.

Значения Q_0 и ω_0 для биквадратных звеньев часто рассчитывает САПР фильтров (для первого звена $\omega_0 = 0.9941$, $Q_0 = 5.556$; для второго звена $\omega_0 = 0.6552$, $Q_0 = 1.399$); но их нетрудно рассчитать вручную, сравнив знаменатели дробей формулы (7) и формулы (11) (см. также формулу (4)). Например, для первого биквадратного звена $\omega_0 = 0.9883^{0.5} = 0.9941$, $Q_0 = \omega_0/0.1789 = 5.557$ (формулы (2), (4), (7), (12)). Заметим, что САПР выполняет расчеты с более высокой точностью, чем выводимые 4 десятичных разряда (этот параметр обычно можно изменять) и округляет их до заданного числа значащих цифр (в данном примере выбрано 4); этим и объясняется различие на единицу младшего разряда при расчете Q_0 .

Поправочные коэффициенты для первого биквадратного звена $K_{\omega n} = 1$ (для нуля; формулы (7), (12)) и $K_{\omega 0} = \omega_0 = 0.9883^{0.5} = 0.9941$; $Q_0 = \omega_0/0.1789 = 5.556$ (для полюса, формулы те же). Для второго звена $K_{\omega n} = 1$, $K_{\omega 0} = \omega_0 = 0.4293^{0.5} = 0.6552$; $Q_0 = \omega_0/0.4684 = 1.399$. Для третьего звена $K_{\omega n} = 1$ (для нуля; формулы (7), (12)), $K_{\omega 0} = \omega_0 = 0.2895$.

Коэффициент перед дробями можно учесть в виде множителя в цифровом фильтре (в примере реализации фильтра так и сделано).

Лучшим вариантом может быть распределение коэффициента между звеньями II порядка, поскольку они усиливают сигнал. Это улучшит масштабирование цифрового сигнала (поддержание примерно постоянного уровня при цифровой фильтрации), что особенно важно для целочисленной арифметики в цифровом фильтре, поскольку снижается возможность выхода цифрового сигнала за пределы разрядной сетки. В этом случае для каждого биквадратного звена будет коэффициент $0.1228^{0.5} = 0.3504$. Сравнив передаточные функции таких биквадратных звеньев с формулой (12), получим $\omega_n = 0.3504^{0.5} = 0.5920 = K_{\omega_n}$. Однако в этом в этом случае вычислительная сложность цифровой фильтрации увеличивается на одно умножение. В примере расчета управляемого фильтра применен отдельный множитель.

Расчет коэффициентов цифрового эквивалента аналогового билинейного звена для ФНЧ-прототипа Чебышева I рода нечетного порядка. Передаточная функция используемого билинейного звена аналогового ФНЧ-прототипа Чебышева I рода нечетных порядков, имеющего только полюсы (формула (3)) запишется в следующем виде:

$$H(p) = \frac{\omega_n}{p + \omega_0}, \quad (8)$$

где

$$\omega_0 = 2\pi f_0 - \text{круговая частота полюса}, \quad (9)$$

f_0 – циклическая частота полюса,

ω_n – круговая частота нуля.

Передаточная функция цифрового звена I порядка

$$H(z) = \frac{a_0 + a_1 z^{-1}}{1 - b_1 z^{-1}}. \quad (10)$$

Применив билинейное преобразование (4) к формуле (8), с учетом предсказаний (формула (6)) и поправочных коэффициентов, определяемых из формулы (7), после преобразований получим

$$\begin{aligned} k_{01} &= \frac{1}{2 + K_{\omega_0} \omega_c^n T_d}; \\ a_0 &= k_{01} K_{\omega_n} \omega_c^n T_d; \\ a_1 &= k_{01} K_{\omega_n} \omega_c^n T_d; \\ b_1 &= k_{01} (2 - K_{\omega_0} \omega_c^n T_d), \end{aligned} \quad (11)$$

где ω_c^n – круговая частота среза управляемого цифрового фильтра с учетом предсказаний (формула (6));

K_{ω_n} и K_{ω_0} – поправочные коэффициенты, численно равные частоте нуля и полюса соответственно передаточной функции звена I порядка ФНЧ-прототипа с частотой среза 1 рад/с (пример передаточной функции для ФНЧ-прототипа Чебышева I рода V порядка с частотой среза 1 рад/с и неравномерностью в полосе пропускания 1 дБ приведен в формуле (7)),

T_d – шаг дискретизации.

Расчет коэффициентов цифрового эквивалента аналогового биквадратного звена для ФНЧ-прототипа Чебышева I рода при каскадной реализации. Передаточная функция аналогового звена II порядка

$$H(p) = \frac{\omega_n^2}{p^2 + \frac{\omega_0}{Q_0} p + \omega_0^2} \quad (12)$$

где ω_0 и Q_0 – частота и добротность полюса ($\Delta\omega = \omega_0 / Q_0$ – полоса пропускания по уровню 0,707); $\omega_n = 1$ рад/с.

Системная функция цифрового фильтра определяется по формуле

$$H(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2}}{1 - b_1 z^{-1} - b_2 z^{-2}}, \quad (13)$$

где

$$k_0 = \frac{1}{4 + 2 \frac{K_{\omega_0} \omega_c^n T_d}{Q_0} + (K_{\omega_0} \omega_c^n T_d)^2};$$

$$k_n = (K_{\omega_n} \omega_c^n T_d)^2;$$

$$a_0 = k_0 k_n; \quad (14)$$

$$a_1 = 2k_0 (K_{\omega_n} \omega_c^n T_d)^2;$$

$$a_2 = k_0 (K_{\omega_n} \omega_c^n T_d)^2;$$

$$b_1 = k_0 [8 - 2(K_{\omega_0} \omega_c^n T_d)^2];$$

$$b_2 = k_0 \left[2 \frac{K_{\omega_0} \omega_c^n T_d}{Q_0} - (K_{\omega_0} \omega_c^n T_d)^2 - 4 \right];$$

ω_c^n – частота среза с предискажениями (для компенсации искажений оси частот метода билинейного преобразования; формула (5));

K_{ω_n} и K_{ω_0} – поправочные коэффициенты, численно равные частоте нуля и полюса соответственно передаточной функции звена ФНЧ-прототипа с частотой среза 1 рад/с (пример передаточной функции для ФНЧ-прототипа Чебышева III рода V порядка с частотой среза 1 рад/с и неравномерностью в полосе пропускания 1 дБ приведен в формуле (7)),

T_d – шаг дискретизации.

Таким образом, получены формулы для синтеза цифровых управляемых фильтров с помощью САПР цифровых фильтров или справочников. Подсистемы расчета цифровых фильтров имеются во многих приложениях, например, в системе MATLAB [23–25].

Алгоритм синтеза управляемого цифрового фильтра для билинейного преобразования на примере ФНЧ Чебышева I рода V порядка с частотой среза 1 рад/с и неравномерностью в полосе пропускания 1 дБ.

1. Задать порядок фильтра N_F .
2. Задать циклическую частоту среза f_c (или круговую частоту среза ω_c).
3. Задать циклическую частоту дискретизации f_d (или круговую частоту дискретизации ω_d).
4. С помощью формулы (6) определить предискаженную частоту среза ω_c^n .
5. Рассчитать или взять из справочника аналоговый ФНЧ-прототип с частотой среза 1 Гц и каскадной реализацией. Он будет иметь $N_{II} = \text{floor}(N_F/2)$ звеньев II порядка и $N_I = N_F - 2 \cdot \text{floor}(N_F/2)$ звеньев I порядка. Если N_F чётное, то $N_I = 0$, если нечётное, то $N_I = 1$. Здесь $\text{floor}(x)$ – наибольшее целое, не большее x .

6. Сравнить полученные передаточные функции звеньев с формулой (8) (звено I порядка; имеет место для нечетного N_F) и формулой (12) (звено или звенья II порядка; имеют место при $N_F > 1$) (если вид формул передаточных функций звеньев прототипа отличается от (8) и/или (12), то формулы этих функций приводят к виду (8) и/или (12), и производят сравнение). В результате для звена I порядка (при его наличии) определяются поправочные коэффициенты: $K_{\omega_n} = \omega_n/\omega_{сп}$, где ω_n – частота нуля звена аналогового прототипа, $\omega_{сп}$ – частота среза этого прототипа; $K_{\omega_0} = \omega_0/\omega_{сп}$, ω_0 – частота полюса биквадратного звена аналогового прототипа. Если частота среза прототипа равна 1 рад/с, поправочные коэффициенты численно равны частотам нулей или полюсов. Для звена II порядка определяется также Q_0 – добротность полюса ($\Delta\omega = \omega_0/Q_0$ – полоса пропускания по уровню 0,707). В некоторых САПР, например, Filter Solutions (бесплатная версия – Filter Free (N_F – до 3-х)), при нажатии на кнопку Transfer Function выдается передаточная функция, сверху которой приведены нули биквадратных звеньев ω_n (обозначены как Wn), снизу – полюса ω_0 (обозначены как W0) и добротности Q_0 (обозначены как Q). Если нажать на кнопку Print, можно передаточную функцию со всеми сопутствующими параметрами напечатать, например, на виртуальном PDF-принтере; например, на виртуальном принтере «Microsoft Print to PDF» PDF-файл получается с текстовым слоем, что позволяет легко копировать полученные параметры в любую программу, например, Mathcad.

7. Аналоговые звенья I и II порядка преобразуются с помощью методики в цифровые (их число равно числу аналоговых звеньев). Методика реализует метод билинейного преобразования. Коэффициенты цифрового звена I порядка (оно имеется, если N_F нечетно) с системной функцией (10) определяются формулой (11). Коэффициенты каждого цифрового звена II порядка с системной функцией (13) определяются формулой (14), в этих формулах ω_c^II – частота среза цифрового управляемого фильтра (в целом), в значении которой введены предсказания (для компенсации искажений оси частот метода билинейного преобразования; формула (6));

8. Строятся цифровые звенья II и I порядка и включаются каскадно. (см. ниже пример расчета характеристик цифрового управляемого фильтра).

Пример расчета цифрового управляемого ФНЧ Чебышева I рода V порядка с частотой среза 1 рад/с и неравномерностью в полосе пропускания 1 дБ в системе Mathcad с помощью пошагового алгоритма на основе билинейного преобразования.

1. $fd := 20 \cdot 10^3$ частота дискретизации.
2. $td := \frac{1}{fd}$ шаг дискретизации.
3. $f_{sr} := 4 \cdot 10^3$ циклическая частота среза фильтра.
4. $w0 := 2 \cdot \pi \cdot f_{sr}$ круговая частота среза фильтра.

Введение предсказаний

5. $w0p := \frac{2}{td} \cdot \tan\left(\frac{w0 \cdot td}{2}\right)$ круговая частота среза с предсказаниями.
6. $\Omega_0 := w0p \cdot td$ относительная частота среза с предсказаниями.
7. $f := 0, \frac{fd}{600} .. \frac{fd}{2}$ диапазон отображаемых на графике АЧХ частот.

Передаточная функция аналогового билинейного звена

$$8. \quad H_{a1}(p, \omega_n, \omega_0) := \frac{\omega_n}{p + \omega_0}$$

Блок расчета коэффициентов билинейного звена

$$9. \quad k_{01}(\omega_{cp}) := \frac{1}{2 + \omega_{cp}}$$

$$10. \quad a_{0_}(\omega_{np}, \omega_{cp}) := k_{01}(\omega_{cp}) \cdot \omega_{np}$$

$$11. \quad a_{1_}(\omega_{np}, \omega_{cp}) := k_{01}(\omega_{cp}) \cdot \omega_{np}$$

$$12. \quad b_{1_}(\omega_{np}, \omega_{cp}) := k_{01}(\omega_{cp}) \cdot (2 - \omega_{cp})$$

$$13. \quad H_{d_}(z, \omega_{np}, \omega_{cp}) := \frac{a_{0_}(\omega_{np}, \omega_{cp}) \cdot z^0 + a_{1_}(\omega_{np}, \omega_{cp}) \cdot z^{-1}}{1 - b_{1_}(\omega_{np}, \omega_{cp}) \cdot z^{-1}}$$

Передаточная функция аналогового биквадратного звена

$$14. \quad H_a(p, \omega_n, \omega_0, Q_0) := \frac{\omega_n^2}{p^2 + p \frac{\omega_0}{Q_0} + \omega_0^2}$$

Блок расчета коэффициентов биквадратного звена

$$15. \quad k_n(\omega_n) := \omega_n^2$$

$$16. \quad k_0(\omega_0, Q_0) := 4 + 2 \cdot \frac{\omega_0}{Q_0} + \omega_0^2$$

$$17. \quad a_0(\omega_n, \omega_0, Q_0) := \frac{k_n(\omega_n)}{k_0(\omega_0, Q_0)}$$

$$18. \quad a_1(\omega_n, \omega_0, Q_0) := \frac{2\omega_n^2}{k_0(\omega_0, Q_0)}$$

$$19. \quad a_2(\omega_n, \omega_0, Q_0) := \frac{\omega_n^2}{k_0(\omega_0, Q_0)}$$

$$20. \quad b_1(\omega_0, Q_0) := \frac{8 - 2 \cdot \omega_0^2}{k_0(\omega_0, Q_0)}$$

$$21. \quad b_2(\omega_0, Q_0) := \frac{-4 + 2 \cdot \frac{\omega_0}{Q_0} - \omega_0^2}{k_0(\omega_0, Q_0)}$$

Системная функция цифрового биквадратного фильтра

$$22. \quad H_d(z, \omega_n, \omega_0, Q_0) := \frac{a_0(\omega_n, \omega_0, Q_0) \cdot z^0 + a_1(\omega_n, \omega_0, Q_0) \cdot z^{-1} + a_2(\omega_n, \omega_0, Q_0) \cdot z^{-2}}{1 - b_1(\omega_0, Q_0) \cdot z^{-1} - b_2(\omega_0, Q_0) \cdot z^{-2}}$$

Передаточные функции аналоговых биквадратных звеньев

23. $BQa1(p) := Ha(p, w0, w0 \cdot 0.9941, 5.556)$.

24. $BQa2(p) := Ha(p, w0, w0 \cdot 0.6552, 1.399)$

25. $BL(p) := Ha1(p, w0, w0 \cdot 0.2895)$.

Передаточная функция аналогового ФНЧ Чебышева I рода V порядка

26. $Ha5(p) := BQa1(p) \cdot BQa2(p) \cdot BL(p) \cdot 0.1228$

Системные функции цифровых звеньев II порядка

27. $BQ1(z) := Hd(z, \Omega 0, \Omega 0 \cdot 0.9941, 5.556)$.

28. $BQ2(z) := Hd(z, \Omega 0, \Omega 0 \cdot 0.6552, 1.399)$.

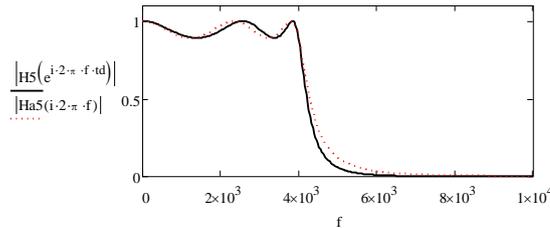
29. $BQ_{-}(z) := Hd_{-}(z, \Omega 0, \Omega 0 \cdot 0.2895)$.

Системная функция цифрового ФНЧ Чебышева I рода V порядка

30. $H5(z) := BQ_{-}(z) \cdot BQ1(z) \cdot BQ2(z) \cdot 0.1228$

АЧХ цифрового ФНЧ Чебышева I рода V порядка (черный цвет) и такого же аналогового фильтра (красные точки)

31.



Эксперименты с моделью управляемого фильтра при перестройке фильтра на лету. Было проведено 4 эксперимента. В этих экспериментах на вход управляемого фильтра подавалась синусоида частотой 1 кГц, 2 кГц, 2.2 кГц и 3 кГц соответственно. В ходе каждого эксперимента фильтр имел начальную частоту среза 4 кГц, которая скачком становилась равной 2 кГц при $n = 800$, где n – дискретное время; частота дискретизации составляла 160 кГц (в приведенной выше Mathcad-программе частота дискретизации была равной 20 кГц).

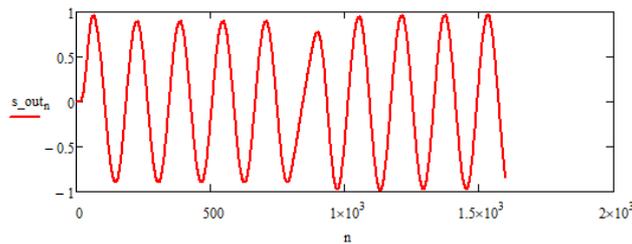


Рис. 1. График выходного сигнала s_{out_n} управляемого фильтра с переключением частоты среза с 4 кГц на 2 кГц при $n = 800$; частота входного сигнала 1 кГц

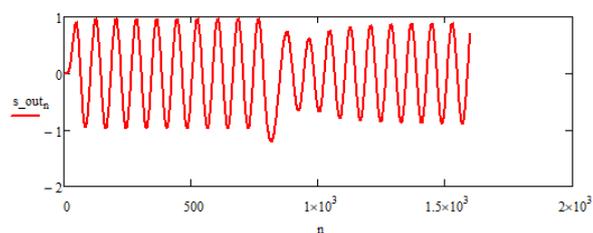


Рис. 2. График выходного сигнала s_{out_n} управляемого фильтра с переключением частоты среза с 4 кГц на 2 кГц при $n = 800$; частота входного сигнала 2 кГц

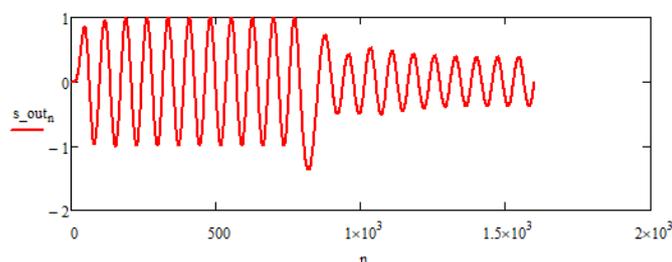


Рис. 3. График выходного сигнала s_{out_n} управляемого фильтра с переключением частоты среза с 4 кГц на 2 кГц при $n = 800$; частота входного сигнала 2.2 кГц

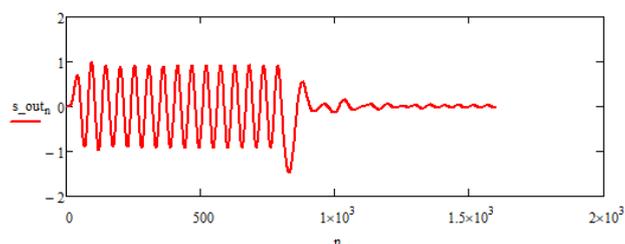


Рис. 4. График выходного сигнала управляемого фильтра с переключением частоты среза с 4 кГц на 2 кГц при $n = 800$; частота входного сигнала 3 кГц

Как видно из рис. 1–4, управляемый фильтр работоспособен.

Однако на рисунке 2 уровень выходного сигнала фильтра отличается от ожидаемого (0,707 от коэффициента передачи фильтра на нулевой частоте). Это связано с тем, что при расчете ФНЧ-прототипа его параметры были проквантованы до 4-х десятичных разрядов, что привело к относительно небольшому смещению АЧХ цифрового управляемого фильтра, которое, однако, стало заметным на графике выходного сигнала из-за большой крутизны переходной полосы АЧХ управляемого фильтра.

Переключение частоты среза управляемого фильтра вызывает выброс на графике выходного сигнала фильтра, в общем случае разный для каждого эксперимента.

Выводы. Таким образом, предлагаемые методика и пошаговый алгоритм синтеза управляемых цифровых рекурсивных фильтров нижних частот Чебышева I рода с бесконечной импульсной характеристикой позволяет синтезировать цифровые управляемые фильтры.

Из-за деформации оси частот при использовании метода билинейного преобразования наряду с отсутствием эффекта наложения АЧХ фильтров (что является положительным моментом) наблюдается увеличение крутизны АЧХ в переходной полосе при переходе от аналогового фильтра к цифровому, что также является положительным эффектом. Сказанное иллюстрируется графиком АЧХ в тексте модели в системе Mathcad.

На базе такого алгоритма можно построить САПР цифровых управляемых БИХ-фильтров нижних частот Чебышева I рода.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Гольденберг Л.М., Матюшкин Б.Д., Поляк М.Н. Цифровая обработка сигналов: справочник. – М.: Радио и связь, 1985.
2. Лэм Г. Аналоговые и цифровые фильтры: расчет и реализация. – М.: Мир, 1982.
3. Рабинер Р., Гоулд Б. Теория и применение цифровой обработки сигналов. – М.: Мир, 1978.
4. Haykin S. Adaptive Filter Theory. – Prentice Hall, 2002.
5. Widrow B., Stearns S.D. Adaptive Signal Processing. – 1st. – Englewood Cliffs, NJ: Prentice-Hall, 1985.
6. Thakor N.V., Yi-Sheng Zhu. Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection // IEEE Transactions on Biomedical Engineering. – August 1991. – Vol. 38, Issue 8.
7. Монзинго Р.А., Миллер Т.У. Адаптивные антенные решетки. – М.: Радио и связь, 1986.
8. Журавлев А.К., Лукошкин А.П., Поддубный С.С. Обработка сигналов в адаптивных антенных решетках. – Л.: Изд-во Ленингр. ун-та, 1983. – 240 с.
9. Турулин И.И., Ткаченко М.Г. Быстроперестраиваемые цифровые фильтры: монография. – Таганрог: Изд-во ТТИ ЮФУ, 2008.
10. Haridas N., Elias E. Design of reconfigurable low-complexity digital hearing aid using farrow structure based variable bandwidth filters // J. Appl. Res. Technol. – 2016. – Vol. 2, No. 14. – P. 154-165.
11. Yu Y.J., Xu W.J. Mixed-radix fast filter bank approach for the design of variable digital filters with simultaneously tunable bandedge and fractional delay // IEEE Trans. Signal Process. – 2011. – Vol. 1, No. 60. – P. 100-111.
12. Kumar A., Suman S., Singh G. A new closed form method for design of variable bandwidth linear phase FIR filter using different polynomials // AEU-Int. J. Electr. Commun. – 2014. – Vol. 4, No. 68. – P. 351-360.
13. Illa A., Haridas N., Elias E. Design of multiplier-less FIR filters with simultaneously variable bandwidth and fractional delay // Engineering science and technology. – 2016. – Vol. 3, No. 19. – P. 1160-1165.
14. George J.T., Elias E. Continuously variable bandwidth sharp fir filters with low complexity // J. Signal Inf. Process. – 2012. – Vol. 3, No. 3. – P. 308-315.
15. Constantinides A.G. Spectral transformations for digital filters // Proceedings of the Institution of Electrical Engineers. – 1970. – Vol. 117. – P. 1585-1590.
16. Введение в цифровую фильтрацию / под ред. Р. Богнера, А. Константинодиса. – М.: Мир, 1976.
17. Опенгейм А.В., Шафер Р.В. Цифровая обработка сигналов. – М.: Связь, 1975.
18. Oppenheim A.V., Willsky A.S., and Young I.T. Signals and Systems. – Englewood Cliffs, NJ: Prentice Hall, 1983.
19. Антонию А. Цифровые фильтры: анализ и проектирование, – М.: Радио и связь, 1983.
20. Proakis J.G. Digital Signal Processing: Principles Algorithms and Applications. – Pearson Education India, 2001. tal Signal Processing: Theory, Analysis and Digital-Filter Design / B.S. Nair. – PHI Learning Pvt. Ltd., 2004.
21. Турулин И.И. Управляемые цифровые фильтры. – Таганрог: ЮФУ, 2009. – 260 с. – Рус. Деп. в ВИНТИ 18.06.09 № 383-B2009.
22. Турулин И.И., Булгакова Ю.И. Методика проектирования управляемых цифровых фильтров на основе метода билинейного преобразования // Авиакосмическое приборостроение. – 2011. – № 6. – С. 30-36.

23. *Turulin I.I., Mogheer H.S.* Analysis of Controlled Digital Recursive High-Pass Filters Structures with Infinite Non-Negative Impulse Response // 2022 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). – IEEE, 2022. – P. 755-759.
24. *Mogheer H.S., Turulin I.I.* Reduction of Signal Overshooting Caused by Cutoff Frequency Changing in the Controlled Digital Butterworth Low Pass Filter // 2022 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). – IEEE, 2022. – P. 783-788.
25. *Mogheer Hussein Sh., Turulin I.I.* Analysis of MATLAB System Applicability for Synthesis of Controlled Butterworth Digital Recursive IIR Filters // Известия ЮФУ. Технические науки. – 2021. – № 3 (220). – С. 72-81.

REFERENCES

1. *Gol'denberg L.M., Matyushkin B.D., Polyak M.N.* Tsifrovaya obrabotka signalov: spravochnik [Digital Signal Processing: handbook]. Moscow: Radio i svyaz', 1985.
2. *Lem G.* Analogovye i tsifrovye fil'try: raschet i realizatsiya [Analog and digital filters: calculation and implementation]. Moscow: Mir, 1982.
3. *Rabiner R., Gould B.* Teoriya i primeneniye tsifrovoy obrabotki signalov [Theory and application of digital signal processing]. Moscow: Mir, 1978.
4. *Haykin S.* Adaptive Filter Theory. Prentice Hall, 2002.
5. *Widrow B., Stearns S.D.* Adaptive Signal Processing. 1st. Englewood Cliffs, NJ: Prentice-Hall, 1985.
6. *Thakor N.V., Yi-Sheng Zhu.* Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection, *IEEE Transactions on Biomedical Engineering*, August 1991, Vol. 38, Issue 8.
7. *Monzingo R.A., Miller T.U.* Adaptivnyye antennnye reshetki [Adaptive antenna arrays]. Moscow: Radio i svyaz', 1986.
8. *Zhuravlev A.K., Lukoshkin A.P., Poddubnyy S.S.* Obrabotka signalov v adaptivnykh antennnykh reshetkakh [Signal processing in adaptive antenna arrays]. Leningrad: Izd-vo Leningr. un-ta, 1983, 240 p.
9. *Turulin I.I., Tkachenko M.G.* Bystroperestraivaemye tsifrovye fil'try: monografiya [Fast-tunable digital filters: a monograph]. Taganrog: Izd-vo TTI YuFU, 2008.
10. *Haridas N., Elias E.* Design of reconfigurable low-complexity digital hearing aid using farrow structure based variable bandwidth filters, *J. Appl. Res. Technol.*, 2016, Vol. 2, No. 14, pp. 154-165.
11. *Yu Y.J., Xu W.J.* Mixed-radix fast filter bank approach for the design of variable digital filters with simultaneously tunable bandedge and fractional delay, *IEEE Trans. Signal Process.*, 2011, Vol. 1, No. 60, pp. 100-111.
12. *Kumar A., Suman S., Singh G.* A new closed form method for design of variable bandwidth linear phase FIR filter using different polynomials, *AEU-Int. J. Electr. Commun.*, 2014, Vol. 4, No. 68, pp. 351-360.
13. *Illa A., Haridas N., Elias E.* Design of multiplier-less FIR filters with simultaneously variable bandwidth and fractional delay, *Engineering science and technology*, 2016, Vol. 3, No. 19, pp. 1160-1165.
14. *George J.T., Elias E.* Continuously variable bandwidth sharp fir filters with low complexity, *J. Signal Inf. Process.*, 2012, Vol. 3, No. 3, pp. 308-315.
15. *Constantinides A.G.* Spectral transformations for digital filters, *Proceedings of the Institution of Electrical Engineers*, 1970, Vol. 117, pp. 1585-1590.
16. *Vvedeniye v tsifrovuyu fil'tratsiyu* [Introduction to digital filtering], ed. by R. Bognera, A. Konstantinidisa. Moscow: Mir, 1976.
17. *Oppengeym A.V., Shafer R.V.* Tsifrovaya obrabotka signalov [Digital signal processing]. Moscow: Svyaz', 1975.
18. *Oppenheim A.V., Willsky A.S., and Young I.T.* Signals and Systems. Englewood Cliffs, NJ: Prentice Hall, 1983.
19. *Anton'yu A.* Tsifrovye fil'try: analiz i proektirovaniye [Digital filters: Analysis and design], Moscow: Radio i svyaz', 1983.
20. *Proakis J.G.* Digital Signal Processing: Principles Algorithms and Applications, Pearson Education India, 2001. tal Signal Processing: Theory, Analysis and Digital-Filter Design, B.S. Nair. PHI Learning Pvt. Ltd., 2004.

21. *Turulin I.I.* Upravlyaemye tsifrovye fil'try [Controlled digital filters]. Taganrog: YuFU, 2009, 260 p. Rus. Dep. v VINITI 18.06.09 № 383-V2009.
22. *Turulin I.I., Bulgakova Yu.I.* Metodika proektirovaniya upravlyaemykh tsifrovyykh fil'trov na osnove metoda bilineynogo preobrazovaniya [The methodology of designing controlled digital filters based on the method of bilinear transformation], *Aviakosmicheskoe priborostroenie* [Aerospace instrumentation], 2011, No. 6, pp. 30-36.
23. *Turulin I.I., Mogheer H.S.* Analysis of Controlled Digital Recursive High-Pass Filters Structures with Infinite Non-Negative Impulse Response, *2022 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*. IEEE, 2022, pp. 755-759.
24. *Mogheer H.S., Turulin I.I.* Reduction of Signal Overshooting Caused by Cutoff Frequency Changing in the Controlled Digital Butterworth Low Pass Filter, *2022 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*. IEEE, 2022, pp. 783-788.
25. *Mogheer Hussein Sh., Turulin I.I.* Analysis of MATLAB System Applicability for Synthesis of Controlled Butterworth Digital Recursive IIR Filters. *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 3 (220), pp. 72-81.

Статью рекомендовал к опубликованию д.т.н., профессор Н.Н. Чернов.

Турулин Игорь Ильич – Южный федеральный университет; e-mail: iiturulin@sfedu.ru; г. Таганрог, Россия; тел.: +78634371632; кафедра теоретических основ радиотехники; д.т.н.; профессор.

Аль-Карави Хусейн Шукор Мукер – e-mail: alkaravi@sfedu.ru; кафедра теоретических основ радиотехники; аспирант.

Turulin Igor Ilich – Southern Federal University; e-mail: iiturulin@sfedu.ru; Taganrog, Russia; phone: +78634371632; the department of theoretical foundations of radio engineering dr. of eng. sc.; professor.

Hussein Shakor Mogheer – e-mail: alkaravi@sfedu.ru; the department of theoretical foundations of radio engineering; postgraduate student.

УДК 004.67

DOI 10.18522/2311-3103-2022-4-182-192

А.А. Чусов, М.А. Копаева

АЛГОРИТМ ПОТОКОВОЙ КОМПРЕССИИ ДАННЫХ С ПЛАВАЮЩЕЙ ЗАПЯТОЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ ОБЕСПЕЧЕНИЯ НАУЧНЫХ ЭКСПЕРИМЕНТОВ

Представлен оригинальный алгоритм и метод реализации однопроходной компрессии потока числовых данных с плавающей запятой. Целью работы является разработка и формализация алгоритмического метода однопроходной потоковой компрессии числовых значений с плавающей запятой, обеспечивающего высокую оперативность кодирования и декодирования, поскольку практика применения для этого существующих реализаций показывает их недостаточную оперативность, неприемлемую ресурсоемкость, низкую применимость для потоковой обработки большого объема данных с плавающей запятой в реальном времени. Для достижения этой цели были решены следующие задачи. Описана математическая модель и алгоритм компрессии потока скалярных числовых значений с плавающей запятой, а также результаты проведенного экспериментального исследования созданного метода реализации компрессии структур одномерных и двумерных научных данных. Модель опирается на распространенный метод реализации расширенного множества действительных чисел, отображаемых на числа с плавающей запятой двойной точности, представление которых регламентировано форматом binary_64 стандарта IEEE-754. Алгоритм может быть использован в составе распределенных высокопроизводительных информационных систем обеспечения научных экспериментов для реализации задач, критичных к производительности

операций ввода-вывода и инфокоммуникационного обмена данными. Производительность и применимость алгоритма в реализациях потоковой обработки данных обусловлены его однопроходным поведением, относительно низкими требованиями к априорно известному и фиксированному размеру истории, на основе которой построена работа предиктора алгоритма компрессии. Действительно, полученные результаты показывают сравнимую с более ресурсоемкими универсальными кодерами результативность компрессии при значительном росте оперативности. При синхронизации параметров компрессии и декомпрессии векторных данных в предположении взаимной корреляции магнитуд скаляров одного измерения, возможно применение параллелизма класса ОКМД (одна команда-множество данных) для дальнейшего увеличения производительности предиктора (а следовательно, компрессии и декомпрессии), если реализующее устройство обладает возможностью обращения к изменяемой памяти, реализующей историю, на основе сдвиговых значений, формируемых параллельно или векторно, например, с помощью инструкций класса VGATHER микропроцессоров Intel. Реализация алгоритма внедрена авторами в параллельно-распределенную систему моделирования волновых полей с целью снижения издержек, обусловленных операциями ввода-вывода при взаимодействии системы узлов через вычислительную сеть. Экспериментально показана более высокая оперативность компрессии при сравнимой результативности универсальных многопроходных кодеров RAR, ZIP и 7Z.

Компрессионное кодирование; кодирование источника; арифметика с плавающей запятой; потоки данных.

A.A. Chusov, M.A. Kopaeva

ALGORITHM FOR COMPRESSION OF FLOATING-POINT DATA IN SCIENTIFIC RESEARCH SUPPORT SYSTEMS

The paper presents an original algorithm and its implementation for single pass real-time compression of streams of numeric floating-point data. The purpose of the research is to develop and formalize a single-pass algorithm of stream floating-point data compression in order to increase performance of both encoding and decoding, because a use of existing implementations provides insufficient speed of compression, are too restrictive on hardware resources and limited in applicability to real-time stream data compression when it comes to floating-point data. For that, the following issues have been addressed. The developed mathematical model and the algorithm for compression of scalar floating-point data are described together with results of experimental research of the compression method applied to single-dimensional and two-dimensional scientific data. The model is based upon the commonly-used binary₆₄ representation, of the IEEE-754 standard, onto which extended real-line values are mapped. The algorithm can be implemented as part of high-performance distributed systems in which performance of input-output operations, as well as internetwork communication, are critical to overall efficiency. The performance and applicability of the algorithm in data stream compression result from its single-pass behaviour, relatively low requirements to a priori known and statically defined size of memory required to implement history of compression, which the predictor, used in compression and decompression, is based on. Indeed, the measured compression ratios are comparable to ones which result from using more resource-intensive universal coders but providing significantly lower latency. Provided synchronization of parameters of both compressor and decompressor applied to a stream of vector values and assuming a correlation between absolute values of scalars of the same dimension within the vectors, further improvement of the predictor performance can be attained by means of SIMD-class parallelism which, in turn, is beneficial for overall performance of compression and decompression, provided that the underlying hardware is capable of addressing random-access memory based on offsets in a vector register, such as by employment of the VGATHER class instructions of Intel processors. In order to reduce the bottlenecks associated with input-output, an implementation of the algorithm is employed by the authors as part of a computing system used for parallel simulation of wave fields which is distributed via a network. The experiments described in the paper demonstrate significant performance increase of the proposed coder compared to well-known universal compressors, RAR, ZIP and 7Z, while the achieved compression factors remain comparable.

Data compression; source compression coding; floating-point arithmetic; data streams.

Введение. При обработке экспериментальных научных данных, представленных в формате чисел с плавающей запятой, их передача в несжатом виде может иметь неприемлемо низкую производительность, например, в системах реального времени. Однако универсальные алгоритмы компрессии направлены на увеличение результативности сжатия, а не оперативности, поэтому в случаях, когда предъявляются требования к оперативности или энергоэффективности инфокоммуникационных систем, алгоритмы необходимо специализировать. Для сжатия данных с плавающей запятой без потерь существуют методы сжатия на основе прогнозирования. Такие методы рассматривают число с плавающей запятой как целое число той же разрядности и предсказывают целочисленное значение, используя целочисленную операцию. Эти алгоритмы вычисляют разность или результат логической операции «исключающее ИЛИ» между предсказанным и исходным значением в каждом наборе данных для кодирования в более короткие битовые строки, чем исходные.

Первыми статьями, в которых были представлены основанные на прогнозировании алгоритмы сжатия научных данных с плавающей запятой, были [1–3]. В [1] разработали метод сжатия без потерь для данных временных рядов с плавающей запятой с использованием дельта-сжатия, и также объясняется проблема с представлением в памяти данных с плавающей запятой: даже если десятичное представление двух чисел может быть очень похожим, это не обязательно будет так для машинного представления в памяти [4, 5]. Подход, предложенный в [2], сжимает любой линейный поток данных, создавая хеш-ключ на основе различий нескольких последних выборок с поиском повторяющихся шаблонов в данных. В [6] предлагается использовать комбинированный интер-экстраполяционный предиктор: экстраполяция оценивает будущие данные, исходя из последовательности прошлых данных, а интерполяция конструирует соответствующие данные как часть последовательности данных. Существует схема, которая содержит два предиктора [7]. Два спрогнозированных значения сравниваются с действительным, и для выполнения последующей операции «исключающее ИЛИ» выбирается тот прогноз, который ближе к действительному значению. В [8] приводится многопроходный алгоритм сжатия без потерь $ndz\dot{p}$ для многомерных регулярных сеток данных с плавающей запятой одинарной и двойной точности. Алгоритм разработан для реализации на многоядерных процессорах с поддержкой ОКМД (одиночный поток команд, множественный поток данных) [9]. В области сжатия данных с плавающей запятой без потерь, существуют исследования аппаратных реализаций [10–15]. В [6] предложили аппаратную реализацию сжатия данных с плавающей запятой двойной точности путем группирования экспонент. В ряде работ [16–18] проведены исследования по синтезу алгоритмов компрессии.

Ниже представлен разработанный авторами настоящей работы алгоритм сжатия данных без потерь на основе прогнозирования с использованием хеш-таблицы.

Математическая модель кодера источника. Абстрактно работа кодеров с предикторами может быть сформулирована в виде функциональной диаграммы, представленной на рис. 1 [19].

В рамках настоящей работы предлагаются новые методы прогнозирования, структурирования и использования хеш-таблицы.

На рис. 2 отображены блок-схемы созданных алгоритмов компрессии и декомпрессии. На вход реализации компрессии поступает число с плавающей запятой двойной точности D . Далее функцией $hash$ производится вычисление хеш на основе имеющихся трех последних разностей, хранимых кодером.



Рис. 1. Функциональная диаграмма алгоритмов компрессии на основе предикторов

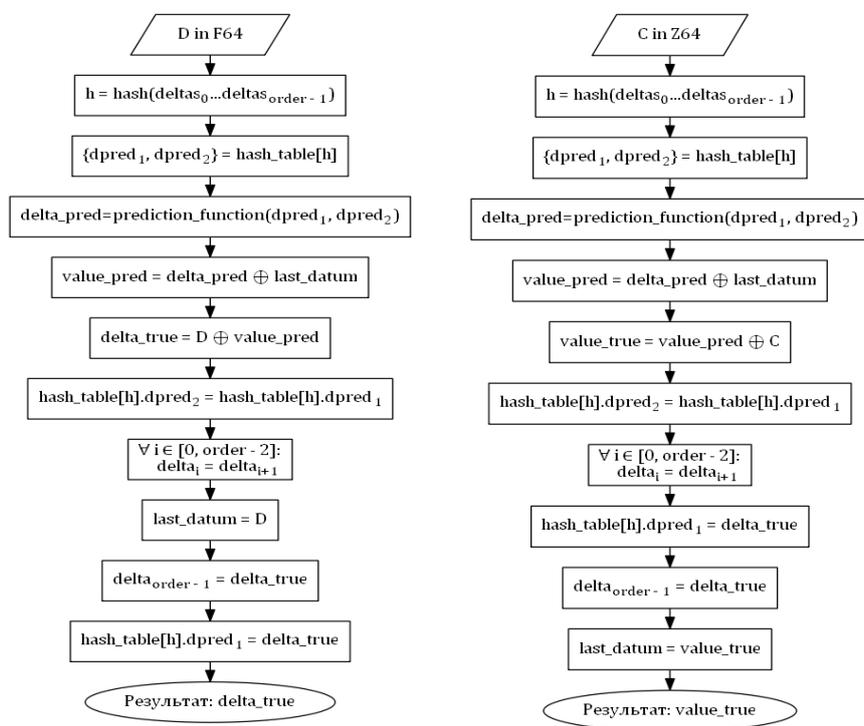


Рис. 2. Блок-схема алгоритмов предсказания при компрессии (слева) и декомпрессии (справа)

Блок-схема вычисления ключа h хеш-таблицы «hash_table» осуществляется на основе нескольких последних вычисленных полиномиальных (в кольце $\mathbb{Z}_2[x]$, то есть посредством операции «исключающее ИЛИ») разностей между предсказанными и вычисленными значениями. На рис. 2 эти разности обозначены переменными $delta_i$, где i – элемент индексного множества $[0, order - 1]$, а число этих разностей, в реализации принятое равным трем, определяет порядок предиктора. Таким образом предсказание осуществляется путем поиска истории процесса компрессии (или, соответственно, декомпрессии) на основе числа, равного порядку кодера, последних обработанных элементов данных.

Это достигается путем вычисления ключа хеш-таблицы, содержащей вероятные полиномиальные разности между истинными и предсказанными значениями. Тогда на этой основе и делается прогноз, который совпадает с последним ранее записанным в таблицу (по этому же ключу) значением. Реализуемость функции вычисления хеш-ключа обусловлена размером хеш-таблицы, который определяет разрядность ключа, а также тем, насколько вероятным должен быть успех поиска существующего элемента хеш-таблицы, так как в пределе, эта вероятность падает экспоненциально с увеличением порядка компрессора. В этом случае компрессор становится эквивалентным компрессору на основе последней, предыдущей, полиномиальной разности (в предположении, что хеш-таблица проинициализирована нулями). Однако вероятность успешного поиска может быть увеличена путем выбора в процессе вычисления ключа лишь наименее часто изменяемых старших бит числа с плавающей запятой.

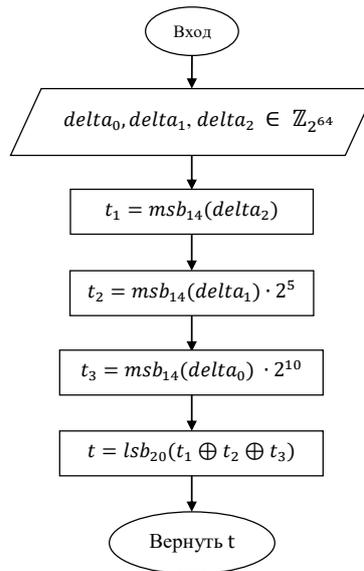


Рис. 3. Блок-схема вычисления hash

По этим причинам были получены удовлетворительные параметры функции $hash$, которая приведена на рис. 3. На вход поступают три целочисленные полиномиальные разности $delta_i$. Для вычислений используются старшие 14 бит каждой разности, возвращаемые функцией msb (most significant bits) как единый целочисленный скаляр. Самая последняя (недавняя) разность $delta_2$ не подвергается операции побитового сдвига. Разности $delta_1$ и $delta_0$ подвергаются операции побитового сдвига влево на 5 бит и 10 бит соответственно. Далее результаты по-

линомиально складываются, и 20 младших бит суммы, реализуемые на рис. 3 функцией *lsb* – least significant bits, возвращаются в качестве значения функции *hash* и ключа хеш-таблицы.

Использование двадцати битового ключа дает возможность реализовать таблицу в виде одномерного массива с 2^{20} элементами. Значением элемента является пара значений $dpred_1$ и $dpred_2$, которые являются последними шестидесяти четырёх битными полиномиальными разностями, изначально равными нулю либо полученными ранее для того же значения хеш-ключа.

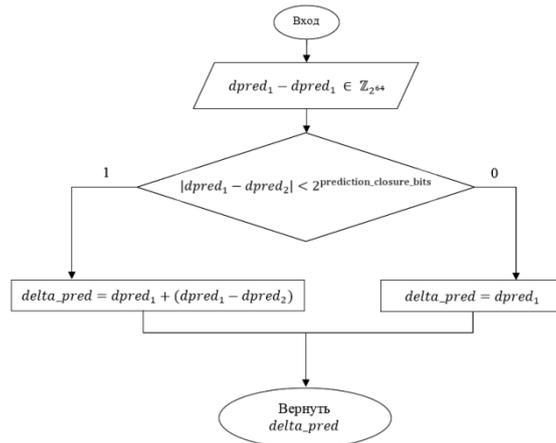


Рис. 4. Блок-схема функции *prediction_function*

В более простом варианте алгоритма в качестве предсказанного значения достаточно использование лишь одного элемента истории в таблицы. Однако лучшим себя показал модифицированный алгоритм (рис. 4) предсказания на основе двух элементов истории, если они достаточно похожи. В реализации алгоритм представлен функцией *prediction_function*, которая возвращает $dpred_1$, если два её аргумента – $dpred_1$ и $dpred_2$ – имеют менее заданного (управляющим параметром *prediction_closure_bits* кодера) количества одинаковых старших бит. Если же количество совпадающих старших бит больше либо равно этому значению, в качестве предсказанной разницы функция *prediction_function* возвращает значение $dpred_1 + (dpred_1 - dpred_2)$.

Выбор значения параметра *prediction_closure_bits* апостериорно обусловлен результативностью сжатия: для выбранных наборов данных экспериментально полученный оптимум составляет 14 бит.

На следующем этапе работы алгоритма компрессии производится обновление значений в хеш-таблице, соответствующих рассчитанному хеш. Также обновляются значения разностей $delta_i$.

Предшествующий элемент сжимаемых данных *last_datum* полиномиально суммируется с результатом операции *prediction_function*, и таким образом рассчитывается прогнозируемое значение *pred*. Результатом предсказания при декомпрессии является полиномиальная разница $delta_true$ между прогнозом *pred* и истинным значением *D*. Результатом компрессии числа с плавающей точкой двойной точности – четырехбитовое число ведущих бинарных нулей в этой разнице, деленное на четыре, и оставшиеся младшие биты разницы $delta_true$ в исходном виде.

Обратимость компрессии обусловлена когерентностью состояния декодера состоянию кодера при обработке каждого элемента данных и соответствующего ему кода переменной длины на стороне декодера. Эта когерентность в свою очередь наглядно наблюдается на рис. 1: действительно предсказанное значение $value_pred$ совпадает на кодере и декодере, поскольку оно рассчитано в результате выполнения идентичных операций, и в свою очередь это же значение детерминировано определяет состояние кодера и декодера при обработке последующих элементов данных.

Эксперименты. На рис. 5 и 6 отображены результаты проведенного экспериментального исследования. Для проведения исследования использованы файлы размером от 16 байт до 256 Мбайт, содержащие, с различными параметрами, одномерную линейную функцию, синусоиду, а также двумерную матрицу значений $\Re(Me^{2\pi i f d - \alpha d})$ (где M – магнитуа, f – пространственная частота, d расстояние до точки, которой соответствует выбранный элемент матрицы, α – коэффициент пространственного затухания, $i^2 = -1$).

На рис. 5 приводится сравнение работы созданного специального алгоритма компрессии с работой таких общепринятых реализаций универсального сжатия как 7Z и ZIP для сжатия синусоиды. ZIP использован с аргументом «-9», задающий максимальное сжатие. В качестве показателя оценки качества работы алгоритмов сжатия используется коэффициент сжатия [20], который устанавливает соотношение объема исходных к объему сжатых данных. Коэффициенты сжатия разработанного алгоритма компрессии варьируются от 1 до 1,6. Рис. 5 сравнивает результативности трех реализаций компрессии.

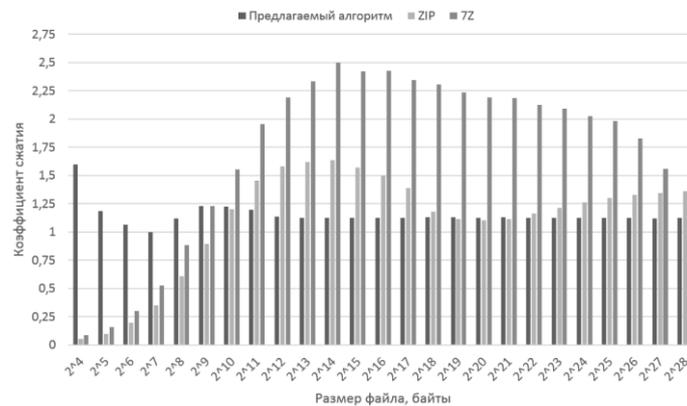


Рис. 5. Зависимость коэффициентов сжатия от размеров файлов для трех методов компрессии синусоидального сигнала

Однако в системах передачи данных, а также в системах с требованиями к оперативности информационного обмена в условиях низких пропускных способностей каналов связи или большого объема передаваемых данных необходимо обеспечение вычислительной эффективности компрессии и декомпрессии, которая определяется временными и связанными энергетическими затратами на обработку данных.

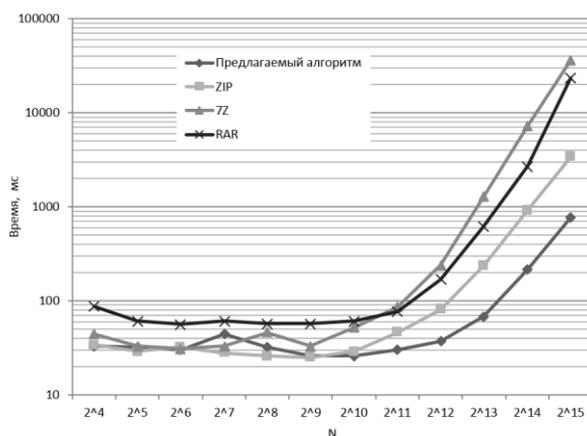


Рис. 6. Оперативность сжатия, измерение которой производилось на высокопроизводительном компьютере, разработанного алгоритма компрессии в сравнении с алгоритмами компрессии 7Z, ZIP, RAR в случае сжатия двумерных данных, из которых составлена квадратная матрица $N \times N$

Поэтому отдельно исследовались возможности по обеспечению представленным однопроходным алгоритмом требований к оперативности. На рис. 6 изображены полученные значения времени сжатия двумерных данных, описанных выше, в зависимости от размеров файлов для четырех методов компрессии. Измерение времени сжатия проводилось на высокопроизводительном компьютере с двумя 24-ядерными центральными процессорами Intel Xeon X12 E5-2695v2 и 128 Гбайт оперативной памяти под управлением операционной системой Microsoft Windows Server 2012 R2. При размерах файла от 32 байт до 128 Мбайт измеренная оперативность разработанного алгоритма варьируется от 26 мс до 763 мс и сравнительно с другими алгоритмами является наиболее высокой. 7Z при обработке больших файлов оказался наиболее медленным. Работа компрессоров ZIP и RAR заняла меньшее время по сравнению с 7Z, но больше, чем работа разработанного алгоритма.

Заключение. Результативность алгоритма, определяемая коэффициентом сжатия, превзошла результативность распространенной программы сжатия данных ZIP, однако уступила реализациям алгоритмов RAR и 7Z.

Вместе с тем наблюдается значительный прирост оперативности компрессии и декомпрессии созданного алгоритма, который является однопроходным, что дает возможность применять его для потоковой обработки данных в реальном времени, например, при передаче результатов измерений с удаленных датчиков, при реализации высокопроизводительного распределенного численного компьютерного моделирования физических процессов и явлений, математические модели которых опираются на вещественные и комплексные числовые множества, которые с достаточной достоверностью аппроксимируются числами с плавающей запятой.

Реализация функций предиктора в виде автомата с конечным состоянием и простота операций синтеза битовых кодов переменной длины позволила авторам работы инкапсулировать реализацию в функции обработки полиморфных классов, моделирующих потоки данных, и, во-первых, добиться указанных в настоящей работе результативности и оперативности компрессии, а также внедрить его в более составную систему имитационного моделирования волновых векторных полей.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Engelson V., Fritzon D., Fritzon P.* Lossless compression of high-volume numerical data from simulations // In Data Compression Conference. – 2000. – P. 574-586.
2. *Ratanaworabhan P., Ke J., Burtcher M.* Fast lossless compression of scientific floating-point data // In Proceedings of Data Compression Conference. – 2006. – P. 133-142.
3. *Lindstrom P., Isenburg M.* Fast and efficient compression of floating-point data // IEEE Transactions on Visual and Computer Graphics. – 2006. – Vol. 12, No. 5. – P. 1245-1250.
4. IEEE 754: Standard for binary floating-point arithmetic, 2008.
5. *Gomez L.A., Cappello F.* Improving floating point compression through binary masks // In Proc. 2013 IEEE Int. Conf. Big Data. – 2013. – P. 326-331.
6. *Imai S., Fukuma S., Mori S.* A Floating Point Data Compression Using Inter-Extrapolative Predictor // IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS). – 2018. – P. 546-549.
7. *Burtcher M., Ratanaworabhan P.* FPC: A High-Speed Compressor for Double-Precision Floating-Point Data // IEEE transactions on computers. – 2009. – Vol. 58, No. 1. – P. 18-31.
8. *Knorr F., Thoman P., Fahringer T.* ndzip: A High-Throughput Parallel Lossless Compressor for Scientific Data // 2021 Data Compression Conference (DCC). – 2021. – P. 103-112.
9. *Hildebrandt J., Habich D. and Lehner W.* BOUNCE: Memory-Efficient SIMD Approach for Lightweight Integer Compression // 2022 IEEE 38th International Conference on Data Engineering Workshops (ICDEW). – 2022. – P. 123-128. – DOI: 10.1109/ICDEW55742.2022.00025.
10. *Tomari H., Inaba M., Hiraki K.* Compressing floating-point number stream for numerical applications // 2010 First International Conference on Networking and Computing. – 2010. – P. 112-119.
11. *Katahira K., Sano K., Yamamoto S.* FPGA-based lossless compressors of floating-point data streams to enhance memory bandwidth // In Proceedings of the International Conference on Application-specific Systems, Architectures and Processors. – 2010. – P. 246-253.
12. *Mondigo A., Ueno T., Tanaka D., Sano K., Yamamoto S.* Design and scalability analysis of bandwidth-compressed stream computing with multiple fpgas // In Proceedings of 2017 12th International Symposium on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC). – 2017. – P. 108-115.
13. *Ueno T., Kono Y., Sano K., Yamamoto S.* Parameterized Design and Evaluation of Bandwidth Compressor for Floating-Point Data Streams in FPGA-Based Custom Computing. – Berlin, Heidelberg: Springer Berlin Heidelberg. 2013. – P. 90-102.
14. *Ueno T., Sano K., Furusawa T.* Performance Analysis of Hardware-Based Numerical Data Compression on Various Data Formats // 2018 Data Compression Conference. – 2018. – P. 345-354.
15. *Ueno T., Sano K., Yamamoto S.* Bandwidth Compression of Floating-Point Numerical Data Streams for FPGA-Based High-Performance Computing // ACM Transactions on Reconfigurable Technology and Systems. – 2017. – Vol. 10, No. 3. – P. 1-22.
16. *Yang A., Mukka H., Hesaaraki F., Burtcher M.* MPC: A Massively Parallel Compression Algorithm for Scientific Data // IEEE International Conference on Cluster Computing. – 2015.
17. *Claggett S., Azimi S., Burtcher M.* SPDP: An Automatically Synthesized Lossless Compression Algorithm for Floating-Point Data // Data Compression Conference. – 2018.
18. *Burtcher M., Hesaaraki F., Mukka H., Yang A.* Real-Time Synthesis of Compression Algorithms for Scientific Data // ACM/IEEE International Conference for High-Performance Computing, Networking, Storage and Analysis. – 2016. – P. 264-275.
19. *Konaeva M.A., Чусов А.А.* Алгоритм и системная реализация компрессии потоков чисел с плавающей точкой при реализации обработки экспериментальных научных данных // Радиоэлектроника. Проблемы и перспективы развития: Сб. трудов Седьмой всероссийской молодежной научной конференции. – Тамбов: Изд. центр ФГБОУ ВО «ТГТУ», 2022.
20. *Ainsworth M., Klasky S., Whitney B.* Compression Using Lossless Decimation: Analysis and Application // SIAM J. Sci. Comput. – 2017. – Vol. 39 (4). – P. B732-B757.

REFERENCES

1. *Engelson V., Fritzson D., Fritzson P.* Lossless compression of high-volume numerical data from simulations, *In Data Compression Conference*, 2000, pp. 574-586.
2. *Ratanaworabhan P., Ke J., Burtcher M.* Fast lossless compression of scientific floating-point data, *In Proceedings of Data Compression Conference*, 2006, pp. 133-142.
3. *Lindstrom P., Isenburg M.* Fast and efficient compression of floating-point data, *IEEE Transactions on Visual and Computer Graphics*, 2006, Vol. 12, No. 5, pp. 1245-1250.
4. IEEE 754: Standard for binary floating-point arithmetic, 2008.
5. *Gomez L.A., Cappello F.* Improving floating point compression through binary masks, *In Proc. 2013 IEEE Int. Conf. Big Data*, 2013, pp. 326-331.
6. *Imai S., Fukuma S., Mori S.* A Floating Point Data Compression Using Inter-Extrapolative Predictor, *IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2018, pp. 546-549.
7. *Burtcher M., Ratanaworabhan P.* FPC: A High-Speed Compressor for Double-Precision Floating-Point Data, *IEEE transactions on computers*, 2009, Vol. 58, No. 1, pp. 18-31.
8. *Knorr F., Thoman P., Fahringer T.* ndzip: A High-Throughput Parallel Lossless Compressor for Scientific Data, *2021 Data Compression Conference (DCC)*, 2021, pp. 103-112.
9. *Hildebrandt J., Habich D. and Lehner W.* BOUNCE: Memory-Efficient SIMD Approach for Lightweight Integer Compression, *2022 IEEE 38th International Conference on Data Engineering Workshops (ICDEW)*, 2022, pp. 123-128. DOI: 10.1109/ICDEW55742.2022.00025.
10. *Tomari H., Inaba M., Hiraki K.* Compressing floating-point number stream for numerical applications, *2010 First International Conference on Networking and Computing*, 2010, pp. 112-119.
11. *Katahira K., Sano K., Yamamoto S.* FPGA-based lossless compressors of floating-point data streams to enhance memory bandwidth, *In Proceedings of the International Conference on Application-specific Systems, Architectures and Processors*, 2010, pp. 246-253.
12. *Mondigo A., Ueno T., Tanaka D., Sano K., Yamamoto S.* Design and scalability analysis of bandwidth-compressed stream computing with multiple fpgas, *In Proceedings of 2017 12th International Symposium on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC)*, 2017, pp. 108-115.
13. *Ueno T., Kono Y., Sano K., Yamamoto S.* Parameterized Design and Evaluation of Bandwidth Compressor for Floating-Point Data Streams in FPGA-Based Custom Computing. Berlin, Heidelberg: Springer Berlin Heidelberg. 2013, pp. 90-102.
14. *Ueno T., Sano K., Furusawa T.* Performance Analysis of Hardware-Based Numerical Data Compression on Various Data Formats, *2018 Data Compression Conference*, 2018, pp. 345-354.
15. *Ueno T., Sano K., Yamamoto S.* Bandwidth Compression of Floating-Point Numerical Data Streams for FPGA-Based High-Performance Computing, *ACM Transactions on Reconfigurable Technology and Systems*, 2017, Vol. 10, No. 3, pp. 1-22.
16. *Yang A., Mukka H., Hesaaraki F., Burtcher M.* MPC: A Massively Parallel Compression Algorithm for Scientific Data, *IEEE International Conference on Cluster Computing*, 2015.
17. *Claggett S., Azimi S., Burtcher M.* SPDP: An Automatically Synthesized Lossless Compression Algorithm for Floating-Point Data, *Data Compression Conference*, 2018.
18. *Burtcher M., Hesaaraki F., Mukka H., Yang A.* Real-Time Synthesis of Compression Algorithms for Scientific Data, *ACM/IEEE International Conference for High-Performance Computing, Networking, Storage and Analysis*, 2016, pp. 264-275.
19. *Kopaeva M.A., Chusov A.A.* Algoritm i sistemnaya realizatsiya kompressii potokov chisel s plavayushchey tochkoy pri realizatsii obrabotki eksperimental'nykh nauchnykh dannykh [An algorithm and systematic approach to compression of floating-point data streams for processing of scientific data], *Radioelektronika. Problemy i perspektivy razvitiya: Sb. trudov Sed'moy vserossiyskoy molodezhnoy nauchnoy konferentsii* [Radioelectronics. Problems and future developments: Seventh Russian Youth Scientific Conference]. Tambov: Izd. tsentr FGBOU VO «TGTU», 2022.
20. *Ainsworth M., Klasky S., Whitney B.* Compression Using Lossless Decimation: Analysis and Application, *SIAM J. Sci. Comput.*, 2017, Vol. 39 (4), pp. B732-B757.

Статью рекомендовала к опубликованию д.т.н., г.н.с. Н.В. Злобина.

Чусов Андрей Александрович – ФГАОУ ВО Дальневосточный федеральный университет; e-mail: chusov.aa@dvfu.ru; г. Владивосток, Россия; тел.: +79147315896; доцент департамента электроники, телекоммуникации и приборостроения; к.т.н.

Кобаева Маргарита Александровна – e-mail: rikopae@yandex.ru; тел.: +79502982261; магистрант департамента электроники, телекоммуникации и приборостроения.

Chusov Andrei Aleksandrovich – Far-Eastern Federal University, e-mail: chusov.aa@dvfu.ru; Vladivostok, Russia; phone: +79147315896; associate professor of the department of electronics, telecommunication and hardware development; cand. of eng. sc.

Кобаева Margarita Aleksandrovna – e-mail: rikopae@yandex.ru; phone: +79502982261; graduate student at the department of electronics, telecommunication and hardware development.

УДК 004.89

DOI 10.18522/2311-3103-2022-4-192-200

В.А. Частикова, С.А. Жерлицын**НЕЙРОСЕТЕВАЯ МЕТОДИКА ИДЕНТИФИКАЦИИ ЛИЧНОСТИ
ПО РИСУНКУ ВЕН ЛАДОНИ**

Описывается работа по созданию нейросетевой методики идентификации личности, основанной на механизме сканирования и анализа рисунка вен ладони, как биометрического параметра. В рамках проведенного исследования описаны предпосылки, цели и причины, по которым разработка надежной системы биометрической идентификации является важным и актуальным направлением деятельности. Сформулирован ряд проблем, присутствующих существующим методам решения поставленной задачи: графовому методу и методу, основанному на вычислении расстояния, выраженного в различных интервальных метриках. Приведено описание принципов их работы. Сформулированы задачи, решаемые системами идентификации личности: сопоставление субъекта идентификации с его идентификатором, однозначно идентифицирующим этого субъекта в информационной системе. Описан механизм считывания рисунка вен с ладони, разработанный для анализа изображения, полученного с восприимчивой к излучению инфракрасного диапазона цифровой камеры. При нахождении в кадре ладони, подсвечиваемой светом ближнего ИК-диапазона, на изображении, полученном с камеры, становится заметен рисунок пролегающих под кожным покровом вен, сосудов и капилляров. В зависимости от организации, система идентификации может на основе предоставленного идентификатора определять соответствующий субъект доступа или проверять принадлежность того же идентификатора предполагаемому субъекту. Приведены 3 метода дальнейшего анализа биометрических данных и идентификации личности: подходы, основанные на категориальной классификации и бинарной классификации, а также комбинированный подход, при котором сначала используется идентификация по первому способу, а затем, по второму, но уже для известного идентификатора доступа, определённого на первом этапе. Приведена результирующая архитектура нейросети для категориальной классификации рисунка вен, описан способ вычисления количества параметров модели в зависимости от числа зарегистрированных субъектов. Представлены основные выводы и экспериментальные замеры точности работы системы при реализации различных методов, а также диаграммы изменения точности моделей во время обучения. Выявлены основные преимущества и недостатки приведённых методов.

Биометрическая идентификация личности; рисунок вен ладони; сверточная нейронная сеть; бинарная классификация; категориальная классификация; информационная безопасность.

V.A. Chastikova, S.A. Zherlitsyn

DEVELOPMENT OF A METHOD FOR PERSONAL IDENTIFICATION BASED ON THE PATTERN OF PALM VEINS

The article describes the work on the creation of a neural network method for identifying a person based on the mechanism of scanning and analyzing the pattern of palm veins as a biometric parameter. As part of the study, the prerequisites, goals and reasons for which the development of a reliable biometric identification system is an important and relevant area of activity are described. A number of problems are formulated that are inherent in existing methods for solving the problem: the graph method and the method based on calculating the distance expressed in various interval metrics. The description of the principles of their work is given. The tasks solved by personal identification systems are formulated: comparison of the subject of identification with its identifier, which uniquely identifies this subject in the information system. A mechanism for reading a pattern of veins from the palm of the hand, developed for analyzing an image obtained with a digital camera sensitive to infrared radiation, is described. When the palm is in the frame, illuminated by the light of the near infrared range, the image obtained from the camera becomes noticeable pattern of veins, vessels and capillaries that lie under the skin. Depending on the organization, the identification system may, based on the provided identifier, determine the appropriate access subject or verify that the same identifier belongs to the intended subject. Three methods for further analysis of biometric data and personal identification are given: approaches based on categorical classification and binary classification, as well as a combined approach, in which identification is first used by the first method, and then, by the second, but already for a known access identifier defined on the first stage. The resulting architecture of the neural network for the categorical classification of the vein pattern is presented, a method for calculating the number of model parameters depending on the number of registered subjects is described. The main conclusions and experimental measurements of the accuracy of the system when implementing various methods are presented, as well as diagrams of changes in the accuracy of models during training. The main advantages and disadvantages of the above methods are revealed.

Biometric personal identification; palm vein pattern; convolutional neural network; binary classification; categorical classification; information security.

Введение. У основных популярных факторов идентификации, таких как, например, ключи и парольная информация, несмотря на простоту и удобство их использования существуют недостатки, в числе которых возможность разглашения, подделки, подбора и утери. Использование биометрической информации минимизирует риск реализации ранее упомянутых угроз [1].

Механизм идентификации достаточно часто является критически важным компонентом системы, на его работу возлагается высокая ответственность. В числе последних тенденций развития механизмов аутентификации нельзя не выделить повышение спроса на методы, использующие биометрические характеристики субъекта [2, 3].

Целью данной работы является разработка методики идентификации личности по рисунку вен ладони на основе нейросетевого аппарата.

Актуальность. В условиях текущего уровня развития современных технологий информационные системы стали критически важными компонентами практически любого вида деятельности. Неограниченный доступ посторонних лиц к некоторым из них способен привести не только к колоссальным экономическим затратам, но в некоторых случаях и к катастрофам техногенного характера. Основным методом борьбы с такого вида угрозами является внедрение систем идентификации, а также контроля доступа [4].

Постановка задачи. На текущий момент набор применяемых для идентификации по рисунку вен ладони технологий базируется в наибольшей степени на двух основных методах: графовом и методе на основе вычисления расстояния (какой-либо интервальной метрики).

Графовый метод. Основан на сопоставлении топологии вен. Существует множество различных реализаций, однако обобщенный алгоритм состоит из следующих этапов:

- ◆ получение изображения ладони с венами;
- ◆ предварительная обработка по уменьшению шумов и выделению вен;
- ◆ построение линий, соответствующих выделенным венам;
- ◆ преобразование набора линий во взвешенный или обыкновенный граф;
- ◆ сравнение длин и весов соответствующих рёбер, углов между рёбрами в вершинах, топологического сходства.

Данный алгоритм при каждом выполнении процедуры идентификации вынужден обращаться к набору идентификационных образцов, что не позволяет изолировать систему идентификации от хранилища биометрических персональных данных.

Также минусом подобного подхода является необходимость сравнивать предоставленный образец с каждым из зарегистрированных в системе, что является крайне ресурсоёмким процессом при растущем количестве субъектов идентификации.

Среди прочих недостатков метода важно отметить следующие: крайне высокая зависимость от качества освещения, угла и изгиба ладони, зашумлённости изображения, фона, общая неустойчивость работы [5].

Метод на основе вычисления расстояния. Основан на вычислении расстояния между анализируемым и зарегистрированным изображениями по различным интервальным метрикам, таким как, косинусное расстояние, евклидово расстояние, и другие. Также в совокупности с рассматриваемым алгоритмом зачастую применяется масштабирование и поворот изображения. Такие меры повышают устойчивость работы алгоритма, однако кратно увеличивают его сложность и ресурсоёмкость. Таким образом, метод сохраняет основные минусы предыдущего – высокая вычислительная сложность при эксплуатации и невозможность разделения системы идентификации и хранилища биометрических персональных данных.

Реализация процесса идентификации на основе классификации при помощи нейронных сетей призвана компенсировать перечисленные выше недостатки существующих подходов.

Получение изображения. Задачей систем идентификации личности является сопоставление субъекта идентификации с его идентификатором, однозначно идентифицирующим этого субъекта в информационной системе. В зависимости от организации, система идентификации может на основе предоставленного идентификатора определять соответствующий субъект доступа или проверять принадлежность того же идентификатора предполагаемому субъекту [6].

Система идентификации личности по рисунку вен ладони анализирует изображение, полученное с восприимчивой к излучению инфракрасного диапазона цифровой камеры. При нахождении в кадре человеческой руки, а именно ладони, подсвечиваемой светом ближнего ИК-диапазона, на изображении, полученном с камеры, становится заметен рисунок пролегающих под кожным покровом вен, сосудов и капилляров.

Полученные таким образом изображения, тем не менее, сильно зависят от качества подсветки и не всегда обладают должной степенью контрастности, однако этот недостаток является устранимым при помощи автоматической постобработки. Пример получаемого изображения приведён на рис. 1.

Каждое изображение хранится в чёрно-белом формате, так как цветовые составляющие не несут полезной информации о рассматриваемых венах.

В данном исследовании предложено несколько вариантов реализации системы идентификации по рисунку вен ладони на основе нейросетевого анализа данных. В качестве математической основы системы выбрана свёрточная нейронная сеть с несколькими полносвязными слоями на выходе, так как именно такая комбинация моделей является наиболее распространённым способом обработки и классификации изображений.



Рис. 1. Пример необработанного снимка вен ладони

Категориальная классификация. Первый из них основан на категориальной классификации. Для прохождения идентификации производится проверка принадлежности предоставленного образца к конкретному классу из выборки [7].

Нейросеть при этом обучается сопоставлению каждого образца обучающей выборки соответствующему строго определённом заранее классу. Выходом нейросети в таком случае при предоставлении ей одного отдельно взятого образца будет являться вектор распределения вероятности принадлежности данного образца между всеми известными нейросети классами. Для минимизации количества ошибок второго рода также рационально ввести пороговое значение вероятности для идентификации личности [8]. Данное значение определяется эмпирически в ходе проведения экспериментов с заданным конкретным набором субъектов идентификации.

В рамках текущего исследования описанный подход был реализован на практике: собрана обучающая выборка, построена и обучена нейросетевая модель.

Для подхода, основанного на категориальной классификации, лучшее значение точности из серии экспериментов составило 0.9149. Графики изменения функций точности и потерь во время обучения для тестовой и валидационной подвыборок приведены на рис. 2.

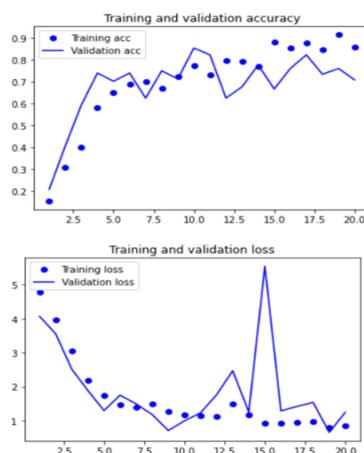


Рис. 2. Графики изменения функций точности и потерь во время обучения категориальной классификации

Результующая нейросетевая архитектура для решения задачи категориальной классификации представлена в табл. 1. Значение n равно количеству субъектов, зарегистрированных в системе, и может варьироваться от 2 до 100 в протестированных конфигурациях, однако может быть изменено в большую сторону.

Таблица 1

Нейросетевая архитектура

Название слоя (тип слоя)	Размерность выходного вектора	Число параметров
conv2d_1 (входной слой)	(1274, 714, 64)	3200
max_pooling2d_1 (субдискретизирующий слой)	(637, 357, 64)	0
conv2d_2 (сверточный слой)	(635, 355, 128)	73856
max_pooling2d_2 (субдискретизирующий слой)	(317, 177, 128)	0
conv2d_3 (сверточный слой)	(315, 175, 128)	147584
max_pooling2d_3 (субдискретизирующий слой)	(157, 87, 128)	0
conv2d_4 (сверточный слой)	(155, 85, 128)	147584
max_pooling2d_4 (субдискретизирующий слой)	(77, 42, 128)	0
flatten_1 (линеаризующий слой)	(413952)	0
Dropout (прореживающий слой)	(413952)	0
Dense_1 (полносвязный слой)	64	26492992
Dense_2 (полносвязный слой)	n	$65 * n$

Бинарная классификация. Вторым способом основан на принципе бинарной классификации. В данном случае задачей нейронной сети является анализ принадлежности предъявляемого идентификатора только лишь одному строго определённому субъекту. Выходом нейронной сети в данной ситуации являются два числа – вероятность соответствия и вероятность несоответствия. Для рассматриваемого способа также необходимо применение порогового значения для подтверждения соответствия [9].

В случае категориальной классификации в системе находится лишь одна единственная нейросеть, которая отвечает за проверку соответствия каждого субъекта доступа своей идентификационной записи. Для добавления нового пользователя в систему будет необходимо переобучение последних слоёв нейросети.

В случае бинарной классификации каждому субъекту доступа соответствует своя легковесная нейросеть, которая быстро обучается и проверяет корректность соответствия пользователя с идентификационной записью в системе. При этом пользователь либо должен указать на запрашиваемую идентификационную запись явным образом, либо использовать дополнительный фактор идентификации, либо система должна линейно провести проверку соответствия полученного изображе-

ния с каждой идентификационной записью. Все перечисленные варианты имеют свои плюсы и минусы и выбор итогового алгоритма зависит от способа применения механизма.

Описанный подход также был реализован на практике с использованием той же обучающей выборки и вычислительной базы.

Для подхода, основанного на бинарной классификации, точность достигала значения 0.9789. Графики изменения функций точности и потерь во время обучения для тестовой и валидационной подвыборок приведены на рис. 3.

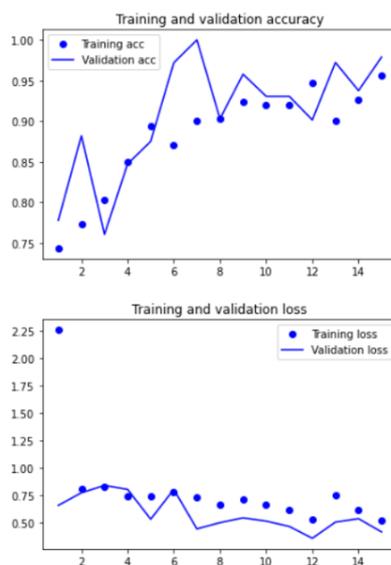


Рис. 3. Графики изменения функций точности и потерь во время обучения бинарной классификации

Комбинированный метод. Помимо вышеописанных возможен третий сценарий применения предложенных способов – комбинированный, двухэтапный. В целях повышения надёжности системы и сведения количества ошибок второго рода к возможному минимуму, используется идентификация по первому способу, а затем, по второму, но уже для известного идентификатора доступа, определённого на первом этапе [13].

Заключение. Представленная в первом способе, основанном на категориальной классификации, архитектура является наиболее ресурсоёмкой из всех используемых в каждой из реализаций методики, что позволяет дать верхнюю оценку вычислительной сложности системы в целом.

Таким образом, разработанная нейронная сеть содержит 12 слоев и от 26 865 346 до 26 871 716 параметров. Интервал количества параметров обусловлен допустимым количеством зарегистрированных субъектов – от 2-х до 100. Так как каждому субъекту соответствует свой выход нейросети, каждый дополнительный субъект добавит в модель 65 новых весов [17].

Применение многокатегориальной классификации с помощью нейросетей для распределения идентификаторов по зарегистрированным в системе пользователям превышает по вычислительной сложности одну проверку соответствия любому из классических методов, но является значительно менее ресурсоёмким [18], чем проверка соответствия с каждым из зарегистрированных субъектов, которая проводится в известных методиках.

Точность определения субъекта идентификации с применением бинарной классификации показывает лучшие результаты, по сравнению с многокатегориальной, однако в изначально описанной конфигурации проигрывает по требовательности к техническим характеристикам устройства-носителя. Предложенный комбинированный подход понижает требовательность к вычислительной мощности устройства и повышает конечную надежность системы.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Частикова В.А., Жерлицын С.А., Воля Я.И.* Нейросетевой метод идентификации личности по неформализованной семантической характеристике // Вестник УрФО. Безопасность в информационной сфере. – 2018. – № 4 (30). – С. 20-26. – DOI 10.14529/secu180403. – EDN YUNKCL.
2. *Частикова В.А., Жерлицын С.А., Воля Я.И., Сотников В.В.* Нейросетевая технология обнаружения аномального сетевого трафика // Прикаспийский журнал: управление и высокие технологии. – № 1 (49). – С. 20-32. – DOI 10.21672/2074-1707.2020.49.4.020-032. – EDN WUCDII.
3. *Jain A.K., Ross A. and Pankanti S.* Biometrics: a tool for information security // Trans Inform Forensics Secur. – No. 1 (2). – P. 125-143.
4. *Malatras A., Geneiatakis D. and Vakalis I.* On the efficiency of user identification: a system-based approach // Int. J. Inf. Secur. – No. 16. – P. 653-671.
5. *Im S., Park H., Kim Y., Han S., Kim S., Kang C., Chung C.* A Biometric identification system by extracting hand vein patterns // J Korean Phys Soc. – Vol. 28 (3). – P. 268-272.
6. *Sarkar A. and Singh B.K.* A review on performance, security and various biometric template protection schemes for biometric authentication systems // Multimed Tools Appl. – Vol. 79. – P. 27721-27776.
7. *Hancock J T and Khoshgoftaar T M* Survey on categorical data for neural networks // Journal of Big Data. – Vol. 7, No. 28.
8. *Chastikova V.A., Zherlitsyn S.A. and Y.I. Volya* Neural network method of identification by unformalized semantic characteristics // News of Volgograd State Technical University no. – Vol. 8 (218). – P. 63-67.
9. *Blokus A. and Krawczyk H.* Systematic approach to binary classification of images in video streams using shifting time windows // SIViP. – No. 13. – P. 341-348.
10. *Hassanat A.B., Albustanji A., Tarawneh A.S., Alrashidi M., Alharbi H., Alanazi M., Alghamdi M., Alkhazi I.S., & Prasath V.* Deep learning for identification and face, gender, expression recognition under constraints // ArXiv, abs/2111.01930, 2021.
11. *Soleymani S., Dabouei A., Taherkhani F., Iranmanesh S.M., Dawson J.M., & Nasrabadi N.M.,* Quality-Aware Multimodal Biometric Recognition // ArXiv, abs/2112.05827, 2021.
12. *Mugalu B.W., Wamala R.C., Serugunda J., & Katumba A.* Face Recognition as a Method of Authentication in a Web-Based System // ArXiv, abs/2103.15144, 2021.
13. *Marattukalam F., Abdulla W.H., & Swain A.K.* 2021 N-shot Palm Vein Verification Using Siamese Networks // 2021 International Conference of the Biometrics Special Interest Group (BIOSIG). – P. 1-5.
14. *Meng Z., Altaf M.U., Juang B.* Active voice authentication // ArXiv, abs/2004.12071, 2020.
15. *Stragapede G., Vera-Rodríguez R., Tolosana R., Morales A., Acien A., & Lan G.L.* Mobile Behavioral Biometrics for Passive Authentication // ArXiv, abs/2203.07300, 2022.
16. *Fuksis R., Pudzs M., Greitans M.* Palm Vein Biometrics Based on Palm Infrared Imaging and Complex Matched Filtering // The 12th ACM Workshop on Multimedia and Security, Rome, 2009. – P. 27.
17. *Chastikova V. A., Zherlitsyn S. A., Volya Y.I.* Analysis of training of deep neural networks with heterogeneous architecture while detecting malicious network traffic // IOP Conference Series: Materials Science and Engineering, Krasnoyarsk, Russian Federation: IOP Publishing Ltd, 2021. – P. 12135. – DOI 10.1088/1757-899X/1047/1/012135.
18. *Частикова В.А., Тутова А.А., Войлова Д.О.* Аналитический обзор методов идентификации личности на основе биометрических характеристик // Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. – 2022. – № 1 (296). – P. 92-112.

19. Промыслов В.Г., Семенов К.В., Менгазетдинов Н.Э. Исследование методов аутентификации операторов в промышленных системах управления // Проблемы управления. – 2022. – № 3. – С. 40-54.
20. Артамонов В.А., Артамонова Е.В. Искусственный интеллект и безопасность: проблемы, заблуждения, реальность и будущее // Россия: тенденции и перспективы развития. – 2022. – № 17-1. – С. 585-594.

REFERENCES

1. Chastikova V.A., Zherlitsyn S.A., Volya Ya.I. Neyrosetevoy metod identifikatsii lichnosti po neformalizovannoy semanticheskoy kharakteristike [Neural network method of identification by unformalized semantic characteristics], *Vestnik UrFO. Bezopasnost' v informatsionnoy sfere* [Urfu Journal. Security in the information sphere no], 2018, No. 4 (30), pp. 20-26. DOI 10.14529/secur180403. EDN YuNKCL.
2. Chastikova V.A., Zherlitsyn S.A., Volya Ya.I., Sotnikov V.V. Neyrosetevaya tekhnologiya obnaruzheniya anomal'nogo setevogo trafika [Neural network technology for detecting anomalous network traffic], *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian journal: Control and High Technologies], No. 1 (49), pp. 20-32. DOI 10.21672/2074-1707.2020.49.4.020-032. EDN WUCDII.
3. Jain A.K., Ross A. and Pankanti S. Biometrics: a tool for information security, *Trans Inform Forensics Secur*, No. 1 (2), pp. 125-143.
4. Malatras A., Geneiatakis D. and Vakalis I. On the efficiency of user identification: a system-based approach, *Int. J. Inf. Secur.*, No. 16, pp. 653-671.
5. Im S., Park H., Kim Y, Han S., Kim S., Kang C., Chung C. A Biometric identification system by extracting hand vein patterns, *J Korean Phys Soc.*, Vol. 28 (3), pp. 268-272.
6. Sarkar A. and Singh B.K. A review on performance, security and various biometric template protection schemes for biometric authentication systems, *Multimed Tools Appl*, Vol. 79, pp. 27721-27776.
7. Hancock J T and Khoshgoftaar T M Survey on categorical data for neural networks, *Journal of Big Data*, Vol. 7, No. 28.
8. Chastikova V.A., Zherlitsyn S.A. and Y.I. Volya Neural network method of identification by unformalized semantic characteristics, *News of Volgograd State Technical University no*, Vol. 8 (218), pp. 63-67.
9. Blokus A. and Krawczyk H. Systematic approach to binary classification of images in video streams using shifting time windows, *SIViP*, No. 13, pp. 341-348.
10. Hassanat A.B., Albustanji A., Tarawneh A.S., Alrashidi M., Alharbi H., Alanazi M., Alghamdi M., Alkhazi I.S., & Prasath V. Deep learning for identification and face, gender, expression recognition under constraints, *ArXiv, abs/2111.01930, 2021*.
11. Soleymani S., Dabouei A., Taherkhani F., Iranmanesh S.M., Dawson J.M., & Nasrabadi N.M., Quality-Aware Multimodal Biometric Recognition, *ArXiv, abs/2112.05827, 2021*.
12. Mugalu B.W., Wamala R.C., Serugunda J., & Katumba A. Face Recognition as a Method of Authentication in a Web-Based System, *ArXiv, abs/2103.15144, 2021*.
13. Marattukalam F., Abdulla W.H., & Swain A.K. 2021 N-shot Palm Vein Verification Using Siamese Networks, *2021 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1-5.
14. Meng Z., Altaf M.U., Juang B. Active voice authentication, *ArXiv, abs/2004.12071, 2020*.
15. Stragapede G., Vera-Rodríguez R., Tolosana R., Morales A., Acien A., & Lan G.L. Mobile Behavioral Biometrics for Passive Authentication, *ArXiv, abs/2203.07300, 2022*.
16. Fuksis R., Pudzs M., Greitans M. Palm Vein Biometrics Based on Palm Infrared Imaging and Complex Matched Filtering, *The 12th ACM Workshop on Multimedia and Security, Rome, 2009*, pp. 27.
17. Chastikova V. A., Zherlitsyn S. A., Volya Y.I. Analysis of training of deep neural networks with heterogeneous architecture while detecting malicious network traffic, *IOP Conference Series: Materials Science and Engineering, Krasnoyarsk, Russian Federation: IOP Publishing Ltd, 2021*, pp. 12135. DOI 10.1088/1757-899X/1047/1/012135.
18. Chastikova V.A., Titova A.A., Voylova D.O. Analiticheskiy obzor metodov identifikatsii lichnosti na osnove biometricheskikh kharakteristik [Analytical review of personal identification methods based on biometric characteristics], *Vestnik Adygeyskogo gosudarstvennogo universiteta. Seriya 4: Estestvenno-matematicheskie i tekhnicheskie nauki* [Bulletin of the Adyghe State University. Series 4: Natural-mathematical and technical sciences], 2022, No. 1 (296), pp. 92-112.

19. Promyslov V.G., Semenov K.V., Mengazetdinov N.E. Issledovanie metodov autentifikatsii operatorov v promyshlennykh sistemakh upravleniya [Research of methods of authentication of operators in industrial control systems], *Problemy upravleniya* [Management Problems], 2022, No. 3, pp. 40-54.
20. Artamonov V.A., Artamonova E.V. Iskusstvennyy intellekt i bezopasnost': problemy, zabluzhdeniya, real'nost' i budushchee [Artificial intelligence and security: problems, delusions, reality and future], *Rossiya: tendentsii i perspektivy razvitiya* [Russia: trends and development prospects], 2022, No. 17-1, pp. 585-594.

Статью рекомендовал к опубликованию д.т.н. Л.А. Видовский.

Частикова Вера Аркадьевна – Кубанский государственный технологический университет; e-mail: chastikova_va@mail.ru; г. Краснодар, Россия; тел.: +79184635536; кафедра компьютерных технологий и информационной безопасности; к.т.н.; доцент.

Жерлицын Сергей Анатольевич – e-mail: kpytooooo@gmail.com; тел.: +79181965775; кафедра компьютерных технологий и информационной безопасности; аспирант.

Chastikova Vera Arkadyevna – Kuban State Technological University; e-mail: chastikova_va@mail.ru; Krasnodar, Russia; phone: +79184635536; the department of computer technologies and information security; cand. of eng. sc.; associate professor.

Zherlitsyn Sergey Anatolyevich – e-mail: kpytooooo@gmail.com; phone: +79181965775; the department of computer technologies and information security; graduate student.

УДК 004.021

DOI 10.18522/2311-3103-2022-4-200-212

К.Н. Алексеев, Д.А. Сорокин, А.Л. Леонтьев

МЕТОДИКА СОЗДАНИЯ ТОПОЛОГИЧЕСКИХ ОГРАНИЧЕНИЙ ПРИ ВЫСОКОЙ УТИЛИЗАЦИИ РЕСУРСОВ ПЛИС

Рассмотрена проблема достижения высокой реальной производительности реконфигурируемых вычислительных систем при решении вычислительно трудоёмких задач различных предметных областей. Величину реальной производительности реконфигурируемых систем определяют параметры выполняемых на них программ, основной компонентой которых являются вычислительные структуры обработки данных, реализованные в виде конфигурационных файлов ПЛИС. При этом одним из ключевых параметров любой вычислительной структуры является тактовая частота ее работы, которая непосредственно влияет на её производительность. Однако достижение высоких тактовых частот сопряжено с рядом проблем, которые современные средства САПР не решают. Причина кроется в неоптимальном топологическом размещении функциональных узлов вычислительной структуры на поле примитивов ПЛИС, особенно при высокой утилизации ресурсов. Это приводит к повышенной нагрузке на коммутационную матрицу ПЛИС и, как следствие, связи между примитивами ПЛИС, имеющими функциональную зависимость, оказываются значительно длиннее, чем это допустимо. Кроме того, излишняя длина связей наблюдается при трассировке соединений между примитивами, которые расположены на разных кремниевых кристаллах ПЛИС или же физически разделены встроенными периферийными устройствами. В настоящей статье описывается методика, которая позволяет рационализировать размещение элементов вычислительной структуры на поле примитивов ПЛИС, минимизировать длину трасс между примитивами, а также минимизировать число трасс между физически разделёнными топологическими областями ПЛИС. Работоспособность предложенной методики показана на примере решения тестовой задачи «КИХ-фильтр» на реконфигурируемом компьютере «Терциус». Проиллюстрированы основные проблемы при достижении целевой тактовой частоты и описан способ их преодоления. Применение методики позволило увеличить тактовую частоту и тем самым поднять производительность «Терциус» на 25% без переработки функциональной схемы вычислительной структуры задачи. Текущие исследования

эффективности предложенной методики позволяют утверждать, что автоматизированные средства создания топологических ограничений на её основе позволят существенно сократить время разработки программ с требуемыми характеристиками для реконфигурируемых вычислительных систем.

Реконфигурируемые вычислительные системы; ПЛИС; САПР; Physical Constraints; Placement Constraints; Timing Closure.

K.N. Alekseyev, D.A. Sorokin, A.L. Leontyev

METHODOLOGY OF TOPOLOGICAL RESTRICTIONS FOR INTENSIVELY USED FPGA RESOURCE

In the paper we consider the problem of achieving high real performance of reconfigurable computer systems in implementing computationally expensive tasks from various problem areas. The parameters of the programs executed on reconfigurable systems determine their real performance. The main component of these programs is the computing data processing structures implemented as FPGA configuration files. At the same time, one of the key parameters of any computing structure is the clock frequency of its operation, which directly affects its performance. However, there are several problems concerning the achievement of high clock rates, and they cannot be solved with the help of modern CAD tools. The reason is the non-optimal topological placement of functional blocks of the computing structure within the field of FPGA primitives, especially with high resource utilization. Due to this, the load on the FPGA switching matrix is increasing, and, as a result, the connections among functionally dependent FPGA primitives turn out to be much longer than is acceptable. In addition, excessive connection length is observed when tracing connections among primitives that are placed on different FPGA chips or are physically separated by on-chip peripherals. In the paper we describe a methodology which provides optimization of the placement of computing structure elements on FPGA primitives, and minimizes the length of traces among primitives, and also minimizes the number of traces among physically separated FPGA topological sections. To prove the proposed methodology, we implemented the test task "FIR-filter" on a reconfigurable computer "Tertius." We have demonstrated the main problems concerning reaching the target clock rate and have described a method for their solution. Owing to our methodology, it is possible to increase the clock rate; hence, the performance of Tertius will increase by 25% without revising the functional circuit of the task's computing structure. According to our current research of the suggested methodology and its efficiency, we claim that CAD tools, used for creating topological restrictions and based on our methodology, will significantly reduce the time for developing programs with the required characteristics for reconfigurable computer systems.

Reconfigurable computer systems; FPGA; CAD; Physical Constraints; Placement Constraints; Timing Closure.

Введение. Основной аппаратной компонентой высокопроизводительных реконфигурируемых вычислительных систем (РВС) являются ПЛИС, объединённые мощной пространственной коммутацией [1]. Именно реконфигурационные возможности современных ПЛИС позволяют соответствующим образом адаптировать архитектуру РВС и решать вычислительно-трудоемкие задачи математической физики, цифровой обработки сигналов, искусственного интеллекта и других областей с производительностью порядка 60–90% от пиковой, тогда как на большинстве задач серверы традиционной архитектуры и кластерные вычислители зачастую работают с производительностью, не превышающей 10% [1, 2].

Однако в отличие от классических вычислительных систем, программирование РВС является более сложным процессом, включающим разработку конфигурационных файлов ПЛИС (англ. *configware*), реализующих вычислительную структуру задачи, разработку программ управления контроллерами внешней памяти и программ управления процессом решения. Проектирование и отладка такой программной триады требуют высокой квалификации специалистов и значительных временных затрат [1, 2]. Конечно, процесс программирования РВС пытаются

ускорить, активно развивая высокоуровневые средства автоматизированного синтеза, однако чаще всего требуемая производительность системы не обеспечивается без низкоуровневой оптимизации вычислительной структуры решаемой задачи [3].

В основном это связано с неспособностью САПР [4, 5], применяемых для получения конфигурационных файлов, обеспечить целевые тактовые частоты при высокой утилизации ресурсов ПЛИС. Только при утилизации менее 40% размещение функциональных узлов вычислительной структуры, выполненное САПР в автоматическом режиме, позволяет проложить трассы оптимальной длины. Это подтверждается экспериментальными исследованиями зависимости тактовых частот от утилизации ресурсов ПЛИС, проведёнными при разработке конфигурационных файлов для таких вычислительно-трудоемких задач, как прогнозирование и исключение кратных волн [6, 7], обратная кинематическая задача сейсморазведки [8], LU-разложение [9], модифицированный попеременно-треугольный метод решения сеточных уравнений [10], молекулярное моделирование [11].

Повышение утилизации естественным образом сокращает число доступных вариантов трассировки соединений, удовлетворяющих требованиям к тактовой частоте. Результаты исследований свидетельствуют о настолько неоптимальном автоматическом размещении функциональных узлов вычислительной структуры с утилизацией логических ресурсов ПЛИС выше 40% и встроенных арифметических блоков и блоков памяти выше 60%, что полученная тактовая частота оказывается до двух раз меньше целевой. В связи с этим, разработчики САПР [4, 5] предлагают оптимизировать процессы топологического размещения вычислительной структуры на поле примитивов ПЛИС и трассировки соединений между ее функциональными узлами путём перебора различных стратегий (алгоритмов размещения и трассировки). Некоторые САПР также предлагают возможность автоматического определения наилучшего сочетания стратегий на основе данных машинного обучения [4], другие ориентированы на наиболее оптимальное размещение элементов функциональных узлов и трассировку связей между ними [12–15].

Однако для получения высокой реальной производительности РВС, помимо прочего, должны учитываться требования по размещению функциональных узлов вычислительной структуры относительно друг друга, а также геометрические и архитектурные особенности ПЛИС. Важно отметить, что на размещение функциональных узлов влияют не только обусловленные математическим аппаратом решаемой задачи функциональные зависимости, но и зависимости от узлов управления ходом вычислений и потоками данных. Из-за отсутствия комплексного анализа обозначенных выше требований, современные САПР не позволяют при высокой утилизации ПЛИС в автоматическом режиме получать конфигурационные файлы с необходимыми характеристиками [3].

В настоящее время удовлетворить такой совокупности требований можно лишь путем «ручного» управления процессом топологического размещения элементов вычислительной структуры на поле примитивов ПЛИС. В качестве инструментария современные САПР предлагают использовать так называемые топологические ограничения [16]. Анализ работ [14, 15, 17, 18], в которых топологические ограничения используются для достижения целевых тактовых частот, позволяет утверждать, что на сегодняшний момент не существует какой-либо методики, нацеленной на достижение реальной производительности РВС, близкой к пику.

В рамках данной статьи описана методика создания топологических ограничений ПЛИС, основанная на подходе, разработанном и используемом специалистами ООО «НИЦ СЭ и НК» [3]. Применение методики обеспечивает достижение целевых тактовых частот при высокой утилизации доступных ресурсов РВС за счет более оптимального топологического размещения вычислительной структуры

на поле примитивов ПЛИС по сравнению со всеми известными САПР. Работоспособность методики проиллюстрирована на примере решения прикладной тестовой задачи.

Методика создания топологических ограничений ПЛИС. Реальная производительность РВС напрямую зависит от степени масштабирования вычислительной структуры и тактовой частоты ее работы.

В свою очередь степень масштабирования и тактовая частота зависят от утилизации аппаратного ресурса ПЛИС. Аппаратный ресурс определяется числом примитивов (таблиц истинности LUT, триггеров FF, встроенных арифметических блоков DSP и блоков памяти BRAM), числом аппаратно-реализованных периферийных устройств (каналы ввода/вывода данных и встроенные стандартные интерфейсы). Так же к аппаратному ресурсу относится коммутационная матрица – внутренний интерконнект, расположенный по горизонтали и вертикали между примитивами ПЛИС и имеющий коммутационные узлы в местах пересечений трасс.

При высокой утилизации ресурсов ПЛИС именно длины трасс между примитивами в большей степени определяют предельную тактовую частоту вычислительной структуры. В свою очередь на длины трасс влияют как мощность коммутационной матрицы, так и технологии изготовления ПЛИС, предполагающие сопряжение нескольких кремниевых кристаллов в одном корпусе. Фирма AMD – Xilinx использует технологию Stacked Silicon Interconnect Technology для объединения больших областей реконфигурируемой логики (Super Logic Regions, SLRs) в единое поле примитивов (к примеру, семейства UltraScale+ и Versal) [19, 20]. Фирма Intel использует технологию 3D packing для сопряжения реконфигурируемой логики с НВМ памятью и аппаратно-реализованными периферийными устройствами (например, семейство Agilex) [21]. Результаты исследований подтверждают, что целевая тактовая частота не может быть достигнута в случае, если вычислительная структура использует множество трасс между примитивами, расположенными на больших расстояниях на разных кремниевых кристаллах, из-за чего реальная производительность РВС может быть ниже предельной на 20% и более. Аналогичная ситуация наблюдается при пересечении трассами областей ПЛИС с аппаратно-реализованными периферийными устройствами.

Помимо прочего, примитивы LUT и FF функциональных узлов концентрируются средствами САПР в непосредственной близости от функционально зависимых примитивов DSP и BRAM. В случае неоптимального размещения функциональных узлов на поле примитивов ПЛИС, их элементы мешают друг другу проложить трассы требуемой длины между собой. Обычно подобная ситуация наблюдается при высокой утилизации ресурсов ПЛИС, когда при размещении DSP и BRAM не учитываются функциональные зависимости между узлами вычислительной структуры и зависимости от узлов управления ходом вычислений и потоками данных.

Для повышения реальной производительности РВС при решении вычислительно-трудоемких задач предлагается нижеизложенная методика создания топологических ограничений ПЛИС. Методика направлена как на минимизацию длин трасс между примитивами за счет рационального размещения элементов функциональных узлов в заданной топологической области ПЛИС, так и на минимизацию числа трасс между топологическими областями, разделенными аппаратно-реализованными периферийными устройствами или представляющими собой отдельные кремниевые кристаллы.

Методика предполагает выполнение следующей совокупности действий:

Во-первых, необходимо разработать траекторию размещения узлов вычислительной структуры, учитывающую как функциональные зависимости между ними, так и зависимости от узлов управления ходом вычислений и потоками данных. При этом траектория должна обеспечить:

- ◆ размещение в непосредственной близости функциональных узлов вычислительной структуры, имеющих функциональную зависимость по приему и передаче данных;
 - ◆ последовательное размещение функциональных узлов вычислительной структуры, относительно их зависимости от узлов управления ходом вычислений и потоками данных;
 - ◆ размещение узлов, отвечающих за организацию связи реконфигурируемой логики ПЛИС с внешними устройствами, в непосредственной близости к встроенным периферийным устройствам;
 - ◆ минимальное число пересечений трассами топологических областей ПЛИС, содержащих аппаратно-реализованные периферийные устройства;
 - ◆ минимальное число пересечений трассами топологических областей сопряжения разных кремниевых кристаллов ПЛИС;
 - ◆ исключение пересечений уже занятых топологических областей ПЛИС.
- Во-вторых, необходимо разместить элементы функциональных узлов на поле примитивов ПЛИС в соответствии с разработанной траекторией:
- ◆ создать топологические ограничения для элементов вычислительной структуры, использующих примитивы аппаратно-реализованных арифметических блоков и блоков памяти ПЛИС;
 - ◆ создать топологические ограничения для элементов вычислительной структуры, использующих примитивы аппаратно-реализованных периферийных устройств;
 - ◆ создать топологические ограничения для элементов вычислительной структуры, размещенных в конкретных CLB.

Проведенные исследования показали, что размещение арифметических блоков и блоков памяти в соответствии с разработанной траекторией позволяет обеспечить более равномерную нагрузку на коммутационную матрицу ПЛИС за счет оптимизации длины трасс между логическими ячейками и в целом между функциональными узлами. Стоит отметить, что иногда этого достаточно для достижения целевой тактовой частоты.

Инструменты топологических ограничений имеются у каждого производителя ПЛИС. В качестве примера приведен инструментарий Physical Constraints фирмы AMD – Xilinx [16]:

- ◆ LUTNM и HLUTNM: размещение двух LUT в одной ячейке с использованием выходов O5 и O6;
 - ◆ PROHIBIT: запрет на размещение элементов функциональных узлов в указанной ячейке или области ПЛИС (clock region; SLR);
 - ◆ PBLOCK (Physical Block): физическое ограничение отдельной области ПЛИС для реализации в ней некоторых функциональных узлов;
 - ◆ LOC: размещение элемента узла в указанной ячейке;
 - ◆ BEL: размещение элемента узла в указанной части CLB;
 - ◆ U_SET и HU_SET: объединение элементов в одну логическую группу;
 - ◆ RLOC: расположение элементов относительно друг друга в рамках одной логической группы;
 - ◆ FIXED_ROUTE: фиксация трассы между двумя конкретными ячейками.
- В-третьих, необходимо уменьшить длины трасс, использующих несколько последовательно подключенных примитивов LUT (трассы с высокой логической нагрузкой, англ. logic levels):
- ◆ применить инструменты ограничения топологической области ПЛИС для размещения элементов функциональных узлов;

- ◆ создать топологические ограничения, обеспечивающие размещение элементов вычислительной структуры относительно их взаимного расположения;
- ◆ создать топологические ограничения, обеспечивающие размещение элементов вычислительной структуры в конкретных примитивах LUT и FF;
- ◆ создать топологические ограничения для фиксации конкретных траекторий трасс между используемыми примитивами ПЛИС.

Подобный вид размещения в некоторых случаях приводит также к минимизации занимаемого функциональным узлом топологического пространства ПЛИС.

В-четвертых, необходимо минимизировать число трасс между примитивами, расположенными на разных кремниевых кристаллах или разделенными аппаратно-реализованными периферийными устройствами:

- ◆ создать топологические ограничения для крупных областей ПЛИС, отталкиваясь от архитектурных особенностей конкретной микросхемы;
- ◆ определить, к какой из определенных топологических областей относится каждый функциональный узел вычислительной структуры;
- ◆ разместить элементы функциональных узлов в конкретной топологической области ПЛИС.

Если после размещения всех функциональных узлов утилизация ресурсов различных топологических областей существенно отличается, возникает неравномерная нагрузка на коммутационную матрицу ПЛИС: в топологической области с меньшей утилизацией ресурсов остается больше вариантов трассировки соединений между элементами, тогда как в физической области с большей утилизацией ресурсов – меньше вариантов. Такое неравномерное использование ресурсов в разных топологических областях может значительно влиять на результирующую тактовую частоту работы вычислительной структуры.

Для того чтобы избежать излишней нагрузки на коммутационную матрицу ПЛИС, необходимо:

- ◆ разделить крупные пограничные функциональные узлы по операциям;
- ◆ разместить пограничные функциональные узлы в обеих смежных топологических областях ПЛИС: часть операций каждого узла разместить в одной области, другую часть – в другой области.

Если доступный аппаратный ресурс топологических областей отличается, иногда рационально изменять саму структуру функционального узла, подстраивая ее под архитектуру конкретной области.

В-пятых, если это возможно, необходимо минимизировать длину трасс между примитивами, расположенными на разных кремниевых кристаллах или разделенными аппаратно-реализованными периферийными устройствами:

- ◆ изменить вычислительную структуру на границах пересечения данных топологических областей путем добавления блоков с нулевой логической нагрузкой ($\text{logic levels} = 0$): связи типа «триггер-триггер»;
- ◆ создать топологические ограничения для элементов функциональных узлов, трассы между которыми пересекают области ПЛИС с аппаратно-реализованными периферийными устройствами или области сопряжения разных кремниевых кристаллов, в непосредственной близости от границ данных топологических областей.

При этом рационально использовать архитектурные решения производителей ПЛИС: фирма AMD – Xilinx реализовала Input multiplexer (IMUX) registers – дополнительные регистры, расположенные на границе Interconnect/CLB (серия Versal) [20]; фирма Intel использует технологию HyperFlex, согласно которой реализовано множество дополнительных регистров в узлах коммутационной матрицы (серии Stratix, Agilex) [21].

В-шестых, необходимо выполнить оценку качества созданного набора топологических ограничений путем анализа значений времени предустановки и удержания сигналов (Setup и Hold) при получении конфигурационных файлов ПЛИС. Если требования целевой тактовой частоты работы вычислительной структуры удовлетворены, стоит завершить выполнение методики, иначе нужно применить методику повторно для дальнейшей оптимизации размещения функциональных узлов в ПЛИС. Необходимо отметить, что оценку можно выполнять после любого из действий, описанных в настоящей методике.

Апробация. Эффективность описанной методики показана на примере создания конфигурационных файлов ПЛИС для решения задачи нагрузочного тестирования и оценки потребляемой мощности реконфигурируемого компьютера (ПК) «Терциус» производства «НИЦ СЭ и НК» [22]. Основным вычислительным элементом рассматриваемой реконфигурируемой системы являются ПЛИС AMD – Xilinx Kintex UltraScale XCKU095 [23]. В качестве целевой тактовой частоты была взята частота 500 МГц.

Функциональные узлы тестовой задачи представляют собой КИХ-фильтр четвертого порядка с симметричными коэффициентами, содержащий семь операций сложения и четыре операции умножения в формате 32FP IEEE 754 (рис. 1) [24, 25].

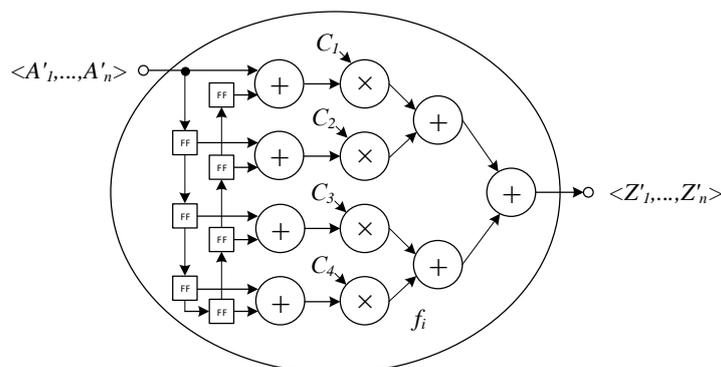


Рис. 1. Схема функционального узла тестовой задачи «КИХ-фильтр»

Для обеспечения максимально возможной утилизации ресурсов ПЛИС реализована вычислительная структура, состоящая из 128 последовательно соединённых функциональных узлов. С помощью САПР Vivado 2020.2 получены конфигурационные файлы ПЛИС, характеристики которых представлены в табл. 1.

Таблица 1

**Характеристики реализации тестовой задачи «КИХ-фильтр»
и занимаемый аппаратный ресурс**

FF	LUT	DSP	BRAM	Frequency	Performance
779344	372611	768	16	400 МГц	563 Gflops
72.48%	69.31%	100%	0.95%		

Как видно из табл. 1, при соответствующей утилизации ресурсов ПЛИС тактовая частота вычислительной структуры на 20% ниже целевой. Визуальный анализ размещения функциональных узлов средствами САПР (рис. 2) подтвердил, что узлы вычислительной структуры расположены в неоптимальном порядке, что дает значительную дополнительную нагрузку на коммутационную матрицу ПЛИС.



Рис. 2. Размещение функциональных узлов на ПЛИС XCKU095, выполненное САПР Vivado 2020.2 при реализации тестовой задачи «КИХ-фильтр»

В соответствии с описанной выше методикой, при создании конфигурационных файлов была разработана траектория размещения по типу «змейка», которая обеспечивает: соседство узлов, имеющих функциональную зависимость или зависимость от процессов плавного запуска решения и плавного сброса; минимальное число пересечений с областями ввода/вывода данных; отсутствие пересечений занятых топологических областей ПЛИС.

Далее было выполнено размещение блоков DSP в соответствии с разработанной траекторией с помощью инструмента AMD – Xilinx Physical Constraints – LOC [16]. Оценка качества созданного набора топологических ограничений показала, что на данном этапе максимальная тактовая частота вычислительной структуры составляет 445 МГц. Визуализация размещения функциональных узлов в ПЛИС (рис. 3) косвенно свидетельствует об оптимизации нагрузки на коммутационную матрицу ПЛИС. В этом случае использовать такие инструменты как BEL, U_SET, NU_SET, RLOC и FIXED_ROUTE [16] не требуется.

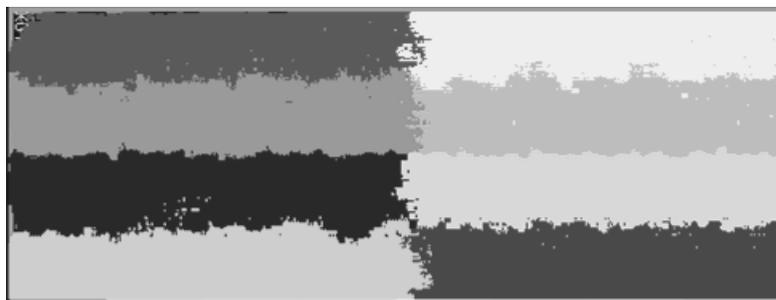


Рис. 3. Результат размещения функциональных узлов на ПЛИС XCKU095 после привязки элементов к примитивам DSP при реализации тестовой задачи «КИХ-фильтр»

Однако при попытке реализации вычислительной структуры на целевой тактовой частоте подавляющее большинство трасс, длина которых не удовлетворяет заданным требованиям, пересекает области ввода/вывода данных (рис. 4).

Для сокращения числа таких трасс поле примитивов ПЛИС с помощью инструмента PBLOCK [16] было разделено на три топологических области, ограниченных областями ввода/вывода данных. Поскольку в ПЛИС XCKU095 два центральных столбца DSP непосредственно граничат с областями ввода/вывода данных, часть операций функциональных узлов, использующих данные примитивы, была размещена в одной из смежных топологических областей, а оставшаяся – в другой.

Такое размещение обеспечило практически одинаковую утилизацию ресурсов в каждой топологической области и равномерную нагрузку на коммутационную матрицу ПЛИС. Для сокращения длины трасс между смежными топологическими областями в соответствующих функциональных узлах были реализованы дополнительные каскады с нулевой логической нагрузкой.

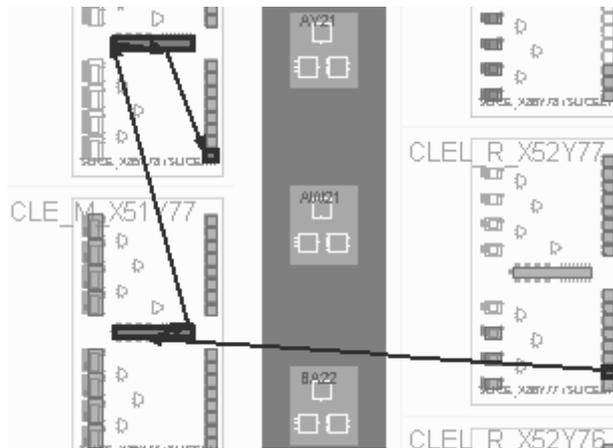


Рис. 4. Трасса, пересекающая области ввода/вывода данных при реализации тестовой задачи «КИХ-фильтр»

Характеристики конфигурационных файлов, полученных после описанных преобразований, приведены в табл. 2.

Таблица 2

Характеристики реализации тестовой задачи «КИХ-фильтр» и занимаемый аппаратный ресурс после применения разработанной методики

FF	LUT	DSP	BRAM	Frequency	Performance
780970	372755	768	16	500 МГц	702 Gflops
72.63%	69.34%	100%	0.95%		

Применение разработанной методики создания топологических ограничений ПЛИС позволило обеспечить работу вычислительной структуры тестовой задачи «КИХ-фильтр» на целевой тактовой частоте 500 МГц при высокой утилизации ресурсов ПЛИС.

Как следствие, при решении тестовой задачи, производительность РК «Терциус» увеличилась в 1.25 раза.

Помимо данной иллюстрации, представленная методика была апробирована при решении тестовой задачи «КИХ-фильтр» на других РВС. Так, без применения предложенной методики, САПР позволил достичь частоты работы 400 МГц на РВС «Сегин» [26], основным вычислительным компонентом которой являются ПЛИС Xilinx семейства Virtex UltraScale Plus. Благодаря применению методики, тактовая частота была увеличена до 525 МГц, за счет чего производительность РВС выросла в 1.31 раза.

Вместе с этим применение методики позволило увеличить производительность РВС при решении ряда других прикладных вычислительно-трудоемких задач разных предметных областей. К примеру, при решении СЛАУ методом LU разложения на РК «Терциус» [9] тактовая частота увеличена с 350 МГц до 435 МГц, за счет чего производительность РВС выросла в 1.24 раза.

Заключение. В рамках данной работы рассмотрено решение задачи многокритериальной оптимизации топологического размещения функциональных узлов вычислительной структуры на поле примитивов ПЛИС. Основными критериями, определяющими сложность данной задачи, являются высокая утилизация ресурсов ПЛИС и высокая частота работы вычислительной структуры.

Показано, что современные САПР зачастую не позволяют получать конфигурационные файлы, удовлетворяющие высоким требованиям к производительности системы. Достижение заданных характеристик системы можно обеспечить путем топологического размещения функциональных узлов вычислительной структуры на поле примитивов ПЛИС.

В результате формализации подхода, разработанного и применяемого специалистами «НИЦ СЭ и НК», была описана методика создания топологических ограничений ПЛИС, нацеленная на достижение высокой реальной производительности РВС при решении вычислительно-трудоемких задач. В отличие от существующих технологий, предлагаемых производителями САПР, разработанная методика позволяет:

- ◆ учитывать функциональные зависимости между узлами вычислительной структуры;
- ◆ учитывать зависимости от узлов управления ходом вычислений и потоками данных;
- ◆ учитывать геометрические и архитектурные особенности ПЛИС путем балансировки степени утилизации во всех топологических областях за счет адаптации функциональных узлов под ресурс области их размещения.

Автоматизация предложенной методики создания топологических ограничений ПЛИС позволит сократить время разработки параллельно-конвейерных программ для высокопроизводительных РВС при решении вычислительно-трудоемких задач. Перспективными являются разработка и применение в автоматическом режиме эвристических алгоритмов оптимального размещения функциональных узлов вычислительной структуры в заданной топологической области ПЛИС, что потенциально может привести к более значительному сокращению времени разработки прикладного программного обеспечения с требуемыми характеристиками по производительности для РВС.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Каляев А.В., Левин И.И.* Модульно-наращиваемые многопроцессорные системы со структурно-процедурной организацией вычислений. – М.: Янус-К, 2003. – 380 с.
2. *Каляев И.А., Левин И.И., Семерников Е.А., Шмойлов В.И.* Реконфигурируемые мультиконвейерные вычислительные структуры. – 2-е изд., перераб. и доп. / под общ. ред. И.А. Каляева. – Ростов-на-Дону: Изд-во ЮНЦ РАН, 2009. – 344 с. – ISBN 978-5-902982-61-6.
3. *Алексеев К.Н., Сорокин Д.А., Леонтьев А.Л.* Метод управления размещением элементов вычислительной структуры при максимальной утилизации ресурсов ПЛИС // XIV Всероссийская мультikonференция по проблемам управления (МКПУ-2021): Матер. XIV мультikonференции (Дивноморское, Геленджик, 27 сентября – 2 октября 2021 г.): в 4 т. Т. 2. Редкол.: И.А. Каляев, В.Г. Пешехонов и др. – Ростов-на-Дону; Таганрог: Изд-во ЮФУ, 2021. – ISBN 978-5-9275-3846-1. – С. 238-240.
4. AMD Xilinx Vivado Overview. – Режим доступа: <https://www.xilinx.com/products/design-tools/vivado.html> (дата обращения: 18.05.2022).
5. Intel Quartus Prime Software Suite Overview. – Режим доступа: <https://www.intel.com/content/www/us/en/software/programmable/quartus-prime/overview.html> (дата обращения: 18.05.2022).
6. *Alekseev K., Levin I., Sorokin D.* Implementation of surface-related multiple prediction task on reconfigurable computer systems // Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming and Computer Software. – 2020. – No. 13 (1). – P. 81-94.

7. *Алексеев К.Н., Сорокин Д.А., Матросов А.Ю., Семерникова Е.Е.* Структурно-процедурная реализация алгоритма прогнозирования кратных волн на ПЛИС // Известия ЮФУ. Технические науки. – 2016. – № 12. – С. 16-28.
8. *Алексеев К.Н., Левин И.И.* Реализация обратной кинематической задачи сейморазведки для микросейсмического мониторинга на реконфигурируемых вычислительных системах в реальном масштабе времени // Известия ЮФУ. Технические науки. – 2018. – № 8 (202). – С. 221-231.
9. *Левин И.И., Пелипец А.В.* Эффективная реализация распараллеливания на реконфигурируемых системах // Вестник компьютерных и информационных технологий. – 2018. – № 8. – С. 11-16.
10. *Левин И.И., Доронченко Ю.И., Сорокин Д.А., Чистяков А.Е.* Моделирование распространения акустических волн в массивной породе с применением реконфигурируемой вычислительной системы // Нефтяное хозяйство. – 2016. – № 3. – С. 50-53.
11. *Сорокин, Д.А., Дордопуло А.И.* Методика сокращения аппаратных затрат в сложных системах при решении задач с существенно-переменной интенсивностью потоков данных // Известия ЮФУ. Технические науки. – 2012. – № 4. – С. 213-219.
12. RippleFPGA. – Режим доступа: <https://github.com/cuhk-eda/ripple-fpga> (дата обращения: 18.05.2022).
13. AMF-Placer. – Режим доступа: <https://github.com/zslwyuan/AMF-Placer> (дата обращения: 18.05.2022).
14. *Marrouff D., Shamli A., Martin T., Grewal G., and Areibi S.* A Deep-Learning Framework for Predicting Congestion during FPGA Placement // in 30th Int'l Conference on Field Programmable Logic and Applications, Sweden, September 2020. – P. 138-144.
15. *Chak-Wa Pui, Gengjie Chen, Yuzhe Ma, Evangeline F.Y. Young, and Bei Yu.* Clock-aware UltraScale FPGA placement with machine learning routability prediction // In IEEE/ACM International Conference on Computer-Aided Design (ICCAD). – 2017. – P. 915-922.
16. Vivado Design Suite User Guide: Using Constraints. – Режим доступа: https://www.xilinx.com/content/dam/xilinx/support/documentation/sw_manuals/xilinx2021_1/ug903-vivado-using-constraints.pdf (дата обращения: 18.05.2022).
17. Vivado Design Suite User Guide: Design Analysis and Closure Techniques. – Режим доступа: https://www.xilinx.com/content/dam/xilinx/support/documents/sw_manuals/xilinx2021_2/ug906-vivado-design-analysis.pdf#nameddest=xPerformingTimingAnalysis (дата обращения: 18.05.2022).
18. *NiansongZ., XiangCh., Nachiket K.* RapidLayout: Fast Hard Block Placement of FPGA-optimized Systolic Arrays using Evolutionary Algorithms. – Режим доступа: <https://arxiv.org/abs/2002.06998> (дата обращения: 18.05.2022).
19. UltraScale Architecture Configurable Logic Block. – Режим доступа: <https://docs.xilinx.com/v/u/en-US/ug574-ultrascale-clb> (дата обращения: 18.05.2022).
20. Versal Architecture and Product Data Sheet: Overview. – Режим доступа: <https://docs.xilinx.com/v/u/en-US/ds950-versal-overview> (дата обращения: 18.05.2022).
21. Intel® Agilex™ FPGAs and SoCs Device Overview. – Режим доступа: <https://www.intel.com/content/www/us/en/docs/programmable/683458/current/fpga-and-soc-device-overview.html> (дата обращения: 18.05.2022).
22. НИЦ СЭ и НК: Терциус. – Режим доступа: <http://superevm.ru/index.php?page=tertsius> (дата обращения: 20.05.2022).
23. UltraScale Architecture and Product Data Sheet: Overview. – Режим доступа: <https://docs.xilinx.com/v/u/en-US/ds890-ultrascale-overview> (дата обращения: 20.05.2022).
24. *Лайонс Р.* Цифровая обработка сигналов: пер. с англ. А.А. Бритова. – 2-е изд. – М.: БИНОМ, 2007. – 652 с.
25. *Левин И.И., Семерников Е.А.* Устойчивость конвейерных рекурсивных фильтров // «Вестник Южного научного центра Российской академии наук». – Ростов-на-Дону: Изд-во ЮНЦ РАН, 2005. – Т. 1. – В. 2. – С. 28-40.
26. НИЦ СЭ и НК: Сегин плата вычислительного модуля. – Режим доступа: <http://superevm.ru/index.php?page=segin-plata-vychislitelnogo-modulya> (дата обращения: 20.05.2022).

REFERENCES

1. *Kalyaev A.V., Levin I.I.* Modul'no-narashchivaemye mnogoprotsessornye sistemy so strukturno-protsedurnoy organizatsiyey vychisleniy [Modular-stackable multiprocessor systems with structural and procedural organization of computing]. Moscow: Yanus-K, 2003. 380 p.
2. *Kalyaev I.A., Levin I.I., Semernikov E.A., Shmoylov V.I.* Rekonfiguriruemye mul'tikonveyernye vychislitel'nye struktury [Reconfigurable multiconveyor computing structures]. 2nd ed., rev. and suppl., under the general ed. of I.A. Kalyaeva. Rostov-on-Don: Izd-vo YuNTS RAN, 2009, 344 p. ISBN 978-5-902982-61-6.
3. *Alekseev K.N., Sorokin D.A., Leont'ev A.L.* Metod upravleniya razmeshcheniem elementov vychislitel'noy struktury pri maksimal'noy utilizatsii resursov PLIS [The method of managing the placement of elements of the computing structure with maximum utilization of FPGA resources], *XIV Vserossiyskaya mul'tikonferentsiya po problemam upravleniya (MKPU-2021): Mater. XIV mul'tikonferentsii (Divnomorskoe, Gelendzhik, 27 sentyabrya – 2 oktyabrya 2021 g.)* [XIV All-Russian Multi-conference on Management Problems (MCPU-2021): Materials of the XIV multi-conference (Divnomorskoe, Gelendzhik, September 27- October 2, 2021)]: In 4 vol. Vol. 2. Ed. board: I.A. Kalyaev, V.G. Peshekhonov, etc. Rostov-on-Don; Taganrog: Izd-vo YuFU, 2021. ISBN 978-5-9275-3846-1, pp. 238-240.
4. AMD Xilinx Vivado Overview. Available at: <https://www.xilinx.com/products/design-tools/vivado.html> (accessed 18 May 2022).
5. Intel Quartus Prime Software Suite Overview. Available at: <https://www.intel.com/content/www/us/en/software/programmable/quartus-prime/overview.html> (accessed 18 May 2022).
6. *Alekseev K., Levin I., Sorokin D.* Implementation of surface-related multiple prediction task on reconfigurable computer systems, *Bulletin of the South Ural State University. Series: Mathematical Modelling, Programming and Computer Software*, 2020, No. 13 (1), pp. 81-94.
7. *Alekseev K.N., Sorokin D.A., Matrosov A.Yu., Semernikova E.E.* Strukturno-protsedurnaya realizatsiya algoritma prognozirovaniya kratnykh voln na PLIS [Structural and procedural implementation of the algorithm for predicting multiple waves on FPGA], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2016, No. 12, pp. 16-28.
8. *Alekseev K.N., Levin I.I.* Realizatsiya obratnoy kinematicheskoy zadachi seysmorazvedki dlya mikrozeismicheskogo monitoringa na rekonfiguriruemyykh vychislitel'nykh sistemakh v real'nom masshtabe vremeni [Implementation of the inverse kinematic problem of seismic exploration for microseismic monitoring on reconfigurable computing systems in real time], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2018, No. 8 (202), pp. 221-231.
9. *Levin I.I., Pelipets A.V.* Effektivnaya realizatsiya rasparallelivaniya na rekonfiguriruemyykh sistemakh [Effective implementation of parallelization on reconfigurable systems], *Vestnik komp'yuternyykh i informatsionnykh tekhnologiy* [Bulletin of Computer and Information Technologies], 2018, No. 8, pp. 11-16.
10. *Levin I.I., Doronchenko Yu.I., Sorokin D.A., Chistyakov A.E.* Modelirovanie rasprostraneniya akusticheskikh voln v massivnoy porode s primeneniem rekonfiguriruemoy vychislitel'noy sistemy [Modeling of acoustic wave propagation in a massive rock using a reconfigurable computing system], *Neftyanoe khozyaystvo* [Oil industry], 2016, No. 3, pp. 50-53.
11. *Sorokin, D.A., Dordopulo A.I.* Metodika sokrashcheniya apparatnykh zatrat v slozhnykh sistemakh pri reshenii zadach s sushchestvenno-peremennoy intensivnost'yu potokov dannykh [Methodology for reducing hardware costs in complex systems when solving problems with significantly variable intensity of data flows], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 4, pp. 213-219.
12. RippleFPGA. Available at: <https://github.com/cuhk-eda/ripple-fpga> (accessed 18 May 2022).
13. AMF-Placer. Available at: <https://github.com/zslwyuan/AMF-Placer> accessed 18 May 2022).
14. *Marrouff D., Shamli A., Martin T., Grewal G., and Areibi S.* A Deep-Learning Framework for Predicting Congestion during FPGA Placement, in *30th Int'l Conference on Field Programmable Logic and Applications, Sweden, September 2020*, pp. 138-144.
15. *Chak-Wa Pui, Gengjie Chen, Yuzhe Ma, Evangeline F.Y. Young, and Bei Yu.* Clock-aware UltraScale FPGA placement with machine learning routability prediction, *In IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2017, pp. 915-922.

16. Vivado Design Suite User Guide: Using Constraints. Available at: https://www.xilinx.com/content/dam/xilinx/support/documentation/sw_manuals/xilinx2021_1/ug903-vivado-using-constraints.pdf (accessed 18 May 2022).
17. Vivado Design Suite User Guide: Design Analysis and Closure Techniques. Available at: https://www.xilinx.com/content/dam/xilinx/support/documents/sw_manuals/xilinx2021_2/ug906-vivado-design-analysis.pdf#nameddest=xPerformingTimingAnalysis (accessed 18 May 2022).
18. NiansongZ., XiangCh., Nachiket K. RapidLayout: Fast Hard Block Placement of FPGA-optimized Systolic Arrays using Evolutionary Algorithms. Available at: <https://arxiv.org/abs/2002.06998> (accessed 18 May 2022).
19. UltraScale Architecture Configurable Logic Block. Available at: <https://docs.xilinx.com/v/u/en-US/ug574-ultrascale-clb> (accessed 18 May 2022).
20. Versal Architecture and Product Data Sheet: Overview. Available at: <https://docs.xilinx.com/v/u/en-US/ds950-versal-overview> (accessed 18 May 2022).
21. Intel® Agilex™ FPGAs and SoCs Device Overview. Available at: <https://www.intel.com/content/www/us/en/docs/programmable/683458/current/fpga-and-soc-device-overview.html> (accessed 18 May 2022).
22. NITS SE i NK: Tertsius [SIC SE and NC: Tertius. – Access mode]. Available at: <http://superevm.ru/index.php?page=tertsius> (accessed 20 May 2022).
23. UltraScale Architecture and Product Data Sheet: Overview. Available at: <https://docs.xilinx.com/v/u/en-US/ds890-ultrascale-overview> (accessed 20 May 2022).
24. Layons R. Tsifrovaya obrabotka signalov [Digital signal processing]: transl. from the engl. by A.A. Britova. 2nd ed. Moscow: BINOM, 2007, 652 p.
25. Levin I.I., Semernikov E.A. Ustoychivost' konveyernykh rekursivnykh fil'trov [Stability of conveyor recursive filters], *Vestnik Yuzhnogo nauchnogo tsentra Rossiyskoy akademii nauk* [Bulletin of the Southern Scientific Center of the Russian Academy of Sciences]. Rostov-on-Don: Izd-vo YuNTS RAN, 2005, Vol. 1, V. 2, pp. 28-40.
26. NITS SE i NK: Segin plata vychislitel'nogo modulya [SIC SE and NC: Seguin computing module board]. Available at: <http://superevm.ru/index.php?page=segin-plata-vychislitel'nogo-modulya> (accessed 20 May 2022).

Статью рекомендовал к опубликованию к.т.н. А.Н. Самойлов.

Алексеев Кирилл Николаевич – Общество с ограниченной ответственностью "НИЦ супер-ЭВМ и нейрокомпьютеров"; e-mail: alexseev91@mail.ru; г. Таганрог, Россия; тел.: +79283536268; научный сотрудник; к.т.н.

Сорокин Дмитрий Анатольевич – e-mail: jotun@inbox.ru; тел.: +79508668253; начальник отдела; к.т.н.

Леонтьев Антон Леонидович – e-mail: leontyev_anton@mail.ru; тел.: +79185102920; программист 2 категории.

Alekseev Kirill Nikolayevich – “Supercomputers and Neurocomputers Research Center” Co Ltd; e-mail: alexseev91@mail.ru; Taganrog, Russia; phone: +79283536268; research assistant; cand. of eng. sc.

Sorokin Dmitriy Anatol'evich – e-mail: jotun@inbox.ru; phone: +79508668253; head of department; cand. of eng. sc.

Leont'ev Anton Leonidovich – e-mail: leontyev_anton@mail.ru; phone: +79185102920; 2nd category programmer.

Раздел III. Электроника, связь и навигация

УДК 621.315.592

DOI 10.18522/2311-3103-2022-4-213-221

Н.М. Богатов, В.С. Володин, Л.Р. Григорьян, А.И. Коваленко, М.С. Коваленко

ИМПУЛЬСНЫЕ ХАРАКТЕРИСТИКИ КРЕМНИЕВЫХ СТРУКТУР С N-P ПЕРЕХОДОМ, ОБЛУЧЕННЫХ ПРОТОНАМИ

В настоящее время активно развиваются методы создания полупроводниковых структур с заданными свойствами с помощью облучения ионизирующими частицами (инженерия радиационных дефектов). Взаимодействие радиационных дефектов с примесями, дислокациями и другими дефектами структуры обуславливает изменение свойств полупроводников и полупроводниковых приборов. Облучение протонами позволяет контролируемо создавать радиационные дефекты с максимумом распределения в заранее рассчитанной области. Цель работы – анализ влияния облучения низкоэнергетическими протонами на импульсные характеристики кремниевых структур с n^+ -р переходом. Задача – определение эффективного времени жизни τ носителей заряда в области пространственного заряда (ОПЗ) n^+ -р перехода. Исследовались n^+ -р-р⁺-структуры из кремния, выращенного методом Чохральского, облучённые со стороны n^+ -слоя потоком низкоэнергетических протонов при температуре образцов 300 К и 83 К. Для измерения импульсных характеристик использовались биполярные прямоугольные импульсы напряжения с постоянной амплитудой 10 мВ и частотой 1 МГц. Экспериментальные данные объясняются с помощью моделей нестационарного переноса носителей заряда в неоднородных полупроводниках и образования радиационных дефектов в кремнии под действием протонов. Рассчитаны распределения по глубине среднего числа первичных радиационных дефектов: междоузельного кремния, вакансий, дивакансий, созданных одним протоном на единице длины проективного пробега. Показано, что облучение протонами с дозой 10^{15} см⁻² и энергией 40 keV не изменяет значение τ , а с энергией 180 keV создает в ОПЗ n^+ -р перехода область с эффективным временем жизни $5.5 \cdot 10^{-8}$ с.

Кремний; n^+ -р переход; время жизни; протон.

N.M. Bogatov, V.S. Volodin, L.R. Grigoryan, A.I. Kovalenko, M.S. Kovalenko

IMPULSE CHARACTERISTICS OF SILICON STRUCTURES WITH N-P JUNCTION IRRADIATED BY PROTONS

Currently, methods are being actively developed to create semiconductor structures with desired properties by irradiation with ionizing particles (radiation defect engineering). The interaction of radiation defects with impurities, dislocations and other structural defects causes a change in the properties of semiconductors and semiconductor devices. Irradiation with protons makes it possible to controllably create radiation defects with a distribution maximum in a pre-calculated region. The aim of this work is to analyze the effect of irradiation with low-energy protons on the impulse characteristics of silicon structures with an n^+ -p junction. The task is to determine the effective lifetime τ of charge carriers in the space charge region (SCR) of the n^+ -p junction. The n^+ -p-p⁺-structures made of silicon grown by the Czochralski method, irradiated from the side of the n+-layer by a low-energy proton flux at sample temperatures of 300 K and 83 K were studied. To measure the impulse characteristics, bipolar rectangular voltage pulses with a constant amplitude of 10 mV and a frequency of 1 MHz were used. The experimental data are explained using models of nonstationary charge carrier transport in inhomogeneous semiconductors and the formation of radiation defects in silicon under

the action of protons. Depth distributions of the average number of primary radiation defects are calculated: interstitial silicon, vacancies, divacancies created by one proton per unit length of the projective path. It is shown that irradiation with protons with a dose of 10^{15} cm^{-2} and an energy of 40 keV does not change the value of τ , but with an energy of 180 keV creates a region with an effective lifetime of $5.5 \cdot 10^{-8} \text{ s}$ in the SCR of the n^+ - p junction.

Silicon; n^+ - p junction; lifetime; proton.

Введение. Создание полупроводниковых структур с заданными свойствами с помощью облучения ионизирующими частицами (инженерия радиационных дефектов) – актуальное направление полупроводниковой технологии [1]. Взаимодействие примесей, радиационных дефектов, дислокаций обуславливает изменение свойств полупроводников и полупроводниковых приборов [2–5].

Свойства радиационных дефектов зависят от типа ионизирующих частиц, дозы облучения, концентрации основных примесей. Приборы микро и нанoeлектроники чувствительны к изменениям свойств поверхностной области, вызванным протонами. Облучение протонами позволяет контролируемо изменять свойства структур на глубине от 0,1 мкм до 1 мм [6].

Технологии протонного облучения полупроводниковых приборов позволяют селективно вводить центры рекомбинации в кремниевые структуры, уменьшать эффективное время жизни носителей заряда в скрытых слоях, создавать слои с водород содержащими центрами [7]. Образовавшийся в результате имплантации ионов водорода нарушенный слой неоднороден, интегральные характеристики нарушенного слоя исследованы в [8].

Таким образом, определение эффективного времени жизни τ носителей заряда в области накопления радиационных дефектов является актуальной задачей.

Для измерения времени жизни неравновесных носителей заряда в кремнии с учетом влияния поверхностной рекомбинации используются измерения спада фотопроводимости [9, 10]. Бесконтактные измерения времени жизни с помощью регистрируемой микроволновым излучением фотопроводимости используются для контроля результатов технологических воздействий [11]. Для определения влияния структурных дефектов на объемную составляющую времени жизни неосновных носителей заряда и скорости рекомбинации в p - n -переходе в двухстороннем солнечном элементе из поликристаллического кремния применяется импульсное освещение [12].

Чтобы исключить временную зависимость фототока и фотопроводимости, использующуюся в методиках [9–12], в [13] измерялось переходное напряжение в неосвещенных кремниевых n^+ - p - p^+ структурах, облученных низкоэнергетическими протонами.

Цель работы – анализ влияния облучения низкоэнергетическими протонами на импульсные характеристики кремниевых структур с n - p -переходом.

Методика исследования. Экспериментально исследовались 10 образцов с n^+ - p - p^+ -структурой из кремния, выращенного методом Чохральского, с удельным сопротивлением базы p -типа $\rho = 10 \text{ }\Omega\text{-cm}$ и концентрацией равновесных дырок $p_0 \approx 10^{15} \text{ cm}^{-3}$, глубиной диффузионных n^+ - p и p - p^+ -переходов $d_n \approx d_p \approx 0,45 \text{ }\mu\text{m}$, толщиной $L \approx 200 \text{ }\mu\text{m}$. Поверхностная концентрация фосфора составляла $N_p \approx 10^{20} \text{ cm}^{-3}$, бора – $N_B \approx 10^{20} \text{ cm}^{-3}$. Образцы площадью $S \approx 1 \text{ cm}^2$ были получены лазерным разделением фотоэлектрических преобразователей площадью $S \approx 100 \text{ cm}^2$, содержащих контактную систему в виде решетки на лицевой и тыльной поверхностях, с помощью твердотельного ИАГ-лазера в импульсном режиме работы. Лазерная обработка торцов минимизирует токи утечки по торцам.

Образцы облучались со стороны n^+ -слоя потоком протонов с энергией $E_p = 40$ keV, $E_p = 180$ keV и дозой $F_p = 10^{15}$ cm^{-2} при температуре образцов $T_p = 300$ К, $T_p = 83$ К на имплантере Extrion/Varian: №1 – $E_p = 180$ keV, $T_p = 83$ К; №2 – $E_p = 40$ keV, $T_p = 83$ К; №3 – $E_p = 40$ keV, $T_p = 300$ К. Контрольная группа из 7 образцов, имеющих очень близкие электрофизические характеристики, №4 не облучалась.

Импульсные характеристики измерялись с помощью цифрового осциллографа DSOX2022A, реализующего функции генератора импульсов напряжения и мультиметра. Принципиальная электрическая схема показана на рис. 1. Для определения времени переключения использовались биполярные прямоугольные импульсы напряжения с постоянной амплитудой $U_m = 10$ mV и частотой $f = 1$ MHz. Зависимость напряжения U от времени для исследуемых образцов измерялась в темноте.

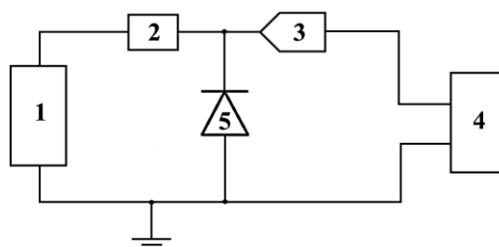


Рис. 1. Принципиальная электрическая схема: 1 – генератор импульсов напряжения, 2 – кабель сопротивлением $R = 0,1$ Ом, 3 – высокочастотный щуп $f_{max} = 300$ МГц, 4 – осциллограф, 5 – исследуемый образец

Паспортные данные осциллографа DSOX2022A: в режиме сбора данных разрешение по напряжению 0,01 mV с погрешностью 4%, максимальное временное разрешение 5 ps, входное сопротивление 1 МОм, входная емкость 18 pF. Длительность импульса составляла $0,5 \cdot 10^{-6}$ s, шаг дискретизации измеряемого сигнала – 2,5 ns. При этих параметрах эксперимента измеряемое время переключения τ должно находиться в пределах $5 \cdot 10^{-9}$ s $< \tau < 0,25 \cdot 10^{-6}$ s. Измерения проводились при температуре $T = 300$ К, так что выполняется неравенство

$$eU_m / kT < 1, \quad (1)$$

где e – элементарный заряд, k – постоянная Больцмана. В этом случае импульсные характеристики исследуемых структур симметричны при переключении напряжения (рис. 2). Зависимости $U(t)$ для образцов №2, №3, №4 отличаются мало. Переходные процессы не успевают завершиться, поэтому функции $U(t)$ не выходят на насыщение.

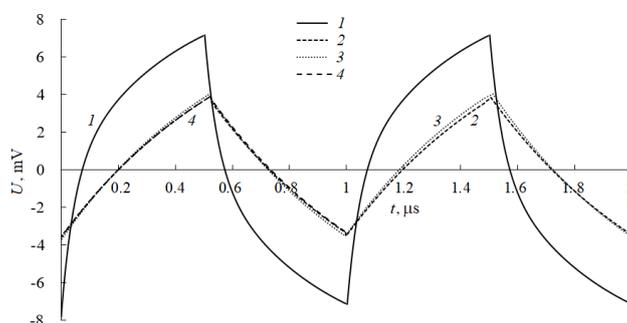


Рис. 2. Импульсные характеристики при частоте импульсов 1 МГц: 1 – образец №1, 2 – образец №2, 3 – образец №3, 4 – образец №4

Темновые стационарные вольтамперные характеристики (ВАХ), измеренные с помощью измерителя параметров полупроводниковых приборов ИППП-1 в диапазоне напряжений $-10 \text{ mV} < U < 10 \text{ mV}$, линейны для всех образцов. Эффективное сопротивление образцов $R_{ef} = dU/dI$, рассчитанное по темновым ВАХ в этом диапазоне напряжений, принимает значения: №1 – 5586 Ом, №2 – 32258 Ом, №3 – 26315 Ом, №4 – 41841 Ом.

В работах [14, 15] в результате анализа темновых ВАХ, измеренных в диапазоне $U < 0,6 \text{ В}$, определены значения шунтирующего сопротивления R_{sh} и сосредоточенного последовательного сопротивления R_s образцов. Значения R_{sh} : №1 – $1.2 \cdot 10^5 \text{ Ом}$, №2 – $1.1 \cdot 10^5 \text{ Ом}$, №3 – $1.1 \cdot 10^5 \text{ Ом}$, №4 – $3.4 \cdot 10^5 \text{ Ом}$. Значения R_s : №1 – 3.1 Ом, №2 – 2.3 Ом, №3 – 2.2 Ом, №4 – 4.2 Ом. Для всех образцов значения $R_{sh} > R_{ef}$, следовательно, время переключения для зависимостей $U(t)$ на рис. 2 не определяется токами утечки по торцам n - p перехода.

Переходные характеристики, измеренные по схеме рис. 1 без исследуемого образца, имеют время нарастания фронта импульса $t_{f0} = 27.5 \text{ ns}$. Если вместо исследуемых образцов включить омические сопротивления с $R = R_{ef}$, то время нарастания фронта импульса $t_f < t_{f0}$. Следовательно, значения времени переключения, характерные для рис. 2, не определяются параметрами измерительной цепи. Наблюдаемые переходные процессы обусловлены явлениями в n^+p -структурах.

Неравенство (1) позволяет полностью линеаризовать фундаментальную систему дифференциальных уравнений Шокли [16] транспорта носителей заряда в неоднородных полупроводниках в одномерном нестационарном случае.

В n области уравнение диффузионного распространения неравновесных дырок δp :

$$\frac{\partial \delta p}{\partial t} - D_p \frac{\partial^2 \delta p}{\partial x^2} + \frac{\delta p}{\tau_p} = 0, \quad (2)$$

где D_p – коэффициент диффузии, τ_p – время жизни дырок.

В p области уравнение диффузионного распространения неравновесных электронов δn :

$$\frac{\partial \delta n}{\partial t} - D_n \frac{\partial^2 \delta n}{\partial x^2} + \frac{\delta n}{\tau_n} = 0, \quad (3)$$

где D_n – коэффициент диффузии, τ_n – время жизни электронов.

В области пространственного заряда (ОПЗ) n - p перехода со стороны n слоя уравнение диффузионно-дрейфового распространения неравновесных дырок:

$$\frac{\partial \delta p}{\partial t} + \frac{e\mu_p}{\varepsilon\varepsilon_0} N_D^+ \left(\delta p + (x + d_n) \frac{\partial \delta p}{\partial x} \right) - D_p \frac{\partial^2 \delta p}{\partial x^2} + \frac{\delta p}{\tau_p} = 0, \quad (4)$$

где μ_p – подвижность дырок, N_D^+ – концентрация ионизованных доноров в этой области, ε – диэлектрическая проницаемость, ε_0 – диэлектрическая постоянная, d_n – левая граница ОПЗ.

В ОПЗ со стороны p области уравнение диффузионно-дрейфового распространения неравновесных электронов:

$$\frac{\partial \delta n}{\partial t} + \frac{e\mu_n}{\varepsilon\varepsilon_0} N_A^+ \left(\delta n + (x - d_p) \frac{\partial \delta n}{\partial x} \right) - D_n \frac{\partial^2 \delta n}{\partial x^2} + \frac{\delta n}{\tau_n} = 0, \quad (5)$$

где μ_n – подвижность электронов, N_A^+ – концентрация ионизованных акцепторов в этой области, d_p – правая граница ОПЗ. В уравнениях (4, 5) использовалась модель резкого n - p перехода [17].

Краевые условия для уравнений (2–5) учитывают рекомбинацию носителей заряда на внешних поверхностях, падение напряжения в n - p переходе. На границах ОПЗ решения в соседних областях сшиваются. Неравенство (1) позволяет линеаризовать задачу, поэтому частное решение уравнений (2–5) ищем в виде

$$\delta n(x, t) = A_n(x) - B_n(x) \cdot \exp\left(-\frac{t-t_0}{\tau}\right), \text{ при } t_0 \leq t \leq t_m, \quad (6)$$

$$\delta p(x, t) = A_p(x) - B_p(x) \cdot \exp\left(-\frac{t-t_0}{\tau}\right), \text{ при } t_0 \leq t \leq t_m, \quad (7)$$

где функции $A_n(x)$, $B_n(x)$, $A_p(x)$, $B_p(x)$ найдем, решая уравнений (2–5) с краевыми условиями методом разделения переменных. Суперпозиция решений вида (6), (7) с различными значениями τ также является решением уравнений (2–5), соответствующим условиям эксперимента. Полученные формулы громоздки, приводить их в статье не целесообразно.

Падение напряжения на сосредоточенном последовательном сопротивлении R_s много меньше U , поэтому падение напряжения в ОПЗ n - p перехода $U_{np}(t) \approx U(t)$. Совместность решений уравнений (2–5) и экспериментальных данных требует, чтобы

$$U_{np}(t) = A_1(a_1 e^{-(t-t_0)/\tau_1} + a_2 e^{-(t-t_0)/\tau_2}) - A_2, \quad (8)$$

где $t_0 \leq t \leq t_0 + 0.5 \cdot 10^{-6}$ s, t_0 – начало отсчета совпадает с фронтом изменения напряжения, постоянные A_1 , A_2 , a_1 , a_2 , τ_1 , τ_2 найдем, аппроксимируя экспериментальные зависимости, представленные на рис. 2.

Результаты исследования и их обсуждение. Результаты измерений обрабатывались с помощью программ Excel и Mathcad. Импульсные характеристики исследуемых структур с достаточной точностью аппроксимируются одно экспоненциальной зависимостью напряжения от времени ($a_2 = 0$ в (8)) для образцов №2, №3, №4, а для образца №1 – двух экспоненциальной зависимостью (8). В результате найдены следующие значения времени переключения для образцов: №2 – $\tau_1 = 6.6 \cdot 10^{-7}$ s, №3 – $\tau_1 = 6.3 \cdot 10^{-7}$ s, №4 – $\tau_1 = 6.4 \cdot 10^{-7}$ s. Для образца №1 найдены два значения $\tau_1 = 4.2 \cdot 10^{-7}$ s, $\tau_2 = 5.5 \cdot 10^{-8}$ s.

Для объяснения полученных значений времени переключения воспользуемся моделью образования первичных радиационных дефектов (ПРД) в кремнии [18–20]. На рис. 3 показаны рассчитанные распределения по глубине среднего числа ПРД: G_{Si} – междоузельного кремния, G_V – вакансий, G_W – дивакансий, созданных одним протоном на единице длины проективного пробега.

Количество ПРД, созданных в n^+ -слое на расстоянии 0.41 μm от поверхности протонами с $E_p = 40$ keV при $T_p = 83$ K и $T_p = 300$ K, отличается в несколько раз. Однако значения τ в облученных №2, №3 и необлученном №4 образце близки и, следовательно, не могут быть временем жизни дырок в n^+ -слое. Эти значения много меньше времени жизни электронов в базе p -типа. Поэтому считаем, что измеренные зависимости $U(t)$ определяются падением напряжения в ОПЗ n^+ - p перехода, а значения τ являются эффективным временем жизни носителей заряда в этой области.

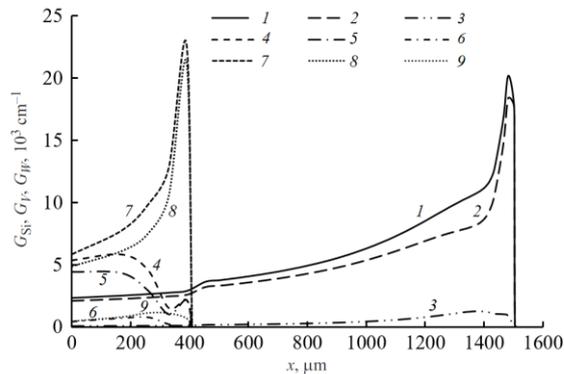


Рис. 3. Распределение среднего числа ПРД, созданных на единицу длины проективного пробега одним протоном: $E_p = 180 \text{ keV}$, $T_p = 83 \text{ K}$: 1 – G_{Si} , 2 – G_V , 3 – G_W ; $E_p = 40 \text{ keV}$, $T_p = 83 \text{ K}$: 4 – G_{Si} , 5 – G_V , 6 – G_W ; $E_p = 40 \text{ keV}$, $T_p = 300 \text{ K}$: 7 – G_{Si} , 8 – G_V , 9 – G_W

Протоны с $E_p = 180 \text{ keV}$ создают ПРД на глубине $1.51 \mu\text{m}$ во всей ОПЗ n^+ -р перехода (рис. 3). Следовательно, структура ОПЗ образца №1 изменилась, в ней существуют две области с различными значениями эффективного времени жизни τ_1 и τ_2 . Значение τ_2 относится к области с высокой концентрацией радиационных дефектов в окрестности пика Брэгга, расположенного при $x = 1.48 \mu\text{m}$ (рис. 3).

Заключение. В кремниевой структуре протоны с энергией 180 keV создают ПРД в слое толщиной $1.51 \mu\text{m}$, а протоны с энергией 40 keV – в слое толщиной $0.41 \mu\text{m}$. При температуре облучения образцов 83 K количество ПРД в пике распределения в конце проективного пробега протона с $E_p = 40 \text{ keV}$ оказывается много меньше, чем для протона с $E_p = 180 \text{ keV}$, что обусловлено различием процессов разделения пар Si_i , V в кремнии n - и p -типа проводимости. Облучение протонами с $E_p = 180 \text{ keV}$ изменяет физические свойства высоколегированного слоя n^+ -типа и всей ОПЗ n^+ -р перехода. Протоны с $E_p = 40 \text{ keV}$ изменяют свойства слоя n -типа, не воздействуя на ОПЗ, если глубина залегания n^+ -р перехода превышает среднюю длину проективного пробега протона. Эти выводы подтверждаются экспериментальными исследованиями вольтамперных характеристик и спектров пропускания кремниевых структур с n^+ -р переходом [14, 15].

Анализ импульсных характеристик показывает, что в результате облучения протонами с энергией 180 keV и дозой 10^{15} cm^{-2} изменились свойства ОПЗ n^+ -р перехода так, что образовались две области с различными значениями эффективного времени жизни $\tau_1 = 4.2 \cdot 10^{-7} \text{ s}$, $\tau_2 = 5.5 \cdot 10^{-8} \text{ s}$. Таким образом, время переключения снизилось на порядок по сравнению с необлученными образцами.

Благодарности. Авторы благодарят сотрудников ФГБУН Институт проблем технологии микроэлектроники и особо чистых материалов Российской академии наук Ю.А. Агафонова, В.И. Зиненко за облучение образцов.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Челябинский А.Р., Комаров Ф.Ф. Дефектно-примесная инженерия в имплантированном кремнии // Успехи физических наук. – 2003. – Т. 173, № 8. – С. 813-846.
2. Lebedev A.A., Davydov V.Yu., Smirnov A.N., Eliseyev I.A., Davydovskaya K.S., Zavarin E.E., Zakheim D.A., Lundin W.V., Nikolaev A.E., Sakharov A.V., Tsatsulnikov A.F., Zubov A.V., Kozlovski V.V. Proton irradiation effects on GaN-based epitaxial structures // Journal of Physics: Conference Series. – 2020. 1697, 012073. – P. 1-6.

3. Брудный В.Н. Радиационные эффекты в полупроводниках // Вестник Томского государственного университета. Серия "Физика". – 2005. – № 285: – С. 95-102.
4. Соболев Н.А. Инженерия дефектов в имплантационной технологии кремниевых светоизлучающих структур с дислокационной люминесценцией // Физика и техника полупроводников. – 2010. – Т. 44, В. 1. – С. 3-25.
5. Pes P.A. Evolution of space solar cells // Solar Energy Materials & Solar Cells. – 2001. – Vol. 68, I. 1. – P. 1-13.
6. Козлов В.А., Козловский В.В. Легирование полупроводников радиационными дефектами при облучении протонами и α -частицами // Физика и техника полупроводников. – 2001. – Т. 35, В. 7. – С. 769-795.
7. Губарев В., Семенов А., Сурма А., Столбунов В. Технология протонного облучения и возможности ее применения для улучшения характеристик силовых диодов и тиристоров // «Протон-Электротекс». – URL: https://proton-electrotex.com/files/project_52/15_Protonnoe_obluchenie_rus.pdf (дата публикации 29.11.2011).
8. Асадочиков В.Е., Дьячкова И.Г., Золотов Д.А., Кривонос Ю.С., Бублик В.Т., Шихов А.И. О связи протонного облучения и термической обработки монокристаллического кремния с его структурой // Известия вузов. Материалы электронной техники. – 2019. – Т. 22, № 1. – С. 18-26.
9. Anfimov I.M., Kobeleva S.P., Pylnev A.V., Schemerov I.V., Egorov D. S., Yurchuk S.Yu. On the Problem of Determining the Bulk Lifetime by Photoconductivity Decay on the Unpassivated Samples of Monocrystalline Silicon // Russian Microelectronics. – 2017. – Vol. 46, No. 8. – P. 585-590.
10. Koshelev O.G., Vasiljev N.G. Separate determination of the photoelectric parameters of $n^+p(n)-p^+$ silicon structure base region by noncontact method based on measurements of quantum efficiency relationships at two wavelengths // Modern Electronic Materials. – 2017. – Vol. 3, No. 3. – P. 127-130.
11. Bscheid C., Engst C.R., Eisele I., Kutter C. Minority Carrier Lifetime Measurements for Contactless Oxidation Process Characterization and Furnace Profiling // Materials. – 2019. – Vol. 12, No. 1. – P. 1-13.
12. Sam R., Zouma B., Zougmore F., Koalaga Z., Zougrana M., Zerbo I. 3D determination of the minority carrier lifetime and the p-n junction recombination velocity of a polycrystalline silicon solar cell // IOP Conf. Series: Materials Science and Engineering. – 2012. – Vol. 29, No. 012018. – P. 1-8.
13. Богатов Н.М., Григорьян Л.Р., Коваленко А.И., Коваленко М.С., Лунин Л.С. Импульсные характеристики кремниевых фотоэлектрических преобразователей, облученных низкоэнергетическими протонами // Письма в ЖТФ. – 2021. – Т. 47, В. 7. – С. 10-12.
14. Агафонов Ю.А., Богатов Н.М., Григорьян Л.Р., Зиненко В.И., Коваленко А.И., Коваленко М.С., Колоколов Ф.А. Влияние радиационных дефектов, созданных низкоэнергетическими протонами в сильнолегированном слое, на характеристики кремниевых n^+p-p^+ -структур // Поверхность. Рентгеновские, синхротронные и нейтронные исследования. – 2018. – № 10. – С. 86-91.
15. Богатов Н.М., Григорьян Л.Р., Коваленко А.И., Коваленко М.С., Колоколов Ф.А., Лунин Л.С. Влияние радиационных дефектов, созданных низкоэнергетическими протонами при температуре 83 К, на характеристики кремниевых фотоэлектрических структур // Физика и техника полупроводников. – 2020. – Т. 54, В. 2. – С. 144-149.
16. Shockley W. The theory of p-n junctions in semiconductors and p-n junction transistors // Bell Syst. Tech. J. – 1949. – Vol. 28, No. 7. – P. 435-439.
17. Pulfrey D.L. Understanding Modern Transistors and Diodes. – Cambridge University Press, 2010. – 335 p.
18. Богатов Н.М. Радиационные дефекты в кремнии, выращенном методом Чохральского // Поверхность. – 1999. – № 3. – С. 72-78.
19. Bogatov N.M., Kovalenko M.S. Calculation of Frenkel Pairs Separation, Formed in Silicon as a Result of Ionizing Particles Irradiation // AASCIT Journal of Physics. – 2017. – Vol. 3, I. 3. – P. 13-17.
20. Богатов Н.М., Григорьян Л.Р., Кленевский А.В., Коваленко М.С. Моделирование областей разупорядочения в процессе радиационного дефектообразования // Экологический вестник научных центров Черноморского экономического сотрудничества. – 2019. – Т. 16, № 1. – С. 59-65.

REFERENCES

1. Chelyadinskiy A.R., Komarov F.F. Defektno-primesnaya inzheneriya v implantirovannom kremnii [Defect-impurity engineering in implanted silicon], *Uspekhi fizicheskikh nauk* [Successes of Physical Sciences], 2003, Vol. 173, No. 8, pp. 813-846.
2. Lebedev A.A., Davydov V.Yu., Smirnov A.N., Eliseyev I.A., Davydovskaya K.S., Zavarin E.E., Zakheim D.A., Lundin W.V., Nikolaev A.E., Sakharov A.V., Tsatsulnikov A.F., Zubov A.V., Kozlovski V.V. Proton irradiation effects on GaN-based epitaxial structures, *Journal of Physics: Conference Series*, 2020, 1697, 012073, pp. 1-6.
3. Brudnyy V.N. Radiatsionnye efekty v poluprovodnikakh [Radiation effects in semiconductors], *Vestnik Tomskogo gosudarstvennogo universiteta. Seriya "Fizika"* [Bulletin of Tomsk State University. Series "Physics"], 2005, No. 285, pp. 95-102.
4. Sobolev N.A. Inzheneriya defektov v implantatsionnoy tekhnologii kremnievykh svetoizluchayushchikh struktur s dislokatsionnoy lyuminestsentsiyey [Engineering of defects in implantation technology of silicon light-emitting structures with dislocation luminescence], *Fizika i tekhnika polu-provodnikov* [Physics and technology of semi-conductors], 2010, Vol. 44, V. 1, pp. 3-25.
5. Iles P.A. Evolution of space solar cells, *Solar Energy Materials & Solar Cells*, 2001, Vol. 68, I. 1, pp. 1-13.
6. Kozlov V.A., Kozlovskiy V.V. Legirovanie poluprovodnikov radiatsionnymi defektami pri obluchanii protonami i α -chastitsami [Doping of semiconductors with radiation defects when irradiated with protons and α -particles], *Fizika i tekhnika poluprovodnikov* [Physics and Technology of semiconductors], 2001, Vol. 35, V. 7, pp. 769-795.
7. Gubarev V., Semenov A., Surma A., Stolbunov V. Tekhnologiya protonnogo oblucheniya i vozmozhnosti ee primeneniya dlya uluchsheniya kharakteristik silovykh diodov i tiristorov [Proton irradiation technology and the possibilities of its application to improve the characteristics of power diodes and thyristors], «Proton-Elektroteks». Available at: https://proton-electrotex.com/files/project_52/15_Protonnoe_obluchenie_rus.pdf (publication date 29 November 2011).
8. Asadchikov V.E., D'yachkova I.G., Zolotov D.A., Krivososov Yu.S., Bublik V.T., SHikhov A.I. O svyazi protonnogo oblucheniya i termicheskoy obrabotki monokristallicheskogo kremniya s ego strukturoy [On the connection of proton irradiation and heat treatment of monocrystalline silicon with its structure], *Izvestiya vuzov. Materialy elektronnoy tekhniki* [Izvestiya vuzov. Materials of electronic equipment], 2019, Vol. 22, No. 1, pp. 18-26.
9. Anfimov I.M., Kobeleva S.P., Pylnev A.V., Schemerov I.V., Egorov D. S., Yurchuk S.Yu. On the Problem of Determining the Bulk Lifetime by Photoconductivity Decay on the Unpassivated Samples of Monocrystalline Silicon, *Russian Microelectronics*, 2017, Vol. 46, No. 8, pp. 585-590.
10. Koshelev O.G., Vasiljev N.G. Separate determination of the photoelectric parameters of $n^+p(n)-p^+$ silicon structure base region by noncontact method based on measurements of quantum efficiency relationships at two wavelengths, *Modern Electronic Materials*, 2017, Vol. 3, No. 3, pp. 127-130.
11. Bscheid C., Engst C.R., Eisele I., Kutter C. Minority Carrier Lifetime Measurements for Contactless Oxidation Process Characterization and Furnace Profiling, *Materials*, 2019, Vol. 12, No. 1, pp. 1-13.
12. Sam R., Zouma B., Zougmore F., Koalaga Z., Zoungrana M., Zerbo I. 3D determination of the minority carrier lifetime and the p-n junction recombination velocity of a polycrystalline silicon solar cell, *IOP Conf. Series: Materials Science and Engineering*, 2012, Vol. 29, No. 012018, pp. 1-8.
13. Bogatov N.M., Grigor'yan L.R., Kovalenko A.I., Kovalenko M.S., Lunin L.S. Impul'snye kharakteristiki kremnievykh fotoelektricheskikh preobrazovateley, obluchennykh nizkoenergeticheskimi protonami [Pulse characteristics of silicon photovoltaic converters irradiated with low-energy protons], *Pis'ma v ZhTF* [Letters to the Journal of Technical Physics], 2021, Vol. 47, V. 7, pp. 10-12.
14. Agafonov Yu.A., Bogatov N.M., Grigor'yan L.R., Zinenko V.I., Kovalenko A.I., Kovalenko M.S., Kolokolov F.A. Vliyanie radiatsionnykh defektov, sozdannykh nizkoenergeticheskimi protonami v sil'nolegirovannom sloe, na kharakteristiki kremnievykh n^+p-p^+ -struktur [The effect of radiation defects created by low-energy protons in a highly alloyed layer on the characteristics of silicon n^+p-p^+ structures], *Poverkhnost'. Rentgenovskie, sinkhrotronnye i neytronnye issledovaniya* [Surface. X-ray, synchrotron and neutron studies], 2018, No. 10, pp. 86-91.

15. *Bogatov N.M., Grigor'yan L.R. Kovalenko A.I., Kovalenko M.S., Kolokolov F.A., Lunin L.S.* Vliyaniye radiatsionnykh defektov, sozdannykh nizkoenergeticheskimi protonami pri temperature 83 K, na kharakteristiki kremnievykh fotoelektricheskikh struktur [The influence of radiation defects created by low-energy protons at a temperature of 83 K on the characteristics of silicon photovoltaic structures], *Fizika i tekhnika poluprovodnikov* [Physics and technology of semiconductors], 2020, Vol. 54, V. 2, pp. 144-149.
16. *Shockley W.* The theory of p-n junctions in semiconductors and p-n junction transistors, *Bell Syst. Tech. J.*, 1949, Vol. 28, No. 7, pp. 435-439.
17. *Pulfray D.L.* Understanding Modern Transistors and Diodes. Cambridge University Press, 2010, 335 p.
18. *Bogatov N.M.* Radiatsionnye defekty v kremnii, vyrashchennom metodom Chokhralskogo [Radiation defects in silicon grown by the Chokhralsky method], *Poverkhnost'* [Surface], 1999, No. 3, pp. 72-78.
19. *Bogatov N.M., Kovalenko M.S.* Calculation of Frenkel Pairs Separation, Formed in Silicon as a Result of Ionizing Particles Irradiation, *AASCIT Journal of Physics*, 2017, Vol. 3, I. 3, pp. 13-17.
20. *Bogatov N.M., Grigor'yan L.R., Klenevskiy A.V., Kovalenko M.S.* Modelirovaniye oblastey razuporyadocheniya v protsesse radiatsionnogo defektoobrazovaniya [Modeling of areas of disordering in the process of radiation defect formation], *Ekologicheskyy vestnik nauchnykh tsentrov Chernomorskogo ekonomicheskogo sotrudnichestva* [Ecological Bulletin of the scientific centers of the Black Sea Economic Cooperation], 2019, Vol. 16, No. 1, pp. 59-65.

Статью рекомендовал к опубликованию д.т.н., профессор А.С. Магомадов.

Богатов Николай Маркович – Кубанский государственный университет; e-mail: bogatov@phys.kubsu.ru; г. Краснодар, Россия; тел.: +79034513106; д.ф.-м.н.; профессор.

Володин Владимир Сергеевич – e-mail: volodinvs1995@mail.ru; тел.: +79898203672; аспирант.

Григорьян Леонтий Рустемович – e-mail: leonmezon@mail.ru; тел.: +79181681719; к.ф.-м.н.; доцент.

Коваленко Алексей Иванович – e-mail: physinf@phys.kubsu.ru. тел.: 88612199501*266; аспирант,

Коваленко Максим Сергеевич – e-mail: m.s.kovalenko@ya.ru; тел.: +79184694954; к.ф.-м.н.; доцент.

Bogatov Nikolay Markovich – Kuban State University; e-mail: bogatov@phys.kubsu.ru; Krasnodar, Russia; phone: +79034513106; dr. of phys. and math.; professor.

Volodin Vladimir Sergeevich – e-mail: volodinvs1995@mail.ru; phone: +79898203672; graduate student.

Grigoryan Leontiy Rustemovich – e-mail: leonmezon@mail.ru; phone: +79181681719; cand. of phys. and math. sc.; associate professor.

Kovalenko Alexey Ivanovich – e-mail: physinf@phys.kubsu.ru; phone: +78612199501*266; graduate student.

Kovalenko Maxim Sergeevich – e-mail: m.s.kovalenko@ya.ru; phone: +79184694954; cand. of phys. and math. sc.; associate professor.

Р.Р. Ибадов, В.П. Федосов, С.Р. Ибадов

АНАЛИЗ ПОДСТИЛАЮЩЕЙ ПОВЕРХНОСТИ ПРИ ФОРМИРОВАНИИ ИЗОБРАЖЕНИЯ В РЕЖИМЕ ДОПЛЕРОВСКОГО ОБУЖЕНИЯ ЛУЧА

Радар на основе сканирования с реальным лучом широко используется как в гражданской, так и в военной сфере. Однако трудно реализовать высокое разрешение по азимуту стационарной платформы или платформы с неравномерным движением с помощью традиционных алгоритмов обработки сигналов. Технология Доплеровского обужения луча (ДОЛ) представляет собой сочетание высокого разрешения и производительностью в реальном времени по сравнению с технологией РЛС с синтезированной апертурой (САР), которая вдоль азимутального направления использует доплеровский сдвиг между эхо-сигналами от объектов на подстилающей поверхности, вызванный движением радиолокационной платформы. К сожалению, традиционный алгоритм визуализации ДОЛ, конструирующий доплеровский фильтр с помощью БПФ, имеет низкое разрешение по азимуту и высокий уровень боковых лепестков, что ограничивает дальнейшее улучшение разрешения по азимуту. В статье исследован алгоритм построения карты подстилающей поверхности в направлении движения носителя РЛС на основе ДОЛ и проведен анализ изображения карты с помощью преобразования Фурье. Показан трехмерный вид карты подстилающей поверхности с распределением значений на изображениях. Предметом исследования являются метод и алгоритм построения карты подстилающей поверхности в режиме доплеровского обужения луча и выявление цепочных структур на основе анализа преобразования Фурье. Объектом исследования является набор тестовых изображений карты местности. Результатом исследования является разработка алгоритма построения карты с целью выявления цепочных структур на подстилающей поверхности. Новизной работы является алгоритм, позволяющий построить карту подстилающей поверхности на основе ДОЛ с учётом слепой зоны в направлении движения носителя РЛС. Полученные результаты позволяют выявлять также цепочные структуры в интересующей области. Проверена возможность оценки периодичности элементов изображения с использованием Фурье-преобразования. В результате решения сформулированных задач можно сделать следующие выводы: – разработан алгоритм построения карты подстилающей поверхности на основе ДОЛ с коррекцией изображения в направлении движения носителя РЛС; – анализ результатов проведённого исследования показал, что предложенный алгоритм позволяет выявить цепочные структуры на карте местности.

Доплеровское обужение луча; разрешение; карта местности; картографирование; преобразование Фурье; модуляция.

R.R. Ibadov, V.P. Fedosov, S.R. Ibadov

ANALYSIS OF UNDERLYING SURFACE IN IMAGE FORMATION IN DOPPLER BEAM SHARPENING MODE

Radar based on real beam scanning is widely used in both civil and military spheres. However, it is difficult to realize high azimuth resolution of a stationary platform or a platform with non-uniform motion using conventional signal processing algorithms. Doppler beam sharpening (DBS) technology is a combination of high resolution and real-time performance compared to Synthetic Aperture Radar (SAR) technology, which uses the Doppler shift between echoes from objects on the underlying surface along the azimuth direction, caused by the movement of the radar platform. Unfortunately, the traditional DBS imaging algorithm, which constructs a Doppler filter using an FFT, has a low azimuth resolution and a high level of side lobes, which limits further improvement in azimuth resolution. In the article, the algorithm for constructing a map of the underlying surface in the direction of movement of the radar carrier based on the DBS was studied and the map image was analyzed using the Fourier transform. A three-dimensional view of the map of the underlying surface is shown with the distribution of values in the images. The subject of the study is the method and algorithm for constructing a map of the underlying surface in the Doppler beam sharpening mode and identifying chain structures based on the analysis of the Fourier transform. The object of the study is

a set of test images of the terrain map. The result of the study is the development of an algorithm for constructing a map in order to identify chain structures on the underlying surface. The novelty of the work is an algorithm that allows you to build a map of the underlying surface based on the DBS, taking into account the blind zone in the direction of movement of the radar carrier. The results obtained also make it possible to reveal chain structures in the region of interest. The possibility of estimating the periodicity of image elements using the Fourier transform has been tested. As a result of solving the tasks set, the following conclusions can be drawn: – an algorithm has been developed for constructing a map of the underlying surface based on DBS with image correction in the direction of movement of the radar carrier. – analysis of the results of the study showed that the proposed algorithm allows you to identify chain structures on the map.

Doppler beam sharpening; resolution; local map; mapping; Fourier transform; modulation.

Введение. В сложной оперативной обстановке городской инфраструктуры беспилотным летательным аппаратам необходимо контролировать его ориентацию относительно объектов на подстилающей поверхности [1–5]. Отображение наземных объектов и обнаружение медленно движущихся целей – две важные функции бортовой РЛС управления класса "воздух-земля". Технология доплеровского облучения луча (ДОЛ) [6] может заставить бортовой радар управления реализовать быстрое изображение земли с небольшими вычислениями в условиях широкого сканирования антенны. Для наземной индикации движущихся целей [7] во избежание конкуренции дальних слабых целей с сильными помехами ближнего действия, вызванными неоднозначностью дальности, обычно используется низкая частота повторения импульсов в режиме «воздух-земля», для бортовой РЛС управления, антенна работает в режиме переднего сектора обзора. Из-за расширения и смещения луча отражения от главного лепестка при движении платформы наземные медленно движущиеся цели легко попадают в помехи от главного лепестка в доплеровской области. Помехи от главного лепестка становятся основным фактором, ограничивающим работу радара. В этом случае даже антенна со сверхмалыми боковыми лепестками не может быть эффективной для улучшения характеристик обнаружения системы обработки, такой как каскадная пространственно-временная обработка (CST). Для решения этой проблемы проводятся обширные исследования по подавлению помех бортовой подвижной платформы [8].

Обработка радиолокационного сигнала – это метод, используемый для извлечения информации из эхо-сигналов от объектов на поверхности. Эхо-сигналы, принимаемые приемником РЛС, обрабатываются в цифровом формате. Радары широко классифицируются на основе разрешающей способности [9].

Разрешение можно определить как способность радара различать близко расположенные цели. Разрешение по дальности и разрешение по азимуту определяются для обнаружения близко расположенных целей по дальности и углу соответственно. Точное разрешение диапазона может быть получено методом сжатия импульсов. Сжатие импульсов включает использование длинного импульса с внутренней модуляцией для получения высокой энергии, что в конечном итоге дает разрешение короткого импульса [10].

Метод повышения резкости доплеровского луча (ДОЛ) может быть выражен как комбинированный продукт теории антенн и цифровой обработки сигналов [11]. Разрешение по азимуту улучшается методом ДОЛ. В методе повышения резкости доплеровского луча полученные сигналы сжимаются импульсами для получения желаемого разрешения [12–14]. Согласованные фильтры представляют собой линейные сети, которые используются для увеличения пикового отношения сигнал-шум (SNR) сигнала, что, в свою очередь, увеличивает вероятность обнаружения цели. ДОЛ использует преимущество поступательного движения радара для создания эквивалентной большой антенны [15]. Увеличение виртуальной длины антенны обеспечивает увеличение разрешающей способности в угловом направлении.

На рис. 1 самолет движется вдоль направления оси X с фиксированной скоростью V на высоте H . Луч радара сканирует область изображения по часовой стрелке с угловой скоростью ω . R_0 представляет собой начальное расстояние между радаром самолета и целью, расположенной в точке $P(x, y)$. В момент времени t дальность может быть выражена как:

$$R(x, y, t) = \sqrt{R_0^2 + (vt)^2 - 2R_0 \cos \theta_0 \cos \varphi_0 vt}, \quad (1)$$

где φ_0 и θ_0 – обозначают начальный угол наклона и поперечный угол между направлением максимума характеристики направленности бортовой сканирующей РЛС и объектом $P(x, y)$. По геометрической конструкции может быть выведено: $\cos \alpha_0 = \cos \varphi_0 \cos \theta_0$.

Поскольку время сканирования луча относительно направления на точечный объект невелико, дальность может быть предварительно обработана на основе следующего выражения, с помощью разложения в ряд Тейлора [21]:

$$R(x, y, t) \cong R_0 - \cos \theta_0 \cos \varphi_0 vt. \quad (2)$$

Радиолокационная система передает сигналы с линейной частотной модуляцией (ЛЧМ) для достижения высокого разрешения по дальности с помощью метода сжатия импульсов. Для области изображения Ω , эхо-сигнал после процесса преобразования с понижением частоты может быть выражен как:

$$y(\tau, t) = \sum_{(\tau, t) \in \Omega} \sigma(\tau, t) A(\theta) \operatorname{rect} \left[\frac{\tau - \tau_d}{T_r} \right] \cdot \exp(j\pi K_r [\tau - \tau_d]^2) \cdot \exp(-j2\pi f_0 \tau_d), \quad (3)$$

где τ и t – быстрое и медленное время; σ – представляет собой коэффициент обратного рассеяния целей; $A(\theta)$ – функция модуляции диаграммы направленности антенны; τ_d и T_r – обозначают задержку эхо-сигнала и длительность передаваемого сигнала.

Затем после процесса сжатия импульса выполняется метод коррекции смещения диапазона, чтобы реализовать высокое разрешение по диапазону, и эхо-сигнал может быть преобразован как [23]:

$$y(\tau, t) = \sum_{(\tau, t) \in \Omega} \sigma(\tau, t) A(\theta) \sin c \left(B \left[\tau - \frac{2R_0}{c} \right] \right) \cdot \exp \left(-j \frac{2\pi f_0}{c} R(t) \right), \quad (4)$$

где B и c – ширина полосы пропускания и скорость света. Анализ эхо-сигнала в дискретной форме может быть выражен следующим образом:

$$y(m, n) = \sum_{k=0}^{K-1} \sigma_k(m) \exp(jnf_k) + e(m, n), \quad (5)$$

где $y(m, n)$ – эхо-сигнал n -го азимута и m -го интервала дальности; $m = 1, \dots, M$; $n = 1, \dots, N$. K – точки дискретизации в доплеровской частотной области; $f_k = \frac{2\pi k}{K}$ – нормированная доплеровская частота. $k = 0, \dots, K-1$; $\sigma_k(m)$ – m -й интервал дальности означает коэффициент отражения в k -ой точке доплеровской частоты; $e(m, n)$ – аддитивный гауссовский белый шум.

Для объекта, расположенного по азимуту θ и углу наклона φ , доплеровская центральная частота равна

$$f_d = \frac{2v \cos \theta \cos \varphi}{\lambda}, \quad (6)$$

где λ – длина волны излучаемого сигнала.

Можно обнаружить, что градиент центральной доплеровской частоты на различных азимутах способствует формированию изображения карты по азимуту. Далее приводится результат моделирования предложенным алгоритмом на тестовых изображениях подстилающей поверхности.

Экспериментальные результаты. На рис. 2 представлен результат моделирования карты подстилающей поверхности в режиме ДОЛ [24].

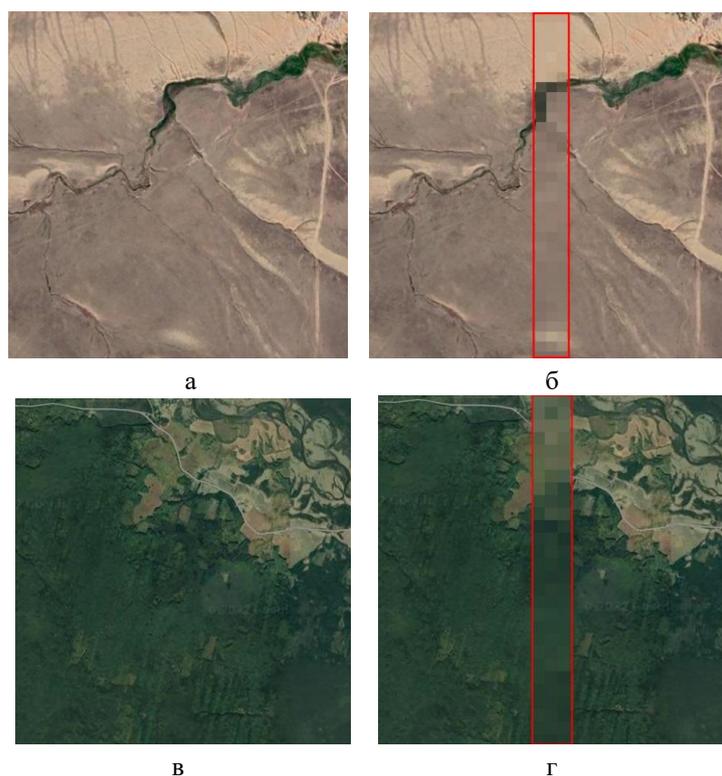


Рис. 2. Результат моделирования карты подстилающей поверхности: а и в – оригинальные изображения; б и г – формирование в режиме ДОЛ с выделенной красными линиями слепой зоной

Из анализа изображений на рис. 2,б,г можно рассмотреть потери в формировании изображения карты подстилающей поверхности в направлении движения носителя РЛС [25].

На рис. 3 представлен прологарифмированный спектр изображения со смещенным центром: изменены расположения низкой и высокой частот изменения изображения вдоль осе координат, а логарифмирование применено для удобства отображения результатов исследования [26].

Линии на рис. 3,а показывают повторяющиеся объекты (элементы разрешения) на изображении карты поверхности (см. рис 3,г). Углы наклона линий на Фурье-образе изображения карты (см. рис 3,е) соответствуют направлениям дорог на поверхности. Фурье-образ имеет элементы сходства со спектром гармонического колебания.

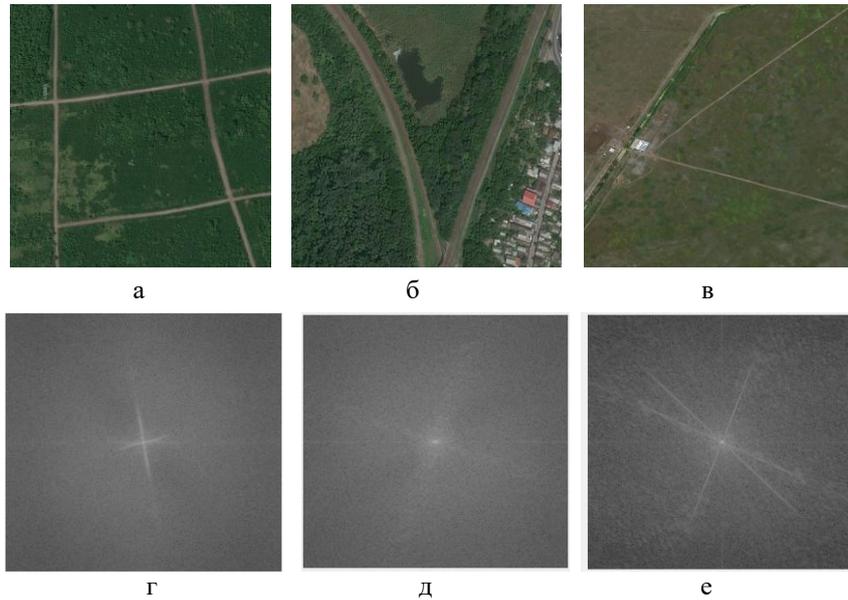


Рис. 3. Результат прологарифмированного спектра изображения карты подстилающей поверхности: а, б, в – оригинальные изображения; г, д, е – Фурье-образ оригинальных изображений

Трехмерный вид изображений карты подстилающей поверхности представлены на рис. 4 для прологарифмированного и не прологарифмированного спектра изображения карты подстилающей поверхности изображений соответственно на рис. 2,б,г.

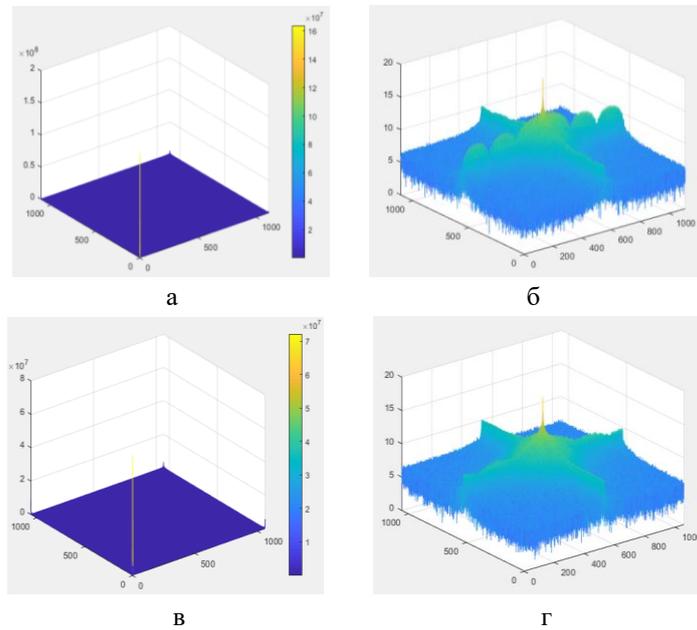


Рис. 4. Результат трехмерного вида изображений карты подстилающей поверхности спектра изображения карты подстилающей поверхности: а, в – для не прологарифмированных массивов; б, г – для прологарифмированных массивов

Графики на рис. 4,а,в показывают 3D-изображение не прологарифмированной карты подстилающей поверхности рис. 2,б,г. Графики на рис. 4,б,г показывают 3D-изображения прологарифмированной карты подстилающей поверхности рис. 2,б,г. На графиках видно что чем выше значения тем ближе к желтому, чем меньше, тем ближе к синему.

Заключение. В работе представлен алгоритм моделирования карты подстилающей поверхности в направлении движения носителя РЛС на основе ДОЛ и выявление периодичности объектов на изображениях с помощью преобразования Фурье. Представлены результаты Фурье образа прологарифмированного и не прологарифмированного спектра изображения со смещенным центром. В ходе экспериментов выявлено, что Фурье-образ имеет элементы сходства со спектром гармонического колебания.

Исследование выполнено при поддержке гранта Российского научного фонда № 22-29-01389 от 21.12.2021 г. в Южном федеральном университете.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Белый Ю.И., Таганцев В.А.* Радиолокационный прицельный комплекс Н001: модернизация продолжается // Радиотехника. – 2005. – №. 2. – С. 28-29.
2. *Матвеев А.М.* Построение модели и предобработка изображения подстилающей поверхности для радиолокационных систем с доплеровским обужением луча на основе информации, получаемой о поверхности в оптическом диапазоне. – М.: «Электронный журнал», 2004.
3. *Величкин А.И., Карпов О.А., Таланцев В.В., Толстов Е.Ф.* Повышение разрешающей способности авиационной РЛС при наблюдении вперед // Радиотехника. – 1998. – № 12. – С. 12.
4. *Козаев А.А., Колтышев Е.Е., Фролов А.Ю., Янковский В.Т.* Алгоритм доплеровского измерения скорости в РЛС с синтезированной апертурой // Радиотехника. – 2005. – №. 6. – С. 13-16.
5. *Антипов В.Н., Суляков Д.Ю.* Картографирование и обнаружение наземных движущихся целей // Радиотехника. – 2005. – №. 6. – С. 10-12.
6. *Витязев В.В., Колодько Г.Н., Витязев С.В.* Способы и алгоритмы формирования радиолокационного изображения в режиме доплеровского обужения луча // Цифровая обработка сигналов. – 2006. – №. 3. – С. 31-41.
7. *Витязев В.В., Витязев С.В.* Методы синтеза узкополосного адаптивного КИХ-фильтра на основе многоскоростной обработки // Цифровая обработка сигналов. – 2007. – №. 4. – С. 13.
8. *Qi L., Zheng M., Yu W., et al.* Super-resolution Doppler beam sharpening imaging based on an iterative adaptive approach // Remote sensing letters. – 2016. – Vol. 7, No. 3. – P. 259-268.
9. *Mao D., Zhang, Y., Kang, Y., Zhang, Y., et al.* Doppler beam sharpening imaging based on beam-recursion fiaa // IEEE Radar Conference (RadarConf). – IEEE, 2019. – P. 1-4.
10. *Xue M., Xu L., Li J.* IAA spectral estimation: fast implementation using the Gohberg–Semencul factorization // IEEE Transactions on Signal Processing. – 2017. – Vol. 59, No. 7. – P. 3251-3261.
11. *Yang H., Mao, D., Zhang, Y., Zhang, Y., Huang, Y., et al.* Doppler beam sharpening imaging based on fast iterative adaptive approach // IEEE Radar Conference (RadarConf). – IEEE, 2017. – P. 1419-1423.
12. *Zhang Y., Jakobsson A., Yang J.* Range-recursive IAA for scanning radar angular super-resolution // IEEE Geoscience and Remote Sensing Letters. – 2017. – Vol. 14, No. 10. – P. 1675-1679.
13. *Mao D., Zhang, Y., Zhang, Y., Huang, Y., et al.* Super-resolution Doppler beam sharpening method using fast iterative adaptive approach-based spectral estimation // Journal of Applied Remote Sensing. – 2018. – Vol. 12, No. 1. – P. 015020.
14. *Huang Y., Zha, Y., Zhang, Y., et al.* Real-beam scanning radar angular super-resolution via sparse deconvolution // IEEE Geoscience and Remote Sensing Symposium. – IEEE – 2014. – P. 3081-3084.
15. *Zhang Y., Jakobsson, A., Zhang, Y., et al.* Wideband sparse reconstruction for scanning radar // IEEE Transactions on Geoscience and Remote Sensing. – 2018. – Vol. 56, No. 10. – P. 6055-6068.
16. *Karlsson J., Rowe, W., Xu, L., et al.* Fast missing-data IAA with application to notched spectrum SAR // IEEE Transactions on Aerospace and Electronic Systems. – 2014. – Vol. 50, No. 2. – P. 959-971.

17. Paulus A.S., Melvin W.L., Williams D.B. Multichannel GMTI techniques to enhance integration of temporal signal energy for improved target detection // IET Radar, Sonar & Navigation. – 2017. – Vol. 11, No. 3. – P. 395-403.
18. Li X., Feng D., Liu H.W., et al. Dimension-reduced space-time adaptive clutter suppression algorithm based on lower-rank approximation to weight matrix in airborne radar // IEEE Transactions on Aerospace and Electronic Systems. – 2014. – Vol. 50, No. 1. – P. 53-69.
19. Xu J., Wang, C., Liao, G., et al. Sum and difference beamforming for angle-doppler estimation with STAP-based radars // IEEE Transactions on Aerospace and Electronic Systems. – 2016. – Vol. 52, No. 6. – P. 2825-2837.
20. Schleher D.C. MTI and pulsed Doppler radar with MATLAB. – Artech House, 2010.
21. Li M., Wei H., Sun J., et al. Keeping sharpening ratio constant for DBS of airborne mechanic scanning radar. – 2009. – P. 497-497.
22. Fedosov V.P., Ibadov R.R., Ibadov S.R. Restoration of the Lost Area of the Underlying Surface Image Using the Saliency Map // Radiation and Scattering of Electromagnetic Waves (RSEMW). – IEEE, 2021. – P. 447-450.
23. Ибадов Р.Р., Федосов В.П., Ибадов С.Р. Восстановление дефектов и слепой зоны на изображениях подстилающей поверхности для бортовых радиолокационных систем картографирования на основе Доплеровского обужения луча // Известия ЮФУ. Технические науки. – 2020. – № 6 (216). – С. 41-52.
24. Ibadov R.R., Gapon N.V., Ibadov S.R., et al. Image reconstruction using the modified texture synthesis algorithm // IOP Conference Series: Materials Science and Engineering. – IOP Publishing, 2021. – Vol. 1029, No. 1. – P. 012117.
25. Ibadov R.R., Fedosov V.P., Ibadov S.R. The method of spatial-temporal reconstruction of dynamic images based on a geometric model with contour and texture analysis // IOP Conference Series: Materials Science and Engineering. – IOP Publishing, 2021. – Vol. 1029, No. 1. – P. 012093.
26. Ибадов Р.Р., Федосов В.П., Воронин В.В., Ибадов С.Р. Исследование метода синтеза текстур изображений поверхности земли на основе нейронной сети // Известия ЮФУ. Технические науки. – 2019. – № 5. – С. 16-25.

REFERENCES

1. Bely Yu.I., Tagantsev V.A. Radiolokatsionnyy pritsel'nyy kompleks N001: modernizatsiya prodolzhaetsya [Radar sighting system N001: modernization continues], *Radiotekhnika* [Radio engineering], 2005, No. 2, pp. 28-29.
2. Matveev A.M. Postroenie modeli i predobrabotka izobrazheniya podstilayushchey poverkhnosti dlya radiolokatsionnykh sistem s doplerovskim obuzheniem luchy na osnove informatsii, poluchaemoy o poverkhnosti v opticheskom diapazone [Construction of a model and pre-processing of the image of the underlying surface for radar systems with Doppler beam narrowing based on information obtained about the surface in the optical range]. Moscow: «Elektronnyy zhurnal», 2004.
3. Velichkin A.I., Karpov O.A., Talantsev V.V., Tolstov E.F. Povyshenie razreshayushchey sposobnosti aviatsionnoy RLS pri nablyudenii vpered [Improving the resolution of an aviation radar during forward observation], *Radiotekhnika* [Radio engineering], 1998, No. 12, pp. 12.
4. Kozaev A.A., Koltyshev E.E., Frolov A.Yu., Yankovskiy V.T. Algoritm doplerovskogo izmereniya skorosti v RLS s sintezirovannoy aperturoy [Algorithm for Doppler velocity measurement in synthetic aperture radars], *Radiotekhnika* [Radio engineering], 2005, No. 6, pp. 13-16.
5. Antipov V.N., Suslyakov D.Yu. Kartografirovaniye i obnaruzheniye nazemnykh dvizhushchikhsya tseley [Mapping and detection of ground moving targets], *Radiotekhnika* [Radio engineering], 2005, No. 6, pp. 10-12.
6. Vityazev V.V., Kolod'ko G.N., Vityazev S.V. Sposoby i algoritmy formirovaniya radiolokatsionnogo izobrazheniya v rezhime doplerovskogo obuzheniya luchy [Methods and algorithms for the formation of a radar image in the mode of Doppler beam narrowing], *Tsifrovaya obrabotka signalov* [Digital signal processing], 2006, No. 3, pp. 31-41.
7. Vityazev V.V., Vityazev S.V. Metody sinteza uzkopolosnogo adaptivnogo KIKh-fil'tra na osnove mnogoskorostnoy obrabotki [Synthesis methods for a narrow-band adaptive FIR filter based on multi-rate processing], *Tsifrovaya obrabotka signalov* [Digital signal processing], 2007, No. 4, pp. 13.
8. Qi L., Zheng M., Yu W., et al. Super-resolution Doppler beam sharpening imaging based on an iterative adaptive approach, *Remote sensing letters*, 2016, Vol. 7, No. 3, pp. 259-268.

9. Mao D., Zhang, Y., Kang, Y., Zhang, Y., et al. Doppler beam sharpening imaging based on beam-recursion fiao, *IEEE Radar Conference (RadarConf)*. IEEE, 2019, pp. 1-4.
10. Xue M., Xu L., Li J. IAA spectral estimation: fast implementation using the Gohberg–Semencul factorization, *IEEE Transactions on Signal Processing*, 2017, Vol. 59, No. 7, pp. 3251-3261.
11. Yang H., Mao, D., Zhang, Y., Zhang, Y., Huang, Y., et al. Doppler beam sharpening imaging based on fast iterative adaptive approach, *IEEE Radar Conference (RadarConf)*. IEEE, 2017, pp. 1419-1423.
12. Zhang Y., Jakobsson A., Yang J. Range-recursive IAA for scanning radar angular super-resolution, *IEEE Geoscience and Remote Sensing Letters*, 2017, Vol. 14, No. 10, pp. 1675-1679.
13. Mao D., Zhang, Y., Zhang, Y., Huang, Y., et al. Super-resolution Doppler beam sharpening method using fast iterative adaptive approach-based spectral estimation, *Journal of Applied Remote Sensing*, 2018, Vol. 12, No. 1, pp. 015020.
14. Huang Y., Zha, Y., Zhang, Y., et al. Real-beam scanning radar angular super-resolution via sparse deconvolution, *IEEE Geoscience and Remote Sensing Symposium*. IEEE, 2014, pp. 3081-3084.
15. Zhang Y., Jakobsson, A., Zhang, Y., et al. Wideband sparse reconstruction for scanning radar, *IEEE Transactions on Geoscience and Remote Sensing*, 2018, Vol. 56, No. 10, pp. 6055-6068.
16. Karlsson J., Rowe, W., Xu, L., et al. Fast missing-data IAA with application to notched spectrum SAR, *IEEE Transactions on Aerospace and Electronic Systems*, 2014, Vol. 50, No. 2, pp. 959-971.
17. Paulus A.S., Melvin W.L., Williams D.B. Multichannel GMTI techniques to enhance integration of temporal signal energy for improved target detection, *IET Radar, Sonar & Navigation*, 2017, Vol. 11, No. 3, pp. 395-403.
18. Li X., Feng D., Liu H.W., et al. Dimension-reduced space-time adaptive clutter suppression algorithm based on lower-rank approximation to weight matrix in airborne radar, *IEEE Transactions on Aerospace and Electronic Systems*, 2014, Vol. 50, No. 1, pp. 53-69.
19. Xu J., Wang, C., Liao, G., et al. Sum and difference beamforming for angle-doppler estimation with STAP-based radars, *IEEE Transactions on Aerospace and Electronic Systems*, 2016, Vol. 52, No. 6, pp. 2825-2837.
20. Schleher D.C. MTI and pulsed Doppler radar with MATLAB. Artech House, 2010.
21. Li M., Wei H., Sun J., et al. Keeping sharpening ratio constant for DBS of airborne mechanic scanning radar, 2009, pp. 497-497.
22. Fedosov V.P., Ibadov R.R., Ibadov S.R. Restoration of the Lost Area of the Underlying Surface Image Using the Saliency Map, *Radiation and Scattering of Electromagnetic Waves (RSEMW)*. IEEE, 2021, pp. 447-450.
23. Ibadov R.R., Fedosov V.P., Ibadov S.R. Vosstanovlenie defektov i slepoy zony na izobrazheniyakh podstilayushchey poverkhnosti dlya bortovykh radiolokatsionnykh sistem kartografirovaniya na osnove Doplerovskogo obuzheniya lucha [Restoration of defects and a blind zone on images of the underlying surface for airborne radar mapping systems based on Doppler beam sharpening], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2020, No. 6 (216), pp. 41-52.
24. Ibadov R.R., Gapon N.V., Ibadov S R., et al. Image reconstruction using the modified texture synthesis algorithm, *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2021, Vol. 1029, No. 1, pp. 012117.
25. Ibadov R.R., Fedosov V.P., Ibadov S. R. The method of spatial-temporal reconstruction of dynamic images based on a geometric model with contour and texture analysis, *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2021, Vol. 1029, No. 1, pp. 012093.
26. Ibadov R.R., Fedosov V.P., Voronin V.V., Ibadov S.R. Issledovanie metoda sinteza tekstur izobrazheniy poverkhnosti zemli na osnove neyronnoy seti [Study of the method for synthesizing textures of images of the earth's surface based on a neural network], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2019, No. 5, pp. 16-25.

Статью рекомендовал к опубликованию д.т.н. В.И. Марчук.

reduces the energy efficiency of the communication system. The aim of the work was to increase the efficiency of using the channel for radio relay and tropospheric communication by using OFDM (orthogonal frequency-division multiplexing) signals in the system using adaptive coding and modulation. In the course of execution, the modulator and demodulator models of the OFDM signal are implemented. When using various signal code structures in various reception/transmission conditions, it is possible to achieve optimal use of the frequency and energy resources, to create systems that adapt to the conditions of signal propagation. To implement this mechanism, a service field was introduced into the transmitted service data, which contains information about the code rate used, the modulation type, and the interleaving depth. This approach allows optimizing the use of energy and frequency resources. Together with the use of channel quality estimation algorithms, it becomes possible to dynamically change the signal-code structure when the reception conditions change. By adjusting the interleaving depth, it is possible to optimize the S/N threshold or the amount of information delay in the channel, depending on the system requirements. The use of adaptive choice of code rate and modulation will allow more efficient use of the channel resource with a constant change in its state. The obtained results will significantly increase the energy efficiency of the OFDM system, lead to stable communication in non-stationary channels and increase the throughput.

Adapting coding and modulation; tropospheric communication; radio relay communication; coding; modulation; OFDM.

Введение. Станции тропосферной и радиорелейной связи могут использоваться для передачи информации между удаленными объектами вне зависимости от рельефа местности в условиях, когда применение кабельной связи невозможно или неэффективно. Тропосферная и радиорелейная связь являются экономически более выгодными видами связи по сравнению со спутниковой связью из-за отсутствия расходов на аренду спутникового ресурса. Несмотря на более высокую стоимость самого оборудования, за один-два года эксплуатации станции тропосферной и/или радиорелейной связи, инвестиции в приобретение аппаратуры окупаются, поскольку пользователь не платит за аренду спутникового канала и не несет никаких других дополнительных расходов.

Основными ограничениями пропускной способности тропосферных и радиорелейных систем связи являются межсимвольная интерференция и частотно-селективные замирания, вызванные многолучевым распространением сигнала [1]. Одним из способов улучшения вышеназванных систем связи является применение цифровой схемы модуляции OFDM [2]. Гибкость систем на основе OFDM позволяет выбирать оптимальные параметры сигнала и динамически их изменять исходя из текущих параметров среды распространения. Данный подход может дать выигрыш в пропускной способности канала. Современные системы связи обладают огромной вычислительной мощностью, что позволяет, используя ресурсоёмкие алгоритмы, достигнуть высоких информационных скоростей и обеспечить максимально эффективное использование канала связи. Кроме того, возможно использование одних и тех же систем связи, как для тропосферной, так и для радиорелейной связи, что увеличивает возможности применения, как в гражданской, так и в военной технике, уменьшает номенклатуру выпускаемых изделий, без потери в функциональности.

Постановка задачи. Выбор оптимальной сигнально-кодовой конструкции является важнейшей задачей при проектировании современных средств связи. Использование OFDM является перспективным направлением развития тропосферной и радиорелейной связи. Гибкость системы на основе OFDM позволяет динамически подстраиваться под условия передачи, требования к задержкам в канале. В связи с этим была поставлена задача разработать приёмник и передатчик для использования в тропосферном и радиорелейном канале связи адаптирующийся под состояние радиоканала. В качестве критериев адаптации использовать вид модуляции, кодовую скорость и глубину перемежения. Предусмотреть передачу

служебных данных содержащих информацию о сигнально-кодовой конструкции. Изучить возможный выигрыш от использования адаптивной системы по сравнению с системой, обеспечивающей гарантированную скорость исходя из наихудшего состояния канала в течении суток.

Решение поставленной задачи. Модель рассмотренная в данной работе, разработана в среде MatLab/simulink, исследованы различные сигнально-кодовые конструкции, проведено сравнение, сделаны выводы о возможности использования тех или иных сигнально-кодовой конструкции (СКК).

Ниже приведено описание приёмной и передающей части рассмотренной системы.



Рис. 1. Схема передающей части

Структура OFDM-модулятора представлена на рис. 1. Турбо-кодер кодирует данные со скоростью $1/3$ с последующим выкалыванием данных, до скоростей $1/3$, $1/2$, $2/3$, $3/4$ или $5/6$. Далее закодированные данные поступают на сверточный перемежитель от 0 до 400 мс с шагом 10 мс. Далее происходит формирование блоков, вставляются синхроследовательности и маркирование блоков. Далее происходит маппинг закодированных данных на созвездия QPSK, 8PSK, 16QAM, 64QAM, 128QAM или 256QAM. Вместе с данными формируется заголовок, который модулируется QPSK модуляцией и кодируется с кодовой скоростью $1/3$, содержащий информацию о текущей СКК, а также информацию об оценке канала с приемника. Далее по длине кадра выполняется матричное перемежение и накладывается ПСП для выполнения скремблирования. Далее происходит вставка пилот тонов для оценки характеристики канала в приёмнике. Перемеженные данные поступают на OFDM модулятор, где поднесущие с данными и пилотами преобразуются во временную область [3].

Далее рассмотрим приёмную часть, на рис. 2 приведена её структурная схема.

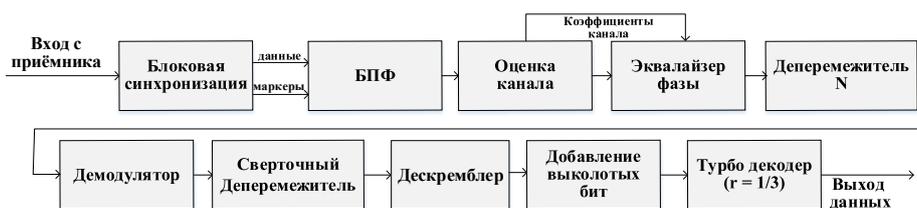


Рис. 2. Схема приёмной части

Блоковый синхронизатор реализован по схеме скользящего коррелятора Система имеет многоуровневую синхронизацию. На начальном этапе осуществляется поиск основных синхронизирующих последовательностей, маркирующих начало блоков. Поиск осуществляется сравнением пиков корреляционной функции с пороговым значением. С выхода блокового синхронизатора сигнал поступает на OFDM-демодулятор, реализованный посредством быстрого преобразования Фурье (БПФ). Далее сигнал поступает на блок оценки канала. Данный блок выполняет

скользящее усреднение сигналов пилот-тонов по времени, меньшему интервала когерентности канала. Затем проводится оценка сдвига фазы и изменения амплитуды каждого пилот-тона. Полученные значения интерполируются по частоте между пилот-тонами для получения оценочных сдвигов фазы и изменений амплитуды по информационным поднесущим, расположенным между пилот-тонами.

Далее сигнал выравнивается по фазе, что необходимо для корректной работы блока умножения на ПСП и поступает на депережежитель длиной N (число несущих OFDM). В многолучевом канале при достаточно большом количестве несущих замирания на соседних несущих являются коррелированными. Это приводит к тому, что символы, несмотря на расширение спектра, могут попасть в замирание целиком, что приведет к усилению шума. Корреляция между несущими приводит к увеличению вероятности ошибки на бит. Чтобы избежать корреляции, в передатчике перемежитель N переставляет местами расширенные символы по поднесущим таким образом, чтобы различные куски одного символа оказались в некоррелированных группах. Депережежитель N выполняет обратную функцию, вновь объединяя части одного символа. Данные с депережежителя поступают на демодулятор, а затем на сверточный депережежитель. После чего к данным добавляются выколотые при увеличении кодовой скорости биты и происходит декодирование сигнала.

В модель были добавлены источник сигнала (генератор ПСП) и контрольная схема счета ошибок позволяющая определить достоверность принимаемой информации. Сигнал был пропущен через канал с аддитивным белым гаусовским шумом. При моделировании были получены кривые достоверности для различных сигнально-кодовых конструкций.

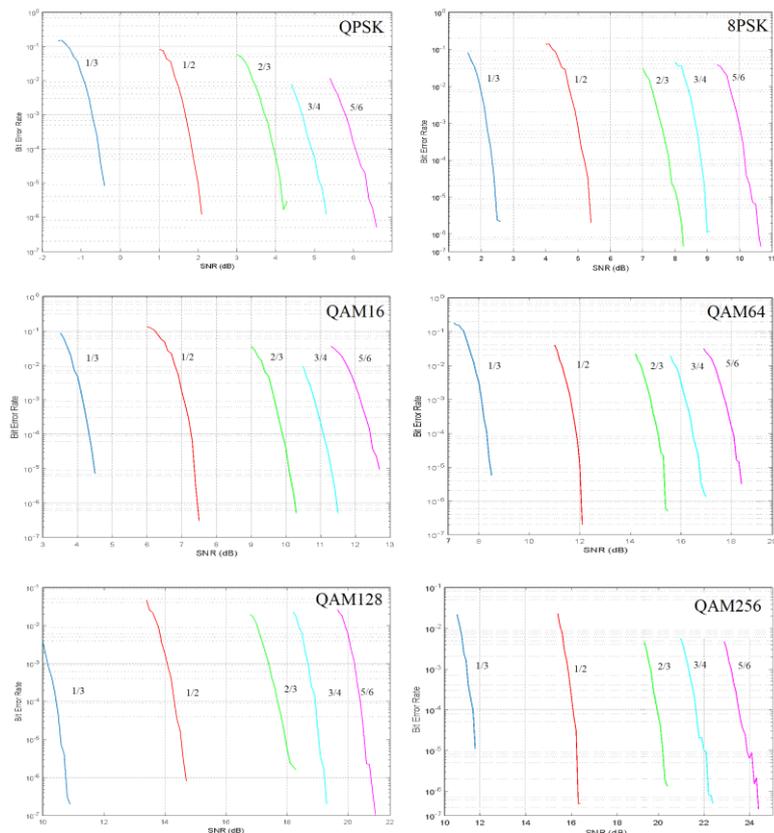


Рис. 3. Зависимость BER от C/Ш для различных сигнально-кодовых конструкций

При использовании различных сигнально кодовых конструкций в различных условиях приёма/передачи возможно достигать оптимального использования частотного ресурса и энергетика, создавать системы адаптирующиеся под условия распространения сигнала. Для реализации данного механизма в передаваемые служебные данные было внедрено служебное поле, содержащее в себе информацию об используемой кодовой скорости и типе модуляции закодированное в соответствии с табл. 1.

Кроме того в сигнал подмешаны пилот сигналы, которые позволяют оценивать качество канала. В связи с тем, что изменения состояния тропосферного канала происходят до ста раз в секунду, и носят случайный характер, возникает необходимость постоянного слежения за каналом и оперативной подстройки. Для увеличения устойчивости алгоритмов к случайным выбросам шумов эквалайзер подстраивается итерационно с постоянной времени подобранной таким образом, чтобы система могла подстроиться под изменения в канале и при этом отфильтровать максимально возможное количество шумов в измерениях уровня сигнала. Эквалайзер спроектирован по критерию минимальное среднеквадратичное отклонение ошибки (MMSE) позволяет взвешенно усиливать шум и полезный сигнал на ослабленных из-за селективности канала поднесущих, что улучшает итоговый BER системы. Математически данный метод можно описать как

$$X(k) = Y(k)H(k)'D[H(k)] + D[w] \times D[x],$$

где $D[w]$ – дисперсия шума, $D[x]$ – дисперсия полезного сигнала. $X(k)$ – переданный сигнал, $Y(k)$ – принятый сигнал, $H(k)$ – передаточная функция канала $D[H(k)]$ – дисперсия передаточной функции. На практике мощность полезного сигнала нормируется к единице и поэтому требуется только оценка мощности шума. Коррекция фазы и амплитуды в промежутках между пилот тонами выполняется путем линейной интерполяции.

Для оценки неравномерности канала происходит вычисление среднеквадратичного отклонения мощности пилот сигналов.

Кроме того, для борьбы с частотно селективными замираниями предусмотрен режим с дублированием информации сразу на несколько поднесущих, что позволяет при приёме просуммировать их и, тем самым увеличить отношение сигнал/шум за счет уменьшения информационной скорости в канале либо увеличения полосы сигнала.

Совокупность предложенных методов позволяет оптимизировать использование энергетического и частотного ресурса, добиться максимально возможной пропускной способности при текущем состоянии канала связи. В совокупности с использованием алгоритмов оценки качества канала возникает возможность динамически изменять сигнально-кодую конструкцию при изменении условий распространения. Регулировкой глубины перемежения возможно оптимизировать пороговое отношение С/Ш или величину задержки информации в канале в зависимости от требований к системе.

Таблица 1

Кодирование видов модуляции и кодовых скоростей

Модуляция	Индекс модуляции	Кодовая скорость	Индекс кодовой скорости	Глубина перемежения, мс	Индекс перемежения
QPSK	1	1/3	1	0	1
8-PSK	2	1/2	2	20	2
16-QAM	3	2/3	3	40	3
64-QAM	4	3/4	4	80	4
128-QAM	5	5/6	5	160	5
256-QAM	6			400	6

Заключение. Проведенные эксперименты позволяют сделать вывод что рассмотренные подходы возможно использовать в системах тропосферной и радиорелейной связи, но следует учитывать высокий пик-фактор OFDM сигнала [4], что делает требования к линейности используемых усилителей более строгими и вызывает необходимость включать в систему алгоритмы уменьшения пик-фактора [5], OFDM сигнала и цифровых предискажений. Моделирование пропускной способности тропосферной линии связи в зависимости от времени суток позволяет сделать вывод, что использование адаптивной системы даст выигрыш в суточной пропускной способности канала не менее 30%, по сравнению с неадаптивной системой, разработанной с учетом наихудшего состояния канала в течении суток. Дальнейшее уточнение полученного выигрыша требует натурных испытаний, запланированных на ближайшее время.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Давыденко Ю.И.* Дальняя тропосферная связь. – М.: Военное издательство министерства обороны СССР, 1968.
2. *Maokai, Chen Xihong 'Shu Tao 'Dong Shaoqiang.* New generation Troposcatter Communication Based on OFDM Modulation // Hu The Ninth International Conference on Electronic Measurement & Instruments ICEMI'2009.
3. *Скляр Б.* Цифровая связь. – М.: Изд. дом Вильямс, 2003.
4. *Baxley R.J., Chunming Zhao, and G. Tong Zhou.* Constrained Clipping for Crest Factor Reduction in OFDM // IEEE Transactions on Broadcasting. – Dec. 2006. – Vol. 52, No. 4. – P. 570-575.
5. *Armstrong J.* Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering // Electronics Letters 28th February 2002. – Vol. 38, No. 5. – P. 246-247.
6. *Arash Soltani, Mehdi Airamlozadeh & Jaber Parchami.* MIMO-OFDM channel estimation based on minimum error entropy criterion under non-Gaussian environment // International Journal of Electronics Letters. – 2022. – DOI: 10.1080/21681724.2022.2087909.
7. *Prabal Gupta, H. Pal Thethi, Ajay Tomer.* An efficient and improved PTS algorithm for PAPR reduction in OFDM system // International Journal of Electronics. – 2022. – 109:7. – P. 1252-1277.
8. *Mohammad Reza Motazedi, Reza Dianat.* An erasure-based scheme for reduction of PAPR in spatial multiplexing MIMO-OFDM using Reed-Solomon codes over GF(216 + 1) // International Journal of Electronics. – 2018. – 105:9. – P. 1583-1597.
9. *Li Y., Wang X. and Zhu A.* "Reducing Power Consumption of Digital Predistortion for RF Power Amplifiers Using Real-Time Model Switching // in IEEE Transactions on Microwave Theory and Techniques. – March 2022. – Vol. 70, No. 3. – P. 1500-1508. – DOI: 10.1109/TMTT.2021.3132347.
10. *Mohammad Reza Motazedi & Reza Dianat.* An erasure-based scheme for reduction of PAPR in spatial multiplexing MIMO-OFDM using Reed-Solomon codes over GF(216 + 1) // International Journal of Electronics. – 2018. – 105:9. – P. 1583-1597. – DOI: 10.1080/00207217.2018.1477192.
11. *Ding L., Ma Z., Morgan D.R., Zierdt M., and Pastalan J.* A least square/Newton method for digital predistortion of wideband signals // IEEE Trans. on Communications. – May 2006. – Vol. 54, No. 5. – P. 833-840.
12. *Gan L.* Adaptive digital predistortion of nonlinear systems: Ph.D. Thesis, Faculty of Electrical and Information Engineering, Graz University of Technology, Graz, Austria. – 2009.
13. *Pawar D.S. and Badodekar H.S.* Review of PAPR Reduction Techniques in Wireless Communication // 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN). – 2018. – P. 204-207. – DOI: 10.1109/GWCN.2018.8668631.
14. *S. D V, T. P, S. B and S. M,* Improved Energy Efficiency of Massive MIMO-OFDM // 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). – 2021. – P. 38-42. DOI: 10.1109/ICICCS51141.2021.9432385.
15. *Jui-Chi Chang, Fang-Biau Ueng and Jiun-Cheng Ning.* Channel shortening and equalization of OFDM/CDMA systems over doubly selective fading channels // International Symposium on Wireless and Pervasive Computing. – 2011. – P. 1-4. – DOI: 10.1109/ISWPC.2011.5751329.
16. *Gross J., Bohge M.* Dynamic Mechanisms in OFDM Wireless Systems: A Survey on Mathematical and System Engineering Contributions // TKN Technical Report TKN-06-001, Berlin, May 2006.

17. Yang W., Li H., Li M., Liu Y. and Liu Q. Channel Estimation for Practical IRS-Assisted OFDM Systems // 2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW). – 2021. – P. 1-6. – DOI: 10.1109/WCNCW49093.2021.9419982.
18. Hussein M.A., Venard O., Feuvrie B. and Wang Y. Digital predistortion for RF power amplifiers: State of the art and advanced approaches // 2013 IEEE 11th International New Circuits and Systems Conference (NEWCAS). – 2013. – P. 1-4. – DOI: 10.1109/NEWCAS.2013.6573671.
19. Jaradat A.M., Hamamreh J.M. and Arslan H. Modulation options for OFDM-based waveforms: Classification comparison and future directions // IEEE Access. – 2019. – Vol. 7. – P. 17263-17278.
20. Tao Jiang, Guanghi Zhu. Complement Block Coding for Reduction in Peak-to-Average Power Ratio of OFDM Signals // IEEE Radio Communications. – Sept. 2005. – P. S17-S22.

REFERENCES

1. Davydenko Yu.I. Dal'naya troposfer'naya svyaz' [Long-range tropospheric communication]. Moscow: Voennoe izdatel'stvo ministerstva oborony SSSR, 1968.
2. Maokai, Chen Xihong 'Shu Tao 'Dong Shaoqiang. New generation Troposcatter Communication Based on OFDM Modulation, *Hu The Ninth International Conference on Electronic Measurement & Instruments ICEMI'2009*.
3. Sklyar B. Tsifrovaya svyaz' [Digital communication]. Moscow: Izd. dom Vil'yams, 2003.
4. Baxley R.J., Chunming Zhao, and G. Tong Zhou, Constrained Clipping for Crest Factor Reduction in OFDM, *IEEE Transactions on Broadcasting*, Dec. 2006, Vol. 52, No. 4, pp. 570-575.
5. Armstrong J. Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering, *Electronics Letters 28th February 2002*, Vol. 38, No. 5, pp. 246-247.
6. Arash Soltani, Mehdi Airamlozadeh & Jaber Parchami. MIMO-OFDM channel estimation based on minimum error entropy criterion under non-Gaussian environment, *International Journal of Electronics Letters*, 2022. DOI: 10.1080/21681724.2022.2087909.
7. Prabal Gupta, H. Pal Thethi, Ajay Tomer. An efficient and improved PTS algorithm for PAPR reduction in OFDM system, *International Journal of Electronics*, 2022, 109:7, pp. 1252-1277.
8. Mohammad Reza Motazedi, Reza Dianat. An erasure-based scheme for reduction of PAPR in spatial multiplexing MIMO-OFDM using Reed-Solomon codes over GF(216 + 1), *International Journal of Electronics*, 2018, 105:9, pp. 1583-1597.
9. Li Y., Wang X. and Zhu A. Reducing Power Consumption of Digital Predistortion for RF Power Amplifiers Using Real-Time Model Switching, in *IEEE Transactions on Microwave Theory and Techniques*, March 2022, Vol. 70, No. 3, pp. 1500-1508. DOI: 10.1109/TMTT.2021.3132347.
10. Mohammad Reza Motazedi & Reza Dianat. An erasure-based scheme for reduction of PAPR in spatial multiplexing MIMO-OFDM using Reed-Solomon codes over GF(216 + 1), *International Journal of Electronics*, 2018, 105:9, pp. 1583-1597. DOI: 10.1080/00207217.2018.1477192.
11. Ding L., Ma Z., Morgan D.R., Zierdt M., and Pastalan J. A least square/Newton method for digital predistortion of wideband signals, *IEEE Trans. on Communications*, May 2006, Vol. 54, No. 5, pp. 833-840.
12. Gan L. Adaptive digital predistortion of nonlinear systems: Ph.D. Thesis, Faculty of Electrical and Information Engineering, Graz University of Technology, Graz, Austria. – 2009.
13. Pawar D.S. and Badodekar H.S. Review of PAPR Reduction Techniques in Wireless Communication, *2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN)*, 2018, pp. 204-207. DOI: 10.1109/GCWCN.2018.8668631.
14. S. D V, T. P, S. B and S. M. Improved Energy Efficiency of Massive MIMO-OFDM, *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2021, pp. 38-42. DOI: 10.1109/ICICCS51141.2021.9432385.
15. Jui-Chi Chang, Fang-Biau Ueng and Jiun-Cheng Ning. Channel shortening and equalization of OFDM/CDMA systems over doubly selective fading channels, *International Symposium on Wireless and Pervasive Computing*, 2011, pp. 1-4. DOI: 10.1109/ISWPC.2011.5751329.
16. Gross J., Bohge M. Dynamic Mechanisms in OFDM Wireless Systems: A Survey on Mathematical and System Engineering Contributions, *TKN Technical Report TKN-06-001, Berlin, May 2006*.
17. Yang W., Li H., Li M., Liu Y. and Liu Q. Channel Estimation for Practical IRS-Assisted OFDM Systems, *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2021, pp. 1-6. DOI: 10.1109/WCNCW49093.2021.9419982.

18. Hussein M.A., Venard O., Feuvrie B. and Wang Y. Digital predistortion for RF power amplifiers: State of the art and advanced approaches, *2013 IEEE 11th International New Circuits and Systems Conference (NEWCAS)*, 2013, pp. 1-4. DOI: 10.1109/NEWCAS.2013.6573671.
19. Jaradat A.M., Hamamreh J.M. and Arslan H. Modulation options for OFDM-based waveforms: Classification comparison and future directions, *IEEE Access*, 2019, Vol. 7, pp. 17263-17278.
20. Tao Jiang, Guanghi Zhu. Complement Block Coding for Reduction in Peak-to-Average Power Ratio of OFDM Signals, *IEEE Radio Communications*, Sept. 2005, pp. S17-S22.

Статью рекомендовал к опубликованию к.т.н., доцент Ф.В. Зандер.

Луферчик Павел Валерьевич – АО «НПП «Радиосвязь»; e-mail: Lufurchikp@gmail.com; г. Красноярск, Россия; тел.: +79232942552; начальник отдела.

Комаров Алексей Александрович – e-mail: komarovalal@gmail.com; тел.: +79831576492; инженер-конструктор.

Штро Павел Викторович – e-mail: faust_256@mail.ru; тел.: +79069712716; главный специалист.

Конев Александр Николаевич – e-mail: FBRLC@ya.ru; тел.: +79082181452; инженер-конструктор.

Lufurchik Pavel Valerievich – JSC SPE Radiosvyaz; e-mail: Lufurchikp@gmail.com; Krasnoyarsk, Russia; phone: +79232942552; head of the department.

Komarov Alexey Alexandrovich – e-mail: komarovalal@gmail.com; phone: +79831576492; design engineer.

Shtro Pavel Viktorovich – e-mail: faust_256@mail.ru; phone: +79069712716; chief specialist.

Konev Alexandr Nikolaevich – e-mail: FBRLC@ya.ru; phone: +79082181452; design engineer.

УДК 621.391.01

DOI 10.18522/2311-3103-2022-4-238-244

П.В. Луферчик, П.В. Штро, А.Н. Конев, А.А. Комаров

РАЗРАБОТКА ЭНЕРГЕТИЧЕСКИ ЭФФЕКТИВНОЙ СИСТЕМЫ СВЯЗИ В ТРОПОСФЕРНОМ РАДИОКАНАЛЕ НА БАЗЕ OFDM-СИГНАЛОВ

Известно, что при передаче данных в системах радиорелейной и тропосферной связи может возникать межсимвольная интерференция. Наличие многолучевого распространения и частотно-селективных замираний в тропосферном, радиорелейном существенно понижает энергетическую эффективность системы связи в целом. Целью работы было добиться увеличения эффективности использования канала для радиорелейной и тропосферной связи путем использования OFDM (orthogonal frequency-division multiplexing – мультиплексирование с ортогональным частотным разделением каналов) при помощи методов уменьшения пик-фактора OFDM сигнала и увеличения линейности передающего тракта. Для оценки алгоритмов цифровых предскажений в среде Matlab/Simulink была разработана модель для методов LMS, NLMS, RLS, RPEM и модель усилителя мощности с реальными характеристиками. По результатам моделирования алгоритмов был выбран RLS. Кроме того, в данной работе был разработан модифицированный вариант алгоритма адаптации на основе рекурсивного метода наименьших квадратов (RLSm). Основным результатом модификации являются: уменьшение количества арифметических операций, необходимых для выполнения одной итерации (более чем в 5 раз), повышение стабильности алгоритмов адаптации, за счёт введения методов регуляризации, уменьшение времени сходимости, за счёт введения экспоненциальной зависимости. Были исследованы различные алгоритмы уменьшения пик-фактора OFDM сигнала, наилучший результат удалось достичь при комбинировании Tone reservation (TR) и Active constellation extension (ACE). Моделирование в среде Matlab/Simulink показало, что комбинация алгоритмов TR и ACE уменьша-

ет пик-фактор OFDM сигналов на ~ 5 дБ для BPSK потока данных и ~ 4.5 дБ для 8-PSK, QAM-16, QAM-64, QAM-128 и QAM-256. Для увеличения линейности передающего тракта был выбран и модернизирован алгоритм ввода цифровых предискажений RLSm, он позволил снизить величину модуля вектора ошибки (EVM) на 13.5дБ, а также увеличить соотношение модуляция/ошибка (MER) на 13.6дБ.

Цифровые предискажения; резервирование тона; расширение активного сигнального созвездия(ACE); пик-фактор; тропосферная связь; радиорелейная связь; OFDM.

P.V. Lufurchik, P.V. Shtro, A.N. Konev, A.A. Komarov

DEVELOPMENT OF ENERGY EFFICIENT COMMUNICATION SYSTEM IN THE TROPOSPHERE RADIO CHANNEL BASED ON OFDM SIGNALS

It is known that inter-symbol interference may occur during data transmission in radio relay and tropospheric communication systems. The presence of multipath propagation and frequency-selective fading in the tropospheric, radio relay channel significantly reduces the energy efficiency of the communication system as a whole. The aim of the work was to increase the efficiency of using the channel for radio relay and tropospheric communication by using OFDM (orthogonal frequency-division multiplexing using the methods of reducing the peak factor of the OFDM signal and increasing the linearity of the transmission path. To evaluate digital predistortion algorithms in the Matlab/Simulink environment, a model was developed for the LMS, NLMS, RLS, RPEM methods and a power amplifier model with real characteristics. Based on the results of modeling algorithms, RLS was chosen. In addition, in this work, a modified version of the adaptation algorithm based on the recursive least squares method (RLSm) was developed. The main result of the modification is: a decrease in the number of arithmetic operations per iteration (more than 5 times), an increase in the stability of adaptation algorithms, due to the introduction of regularization methods, a decrease in the convergence time, due to the introduction of an exponential dependence. Various algorithms for reducing the peak factor of an OFDM signal were investigated, the best result was achieved by combining Tone reservation (TR) and Active constellation extension (ACE). Simulation in the Matlab/Simulink environment showed that the combination of TR and ACE algorithms reduces the peak factor of OFDM signals by ~ 5 дБ for BPSK data stream and ~ 4.5 дБ for 8-PSK, QAM-16, QAM-64, QAM-128 and QAM- 256. To increase the linearity of the transmission path, the RLSm digital pre-distortion algorithm was selected and upgraded, it made it possible to reduce the magnitude of the error vector modulus (EVM) by 13.5дБ, and also increase the modulation/error ratio (MER) by 13.6дБ.

Digital predistortion; tone reservation; active signal constellation expansion (ACE); peak factor; tropospheric communication; radio relay communication; OFDM.

Введение. На сегодняшний день использование тропосферной связи широко распространено для задач, в которых применение проводной связи является невозможным, а спутниковые системы связи не обладают требуемыми тактико-техническими характеристиками, либо слишком дороги. В настоящее время тропосферная и радиорелейная связь активно используется в вооруженных силах России. Основными преимуществами тропосферного канала связи является надёжность и защита от преднамеренных помех. Такой вид связи слабо подвержен влияниям средств радиоэлектронной борьбы. Другим весомым преимуществом тропосферной связи является тот факт, что прием сигналов возможен только в непосредственной близости от падения тропосферного луча. Это означает, что несанкционированный перехват информации, передаваемой с помощью такого типа связи, является практически невыполнимой задачей. Большой потенциал у тропосферных и радиорелейных систем связи имеется и на гражданском рынке. Потребителями таких систем являются компании преимущественно из добывающих отраслей, в том числе нефтегазовой отрасли. Также, на сегодняшний день актуальная задача борьбы с так называемым «цифровым неравенством» граждан РФ. Тропосферные и радиорелейные виды связи могут быть использованы для обеспечения

высокоскоростным интернетом малонаселенных и удаленных поселений. Благодаря описанным выше защитным свойствам, допустимо применение таких систем связи в качестве связи двойного назначения в новопринятых субъектах РФ. Однако, на текущий момент, пропускная способность серийных ответственных станций тропосферной и радиорелейной связи весьма ограничена. Но современные, вычислительно сложные алгоритмы позволяют существенно улучшить эффективность использования канала, тем самым сделав его более привлекательным. Вместе с тем тропосферные системы связи имеют ряд недостатков. В первую очередь стоит помнить, что тропосферный канал связи не стационарен и его состояние может изменяться до ста раз в секунду, а многолучевость приводит к межсимвольной интерференции и частотно-селективным замираниям, что существенно снижает энергетическую эффективность системы связи. Данные недостатки возможно обойти с использованием модуляции OFDM. Данный вид модуляции имеет высокий пик-фактор, что в свою очередь делает очень строгими требования к используемым усилителям. Таким образом разработка и комбинирование алгоритмов уменьшения пик-фактора является актуальной для современной радиотехники задачей.

Постановка задачи. На сегодняшний день серийно выпускаются станции тропосферной связи Р-423-ПМ и Р-423-АМК со скоростью передачи данных не превышающей 2048кбит/с. Увеличение скорости передачи информации является актуальной задачей для тропосферной связи. Препградой для увеличения пропускной способности тропосферного радиоканала выступает многолучевое распространение и частотно-селективные замирания сигнала. Для получения максимальной пропускной способности тропосферного радиоканала необходимо плавное изменение скоростей передачи данных, чтобы приблизиться к истинной пропускной способности радиоканала. Для того чтобы максимально эффективно обойти имеющиеся ограничения было принято решение использовать OFDM сигналов в тропосферном канале связи.

Решение поставленной задачи. Мультиплексирование множества потоков передачи данных различных типов модуляторов (BPSK, QPSK, QAM) позволяет реализовать плавное изменение скоростей передачи данных, а также эффективно бороться с частотно-селективными замираниями.

Станции тропосферной связи зачастую используют нелинейные усилители в передающем тракте, для уменьшения габаритов и энергопотребления станции связи. Для применения OFDM сигналов в существующих станциях тропосферной связи необходимо использовать различные алгоритмы линеаризации передающего тракта, а также использовать алгоритмы уменьшения пик-фактора OFDM сигнала.

В ходе обзора отечественной и зарубежной литературы было обнаружено множество методов уменьшения пик-фактора OFDM-сигналов. Наилучшие результаты уменьшения пик-фактора дала комбинация алгоритмов Tone reservation(TR) [1] и Active constellation extension(ACE) [2]. Моделирование в среде matlab/simulink показало, что комбинация алгоритмов TR и ACE уменьшает пик-фактор OFDM сигналов на ~5.5дБ для BPSK потока данных и ~4.5дБ для 8-PSK, QAM-16, QAM-64, QAM-128 и QAM-256.

Для дальнейшего увеличения энергетической эффективности системы тропосферной связи с OFDM сигналами было рассмотрено несколько способов реализации функционального блока ввода предсказаний:

Аналитический способ, на основе нелинейной функциональной зависимости (полиномиальная модель, полиномиальная модель с памятью, модель на основе рядов Вольтера и др.) [3–5].

Комбинированный способ, на основе табличной аппроксимации известных нелинейных функций.

К преимуществам табличного способа относится невысокая трудоёмкость вычисления значений таблиц аппроксимации. Для табличного способа применяется линейная, кубическая, кусочно-линейная или сплайновая аппроксимация. Однако, табличный способ малоприменим для описания инерционных свойств усилителей мощности, так как для их описания требуется увеличение количества таблиц и существенно усложняется процедура вычисления значений этих таблиц. То же самое относится и к процедуре адаптации таблиц с течением времени, так как адаптировать необходимо каждый элемент таблиц.

Аналитический способ предполагает использование нелинейных функциональных зависимостей, как правило, степенных функций (полиномов), для расчёта коэффициентов полиномов решается линейная задача наименьших квадратов через процедуру псевдо-обращения Мура-Пенроуза, сингулярное разложение, QR-разложение и др. [6, 7]. Преимуществом аналитического способа является возможность использования рядов Вольтера для полноценного описания нелинейных свойств усилителя мощности с учётом влияния инерционных свойств в различные моменты времени, а также упрощённых вариантов рядов Вольтера: модель Вольтера со снижением девиации динамики, полиномиальная модель с памятью, обобщённая полиномиальная модель с памятью и др. [8, 9].

Адаптацию коэффициентов блока ввода предсказаний, выполненного на основе аналитической модели целесообразно выполнять с помощью алгоритма косвенного обучения. Основным преимуществом алгоритма косвенного обучения является то, что он способен начать работу с «нуля», не имея начальных (заранее рассчитанных) параметров для блока ввода предсказаний. Структурная схема системы ввода предсказаний с косвенным обучением представлена на рис. 1.

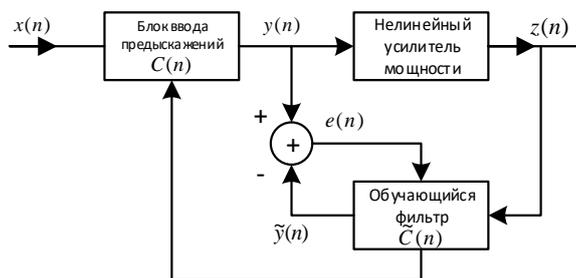


Рис. 1. Структурная схема системы ввода предсказаний с косвенным обучением

Адаптация параметров системы ввода предсказаний происходит по сигналу ошибки, который является разностью между сигналом на выходе блока ввода предсказаний в прямом канале и сигнала на выходе обучающегося фильтра в обратном канале [9]. Коэффициенты системы ввода предсказаний постоянно подстраиваются (рекурсивный алгоритм) при сравнении двух сигналов с выхода блока ввода предсказаний и с выхода обучающегося фильтра. Основными алгоритмами адаптации для систем с косвенным обучением являются: метод наименьших квадратов (LMS), нормализованный метод наименьших квадратов (NLMS), рекурсивный метод наименьших квадратов (RLS), рекурсивный метод прогноза ошибки (RPEM).

Для оценки алгоритмов цифровых предсказаний в среде matlab/simulink была разработана модель для методов LMS, NLMS, RLS, RPEM и модель усилителя мощности с реальными характеристиками. Параметры модели усилителя мощности: диапазон частот 4,4 – 5 ГГц, максимальная излучаемая мощность – 100 Вт, интермодуляционные искажения третьего порядка – минус 27 дБ.

По результатам моделирования алгоритмов был выбран RLS. Кроме того, в рамках данной работы был разработан модифицированный вариант алгоритма адаптации на основе рекурсивного метода наименьших квадратов (RLSm). Основным результатом модификации являются: уменьшение количества арифметических операций, необходимых для выполнения одной итерации (более чем в 5 раз), повышение стабильности алгоритмов адаптации, за счёт введения методов регуляризации, уменьшение времени сходимости, за счёт введения экспоненциальной зависимости фактора «забывания».

Для оценки эффективности алгоритмов адаптации была разработана модель усилителя мощности. Данная модель разработана на основе сигналов входа и выхода, полученных от реального усилителя мощности. Основой модели является аналитическое описание на базе обобщённой полиномиальной модели с памятью [7]:

$$y_{gmp} = \sum_{k \in K_a} \sum_{l \in L_a} a_{kl} x(n-l) |x(n-l)|^k + \sum_{k \in K_b} \sum_{l \in L_b} \sum_{m \in M_b} b_{klm} x(n-l) |x(n-l-m)|^k + \sum_{k \in K_c} \sum_{l \in L_c} \sum_{m \in M_c} c_{klm} x(n-l) |x(n-l+m)|^k \quad (1)$$

В данной модели $K_a L_a$ – число коэффициентов для сигнала и огибающей записанных без временной задержки, $K_b L_b M_b$ – число коэффициентов для сигнала и огибающей записанных с задержкой, $K_c L_c M_c$ – число коэффициентов для сигнала и огибающей записанных с опережением. Таким образом, в данной модели учитывается влияние перекрёстных членов, ограниченных величиной задержки M_b .

Разработанная модель позволила оценить спектральные характеристики сигналов (ACPR), скорость сходимости алгоритмов и модуль вектора ошибки (EVM) относительно входного сигнала. На рис. 2 представлены сигнальные созвездия без использования предсказаний и с использованием предсказаний с алгоритмом адаптации RLS на выходе модели для полосы 28 МГц, частоты несущего сигнала 5 ГГц и мощности 100 Вт.

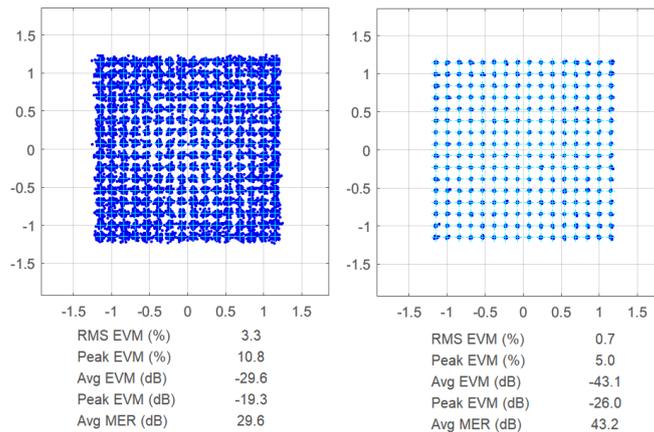


Рис. 2. Сигнальные созвездия с использованием предсказаний и без использования предсказаний с алгоритмом адаптации RLS на выходе модели для полосы 28 МГц, частоты несущего сигнала 5 ГГц и мощности 100 Вт

Как видно из данных, представленных на рис. 2, использование предсказаний позволяет снизить величину модуля вектора ошибки (EVM), а также существенно увеличить соотношение модуляция/ошибка (MER). На рис. 3 представлен спектр сигнала на выходе модели для полосы 28 МГц и частоты несущего сигнала 5 ГГц.

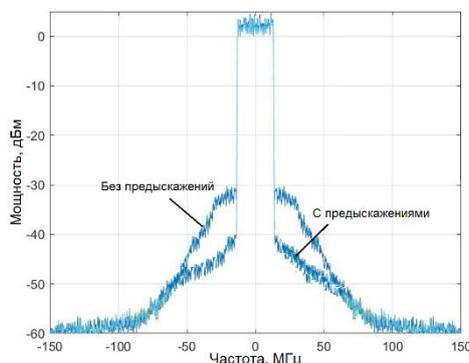


Рис. 3. Спектр сигнала на выходе модели до и после алгоритмов предсказаний

Заключение. Использование предсказаний позволяет существенно снизить уровень нелинейных искажений (уровень помехи в соседнем канале снизился на 10 дБ). Также уменьшается величина модуля вектора ошибки (EVM) на 13.5дБ и существенно увеличивается соотношение модуляция/ошибка (MER) на 13.6дБ.

В результате анализа и моделирования были выработаны методы, позволившие увеличить энергетическую эффективность режима OFDM в тропосферных станциях связи. Проведено моделирование различных алгоритмов понижения пик-фактора. Реализованы два наилучших алгоритма снижения PAPR наиболее подходящих для применения в тропосферном канале. Комбинация алгоритмов Active Constellation Extension (ACE) и Tone Reservation (TR) позволяет существенно уменьшить пик-фактор OFDM сигналов. Для увеличения линейности передающего тракта был выбран и модернизирован алгоритм ввода цифровых предсказаний RLSm, он позволил снизить величину модуля вектора ошибки (EVM) на 13.5дБ, а также увеличить соотношение модуляция/ошибка (MER) на 13.6дБ. Полученные результаты позволят значительно повысить энергетическую эффективность режима OFDM, приведут к устойчивой связи в нестационарном тропосферном канале и увеличению пропускной способности.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Filippo Tosato, Magnus Sandell, Makoto Tanahashi. Tone Reservation for PAPR Reduction: an Optimal Approach through Sphere Encoding, *IEEE ICC 2016 - Signal Processing for Communications Symposium*.
2. Sandeepkumar Vangalaa, Anuradha Sundrub. Adaptive Clipping Active Constellation Extension for PAPR Reduction of OFDM/OQAM System, *6th International Conference on Advances In Computing & Communications, ICACC 2016, 6-8 September 2016, Cochin, India*.
3. Ir. Johan Paduart, Identification of Nonlinear systems using Polynomial Nonlinear State Space Model. Vrije Universiteit Brussel, 2008, pp. 102-104.
4. Zhu A., Brazil T.J. An overview of Volterra series based behavioral modeling of RF/microwave power amplifiers, *Wireless Microwave Technol. Conf.*, 2006, pp.101-107.
5. Jeon W.G., Chang K.H., and Cho Y.S. An Adaptive Data Predistorter for Compensation of Nonlinear Distortion in OFDM systems, *IEEE Transactions on Communications*, 1997, 45:11671171.
6. Ding L., Ma Z., Morgan D.R., Zierdt M., and Pastalan J. A least square/Newton method for digital predistortion of wideband signals, *IEEE Trans. on Communications*, May 2006, Vol. 54, No. 5, pp. 833-840.

7. *Gan L.* Adaptive digital predistortion of nonlinear systems: Ph.D. Thesis, Faculty of Electrical and Information Engineering, Graz University of Technology, Graz, Austria. – 2009.
8. *Morgan D., Ma Zh., Kim Ja., Zierdt M.G., and Pastalan J.* A generalized memory polynomial model for digital predistortion of RF power amplifiers, *IEEE Trans. Sig. Proc.*, 2006, Vol. 54, pp. 3852-3860.
9. *Zhu A., Pedro J.C., Brazil T.J.* Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers, *IEEE Trans. On Microwave Theory and Techniques*, 2006, Vol. 54, No. 12, pp. 4323-4332.
10. *Iqbal H. and Khan S.A.* Selective Mapping: Implementation of PAPR Reduction Technique in OFDM on SDR Platform, *2018 24th International Conference on Automation and Computing (ICAC)*, 2018, pp. 1-6. DOI: 10.23919/ICAC.2018.8749039.
11. *Helaly T.K., Dansereau R.M. and El-Tanany M.* "A Low Complexity PAR Reduction Technique Using Cyclic Shifted Data Sequences in DS-CDMA Signals," *2010 Sixth Advanced International Conference on Telecommunications*, 2010, pp. 196-200. DOI: 10.1109/AICT.2010.54.
12. *Xiao Y., Zhang L. and Imran M.* Active Constellation Extension for Peak Power Reduction Based on Positive and Negative Iterations in OFDM Systems, *2019 UK/ China Emerging Technologies (UCET)*, 2019, pp. 1-5. DOI: 10.1109/UCET.2019.8881859.
13. *Lin W.-L. and Tseng F.-S.* Theory and Applications of Active Constellation Extension, in *IEEE Access*, 2021, Vol. 9, pp. 93111-93118. DOI: 10.1109/ACCESS.2021.3093103.
14. *Fang L. and De Figueiredo R.J.P.* Performance of OFDM-CDMA System with Papr Reduction in Nonlinear Rayleigh Fading Channel, *MILCOM 2006 - 2006 IEEE Military Communications conference*, 2006, pp. 1-6. DOI: 10.1109/MILCOM.2006.302544.
15. *Prabal Gupta, H. Pal Thethi, Ajay Tomer.* An efficient and improved PTS algorithm for PAPR reduction in OFDM system, *International Journal of Electronics*, 2022, 109:7, pp. 1252-1277.
16. *Yang W., Li H., Li M., Liu Y. and Liu Q.* Channel Estimation for Practical IRS-Assisted OFDM Systems, *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2021, pp. 1-6. DOI: 10.1109/WCNCW49093.2021.9419982.
17. *Lu X., Shi Y., Li W., Lei J. and Pan Z.* A Joint Physical Layer Encryption and PAPR Reduction Scheme Based on Polar Codes and Chaotic Sequences in OFDM System, in *IEEE Access*, 2019, Vol. 7, pp. 73036-73045. DOI: 10.1109/ACCESS.2019.2919598.
18. *Mohammad Reza Motazedi, Reza Dianat.* An erasure-based scheme for reduction of PAPR in spatial multiplexing MIMO-OFDM using Reed-Solomon codes over GF(216 + 1), *International Journal of Electronics*, 2018, 105:9, pp. 1583-1597.
19. *Zhang H., Wu H.-C., Jiang H. and Huang S.C.-H.* Robust pilot detection techniques for channel estimation and symbol detection in OFDM systems, *2014 IEEE Global Communications Conference*, 2014, pp. 3025-3031. DOI: 10.1109/GLOCOM.2014.7037269.
20. *Jaradat A.M., Hamamreh J.M. and Arslan H.* Modulation options for OFDM-based waveforms: Classification comparison and future directions, *IEEE Access*, 2019, Vol. 7, pp. 17263-17278.

Статью рекомендовал к опубликованию к.т.н., доцент Ф.В. Зандер.

Луферчик Павел Валерьевич – АО «НПП «Радиосвязь»; e-mail: Lufchikp@gmail.com; г. Красноярск, Россия; тел.: +79232942552; начальник отдела.

Штро Павел Викторович – e-mail: faust_256@mail.ru; тел.: +79069712716; главный специалист.

Конев Александр Николаевич – e-mail: FBRLC@ya.ru; тел.: +79082181452; инженер-конструктор.

Комаров Алексей Александрович – e-mail: komarovalal@gmail.com; тел.: +79831576492; инженер-конструктор.

Lufchik Pavel Valerievich – JSC SPE Radiosvyaz; e-mail: Lufchikp@gmail.com; Krasnoyarsk, Russia; phone: +79232942552; head of the department.

Shtro Pavel Viktorovich – e-mail: faust_256@mail.ru; phone: +79069712716; chief specialist.

Konev Alexandr Nikolaevich – e-mail: FBRLC@ya.ru; phone: +79082181452; design engineer.

Komarov Alexey Alexandrovich – e-mail: komarovalal@gmail.com; phone: +79831576492; design engineer.

О.В. Белоусова, Г.В. Куповых, А.Г. Клово, В.В. Гривцов

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ ЭЛЕКТРОДИНАМИЧЕСКОЙ СТРУКТУРЫ ТУРБУЛЕНТНОГО ПРИЗЕМНОГО СЛОЯ

Представлены результаты математического моделирования электродинамической структуры турбулентного приземного слоя атмосферы. Использована модель стационарного турбулентного электродного эффекта, действующего вблизи поверхности земли. Анализ уравнений методами теории подобия позволил сделать ряд обоснованных физических допущений, позволивших получить аналитические решения. Получены аналитические формулы для расчетов профилей концентраций легких ионов (аэроионов), плотности объемного электрического заряда и напряженности электрического поля в турбулентном электродном слое. В результате математического моделирования исследованы зависимости распределения электрических характеристик приземного слоя от значений электрического поля, степени турбулентного перемешивания и аэрозольного загрязнения атмосферы. Показано, что параметр электродного эффекта (отношение значений напряженности электрического поля на поверхности земли и на верхней границе электродного слоя) практически не зависит от атмосферных условий, тогда как высота электродного слоя и, соответственно, масштаб распределения электрических характеристик приземного слоя меняются в значительной степени. Усиление турбулентного перемешивания в приземном слое приводит к увеличению высоты электродного слоя и, как следствие, масштабов распределения его параметров. Усиление электрического поля или загрязнение воздуха аэрозольными частицами достаточной концентрации приводит к уменьшению его высоты. Увеличение концентрации аэрозольных частиц в атмосфере уменьшает значения плотности электрического заряда у поверхности земли. Теоретические расчеты хорошо согласуются с экспериментальными данными и результатами численного моделирования электродинамической структуры приземного слоя. Полученные в работе аналитические формулы для расчетов электрических характеристик приземного слоя и результаты вычислений могут быть полезны при решении ряда прикладных задач геофизики, в частности для мониторинга электрического состояния атмосферы.

Математическое моделирование; электродинамика; приземный слой; атмосфера; турбулентный электродный эффект; электрическое поле; аэроионы; объемный заряд; аэрозоль.

O.V. Belousova, G.V. Kupovykh, A.G. Klovo, V.V. Grivtsov

MODELING RESULTS OF THE TURBULENT SURFACE LAYER ELECTRODYNAMIC STRUCTURE

The article presents the results of mathematical modeling of turbulent surface layer electrodynamic structure. A model of a stationary turbulent electrode effect operating near the earth's surface is used. The analysis of equations by methods of similarity theory allowed us to make a number of reasonable physical assumptions that allowed us to obtain analytical solutions. Analytical formulas have been obtained for calculating the profiles of concentrations of small ions (aeroions), the density of the space electric charge and the electric field strength in a turbulent electrode layer. As a result of mathematical modeling, the dependences electrical characteristics in surface layer on the values of the electric field, the turbulent mixing degree and aerosol pollution of the atmosphere are investigated. It is shown that the parameter of the electrode effect (the ratio of the values of the electric field strength on the earth's surface and at the upper boundary of the electrode layer) practically does not depend on atmospheric conditions, whereas the height of the electrode layer and, accordingly, the scale of the distribution of the electrical characteristics of the surface layer vary significantly. The intensification of turbulent mixing in the surface layer leads to an increase in the height of the electrode layer and, as a consequence, the scale of distribution of its parameters. The strengthening of the electric field or air pollution by aerosol parti-

cles of sufficient concentration leads to a decrease in its height. An increase in the concentration of aerosol particles in the atmosphere reduces the values of the electric charge density at the earth's surface. Theoretical calculations are in good agreement with experimental data and the results of numerical modeling of the surface layer electrical structure. The analytical formulas obtained in the work for calculating the electrical characteristics of the surface layer and the results of calculations can be useful in solving a number of applied problems of geophysics, in particular for monitoring the electrical state of the atmosphere.

Mathematical modeling; electrostatics; surface layer; atmosphere; turbulent electrode effect; electric field; aeroions; bulk charge; aerosol.

Введение. Электрическое поле атмосферы, обусловленное действием глобальных токовых генераторов [1], подвергается воздействию электродного эффекта вблизи поверхности земли [1–4]. Структура образующегося электродного слоя высотой от нескольких до нескольких десятков метров зависит от значений напряженности электрического поля, ионизации воздуха, турбулентного перемешивания, а также степени аэрозольного и радиоактивного загрязнений в приземном слое. Известные электродинамические модели приземного слоя основываются на теории классического (нетурбулентного) или турбулентного электродного эффекта в атмосфере [1–4]. Имеется достаточно большое количество результатов численных расчетов электрической структуры приземного слоя [3–7]. Однако, для практических расчетов электрических характеристик атмосферы представляют интерес аналитические выражения, получаемые методами математического моделирования.

В работах [8–10] представлен анализ общей системы электродинамических уравнений для приземного слоя атмосферы методами теории подобия. Определены условия стационарности процессов. Получены критерии подобия для различных электрогидродинамических условий в атмосфере, позволяющие обосновано применять те или иные модели электродного эффекта. Определены характерные масштабы электродинамических процессов в приземном слое атмосферы. В рамках модели классического электродного эффекта получены аналитические выражения для распределений концентраций полярных аэроионов ($n_{1,2}$) и электрического поля (E) в различных условиях, а также рассчитаны значения плотности объемного заряда и плотности электрического тока [10].

Цель настоящей работы – математическое моделирование электродинамического состояния турбулентного приземного слоя с учетом влияния аэрозольного загрязнения воздуха.

1. Постановка задачи. Электродинамическая модель стационарного горизонтально-однородного свободного от аэрозоля приземного слоя атмосферы в приближении турбулентного электродного эффекта [3, 4, 6, 7] состоит из ионизационно-рекомбинационных уравнений и уравнения Пуассона:

$$\frac{\partial n_{1,2}}{\partial t} \pm \frac{\partial}{\partial z} (b_{1,2} \cdot n_{1,2} E) - \frac{\partial}{\partial z} (D_T(z) \frac{\partial n_{1,2}}{\partial z}) = q(z) - \alpha n_1 n_2, \quad (1)$$

$$\frac{\partial E}{\partial z} = \frac{e}{\epsilon_0} (n_1 - n_2)$$

где $b_{1,2}$ – подвижности полярных аэроионов, α – коэффициент их рекомбинации, $D_T = D_j z$ – коэффициент турбулентной диффузии аэроионов, q – скорость ионизации воздуха, e – элементарный заряд, ϵ_0 – электрическая постоянная.

Введем в (1) безразмерные переменные, используя характерные значения и масштабы входящих в них физических величин:

$$z' = z/l_1, \quad n'_{1,2} = n_{1,2}/n_\infty, \quad E' = E/E_\infty, \quad n_\infty = \sqrt{q_\infty/\alpha}, \quad l_1 = D_1 \cdot \tau, \quad \tau = (q_\infty \alpha)^{-1/2}$$

Тогда получаем уравнения в безразмерной форме:

$$-\frac{\partial}{\partial z'} \left(z' \frac{\partial n'_{1,2}}{\partial z'} \right) \pm \xi_{1,2} \frac{\partial}{\partial z'} (n'_{1,2} E') = \frac{q}{q_\infty} - n'_1 n'_2, \quad \xi_{1,2} = \frac{|b_{1,2}| \cdot E_\infty \cdot \tau}{l_1}, \quad (2)$$

$$\frac{\partial E}{\partial z} = \gamma (n'_1 - n'_2), \quad \gamma = \frac{e}{\varepsilon_0} \frac{n_\infty}{E_\infty} l_1.$$

Знак «∞» означает верхнюю границу (Z_∞) электродного слоя. Систему (2) характеризуют два безразмерных параметра (критерии подобия) $\xi_{1,2}$ и γ , которые определяют тип модели электродного эффекта [1–3].

2. Модель турбулентного электродного слоя в чистой атмосфере. Рассмотрим случай $\xi_{1,2} < 1$, когда имеет место турбулентный электродный эффект, т.е. перенос аэроионов осуществляется, наряду с электрическими силами, турбулентными потоками воздуха [2, 3, 4–8]. Также будем предполагать выполненным условие $|\gamma| \ll 1$, тогда объёмным зарядом можно пренебречь, т.е. электрическое поле можно считать его постоянным.

Итак, для получения аналитического решения из уравнений системы (1) сделаем следующие допущения: $E = E_0 = \text{const}$, $q = \text{const}$, $D_T(z, t) = D_0 = \text{const}$. Используя идею расщепления по физическим процессам [9–11], линеаризуем правые части ионизационно-рекомбинационных уравнений, положив в первом уравнении $n_2 = n_\infty$, а во втором – $n_1 = n_\infty$, и тогда получаем:

$$\pm b_{1,2} E_0 \frac{dn_{1,2}}{dz} - D_0 \frac{d^2 n_{1,2}}{dz^2} = q - \alpha n_\infty n_{1,2}. \quad (3)$$

Зададим граничные условия как:

$$n_{1,2}(z = z_0) = 0, \quad n_{1,2}(z = z_\infty) = \sqrt{q_\infty / \alpha} = n_\infty, \quad (4)$$

где z_0 – масштаб шероховатости земной поверхности [1–4].

Решая совместно (3) и (4), получаем аналитические выражения для высотных профилей полярных аэроионов.

$$n_1(z) = \frac{q}{\alpha n_\infty} \left(1 - e^{k_{12}(z-z_0)} \right), \quad (5)$$

$$n_2(z) = \frac{q}{\alpha n_\infty} \left(1 - e^{k_{22}(z-z_0)} \right),$$

где

$$k_{12} = \frac{b_1 E_0 - \sqrt{4D_0 \alpha n_\infty + (b_1 E_0)^2}}{2D_0}, \quad k_{22} = \frac{-b_2 E_0 - \sqrt{4D_0 \alpha n_\infty + (b_2 E_0)^2}}{2D_0}.$$

Плотность электрического заряда может быть рассчитана по формуле:

$$\rho(z) = e(n_1(z) - n_2(z)). \quad (6)$$

Тогда получаем:

$$\rho(z) = \frac{eq}{\alpha n_\infty} \left(e^{k_{22}(z-z_0)} - e^{k_{12}(z-z_0)} \right). \quad (7)$$

Для расчёта профиля напряжённости электрического поля воспользуемся уравнением для стационарного полного электрического тока, вытекающего из системы (1) [3, 4].

Значение полного тока можно рассчитать по закону Ома на верхней границе электродного слоя:

$$j_0 = \lambda_\infty \cdot E_0, \quad \lambda_\infty = e(b_1 n_1(\infty) + b_2 n_2(\infty)), \quad (8)$$

где $n_{1,2}(\infty) = (q/\alpha)^{1/2}$ – асимптотические значения концентраций аэроионов на верхней границе электродного слоя.

С учётом сделанных выше допущений, а также полагая постоянным значение электрической проводимости воздуха (λ), уравнение для полного электрического тока примет вид:

$$-\varepsilon_0 D_0 \frac{d^2 E}{dz^2} + \lambda E = j_0. \quad (9)$$

Граничные условия зададим в виде:

$$E(z = z_0) = E_0, \quad E(z \rightarrow \infty) = E_\infty. \quad (10)$$

Решением системы (9) и (10) является профиль электрического поля в турбулентном электродном слое:

$$E(z) = (E_0 - E_\infty) e^{-\frac{(z-z_0)\sqrt{\lambda}}{\varepsilon_0 D_0}} + E_\infty. \quad (11)$$

3. Модель турбулентного электродного слоя в условиях аэрозольного загрязнения. Рассмотрим модель турбулентного электродного слоя, заполненный монодисперсными аэрозольными частицами концентрацией N . Тогда в правой части ионизационно-рекомбинационных уравнений в системе (1), можно добавить член $BNn_{1,2}$, описывающий взаимодействие аэроионов с аэрозолем [12–18]. Параметр $B = \frac{2\eta_1\eta_2}{\eta_1 + 2\eta_2}$ зависит от коэффициентов воссоединения полярных аэроионов ($\eta_{1,2}$) легких ионов с аэрозольными частицами [14, 17]. Очевидно, что для достаточно больших концентраций аэрозоля ($N \geq n_{1,2}$) рекомбинационным членом в правых частях этих уравнений можно пренебречь.

Используя сделанные выше допущения, перейдем к следующей системе уравнений с учетом аэрозольных частиц в воздухе с соответствующими граничными условиями:

$$-D_0 \frac{d^2 n_{1,2}}{dz^2} \pm b_{1,2} E_0 \frac{dn_{1,2}}{dz} = q - BNn_{1,2}, \quad (12)$$

$$n_{1,2}(z = z_0) = 0, \quad n_{1,2}(z \rightarrow \infty) = \frac{q}{BN}.$$

Решением (12) являются профили концентраций полярных аэроионов:

$$n_1(z) = \frac{q}{BN} (1 - e^{k_{11}(z-z_0)}), \quad (13)$$

$$n_2(z) = \frac{q}{BN} (1 - e^{k_{21}(z-z_0)}).$$

где

$$k_{11} = \frac{b_1 E_0 - \sqrt{4D_0 BN + (b_1 E_0)^2}}{2D_0}, \quad k_{21} = \frac{-b_2 E_0 - \sqrt{4D_0 BN + (b_2 E_0)^2}}{2D_0}.$$

Тогда для плотности объемного заряда получаем:

$$\rho(z) = \frac{eq}{BN} (e^{k_{21}(z-z_0)} - e^{k_{11}(z-z_0)}). \quad (14)$$

Расчет профиля напряженности электрического поля осуществляется по формуле (11) с учетом изменённых значений электрической проводимости λ и λ_∞ , а, следовательно, и плотности тока j_0 , за счет присутствия аэрозоля в атмосфере.

4. Результаты вычислений и их анализ. Расчёты электрических характеристик приземного слоя атмосферы проводились по формулам (7), (11), (14) для чистой и загрязнённой аэрозольными частицами атмосферы. Физические параметры, входящие в аналитические выражения, задавались следующими: $b_1 = 1,2 \cdot 10^{-4} \text{ м}^2 \text{ В}^{-1}$, $b_2 = 1,4 \cdot 10^{-4} \text{ м}^2 \text{ с}^{-1} \text{ В}^{-1}$, $\alpha = 1,2 \cdot 10^{-12} \text{ м}^3 \text{ с}^{-1}$, $e = 1,6 \cdot 10^{-19} \text{ Кл}$, $q = 7 \cdot 10^6 \text{ м}^{-3} \text{ с}^{-1}$, $|E_0| = 50; 100; 200 \text{ В} \cdot \text{м}^{-1}$, $j_0 \sim 10^{-12} \text{ А} \cdot \text{м}^{-2}$, $\lambda \sim 10^{-14} \text{ См} \cdot \text{м}^{-1}$, $N \sim 10^{10} \text{ м}^{-3}$, $\varepsilon_0 = 8,85 \cdot 10^{-12} \text{ Ф} \cdot \text{м}^{-1}$, $D_0 = 0,01; 0,1 \text{ м}^2 / \text{с}$, $\eta_1 = 1,4 \cdot 10^{-12} \text{ м}^{-3} \text{ с}^{-1}$, $\eta_2 = 4 \cdot 10^{-12} \text{ м}^{-3} \text{ с}^{-1}$, $z_0 = 2,5 \cdot 10^{-3} \text{ м}$.

Для расчетов напряженности электрического поля использовались значения параметров, следующие из решений ионизационно-рекомбинационных уравнений (3), (4) и (12).

В чистой атмосфере значения электрической проводимости и тока: $\lambda = (3-9) \cdot 10^{-14} \text{ См} \cdot \text{м}^{-1}$, $\lambda_\infty = 9,25 \cdot 10^{-14} \text{ См} \cdot \text{м}^{-1}$, откуда значение полного тока равно $j_0 = \lambda_\infty \cdot E_0 \approx 9 \cdot 10^{-12} \text{ А} \cdot \text{м}^{-2}$. Тогда для расчета профиля напряженности электрического поля по формуле (11) получаем следующие значения параметров $\lambda \approx 6 \cdot 10^{-14} \text{ См} \cdot \text{м}^{-1}$, $j_0 \approx 9 \cdot 10^{-12} \text{ А} \cdot \text{м}^{-2}$.

На рис. 1-2 построены профили плотности объемного заряда и напряженности электрического поля в приземном слое в чистой атмосфере.

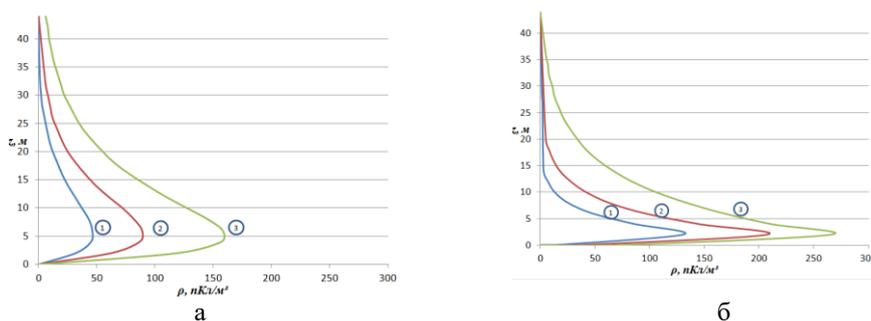


Рис. 1. Профили плотность объемного заряда в приземном слое без аэрозоля: (для $D_0 = 0,1 \text{ м}^2 / \text{с}$ (а) и $D_0 = 0,01 \text{ м}^2 / \text{с}$ (б); кривые (1,2,3) при $E_0 = -50 ; -100 ; -200 \text{ , [В / м]}$)

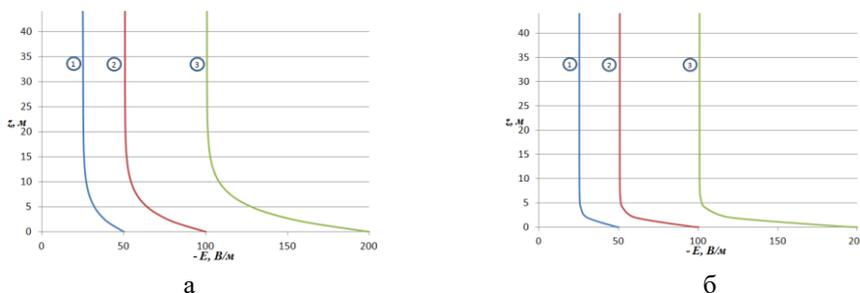


Рис. 2. Профили электрического поля в приземном слое без аэрозоля: (для $D_0 = 0,1 \text{ м}^2 / \text{с}$ (а) и $D_0 = 0,01 \text{ м}^2 / \text{с}$ (б); кривые (1,2,3) при $E_0 = -50 ; -100 ; -200 \text{ , [В / м]}$)

Аналогично, при наличии аэрозольных частиц в воздухе концентрацией $N \sim 10^{10} \text{ м}^{-3}$ имеем $\lambda = (1,3 - 2,6) \cdot 10^{-14} \text{ См} \cdot \text{м}^{-1}$, $\lambda_{\infty} = 2,65 \cdot 10^{-14} \text{ См} \cdot \text{м}^{-1}$, и, соответственно, значение полного тока принимает значение $j_0 = \lambda_{\infty} \cdot E_0 \approx 2,5 \cdot 10^{-12} \text{ А} \cdot \text{м}^{-2}$, а параметры в формуле (11) равны $\lambda \approx 2 \cdot 10^{-14} \text{ См} \cdot \text{м}^{-1}$, $j_0 \approx 2,5 \cdot 10^{-12} \text{ А} \cdot \text{м}^{-2}$.

На рис. 3-4 приведены профили плотности заряда и напряжённости электрического поля в приземном слое, заполненного аэрозольными частицами.

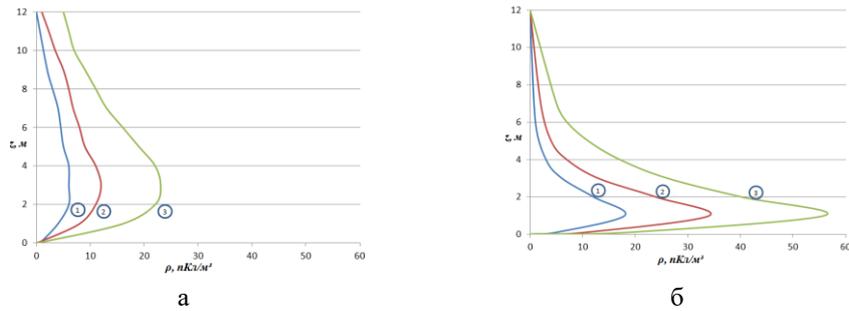


Рис. 3. Профили плотность объемного заряда в условиях аэрозольного загрязнения (для $D_0 = 0,1 \text{ м}^2 / \text{с}$ (а) и $D_0 = 0,01 \text{ м}^2 / \text{с}$ (б); кривые (1,2,3) при $E_0 = -50 ; -100 ; -200 \text{ , [В / м]}$)

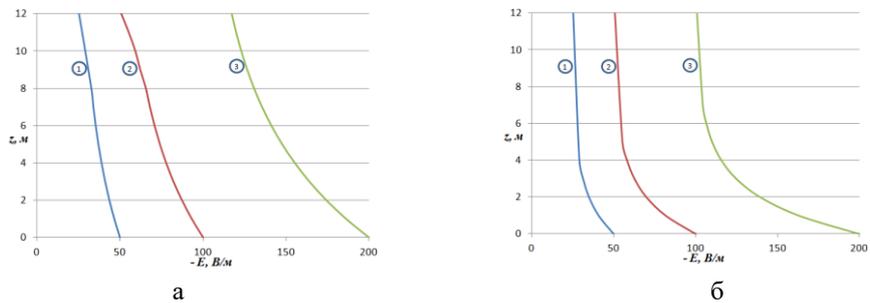


Рис. 4. Профили электрического поля в условиях аэрозольного загрязнения: (для $D_0 = 0,1 \text{ м}^2 / \text{с}$ (а) и $D_0 = 0,01 \text{ м}^2 / \text{с}$ (б); кривые (1,2,3) при $E_0 = -50 ; -100 ; -200 \text{ , [В / м]}$)

В ходе вычислительных экспериментов получены следующие результаты. В условиях турбулентного перемешивания в приземном слое атмосферы высота электродного слоя увеличивается по сравнению с аналогичными расчетами для не-турбулентного случая [11], а его электрическая структура зависит от значений коэффициента турбулентной диффузии. Знак объемного электрического заряда положительный, а максимальное значение его плотности наблюдается у поверхности земли.

Как и в случае классического электродного эффекта [11], появление в приземном воздухе аэрозольных частиц концентрацией сопоставимой с числом аэроионов приводит к уменьшению высоты турбулентного электродного слоя и значений плотности электрического заряда.

С усилением турбулентного перемешивания максимальные значения плотности заряда уменьшаются, значения напряженности электрического поля увеличиваются вследствие увеличения характерных масштабов распределения величин.

Заключение. Методами математического моделирования построены и исследованы аналитические модели турбулентного электродного слоя в чистой атмосфере и в условиях аэрозольного загрязнения. На основании теории подобия сделаны обоснованные физические допущения и получены решения электродинамических уравнений, расщепленных по физическим процессам. Получены распределения концентраций полярных аэроионов, электрической проводимости, плотности объемного заряда и электрического поля. Исследованы их зависимости от значений электрического поля, степени турбулентного перемешивания и аэрозольного загрязнения воздуха. Полученные аналитические профили электрических характеристик приземного слоя хорошо согласуются с известными теоретическими и экспериментальными данными [3–5, 19, 20] и могут быть полезны при решении ряда прикладных задач геофизики и физики атмосферы [21, 22].

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Морозов В.Н.* Атмосферное электричество // Атмосфера. Справочник (справочные данные, модели). – Л.: Гидрометеониздат, 1991. – С. 394-408.
2. *Куповых Г.В.* Электричество приземного слоя // Известия высших учебных заведений. Северо-Кавказский регион. Естественные науки. – 1995. – № 4. – С. 32-34.
3. *Куповых Г.В., Морозов В.Н., Шварц Я.М.* Теория электродного эффекта в атмосфере. – Таганрог: Изд-во ТРТУ, 1998. – 123 с.
4. *Куповых Г.В.* Электродинамические процессы в приземном слое атмосферы. – Таганрог: Изд-во ТТИ ЮФУ, 2009. – 114 с.
5. *Hoppel W.A.* Electrode effect: comparison of the theory and measurement // In: Planetary Electrodynamics, 2. S.C. Coroniti and J. Hughes; editors. – Gordon and Breach Science Publishers, New York, 1969. – P. 167-181.
6. *Куповых Г.В., Морозов В.Н.* Турбулентный электродный эффект в приземном слое // Известия высших учебных заведений. Северо-Кавказский регион. Естественные науки. Приложение № 3. – 2003. – С. 55-62.
7. *Редин А.А., Клово А.Г., Куповых Г.В.* Электродинамическая модель атмосферного приземного слоя // Известия ЮФУ. Технические науки. – 2009. – № 8. – С. 93-106.
8. *Свидельский С.С., Литвинова В.С., Куповых Г.В., Клово А.Г.* Формирование структуры атмосферного электродного слоя // Известия ЮФУ. Технические науки. – 2020. – № 5. – С. 130-141.
9. *Куповых Г.В., Клово А.Г., Кудринская Т.В., Тимошенко Д.В.* Математическое моделирование в электродинамике приземного слоя атмосферы // Матер. XIV Международной конференции «Фундаментальные и прикладные проблемы математики и информатики». – Изд-во ДГУ, 2021. – С. 141-144.
10. *Куповых Г.В., Клово А.Г., Кудринская Т.В., Тимошенко Д.В.* Электродинамические модели приземного слоя атмосферы // X Всероссийская научная конференция «Системный синтез и прикладная синергетика»: Сб. научных трудов (п. Нижний Архыз, Россия 28 сентября – 02 октября 2021 г.). – Ростов-на-Дону; Таганрог: Изд-во ЮФУ, 2021. – С. 190-195.
11. *Куповых Г.В., Клово А.Г., Гривцов В.В., Белоусова О.В.* Моделирование электродинамической структуры нетурбулентного приземного слоя // Известия ЮФУ. Технические науки. – 2022. – № 3 (227). – С. 234-243.
12. *Hoppel W.A., Frick G.M.* Ion-Aerosol Attachment Coefficients and the Steady-State Charge Distribution on Aerosols in a Bipolar Ion Environment // Aerosol Science and Technology. – 1986. – 5:1. – P. 1-21.
13. *Hoppel W.A.* Ionaerosol attachment, ion depletion and charge distribution on aerosols // J. Geoph. Res. – 1985. – Vol. 90, No. D4. – P. 5917-5923.
14. *Редин А.А., Клово А.Г., Куповых Г.В.* Математическое моделирование электродинамической структуры приземного слоя атмосферы в условиях аэрозольного загрязнения // Известия ЮФУ. Технические науки. – 2009. – № 7 (96). – С. 192-200.

15. Редин А.А., Клово А.Г., Куповых Г.В., Морозов В.Н. Генерация объемного заряда вблизи поверхности земли с учетом взаимодействия аэрозольных частиц с аэроионами // Известия высших учебных заведений. Северо-Кавказский регион. Естественные науки. Спецвыпуск. Физика атмосферы. – 2010. – С. 81-85.
16. Kupovykh G., Redin A., Boldyreff A. Modeling of ionization-recombination processes in the atmospheric surface layer // Journal of Electrostatics 71. Elsevier B.V. – 2013. – P. 305-311.
17. Куповых Г.В., Клово А.Г., Тимошенко Д.В. Оценка влияния аэрозольного загрязнения на электрические характеристики приземной атмосферы // Матер. V Всероссийской научной конференции «Проблемы военноприкладной геофизики и контроля состояния природной среды». – СПб.: ВКА им. А.Ф. Можайского, 2018. – С. 483-486.
18. Kupovykh G.V., Timoshenko D.V., Klovo A. G., Kudrinskaya T.V. Electrodynamic processes models in atmospheric surface layer // IOP Conference Series: Materials Science and Engineering. – 2019. – Vol. 698. – 044034. – 8 p.
19. Kupovykh G.V., Timoshenko D.V., Kudrinskaya T.V., Klovo A.G. Modeling of the atmospheric electric field local variations in the turbulent surface layer // Journal of Physics: IOP Conference Series. VIII All-Russian Conference on Atmospheric Electricity. – 2020. – 1604. – 012003. – 5 p.
20. Кудринская Т.В., Куповых Г.В., Редин А.А. Исследования ионизационного состояния приземного слоя атмосферы в разных геофизических условиях // Метеорология и гидрология. – 2018. – № 4. – С. 77-85.
21. Шулейкин В.Н., Шукин Г.Г., Куповых Г.В. Развитие методов и средств прикладной геофизики – атмосферно-электрический мониторинг геологических неоднородностей и зон геодинамических процессов. – СПб.: РГГМУ, 2015. – 206 с.
22. Kupovykh G.V., Kudrinskaya T.V., Grivtsov V.V. The atmosphere electrical characteristics' monitoring as an element of technosphere safety // IOP Conference Series: Materials Science and Engineering. – 2020. – 913. – 052041. – 6 p.

REFERENCES

1. Morozov V.N. Atmosfernoe elektrichestvo [Atmospheric electricity], *Atmosfera. Spravochnik (spravochnye dannye, modeli)* [Atmosphere. Handbook (reference data, models)]. Leningrad: Gidrometeoizdat, 1991, pp. 394-408.
2. Kupovykh G.V. Elektrichestvo prizemnogo sloya [Electricity of the surface layer], *Izvestiya vysshikh uchebnykh zavedeniy. Severo-Kavkazskiy region. Estestvennye nauki* [News of higher educational institutions. The North Caucasus region. Natural sciences], 1995, No. 4, pp. 32-34.
3. Kupovykh G.V., Morozov V.N., Shvarts Ya.M. Teoriya elektrodnoogo effekta v atmosfere [Theory of the electrode effect in the atmosphere]. Taganrog: Izd-vo TRTU, 1998, 123 p.
4. Kupovykh G.V. Elektrodinamicheskie protsessy v prizemnom sloe atmosfery [Electrodynamic processes in the surface layer of the atmosphere]. Taganrog: Izd-vo TTI YuFU, 2009, 114 p.
5. Hoppel W.A. Electrode effect: comparison of the theory and measurement, *In: Planetary Electrodynamics*, 2. S.C. Coroniti and J. Hughes; editors. Gordon and Breach Science Publishers, New York, 1969, pp. 167-181.
6. Kupovykh G.V., Morozov V.N. Turbulentnyy elektrodnyy effekt v prizemnom sloe [Turbulent electrode effect in the surface layer], *Izvestiya vysshikh uchebnykh zavedeniy. Severo-Kavkazskiy region. Estestvennye nauki. Prilozhenie № 3* [Izvestia of higher educational institutions. The North Caucasus region. Natural sciences. Appendix No. 3], 2003, pp. 55-62.
7. Redin A.A., Klovo A.G., Kupovykh G.V. Elektrodinamicheskaya model' atmosfernogo prizemnogo sloya [Electrodynamic model of the atmospheric near-earth layer], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2009, No. 8, pp. 93-106.
8. Svidel'skiy S.S., Litvinova V.S., Kupovykh G.V., Klovo A.G. Formirovanie struktury atmosfernogo elektrodnoogo sloya [Formation of the structure of the atmospheric electrode layer], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2020, No. 5, pp. 130-141.
9. Kupovykh G.V., Klovo A.G., Kudrinskaya T.V., Timoshenko D.V. Matematicheskoe modelirovanie v elektrodinamike prizemnogo sloya atmosfery [Mathematical modeling in the electrostatics of the surface layer of the atmosphere], *Mater. XIV Mezhdunarodnoy konferentsii «Fundamental'nye i prikladnye problemy matematiki i informatiki»* [Proceedings of the XIV International Conference "Fundamental and Applied problems of Mathematics and Computer Science"]. Izd-vo DGU, 2021, pp. 141-144.

10. Kupovykh G.V., Klovo A.G., Kudrinskaya T.V., Timoshenko D.V. Elektrodynamic models of the surface layer of the atmosphere, *X Vserossiyskaya nauchnaya konferentsiya «Sistemnyy sintez i prikladnaya sinergetika»: Sb. nauchnykh trudov (p. Nizhny Arkhyz, Rossiya 28 sentyabrya – 02 oktyabrya 2021 g.)* [Collection of scientific papers (Nizhny Arkhyz, Russia September 28 – October 02, 2021)]. Rostov-on-Don; Taganrog: Izd-vo YuFU, 2021, pp. 190-195.
11. Kupovykh G.V., Klovo A.G., Grivtsov V.V., Belousova O.V. Modelirovanie elektrodinamicheskoy struktury neturbulentnogo prizemnogo sloya [Modeling of the electrodynamic structure of a non-turbulent surface layer], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 3 (227). pp. 234-243.
12. Hoppel W.A., Frick G.M. Ion-Aerosol Attachment Coefficients and the Steady-State Charge Distribution on Aerosols in a Bipolar Ion Environment, *Aerosol Science and Technology*, 1986, 5:1, pp. 1-21.
13. Hoppel W.A. Ionaerosol attachment, ion depletion and charge distribution on aerosols, *J. Geoph. Res.*, 1985, Vol. 90, No. D4, pp. 5917-5923.
14. Redin A.A., Klovo A.G., Kupovykh G.V. Matematicheskoe modelirovanie elektrodinamicheskoy struktury prizemnogo sloya atmosfery v usloviyakh aerazol'nogo zagryazneniya [Mathematical modeling of the electrodynamic structure of the surface layer of the atmosphere under conditions of aerosol pollution], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2009, No. 7 (96), pp. 192-200.
15. Redin A.A., Klovo A.G., Kupovykh G.V., Morozov V.N. Generatsiya ob'emnogo zaryada vblizi poverkhnosti zemli s uchedom vzaimodeystviya aerazol'nykh chastits s aeroionami [Generation of a volumetric charge near the Earth's surface, taking into account the interaction of aerosol particles with aeroions], *Izvestiya vysshikh uchebnykh zavedeniy. Severo-Kavkazskiy region. Estestvennyye nauki. Spetsvypusk. Fizika atmosfery* [Izvestiya vyshevykh uchebnykh uchebnykh iuzov. The North Caucasus region. Natural sciences. Special edition. Physics of the atmosphere], 2010, pp. 81-85.
16. Kupovykh G., Redin A., Boldyreff A. Modeling of ionization-recombination processes in the atmospheric surface layer, *Journal of Electrostatics* 71. – Elsevier B.V., 2013. – P. 305-311.
17. Kupovykh G.V., Klovo A.G., Timoshenko D.V. Otsenka vliyaniya aerazol'nogo zagryazneniya na elektricheskie kharakteristiki prizemnoy atmosfery [Assessment of the effect of aerosol pollution on the electrical characteristics of the surface atmosphere], *Mater. V Vserossiyskoy nauchnoy konferentsii «Problemy voennoprikladnoy geofiziki i kontrolya sostoyaniya prirodnoy sredy»* [Materials of the V All-Russian Scientific Conference "Problems of military geophysics and control of the state of the natural environment"]. Saint Petersburg: VKA im. A.F. Mozhayskogo, 2018, pp. 483-486.
18. Kupovykh G.V., Timoshenko D.V., Klovo A. G., Kudrinskaya T.V. Electrodynamic processes models in atmospheric surface layer, *IOP Conference Series: Materials Science and Engineering*, 2019, Vol. 698, 044034, 8 p.
19. Kupovykh G.V., Timoshenko D.V., Kudrinskaya T.V., Klovo A.G. Modeling of the atmospheric electric field local variations in the turbulent surface layer, *Journal of Physics: IOP Conference Series. VIII All-Russian Conference on Atmospheric Electricity*, 2020, 1604, 012003, 5 p.
20. Kudrinskaya T.V., Kupovykh G.V., Redin A.A. Issledovaniya ionizatsionnogo sostoyaniya prizemnogo sloya atmosfery v raznykh geofizicheskikh usloviyakh [Studies of the ionization state of the surface layer of the atmosphere in different geophysical conditions], *Meteorologiya i gidrologiya* [Meteorology and hydrology], 2018, No. 4, pp. 77-85.
21. Shuleykin V.N., Shchukin G.G., Kupovykh G.V. Razvitie metodov i sredstv prikladnoy geofiziki – atmosferno-elektricheskiy monitoring geologicheskikh neodnorodnostey i zon geodinamicheskikh protsessov [Development of methods and means of applied geophysics – atmospheric-electrical monitoring of geological inhomogeneities and zones of geodynamic processes]. Saint Petersburg: RGGMU, 2015, 206 p.
22. Kupovykh G.V., Kudrinskaya T.V., Grivtsov V.V. The atmosphere electrical characteristics' monitoring as an element of technosphere safety, *IOP Conference Series: Materials Science and Engineering*, 2020, 913, 052041, 6 p.

Статью рекомендовал к опубликованию д.ф.-м.н., профессор А.И. Жорник.

Белусова Ольга Васильевна – Южный федеральный университет; e-mail: obelousova@sfedu.ru; г. Таганрог, Россия; тел.: +78634371705; кафедра высшей математики; ассистент.

Куповых Геннадий Владимирович – e-mail: kupovykh@sfedu.ru; тел.: +78634371636; кафедра высшей математики; зав. кафедрой; д.ф.-м.н.; профессор.

Клово Александр Георгиевич – e-mail: agklovo@sfedu.ru; тел.: +78634371705; кафедра высшей математики; к.ф.-м.н.; доцент.

Гривцов Владимир Владиславович – e-mail: agklovo@sfedu.ru; тел.: +78634371794; кафедра инженерной графики и компьютерного дизайна; к.т.н.; доцент.

Belousova Olga Vasilievna – Southern Federal University; e-mail: obelousova@sfedu.ru; Taganrog, Russia; phone: +78634371705; the department of the higher mathematician; assistant.

Kupovkh Gennady Vladimirovich – e-mail: kupovykh@sfedu.ru; phone: +78634371636; the department of higher mathematics; head of department; dr. of phys. and math. sc.; professor.

Klovo Alexander Georgievich – e-mail: agklovo@sfedu.ru; phone: +78634371705; the department of higher mathematics; cand. of phys. and math. sc.; associate professor.

Gritvsov Vladimir Vladislavovich – e-mail: agklovo@sfedu.ru; phone: +78634371794; the department of engineering graphics and computer design; cand. of eng. sc.; associate professor.

ПРАВИЛА ОФОРМЛЕНИЯ РУКОПИСЕЙ

1. Объем статьи должен быть не менее 12 и не более 18 страниц. Формат (А 4). Редактор **Word 7 for Windows**, шрифт Times New Roman, размер 14, интервал 1,5. Авторы представляют в редакцию 1 экз. статьи и идентичный электронный вариант.

2. Названию статьи предшествует индекс УДК, соответствующий заявленной теме.

3. Текст статьи начинается с названия статьи (на русском и английском языках), фамилии, имени и отчества автора (полностью) и снабжается аннотацией на русском и английском языках объемом **не менее 250-300 слов**. В тексте аннотации указывается цель, задачи исследования и краткие выводы. В аннотации **не следует** давать ссылку на номер публикации в списке литературы к статье. После аннотаций приводятся ключевые слова (словосочетания), несущие в тексте основную смысловую нагрузку (на русском и английском языках).

4. В тексте статьи следует использовать минимальное количество таблиц и иллюстраций. Рисунок должен иметь объяснения значений всех компонентов, порядковый номер, название, расположенное под рисунком. В тексте на рисунок дается ссылка. Таблица должна иметь порядковый номер, заголовок, расположенный над ней. Данные таблиц и рисунков не должны дублировать текст. Формулы должны быть набраны **в редакторе формул Word 7 for Windows**.

5. Цитаты тщательно сверяются с первоисточником и визируются автором на обратной стороне последней страницы: "Цитаты и фактический материал сверены". Подпись, дата.

6. Наличие пристатейного библиографического списка на русском и английском языках обязательно. **Ссылок должно быть не менее 20-ти**, из них на зарубежные источники – не менее 35 %. В тексте ссылки должны быть в квадратных скобках.

Примеры оформления литературы: а) для книг: фамилия, инициалы автора(ов), полное название книги, место, год издания, страницы; б) для статей: фамилия и инициалы автора(ов), полное название сборника, книги, газеты, журнала, где опубликована статья, место и год издания (сборника, книги), номер (для журнала), год и дата (для газеты), выпуск, часть (для сборника), страницы, на которых опубликована статья. Иностранная литература оформляется по тем же правилам.

Ссылки на неопубликованные работы не допускаются.

7. Рукопись должна быть тщательно вычитана. Редакционная коллегия оставляет за собой право при необходимости сокращать статьи, редактировать и отсылать авторам на доработку.

8. Статьи сопровождаются сведениями об авторе(ах) (фамилия, имя, отчество, ученое звание, должность, место работы, адрес, электронный адрес и номер телефона) на русском и английском языках.

9. Плата с аспирантов за публикацию рукописей не взимается.

Адрес журнала в Интернете: <http://izv-tn.tti.sfedu.ru/>.