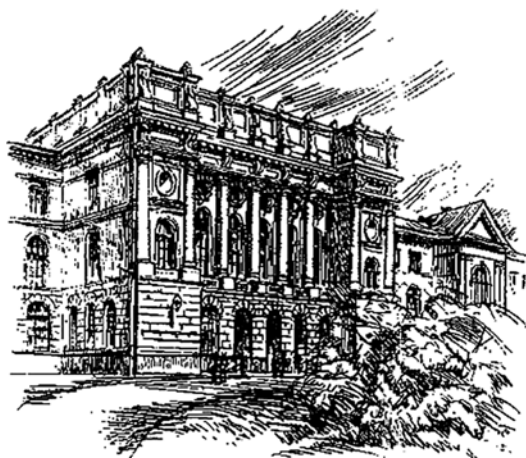


МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ



НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ

САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО
ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

Информатика. Телекоммуникации.
Управление

4(176) 2013

Издательство Политехнического университета
Санкт-Петербург
2013

НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ САНКТ-ПЕТЕРБУРГСКОГО ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА

РЕДАКЦИОННЫЙ СОВЕТ

Васильев Ю.С., академик РАН (председатель); *Алферов Ж.И.*, академик РАН;
Костюк В.В., академик РАН; *Лопота В.А.*, чл.-кор. РАН;
Окрепилов В.В., академик РАН; *Рудской А.И.*, чл.-кор. РАН;
Патон Б.Е., академик НАН Украины и РАН;
Федоров М.П., академик РАН.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Васильев Ю.С., академик РАН (главный редактор); *Арсеньев Д.Г.*, д-р техн. наук, профессор;
Бабкин А.В., д-р экон. наук, профессор (зам. гл. редактора);
Боронин В.Н., д-р техн. наук, профессор; *Глухов В.В.*, д-р экон. наук, профессор;
Дегтярева Р.В., д-р ист. наук, профессор; *Иванов А.В.*, д-р техн. наук, профессор;
Иванов В.К., д-р физ.-мат. наук, профессор; *Козловский В.В.*, д-р физ.-мат. наук, профессор;
Рудской А.И., чл.-кор. РАН (зам. гл. редактора); *Юсупов Р.М.*, чл.-кор. РАН.

ИНФОРМАТИКА. ТЕЛЕКОММУНИКАЦИИ. УПРАВЛЕНИЕ

РЕДАКЦИОННЫЙ СОВЕТ ЖУРНАЛА

Юсупов Р.М., чл.-кор. РАН – председатель;
Абрамов С.М., чл.-кор. РАН;
Воеводин В.В., чл.-кор. РАН;
Заборовский В.С., д-р техн. наук, профессор;
Козлов В.Н., д-р техн. наук, профессор;
Фотиади А.Э., д-р физ.-мат. наук, профессор;
Черноруцкий И.Г., д-р техн. наук, профессор.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА

Юсупов Р.М., чл.-кор. РАН – председатель;
Арсеньев Д.Г., д-р техн. наук, профессор – зам. председателя;
Бабкин А.В., д-р экон. наук, профессор – зам. председателя;
Антонов В.И., д-р техн. наук, профессор;
Голландцев Ю.А., д-р техн. наук, профессор;
Карпов Ю.Г., д-р техн. наук, профессор;
Коротков А.С., д-р техн. наук, профессор;
Макаров С.Б., д-р техн. наук, профессор;
Устинов С.М., д-р техн. наук, профессор;
Цикин И.А., д-р техн. наук, профессор;
Шкодыврев В.П., д-р техн. наук, профессор;
Клавдиев В.Е., канд. техн. наук, доцент.

Журнал с 1995 года издается под научно-методическим руководством Российской академии наук. С 2008 года выпускается в составе сериального периодического издания «Научно-технические ведомости СПбГПУ» ISSN 1994-2354.

Журнал с 2002 года входит в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

Сведения о публикациях представлены в Реферативном журнале ВИНИТИ РАН, в международной справочной системе «Ulrich`s Periodical Directory».

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-51457 от 19.10.2012 г.

Подписной индекс **47517** в каталоге «Газеты. Журналы» Агентства «Роспечать».

Журнал включен в базу данных «Российский индекс научного цитирования» (РИНЦ), размещенную на платформе Национальной электронной библиотеки на сайте <http://www.elibrary.ru>

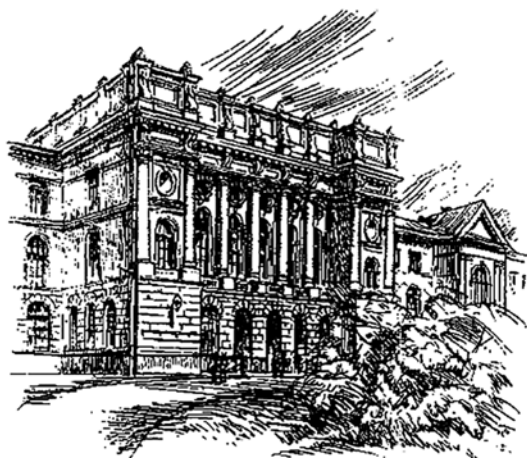
При перепечатке материалов ссылка на журнал обязательна.

Точка зрения редакции может не совпадать с мнением авторов статей.

Адрес редакции и издательства: Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29.
Тел. редакции (812) 552-62-16.

© Санкт-Петербургский государственный политехнический университет, 2013

THE MINISTRY OF EDUCATION AND SCIENCE OF THE RUSSIAN FEDERATION



ST. PETERSBURG STATE
POLYTECHNICAL UNIVERSITY
JOURNAL

Computer Science.
Telecommunications and Control Systems

4(176) 2013

Polytechnical University Publishing House
Saint Petersburg
2013

ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL

EDITORIAL COUNCIL

Y.S. Vasiliev – full member of the Russian Academy of Sciences, President of the St. Petersburg State Polytechnical University, editor-in-chief; *Zh.I. Alferov* – full member of the Russian Academy of Sciences; *V.V. Kostiuk* – full member of the Russian Academy of Sciences; *V.V. Lopota* – corresponding member of the Russian Academy of Sciences; *V.V. Okrepilov* – full member of the Russian Academy of Sciences; *B.E. Paton* – full member of the Russian Academy of Sciences and the National Academy of Sciences of Ukraine; *A.I. Rudskoy* – corresponding member of the Russian Academy of Sciences; *M.P. Fedorov* – full member of the Russian Academy of Sciences.

EDITORIAL BOARD

Y.S. Vasiliev – full member of the Russian Academy of Sciences, President of the St. Petersburg State Polytechnical University, editor-in-chief; *D.G. Arseniev* – Dr.Sc.(tech.), prof.; *A. V. Babkin* – Dr.Sc. (econ.), prof., deputy editor-in-chief; *V.N. Boronin* – Dr.Sc.(tech.), prof.; *V.V. Glukhov* – Dr.Sc. (econ.), prof.; *R. V. Degtyareva* – Dr.Sc. (history), prof.; *A. V. Ivanov* – Dr.Sc.(tech.); *V.K. Ivanov* – Dr.Sc.(phys.-math.), prof.; *V.V. Kozlovsky* – Dr.Sc.(phys.-math.), prof.; *A.I. Rudskoy* – corresponding member of the Russian Academy of Sciences, deputy editor-in-chief; *R.M. Yusupov* – corresponding member of the Russian Academy of Sciences.

COMPUTER SCIENCE. TELECOMMUNICATIONS AND CONTROL SYSTEMS

JOURNAL EDITORIAL COUNCIL

R.M. Yusupov – corresponding member of the Russian Academy of Sciences, head of the editorial council; *S.M. Abramov* – corresponding member of the Russian Academy of Sciences; *V.V. Voevodin* – corresponding member of the Russian Academy of Sciences; *V.S. Zaborovsky* – Dr.Sc.(tech.), prof.; *V.N. Kozlov* – Dr.Sc.(tech.), prof.; *A.E. Fotiadi* – Dr.Sc.(phys.-math.), prof.; *I.G. Chernorutsky* – Dr.Sc.(tech.), prof.

JOURNAL EDITORIAL BOARD

R.M. Yusupov – corresponding member of the Russian Academy of Sciences, head of the editorial board;
D.G. Arseniev – Dr.Sc.(tech.), prof., deputy head of the editorial board;
A. V. Babkin – Dr.Sc. (econ.), prof., deputy head of the editorial board;
V.I. Antonov – Dr.Sc.(tech.), prof.;
Y.A. Gollandtsev – Dr.Sc.(tech.), prof.;
Y.G. Karpov – Dr.Sc.(tech.), prof.;
A.S. Korotkov – Dr.Sc.(tech.), prof.;
S.B. Makarov – Dr.Sc.(tech.), prof.;
S.M. Ustinov – Dr.Sc.(tech.), prof.;
I.A. Tsikin – Dr.Sc.(tech.), prof.;
V.P. Shkodyrev – Dr.Sc.(tech.), prof.;
V.Ye. Klavdiev – Candidate of Technical Sciences, associate prof.

The journal is published under scientific and methodical guidance of the Russian Academy of Sciences since 1995. The journal is published since 2008 as part of the periodical edition «Nauchno-tekhnicheskie vedomosti SPbGPU» (ISSN 1994-2354).

The journal is included in the List of Leading Peer-Reviewed Scientific Journals and other editions to publish major findings of PhD theses for the research degrees of Doctor of Sciences and Candidate of Sciences.

The publications are presented in the VINITI RAS Abstract Journal and Ulrich's Periodical Directory International Database.

The journal is registered with the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR). Certificate ПИ № ФС77-51457 issued Oct. 19, 2012.

Subscription index **47517** in the «Journals and Magazines» catalogue, Rospechat agency.

The journal is on the Russian Science Citation Index (RSCI) database

© Scientific Electronic Library (<http://elibrary.ru/>).

No part of this publication may be reproduced without clear reference to the source.

The views of the authors can contradict the views of the Editorial Board.

The address: 195251 Polytekhnicheskaya Str. 29, St. Petersburg, Russia.

Содержание

Проблемы передачи и обработки информации

Иванков А.А., Шишагин А.Л. <i>Расширение объектной модели стэнфордского парсера для решения задачи идентификации семантических триплетов</i>	9
Глушков П.В., Михайлов Н.В., Юдакин Д.Е. <i>Измерение радионавигационных параметров сигнала СРНС слеящими системами без обратной связи в условиях многолучевого распространения.....</i>	18
Бендерская Е.Н., Никитин К.В. <i>Рекуррентная нейронная сеть как динамическая система и подходы к ее обучению</i>	29

Радиотехника, антенны, СВЧ-устройства

Сухов И.А., Акимов В.П. <i>Применение алгоритмов «сверхразрешения» к радиопеленгаторной антенной решетке из направленных элементов.....</i>	41
--	----

Системный анализ и управление

Кедрин В.С., Кузьмин О.В. <i>Частотный анализ временных рядов периодических функций с помощью оценки численного ранга</i>	47
Писарев А.И., Суляев И.И. <i>Нейроэмулятор теплового режима плавки медно-никелевого сырья в печи Ванюкова</i>	55

Вычислительные машины и программное обеспечение

Петросян Г.С. <i>Языковые средства поддержки систематической обработки ошибок.....</i>	65
---	----

Математическое моделирование: методы, алгоритмы, технологии

Потлов А.Ю., Проскурин С.Г. <i>Алгоритм цветового доплеровского картирования направлений потоков биологических жидкостей в оптической когерентной томографии</i>	71
Черноруцкий И.Г. <i>Практическая оптимизация и невыпуклые задачи.....</i>	79
Бортяков Д.Е., Мещеряков С.В., Солодилова Н.А. <i>Обеспечение качества автоматизированного проектирования металлоконструкций технологических систем на основе распределенной базы данных эксплуатационных дефектов.....</i>	87

Электроника, технологии производства материалов электронной техники

Пилюшко М.М. <i>Дельта-сигма модулятор аналого-цифрового преобразователя с преобразованием во времени.....</i>	95
---	----

Управление в социальных и экономических системах

Бостонов О.Х., Зверев Г.Н. <i>Семиотическая интеграция и управление бизнес-процессами организаций нефтяной отрасли.....</i>	103
--	-----

Высокопроизводительные вычисления

Макоха А.Н. <i>Исследование канонических тривекторов восьмого ранга с помощью теории графов и групп подстановок</i>	112
Мнухин В.Б. <i>Интегральное преобразование для распознавания симметрии изображений</i>	123
Червяков Н.И., Бабенко М.Г., Ляхов П.А., Лавриненко И.Н. <i>Приближенный метод определения знака числа в системе остаточных классов и его техническая реализация</i>	131
Диченко С.А., Елисеев Н.И., Финько О.А. <i>Контроль ошибок функционирования генераторов двоичных ПСП, реализованных на арифметических полиномах</i>	142
Малофей О.П., Щелкунова Ю.О. <i>Подход к созданию структурного кода для записи и считывания данных в запоминающих устройствах инфотелекоммуникационных систем</i>	150
Корнеев П.К., Журавлёва И.А. <i>Построение наилучших среднеквадратических полиномов, приближающих функцию и ее производные</i>	156

Contents

Information Transfer and Processing

Ivankov A.A., Shishagin A.L. <i>Stanford NLP parsers and semantic triplets identification</i>	9
Glushkov P.V., Mikhailov N.V., Yudakin D.E. <i>Open-loop tracking of GNSS signals under condition of multipath propagation</i>	18
Benderskaya E.N., Nikitin K.V. <i>Recurrent neural network as dynamical system and approaches to its training</i>	29

Radio Engineering, Aerials, SHF-devices

Sukhov I.A., Akimov V.P. <i>Super resolution techniques for direction-finder antenna array with directional elements</i>	41
---	----

System Analysys and Control

Kedrin V.S., Kuzmin O.V. <i>Frequency analysis of time series periodic functions by means assessment of the number of ranks</i>	47
Pisarev A.I., Sulyaev I.I. <i>Neural emulator temperature mode of copper-nickel raw materials smelting in Vanyukov furnace</i>	55

Computer Systems and Software

Petrosyan G.S. <i>Language support for systematic error handling</i>	65
---	----

Mathematical Modelling: Methods, Algorithms, Technologies

Potlov A.Yu., Proskurin S.G. <i>Color Doppler mapping algorithm of biological liquids' directed flows using optical coherence tomography</i>	71
Chernorutskiy I.G. <i>Practical optimization and nonconvex problems</i>	79
Bortyakov D.E., Mescheryakov S.V., Solodilova N.A. <i>Quality assurance of computer-aided design for production metalware based on distributed database of operating defects</i>	87

Electronics, Electronic Equipment Material Production Technologies

Pilipko M.M. <i>Time-mode delta-sigma modulator for an analog-to-digital converter</i>	95
---	----

Management in Social and Economic Systems

Bostonov O.H., Zverev G.N. <i>Semiotic integration and control of business processes in petroleum organization</i>	103
---	-----

High-performance Computing

Makoha A.N. <i>Research of canonical trivectors of eighth grade by means of theory of the graphs and groups of substitutions</i>	112
---	-----

Mnukhin V.B. <i>Integral transform for symmetry recognition of gray-level images</i>	123
Chervyakov N.I., Babenko M.G., Lyakhov P.A., Lavrinenko I.N. <i>Approximate method for determining the number sign in residue number system and it's technical sales</i>	131
Dichenko S.A., Eliseev N.I., Finko O.A. <i>Error function generator binary PRS control implemented on arithmetic polynomials</i>	142
Malofey O.P., Shchelkunova Yu.O. <i>Approach to the creation of a structural code to write and include the deployment of data in a storage device infotelecommunication systems</i>	150
Korneyev P.K., Zhuravleva I.A. <i>Constructing best mean-square polynomials approximating a function and its derivatives</i>	156

УДК 681.3.07

А.А. Иванков, А.Л. Шишагин

**РАСШИРЕНИЕ ОБЪЕКТНОЙ МОДЕЛИ СТЭНФОРДСКОГО ПАРСЕРА
ДЛЯ РЕШЕНИЯ ЗАДАЧИ ИДЕНТИФИКАЦИИ
СЕМАНТИЧЕСКИХ ТРИПЛЕТОВ**

А.А. Ivankov, A.L. Shishagin

STANFORD NLP PARSERS AND SEMANTIC TRIPLETS IDENTIFICATION

Изучен информационный поиск семантических триплетов в текстах англоязычного корпуса Википедии. Для идентификации триплетов использован стэнфордский парсер, объектная модель которого расширена авторами. Приведены оценки вероятности ошибки идентификации триплета.

ИНФОРМАЦИОННЫЙ ПОИСК. СЕМАНТИЧЕСКИЙ АНАЛИЗ. СТЭНФОРДСКИЙ ПАРСЕР. СЕМАНТИЧЕСКИЙ ТРИПЛЕТ. ОШИБКИ ИДЕНТИФИКАЦИИ.

We developed some own classes for the object model of Stanford NLP parsers to implement our algorithm of semantic triplet identification. Results of information retrieving are discussed for en.Wikipedia articles.

INFORMATION RETRIEVING. SEMANTIC ANALYSIS. STANFORD NLP PARSERS. SEMANTIC TRIPLET. RETRIEVING ERRORS.

В работе [1], посвященной семантическому анализу и оцениванию различных характеристик документов на основе результатов такого анализа, алгоритмы ориентированы преимущественно на использование элементов метаданных, т. е. элементов разметки, определяющих структуру документа. Мы анализировали текст на уровне заголовков, подзаголовков и т. п., используя соответствующую грамматику языка разметки конкретного документа (в нашей работе это были подмножества грамматики HTML/XML) и модель «мешка слов».

Между тем структурообразующие элементы содержат существенно меньшую часть информации, чем предложения дискурсов в тексте самого документа. Задача семантического анализа не может быть решена в полном объеме, если не выполня-

ется семантический анализ каждого предложения дискурса.

Большинство лингвистов разделяют следующую точку зрения: актуальное членение предложения на естественном языке (ЕЯ) предполагает выделение особого уровня (отличного от синтаксического), на котором грамматическими средствами выражается субъектно-предикатная структура соответствующей мысли. На этом, особом, логико-грамматическом уровне предложения мы имеем дело с логико-грамматическим субъектом и логико-грамматическим предикатом, которые не обязательно должны совпадать с синтаксическими подлежащим и сказуемым. Но в то же время давно замечено [2], что в научно-технической литературе такое соответствие имеет место почти всегда, если привлекать язык статистики.

В пределах отдельного предложения наибольший вклад в семантику этой элементарной части дискурса вносится триплетом (тройкой): подлежащее — сказуемое — дополнение. Научно-технические тексты характеризуют преимущественно как номинативные, но особую роль в них играет сказуемое. Глагол или существительное как часть составного сказуемого — это смыслообразующие элементы фразы (главные ее элементы) [3–5]. Более того, глагол-сказуемое поддерживает тему-рематическую прогрессию текста, что принципиально важно в рамках коммуникативно-функционального подхода к построению модели языка. Напомним также, что язык RDF (Resource Definition Framework) создан консорциумом W3 как средство семантического описания интернет-ресурсов, ориентированное на триплеты: субъект — предикат — объект.

Следовательно, одной из важнейших задач семантического анализа предложений является идентификация семантических триплетов.

Хотелось бы еще заметить, что эту задачу ни в коем случае не следует рассматривать как идентификацию триграмм в вероятностных порождающих моделях документов. В последнем случае для решения, вообще говоря, не требуется ни синтаксический анализ предложения с построением дерева разбора, ни последующий семантический анализ с идентификацией грамматических отношений.

Попытки решить задачу идентификации семантических триплетов в пределах предложения документа произвольного корпуса предпринимались неоднократно. По мере развития методов информационного поиска эта проблема вызывает все больший интерес. Информация о структуре и местоположении триплета в пределах предложения может использоваться для решения самых актуальных проблем обработки текстов на ЕЯ: их автоматического аннотирования, машинного перевода и т. п. Например, описание корпоративных хранилищ информации предлагают в форме семантических графов (на основе дерева синтаксического разбора, снабженного разметкой грамматических отношений лексем) как одного из

представлений семантической сети. Такое представление является весьма наглядным. Другое место приложения результатов решения этой задачи — автоматическое построение онтологий предметных областей и, разумеется, семантическое описание интернет-ресурсов с использованием языка RDF.

Основные трудности в решении связаны с неоднозначностью результатов синтаксического и семантического анализа предложения. Операторы синтаксического разбора порождают нам дерево (граф) структуры предложения. Дальнейший анализ выполняется с помощью оператора семантического анализа, который снабжает дерево синтаксического разбора разметкой грамматических отношений лексем, входящих в него.

Результаты, доставляемые нам этими операторами в конечном итоге, — это лес случайных деревьев. Математические модели на классе случайных помеченных деревьев на практике приводят к ресурсоемким алгоритмам оценивания. Наш опыт работы с подобными моделями позволяет нам утверждать, что получение отдельных оценок на корпусе неаннотированных документов может потребовать до нескольких недель работы вычислительной системы.

Тем самым становится очевидным, что выбор инструментария для реализации алгоритмов решения задачи — ключевой момент в таких исследованиях.

Множество основных критериев выбора содержит по меньшей мере следующие:

- форматы хранения входных данных и результатов;
- портируемость решения на другие платформы;
- язык реализации.

Большинство синтаксических анализаторов для представления результатов синтаксического анализа предложений используют формат Penn TreeBank [6] (рис. 1). Такая форма представления дерева синтаксического разбора стала стандартом де-факто для парсеров, работающих с текстами на английском языке. Это весьма удивительный факт, принимая во внимание, что в отношении других форматов эти

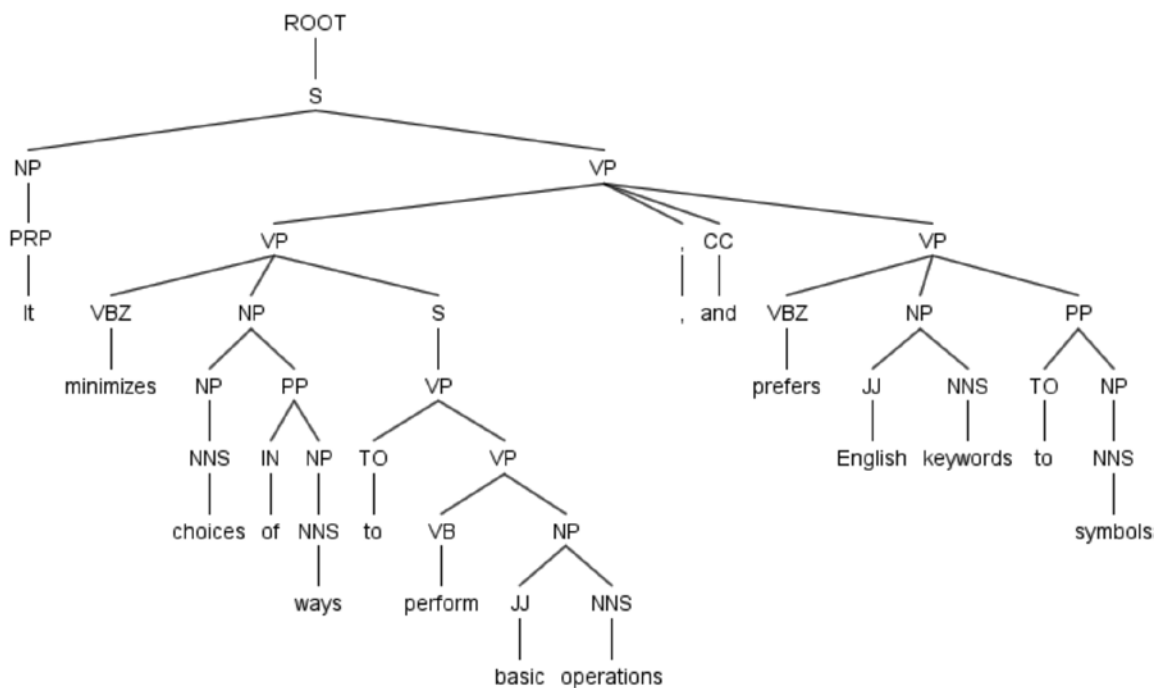


Рис. 1. Графическое представление дерева синтаксического разбора

программные системы отличаются весьма значительно. Отметим, что существует некоторое небольшое подмножество парсеров, где выходной формат ориентирован на представление грамматических зависимостей.

Основное системное требование – программное обеспечение должно легко разворачиваться на платформе WINTEL или в одной из популярных Unix-систем. С точки зрения портируемости программного обеспечения исследователи отдают предпочтение решениям на объектно-ориентированном языке программирования JAVA. Наиболее доступные решения с открытым JAVA-кодом уже более 10 лет предлагаются на сайтах <http://opennlp.apache.org> и <http://sourceforge.net/projects/opennlp>.

Принимая во внимание указанные выше критерии, наш выбор осуществлялся из небольшого множества анализаторов с открытым кодом: LinkParser, MSTParser, MultiParser, Stanford Core NLP.

Необходимость идентификации грамматических отношений в пределах отдельного дерева синтаксического разбора – это

еще одно функциональное требование к инструментарию. Номенклатура грамматических отношений (отношений зависимости, определяемых на узлах дерева синтаксического разбора) существенно отличается от парсера к парсеру. Наименьшая мощность множества грамматических отношений у консервативной схемы [7]. Парсер LinkParser напротив предлагает исследователю более тонкую структуру этого множества. Мощность множества и качество идентификации ее элементов в тексте на ЕЯ – еще два критерия выбора инструмента. Они не взаимодополняющие: выбор в пользу одного может привести к потерям в отношении другого.

В литературе отмечается, что более высокое качество идентификации грамматических отношений демонстрируют парсеры, реализующие алгоритм поиска соответствий между отдельными подграфами дерева синтаксического разбора и абстрактными шаблонами, хранимыми как структуры данных в словарях этой программной системы.

Помимо критериев портируемости и

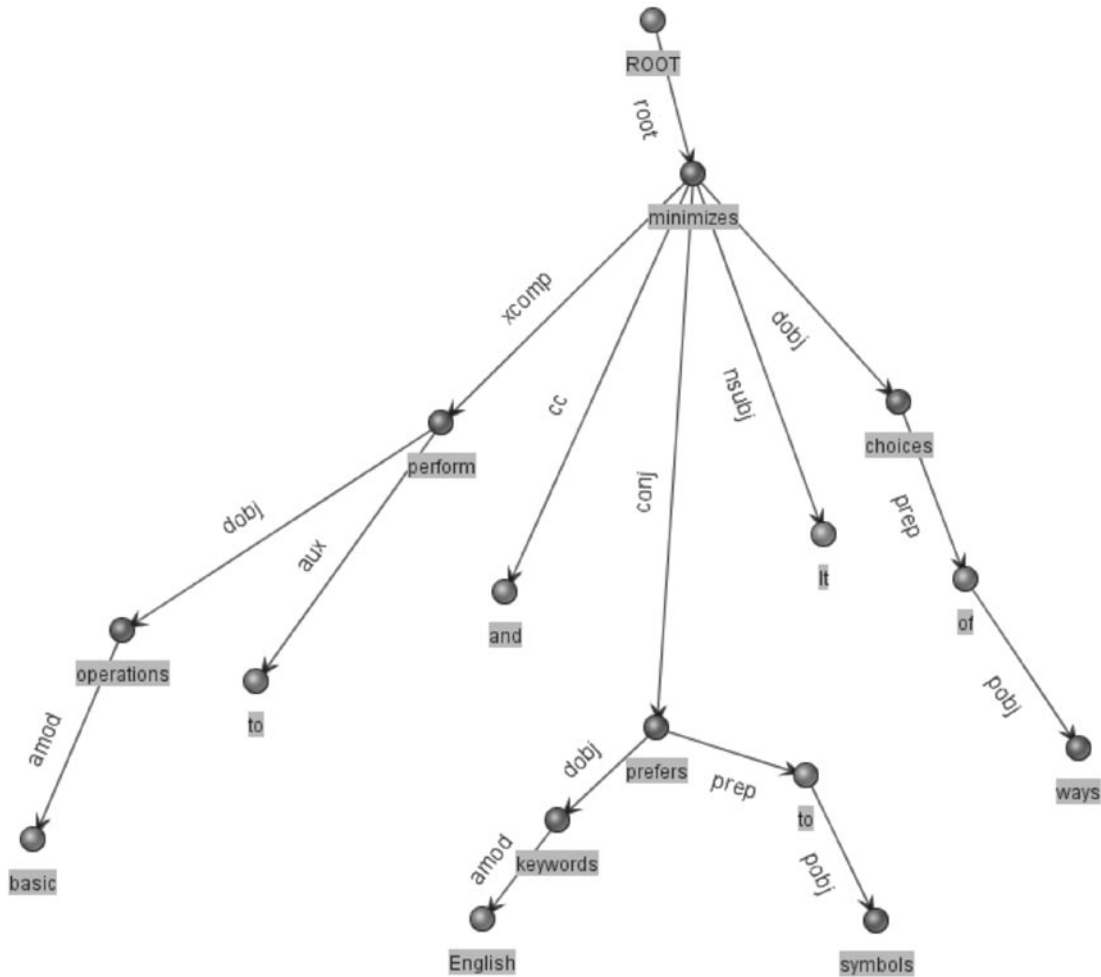


Рис. 2. Графическое представление результатов семантического анализа отдельного предложения

языка реализации, предпочтение отдавали хорошо документированному исходному коду.

Мы остановили свой выбор на пакете Stanford Core NLP [8] по следующим соображениям:

- структуры дерева синтаксического разбора совпадают с классическим форматом Penn TreeBank;
- множество идентифицируемых грамматических отношений расширено по сравнению с классической схемой [7], добавлены дополнительные отношения (см. рис. 2), играющие важную роль, обеспечивая успех последующего семантического анализа;
- форматы этого анализатора уже реализованы разработчиками других семантических анализаторов, т. е. апробированы и

хорошо зарекомендовали себя на практике;

- семантический анализ можно параметризовать с помощью шаблонов, определяющих правила разбора;
- исходный код доступен под лицензией GPL;
- кроссплатформенный язык разработки – JAVA, на этот язык уже портировали и ряд других анализаторов.

Наибольшие сложности породило отсутствие качественной документации. Единственно эффективным инструментом для изучения пакета как всегда оказался отладчик. Библиотеку из 467 классов, около 150 000 строк исходного кода, мы подвергли рефакторингу, исключив «грязную» сборку, и дополнительно разработали 45 классов, реализующих недостающую в

исходном пакете функциональность.

Все другие модификации кода были вынужденной мерой в условиях ограниченности аппаратных ресурсов, которыми мы располагали для проведения исследования. Заметим, что одни лишь словари парсера требовали более 7Г адресного пространства, не говоря уже об объеме собственно байт-кода.

Идентификация триплетов

В ходе анализа публикаций, посвященных семантическому анализу текстов на английском языке, наше внимание привлекли работы, упоминаемые в [9]. Это, пожалуй, один из немногих апробированных подходов к решению задачи идентификации триплетов, алгоритм которого опубликован.

Решение задачи авторы излагают в виде списка правил, применяемых к дереву синтаксического разбора предложения. Структура этого дерева (S в нотации Penn TreeBank) содержит два дочерних элемента: NP – именное словосочетание; VP – глагольное словосочетание. Поиск подлежащего (SUBJ) выполняется в пределах NP – первый дочерний элемент NP, грамматическая категория «часть речи» которого указана как «существительное», атрибутируется как подлежащее (SUBJ). Поиск сказуемого (PRED) осуществляется в пределах VP – наиболее удаленный по уровню иерархии терминальный элемент, грамматическая категория «часть речи» которого указана как «глагол», атрибутируется как сказуемое (PRED).

Наконец, поиск дополнения осуществляется в элементах того же уровня иерархии (по отношению к S), что и VP. По меньшей мере три различных поддерева рассматриваются как потенциальные контейнеры дополнения ((OBJ)):

словосочетание, содержащее предлог (PP);

именное словосочетание (NP);

словосочетание с прилагательным (ADJP).

В первых двух типах словосочетаний осуществляют поиск первого вхождения существительного, в последнем случае – вхождение первого прилагательного.

К сожалению, авторы не привели количественных оценок качества идентификации, ограничившись иллюстрацией работы парсеров на десяти английских предложениях.

Мы реализовали алгоритм идентификации, предложенный в [9] и последующих работах автора, в частности, опубликованных в журнале *Informatica*. Разумеется, наш программный код может отличаться от авторской версии, а в личной переписке с авторами мы не смогли получить от них необходимых нам пояснений, позволяющих утверждать, что наша реализация идентична их реализации.

Тестирование выполнялось на статьях англоязычного корпуса *Wikipedia*, срез на 01.01.2011. Наш выбор обусловлен тем, что в большинстве работ, где обсуждается качество идентификации грамматических отношений и синтаксического анализа, оценки получают на текстах, составленных профессионалами, журналистами СМИ. Мы же хотим получить оценки на текстах «живого» языка, на текстах, которые практически не подвергались редакторской правке.

Результаты тестирования указывают на то, что предложенные авторами [9] приводят к потере большей части семантических триплетов. Рассмотрим итоги разбора первых предложений статьи «Ada (programming language)».

Применение суперпозиции операторов синтаксического и семантического разбора порождает результат, который выглядит следующим образом (скобочная запись):

(ROOT (S (NP (NNP Ada)) (VP (VBZ is) (NP (NP (DT a) (ADJP (JJ structured) (,) (JJ statically)) (JJ typed) (,) (ADJP (JJ imperative) (,) (JJ wide-spectrum) (,) (CC and) (JJ object-oriented) (JJ high-level)) (NN computer) (VBG programming) (NN language)) (,) (VP (VBN extended) (PP (IN from) (NP (NP (NNP Pascal)) (CC and) (NP (JJ other) (NNS languages)))))) (.)

Подлежащее идентифицировано успешно: (NNP Ada). В нотации [9] – subj, в оригинальной стэнфордской нотации – nsubj. Однако сказуемое (NN language) было утеряно, а вместо него в качестве предиката предложено причастие (VBN extended).

В качестве дополнения – (NNP Pascal) – ошибочная интерпретация предложного дополнения из причастного оборота.

Еще один пример:

(ROOT (S (NP (NNP Ada)) (VP (VBD was) (ADVP (RB originally)) (VP (VBN designed) (PP (IN by) (NP (NP (DT a) (NN team)) (VP (VBN led) (PP (IN by) (NP (NP (NNP Jean) (NNP Ichbiah)) (PP (IN of) (NP (NP (NNP CII) (NNP Honeywell) (NNP Bull)) (PP (IN under) (NP (NN contract)))))))) (PP (TO to) (NP (NP (DT the) (NNP United) (NNPS States) (NNP Department)) (PP (IN of) (NP (NP (NNP Defense)) (PRN (-LRB- -LRB-) (NP (NNP DoD)) (-RRB- -RRB-)))))) (PP (IN from) (NP (CD 1977) (TO to) (CD 1983))) (S (VP (TO to) (VP (VB supersede) (NP (NP (DT the) (NNS hundreds)) (PP (IN of) (NP (NP (NN programming) (NNS languages)) (VP (ADVP (RB then)) (VBN used) (PP (IN by) (NP (DT the) (NNP DoD)))))))))))))) (.)

Подлежащее и в этом предложении было успешно идентифицировано: (NNP Ada). Сказуемое (VBN designed) утеряно, а в его качестве предложено (VBN used) из придаточного предложения.

Мы убедились, что правила идентификации семантических триплетов из [9] приводят к потере значимой информации.

Приняв во внимание результаты тестирования правил идентификации триплетов, изложенных в [9], был предложен собственный алгоритм. Он (алгоритм идентификации триплетов) определяется набором структурных и грамматических правил, которые на уровне реализации хранятся во внешнем файле-шаблоне. Эти правила последовательно применяются ко всему дереву синтаксического разбора или отдельным его частям.

Правила идентификации семантических триплетов

Алгоритм реализован на основе набора правил, которые применяются к дереву, полученному по итогам разбора структуры предложения, и к семантическому дереву отношений:

SR (Structure Rules) – структурные правила;

GR (Grammar Rules) – грамматические правила.

Структурные правила служат для идентификации в пределах одного предложения его придаточных предложений, причастных и деепричастных оборотов, отдельных фраз. Таким образом мы определим, в каких частях предложения содержатся объекты «семантический триплет» (далее – ST) из одной группы.

К структурным правилам мы отнесли следующие:

1. Одной семантической группе принадлежат все терминальные узлы дерева структурного разбора и соответствующие им объекты ST, которые имеют одинаковый ближайший вверх по дереву узел S (Sentence) или (Root).

2. Одной семантической группе принадлежат все терминальные узлы дерева структурного разбора и соответствующие им объекты ST, которые имеют одинаковый ближайший вверх по дереву узел VP или NP. В этом случае на узлы VP и NP накладываются дополнительные ограничения: их родительский узел должен быть S и среди дочерних узлов S исходные узлы должны следовать после первого вхождения узла VP.

3. Если в поддереве с корневым узлом PP обнаружено поддерево PP, то в нем игнорируются все поддерева, кроме тех, у которых корневой узел S.

Основное грамматическое правило определяет обобщения и фильтрации грамматических отношений:

1. Из всех типов грамматических зависимостей, которые имеются в иерархическом дереве стэнфордского парсера, для построения наших правил оставлены и обобщены следующие грамматические отношения:

PRED = {pred, aux, auxpass, cop }

SUBJ = { subj, nsubj, nsubjpass, csubj, csubjpass, xsubj }

OBJ = { conj, prep, partmod, obk, iobj, dobj, pobj }

Помимо этого, определяются несколько правил включения очередного объекта ST на основе уже определенных в группе объектов. Эти аспекты поиска триплетов будут изложены в другой нашей работе.

Результаты такого обобщения положительные. В первом из указанных выше предложений подлежащее – (NNP Ada), сказуемое – (NN language). Во втором примере подлежащее – (NNP Ada), обобщенное сказуемое – (VBN designed) совместно с (VBD was), предложное дополнение – (NN team). Но не все результаты так же хорошо согласуются с экспертной оценкой. Приведем еще один характерный пример (в нашей XML нотации):

```

<sentence>
  <text>It has strong built-in language
  support for explicit concurrency, offering tasks,
  synchronous message passing, protected objects
  and nondeterminism. </text>
  <tokens>
    <token> <value > It</value>
  <role>nsubj</role> </token>
    <token> <value > has</value>
  <role>pred</role> </token>
    <token> <value > strong</value> </
  token>
    <token> <value > built-in</value>
  </token>
    <token> <value > language</value>
  </token>

```

```

  <token> <value > support</value>
  <role>dobj</role> </token>
    <token> <value > for</value>
  <role>prep</role> </token>
    <token> <value > explicit</value>
  </token>
    <token> <value > concurrency</
  value> <role>pobj</role> </token>
    <token> <value > ,</value> </to-
  ken>
    <token> <value > offering</value>
  </token>
    <token> <value > tasks</value>
    <role err=Ошибочная иденти-
    фикация грамматического отношения >conj</
    role>
  </token>
  ...
    <token> <value > synchronous</
  value> </token>
    <token> <value > message</value>
  </token>
    <token> <value > passing</value>
    <role err=Ошибочная иденти-
    фикация грамматического отношения >conj</
    role>
  </token>

```

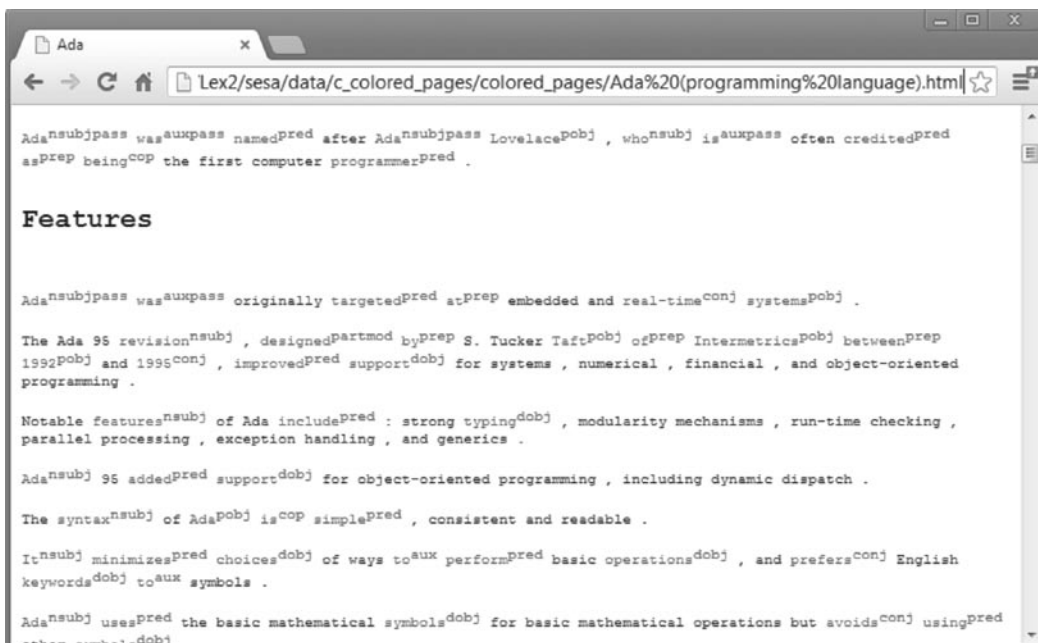


Рис. 3. Форма представления результатов синтаксического и семантического анализа статьи для выполнения ручной разметки экспертом

Оценивание уровня ошибок в результатах

Обобщенное грамматическое отношение	Общее количество	Количество ошибочно идентифицированных	Процент ошибок, %
OBJ	3214	607	18,89
PRED	1835	222	12,1
SUBJ	857	114	13,3
Сумма	5906	943	15,97

```

...
<token> <value > protected</value>
</token>
<token> <value > objects</value>
<role err=Ошибочная иденти-
фикация грамматического отношения > conj</
role>
</token>
<token> <value > and</value> </
token>
<token> <value > nondeterminism</
value>
<role err=Ошибочная иденти-
фикация грамматического отношения > conj</
role>
</token>
...
</tokens>
</sentence>

```

Обеспечив результаты синтаксического и семантического анализа HTML разметкой, мы предоставили эти материалы экспертам в виде (рис. 3), пригодном для выполнения ручной разметки этих же статей.

Ошибки в идентификации грамматических отношений лексемы очень критичны для последующего семантического анализа. Следует заметить, что в данном примере мы выбрали из случайного леса деревьев, порождаемых стэнфордским парсером, дерево, оптимальное по критерию согласия, построенному авторами пакета на основе модели порождающих грамматик, они в бинарной форме хранятся в этой программной системе (файл объемом более 7Г, загружаемый в адресное пространство приложения в ходе его работы).

В нашей работе предпринята попытка провести семантический анализ текстов англоязычной Википедии, т. е. текстов, которые подготовлены различными специалистами, не являющимися профессиональными журналистами или писателями. Мы подчеркиваем это обстоятельство, т. к. в большинстве других работ (в т. ч. в [9]) использовались размеченные экспертами журнальные статьи. Расхождения результатов идентификации семантических триплетов, полученных с помощью нашего алгоритма и алгоритма из [9] в нашей интерпретации, позволяют предположить, что в [9] и последующих работах этой группы исследователей изложен подход, нуждающийся в существенном дополнении как с точки зрения статистической обработки текстов на ЕЯ, так и с точки зрения лингвистики.

Другую причину высокого процента ошибок идентификации триплетов и, как следствие, определения семантики текста, мы видим в том, что в стэнфордском парсере грамматические отношения определяются с помощью порождающих грамматик, но с минимальным привлечением статистических методов. Первоочередные задачи в данном контексте могут быть сформулированы как выработка решающих правил (в статистическом смысле) идентификации грамматических отношений лексем, которые позволили бы минимизировать ошибку в результатах идентификации триплетов. Возможно, в поисках смысла текстов на ЕЯ более обнадеживающими будут результаты привлечения модели случайного леса деревьев. Это направление исследований представляется нам одним из наиболее перспективных.

СПИСОК ЛИТЕРАТУРЫ

1. **Иванков, А.А.** Динамическое отслеживание модулем информационно-поисковой системы изменений в структуре или тексте интернет-

ресурса [Текст] / А.А. Иванков, Д.С. Елисеев // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управле-

ние. –СПб.: Изд-во Политехн. ун-та, 2010. –№ 3 (101). –С. 86–93.

2. **Пумпянский, А.Л.** Информационная роль порядка слов в научной и технической литературе [Текст] / А.Л. Пумпянский. –М.: Наука, 1974. – 247 с.

3. **Апресян, Ю.Д.** Лексическая семантика. Синонимические средства языка [Текст] / Ю.Д. Апресян. – М.: Наука, 1974. – 367 с.

4. **Филлмор, Ч.** Основные проблемы лексической семантики [Текст] / Ч. Филлмор // Новое в зарубежной лингвистике. – М.: Радуга, 1983. –Вып. XII. Прикладная лингвистика. – С. 74–122.

5. **Сильницкий, Г.Г.** Соотношение глагольных признаков различных уровней в английской

речи [Текст] / Г.Г. Сильницкий, С.Е. Андреев, Л.А. Кузьмин, М.И. Кусков. – Минск: Наука и техника, 1990. – 182 с.

6. [Электронный ресурс] / Режим доступа: <http://www ldc.upenn.edu/Catalog/docs/LDC99T42/prsguid1.pdf>

7. Treebanks: Building and Using Parsed Corpora [Text] / Ed. A. Abeille. – Springer, 2003. – P. 302–312.

8. [Электронный ресурс] / Режим доступа: <http://nlp.stanford.edu>

9. **Rusu, Delia.** Semantic Graphs Derived from Triplets with Application in Document Summarization [Text] / Delia Rusu [et al.] // Informatica. – 2009. –Vol. 33. P. 357–362.

REFERENCES

1. **Ivankov A.A., Eliseev D.S.** Dinamicheskoe otslezhivanie modulem informatsionno-poiskovoi sistemy izmenenii v strukture ili tekste Internet-resursa / Nauchno-tehnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie. –St-Petersburg: Izd-vo Politehn. un-ta, 2010. –№ 3 (101). –S. 86–93. (rus)

2. **Pumpianskii A.L.** Informatsionnaia rol' poriadka slov v nauchnoi i tekhnicheskoi literature. – Moscow: Nauka, 1974. – 247 s. (rus)

3. **Apresian Iu.D.** Leksicheskaiia semantika. Sinonimicheskie sredstva iazyka. – Moscow: Nauka, 1974. – 367 s. (rus)

4. **Fillmor Ch.** Osnovnye problemy leksicheskoi semantiki / Novoe v zarubezhnoi lingvistike.

– Moscow: Raduga, 1983. –Vyp. XII. Prikladnaia lingvistika. – S. 74–122. (rus)

5. **Sil'nitskii G.G., Andreev S.E., Kuz'min L.A., Kuskov M.I.** Sootnoshenie glagol'nykh priznakov razlichnykh urovnei v angliiskoi rechi. – Minsk: Nauka i tekhnika, 1990. – 182 s.

6. <http://www ldc.upenn.edu/Catalog/docs/LDC99T42/prsguid1.pdf>

7. Treebanks: Building and Using Parsed Corpora; Ed. A. Abeille. – Springer, 2003. – P. 302–312.

8. <http://nlp.stanford.edu>

9. **Delia Rusu et al.** Semantic Graphs Derived from Triplets with Application in Document Summarization / Informatica. –2009. –Vol. 33. –P. 357–362.

ИВАНКОВ Алексей Александрович – доцент кафедры прикладной математики Санкт-Петербургского государственного политехнического университета, кандидат физико-математических наук.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: ivankov@stu.neva.ru

IVANKOV, Alexei A. St. Petersburg State Polytechnical University.

195251, Politekhnikeskaya Str. 29, St.-Petersburg, Russia.

E-mail: ivankov@stu.neva.ru

ШИШАГИН Алексей Леонидович – выпускник кафедры прикладной математики Санкт-Петербургского государственного политехнического университета, магистр.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

SHISHAGIN, Alexei L. St. Petersburg State Polytechnical University.

195251, Politekhnikeskaya Str. 29, St.-Petersburg, Russia.

УДК 621.396.98:629.195

П.В. Глушков, Н.В. Михайлов, Д.Е. Юдакин

ИЗМЕРЕНИЕ РАДИОНАВИГАЦИОННЫХ ПАРАМЕТРОВ СИГНАЛА СРНС СЛЕДЯЩИМИ СИСТЕМАМИ БЕЗ ОБРАТНОЙ СВЯЗИ В УСЛОВИЯХ МНОГОЛУЧЕВОГО РАСПРОСТРАНЕНИЯ

P.V. Glushkov, N.V. Mikhailov, D.E. Yudakin

OPEN-LOOP TRACKING OF GNSS SIGNALS UNDER CONDITION OF MULTIPATH PROPAGATION

Предложен метод оценки радионавигационных параметров (задержка и частота) сигнала системы GPS, отличный от классических систем слежения за задержкой и фазовой/частотной автоподстройки. Метод, основанный на расчете корреляционной функции узкополосного сигнала в частотно-временной области, позволяет производить оценки параметров не только прямого, но и вероятных отраженных сигналов, значительно снижающих качество позиционирования в городской среде. Метод применим для случаев с движущимся пользователем в присутствии отраженных сигналов высокой мощности (мощности, сравнимой с мощностью прямого сигнала).

СРНС. МНОГОЛУЧЕВОЕ РАСПРОСТРАНЕНИЕ. СЛЕЖЕНИЕ ЗА СИГНАЛОМ.

A new method for GPS signal tracking is being proposed. The method differs from conventional delay-locked loops, frequency/phase-locked loops and is based on calculation of a narrow-band signal correlation function in time-frequency domain. The method allows to track both line-of-sight signal and reflected GNSS signals, which cause serious degradation of positioning accuracy in urban environments. The method is targeted at mobile user under severe multipath propagation conditions, when reflected signal power is comparable to the power of line-of-sight signal.

GNSS. GPS. MULTIPATH PROPAGATION. SIGNAL TRACKING. OPEN-LOOP TRACKING.

Современные спутниковые радионавигационные системы (СРНС) широко применяются для решения задач позиционирования и навигации в самых различных условиях. Отдельным классом СРНС приемников являются приемники, встраиваемые в портативные устройства массового применения. Особенность устройств данного класса – строгие требования к их стоимости, габаритам и энергопотреблению. Подобные устройства используются повсеместно, в т. ч. и в городской среде, что создает дополнительные сложности для приема сигнала: низкий уровень принимаемых сигналов; частые пропадания прямого сигнала; большое количество сильных отраженных сигналов. Данные факторы ухудшают точность измерения радионавигационных параметров (РНП) сигнала (задержка, частота/фаза), требуемых для

решения навигационной задачи.

В статье рассматривается случай приема многолучевого сигнала навигационных космических аппаратов (НКА) движущимся приемником. Предлагается метод уменьшения ошибки измерений РНП в подобных случаях при использовании следящей системы «без обратной связи».

Методы уменьшения влияния многолучевого распространения

Из большого количества известных методов¹ для массового рынка СРНС аппаратуры подходят далеко не все. Эффективные

¹ Здесь и далее, если не приводятся уточнения, под обобщением «методы» понимаются именно методы ослабления влияния многолучевого распространения на качество измерения РНП и позиционирования.

методы, применимые для случая со статичным пользователем [12, 13], несмотря на свою высокую точность, используют дополнительные аппаратные средства. Это неминуемо ведет к увеличению габаритов, энергопотребления и цены конечной системы, что неприемлемо для сравнительно простых портативных устройств. Таким образом, наибольший интерес в данном контексте представляют программные методы, работающие в реальном времени, применимые для статичного и подвижного потребителей. Подавляющее большинство подобных методов можно разделить на две основные группы:

- 1) методы, напрямую рассчитывающие оценки РНП прямого сигнала (наиболее известные представители – [1–3]);
- 2) методы, решающие задачу оптимизации для поиска оптимальных значений РНП как прямого, так и предполагаемых отраженных сигналов (наиболее известные представители – [4–6]).

К основным недостаткам методов первой группы можно отнести их высокие требования к полосе обработки сигнала (не менее 10–12 МГц), а также потери мощности, возникающие при обработке сигнала с применением модифицированной опорной реплики. К недостаткам методов второй группы можно отнести рост количества требуемых вычислительных ресурсов с числом оцениваемых РНП [7], а также проблему выбора адекватных начальных приближений РНП для решения задачи оптимизации.

На сегодняшний день рост вычислительной мощности микропроцессоров при одновременном снижении их цены и размеров является естественным процессом. В связи с этим применение вычислительно сложных алгоритмов улучшения качества позиционирования в сравнительно простых приемниках для массового рынка становится естественной тенденцией. С учетом данного обстоятельства в статье предлагается очередной метод повышения качества измерений РНП сигнала при многолучевом распространении. Технически, предлагаемый метод относится к второй группе. Назначение метода – улучшение качества

измерений РНП для движущегося пользователя.

Описание предлагаемого метода

Предлагаемый метод применяет модель частотно-временного корреляционного пика, обрабатываемого следящей системой «без обратной связи» (Open-Loop Tracking, [8]), в качестве аппроксимирующей функции для поиска оптимальных оценок РНП прямого и отраженного сигналов. Оптимизируемая функция представляет собой сумму квадратов невязок аппроксимации. Предлагается алгоритм выбора начальных оценок искомых параметров. Описывается применение алгоритма в случае с одним отраженным сигналом большой мощности.

Следящая система «без обратной связи». Метод, предлагаемый в данной статье, был разработан для двухсистемного (GPS+GLONASS) СРНС приемника компании MStar Semiconductor (Тайвань) и апробирован на нем же. Отличительной особенностью данного приемника является применение следящей системы «без обратной связи» в качестве основного режима работы [11]. Вместо классических систем слежения за задержкой, фазовой и частотной автоподстройки, при помощи решетки корреляторов и Быстрого Преобразования Фурье (ПБФ), система рассчитывает множество значений мощности сигнала и суммирует их некогерентно. Искомые оценки РНП формируются по глобальному максимуму рассчитанной сетки значений мощности сигнала в частотно-временной области. Опишем данный механизм более подробно. Первым шагом вычисляется последовательность когерентных накоплений сигнала на выходе каждого коррелятора:

$$C_k(f', \tau_l) = \sum_{n=0}^{G-1} (Y_n(f', t_k + nt_s) \times x'(t_k + nt_s + \tau_l)); \quad (1.1)$$

$$(t_k - t_{k-1}) = T_s G = \text{const} = T_{\text{сон}},$$

где Y – комплексная огибающая сигнала²

² Значение битов навигационных данных не учитывается. Считаем, что граница бита данных уже найдена и знак текущего бита известен.

на выходе частотного миксера при некоторой фиксированной оценке доплеровского частотного сдвига f' ; T_s – период дискретизации; τ_l – временная задержка в l -м корреляторе; x^l – локальная реплика дальномерного кода; T_{COH} – интервал когерентного накопления; G – число отсчетов сигнала в T_{COH} . $C_k(f, \tau_l)$; $k = 1:M$ – результирующая последовательность из M когерентных накоплений на выходе l -го коррелятора. Величина T_{COH} выбирается таким образом, чтобы M когерентных накоплений покрывали интервал в 20 мс (длительность бита навигационных данных).

Затем с помощью БПФ рассчитывается спектральная плотность мощности (СПМ) последовательности $C_k(f', \tau_l)$. Обозначим через $S_k(f', \tau_l)$ результат расчета БПФ для последовательности из M когерентных накоплений на выходе l -го коррелятора. В последовательности S , в отличие от C , индекс k принимает значения от нуля до $M - 1$ и обозначает не начальное время когерентного накопления, а частотный индекс в БПФ.

Z последовательных во времени значений СПМ, относящихся к одним и тем же частотным позициям, далее суммируются

некогерентно. Для этого рассчитывается БПФ от $C_k(f', \tau_l)$, принадлежащим к Z следующим друг за другом 20-миллисекундным интервалам:

$$P_k(f', \tau_l) = P_{k,l} = \sum_{i=0}^Z |S_k(f', t_i + \tau_l)|^2; \quad (1.2)$$

$$Z(MT_{\text{COH}}) = T_{\text{NCOH}}; \quad t_{i+1} - t_i = MT_{\text{COH}}.$$

Индексы k и l – частотный индекс в БПФ и индекс коррелятора. Значение $P_{k,l}$ будем называть некогерентным накоплением в k -й частотной позиции для l -го коррелятора. Всю совокупность значений P будем называть буфером некогерентных накоплений (БНН). Пример БНН представлен на рис. 1.

В отсутствие активных помех и отраженных сигналов, искомые РНП (задержка τ и частота f) могут быть оценены по максимуму БНН (оценки будем обозначать τ' и f').

Применение некогерентного накопления совместно со следящей системой «без обратной связи» отлично зарекомендовало себя при навигации в городской среде, где часто происходят кратковременные пропадания сигнала и резкие изменения его мощ-

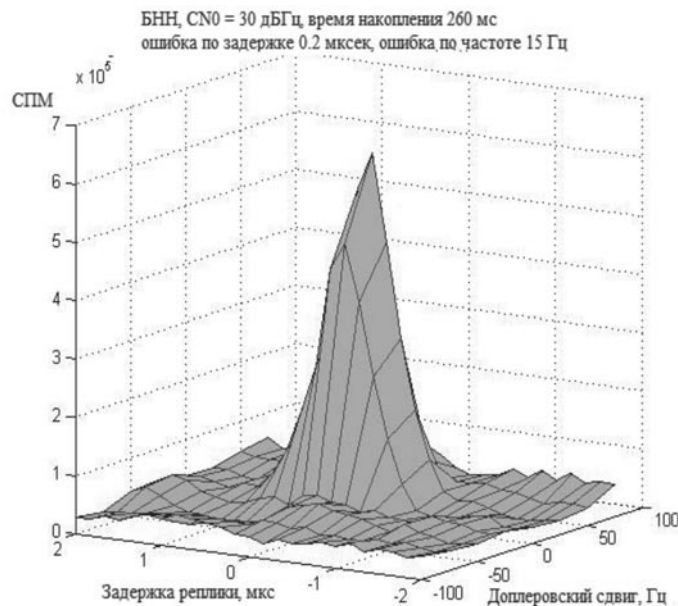


Рис. 1. Пример корреляционного пика в БНН

ности. Становится возможным оперативно продолжить слежение за нестабильным сигналом без необходимости заново производить процесс поиска и захвата, как это делается в классической реализации. Второе весомое преимущество данной системы – слежение за сигналами малой мощности ($C/N_0 < 25$ дБГц [8]), когда классическая система фазовой автоподстройки теряет способность следить за сигналом. Тем не менее, как и все другие, данная система подвержена негативному влиянию многолучевого распространения, в случае, когда задержка отраженного сигнала составляет менее одного чипа дальномерного кода (~ 1 мкс). Для ослабления данного эффекта предлагается использовать алгоритм, описанный далее.

Модель корреляционного пика. Корреляционный пик в БНН, изображенный на рис. 1, можно аппроксимировать соотношением [9]:

$$P'(f', \tau', C') = P'(\beta) = I^2 + Q^2 \approx C'R^2(\tau')\text{sinc}^2(f'T_{\text{COH}}), \quad (2.1)$$

где I/Q – сумма отсчетов сигнала в синфазном/квадратурном каналах обработки; R – значение корреляции сигнала и локальной реплики; τ' , f' , C' – оценки задержки, доплеровской частоты и мощности.

Значения модели (2.1) в частотных (f_k) и временных (τ_l) позициях, соответствующих процедуре формирования БНН, может быть выражено:

$$P'(\beta, f_k, \tau_l) = C'R^2_{\text{LUT}}(\tau' + \tau_l)\text{sinc}^2((f' + f_k)T_{\text{COH}}), \quad (2.2)$$

где β – вектор текущих оценок параметров сигнала; R_{LUT} – табличное значение корреляции сигнала и локальной реплики для имеющейся системы обработки сигнала.

Ошибка аппроксимации будет распределена нормально благодаря многократному некогерентному суммированию³ возведенных в квадрат шумовых составляющих в каналах обработки. При движении антенны приемника относительно отражающей по-

верхности, замирание сигнала, вызванное многолучевым распространением, будет иметь циклический характер. При некогерентном накоплении колебания мощности сигнала будут усредняться. При достаточной скорости движения и T_{COH} , значения в ячейках БНН будут стремиться к суперпозиции мощностей прямого и отраженных сигналов, умноженной на число Z из (1.2).

Обнаружение многолучевого распространения. Имея теоретическую модель корреляционного пика в БНН, возможно применить методы регрессионного анализа для оценки параметров принимаемого сигнала. Поскольку модель сигнала является нелинейной, предлагается выбрать начальные оценки для искомых параметров по глобальному максимуму БНН и затем уточнять их итеративно, применяя метод Ньютона–Гаусса:

$$\beta_{k+1} = \beta_k + (JJ^T)^{-1} Jr, \quad (3.1)$$

$$r_{k,l}(\beta) = P(f_k, \tau_l) - P'(\beta, f_k, \tau_l), \quad (3.2)$$

$$F(\beta) = \sum_{k=0}^{M-1} \sum_{l=1}^L r_{k,l}^2, \quad (3.3)$$

где β – вектор, состоящий из оценок РНП сигнала (задержка τ' , доплеровская частота f') и мощности C' ; J – матрица Якоби, вычисляемая с помощью конечных разностей первого порядка; P/P' – БНН и модель из формул (1.2) и (2.2) соответственно; F – оптимизируемая функция; M – число частотных позиций в БПФ; L – число корреляторов.

В случае если оценки РНП сигнала близки к истинным ($\tau' \approx \tau$, $f' \approx f$, $C' \approx C$), величину невязки аппроксимации в одной ячейке БНН можно представить как

$$P = |I + \eta_I|^2 + |Q + \eta_Q|^2 = I^2 + 2I\eta_I + \eta_I^2 + Q^2 + 2Q\eta_Q + \eta_Q^2, \quad (4.1)$$

$$r = P - P'(\beta) \approx 2I\eta_I + 2Q\eta_Q + \eta_I^2 + \eta_Q^2, \quad (4.2)$$

$$r^2 \approx f(A, \sigma_n) = k_1 A^2 + k_1 \sigma_n^4, \quad (4.3)$$

где I/Q – суммы отсчетов сигнала в синфазном/квадратурном каналах на интервале T_{COH} ; η_I/η_Q – шум в данных каналах; A – амплитуда; σ – дисперсия шума в кон-

³ В соответствии с Центральной Предельной Теоремой.

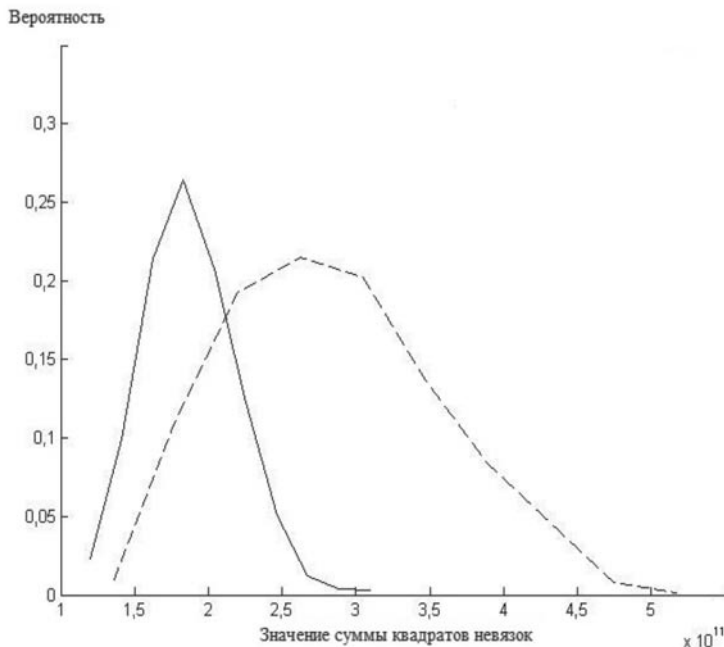


Рис. 2. Плотность вероятности распределения суммы квадратов (—) многолучевой сигнал отсутствует; (- - -) многолучевой сигнал присутствует

кретной системе обработки сигнала.

При появлении в сигнале многолучевой составляющей, сумма квадратов невязок значительно вырастает по сравнению со случаем без таковой. Пример распределения сумм квадратов невязок при наличии многолучевого сигнала и в его отсутствие представлена на рис. 2. Параметры сигнала: $C/N_0 = 30$ дБГц, задержка отраженного сигнала относительно прямого ~ 1 мкс, частотный сдвиг отраженного сигнала относительно прямого ~ 15 Гц, $SMR = 6$ дБ.

При более высоком отношении сигнал/шум (ОСШ) распределения сумм квадратов невязок будут еще более отстоять друг от друга. При снижении ОСШ, а также при приближении РНП отраженного сигнала к РНП прямого распределения перекрываются.

На практике возможно аппроксимировать верхнюю границу распределения суммы квадратов невязок (3.3) в отсутствие многолучевости с помощью соотношения (4.3). Для этого собрана большая статистика значений суммы квадратов невязок при различных значениях амплитуды входного сигнала и коэффициенты k_1/k_2 из (4.3) оце-

нены по методу Ньютона–Гаусса.

Отраженный сигнал считается обнаруженным, если сумма квадратов невязок аппроксимации БНН с помощью модели (2.2) превышает оценку для упомянутой верхней границы распределения.

Моделирование и статистическая обработка результатов позволили определить границы применимости данного метода детектирования многолучевого распространения⁴ (см. табл. 1).

Определение РНП прямого и отраженного сигналов. При положительном результате детектирования отраженной составляющей в сигнале, алгоритм обработки БНН предпринимает попытку оценить параметры прямого и отраженного сигналов. Для этого по аналогии с предыдущим подразделом оцениваются параметры модели, представленной суперпозицией двух корреляционных пиков

⁴ При увеличении времени некогерентного накопления возможно более эффективно обнаруживать многолучевость и на более низких ОСШ, однако в данной статье этот вопрос не рассматривается подробно.

Таблица 1

Границы применимости предложенного метода детектирования

Разность частот прямого/отраженного сигналов (Δf)	Не менее 15 Гц (~ 3 м/с)
Разность временных задержек прямого/отраженного сигналов ($\Delta \tau$)	Не менее 0,1 мкс
Разность мощностей прямого/отраженного сигналов (ΔC)	Не более 6 дБ
Мощность наиболее сильного из двух сигналов (C/N_0)	Не менее 30 дБГц

$$P_{MP}'(\beta_0, \beta_1) = C_0' R^2(\tau_0') \sin^2(f_0' T_{\text{COH}}) + C_1' R^2(\tau_1') \sin^2(f_1' T_{\text{COH}}), \quad (5.1)$$

где R – значение корреляции сигнала и локальной реплики; β_0/β_1 – векторы параметров прямого/отраженного сигналов⁵, состоящие из (τ_0', f_0', C_0') и (τ_1', f_1', C_1') соответственно.

Значения модели (5.1) в частотных (f_k) и временных (τ_l) позициях, соответствующих процессу формирования БНН, можно выразить как

$$P_{MP}'(\beta_0, \beta_1, f_k, \tau_l) = C_0' R_{\text{LUT}}^2(\tau_0' + \tau_l) \times \sin^2((f_0' + f_k) T_{\text{COH}}) + C_1' R_{\text{LUT}}^2(\tau_1' + \tau_l) \sin^2((f_1' + f_k) T_{\text{COH}}), \quad (5.2)$$

где R_{LUT} – табличное значение корреляции сигнала и локальной реплики для имеющейся системы обработки сигнала.

В данном случае достаточно остро встает вопрос выбора начальных условий β_0 и β_1 для модели. Предлагается следующий алгоритм решения данной задачи.

Шаг 1. Нахождение глобального максимума БНН, оценка параметров первого сигнала по найденному максимуму (обозначим его P1).

Шаг 2. Анализ разности БНН и P1 (обозначим ее D1). В случае наличия в D1 нескольких локальных максимумов с большой мощностью (более чем $(C/N_0 - 6)$ дБГц), задержка и мощность P1 уменьшаются, повторяется Шаг 2.

Шаг 3. Нахождение глобального максимума разности D1. Оценка параметров второго сигнала по найденному максимуму

(обозначим его P2).

Шаг 4. Анализ разности (D1–P2) (обозначим ее D2). В случае наличия в D2 глобального максимума большой мощности (более чем $(C/N_0 - 6)$ дБГц), делается предположение о наличии третьего отраженного сигнала⁶ и выбор начальных параметров считается неудавшимся. В подобном случае необходимо использовать другую модель для аппроксимации, его рассмотрение выходит за рамки данной статьи.

При успешном выборе начальных оценок РНП предполагаемых сигналов, их уточнение производится с помощью метода Ньютона–Гаусса, аналогично с данным подразделом. Отличием является оптимизируемая функция:

$$F(\beta_0, \beta_1) = \sum_{k=0}^{M-1} \sum_{l=1}^L r_{k,l}^2; \quad (6)$$

$$r_{k,l} = (P(f_k, \tau_l) - P_{MP}'(\beta_0, \beta_1, f_k, \tau_l)),$$

где M – число частотных позиций в БПФ; L – число корреляторов.

Полученные значения параметров сигналов проходят дополнительную формальную проверку, состоящую из следующих условий:

1. Разность мощностей сигналов не должна превышать 6 дБ.
2. Разность задержек сигналов должна быть не более 0,9 мкс.
3. Разность частот сигналов должна быть не более 45 Гц.
4. Мощности обоих сигналов должны лежать в промежутке [24:48] дБГц.

В случае нарушения условий 1–3, счи-

⁵ Индекс ноль соответствует прямому, единица – отраженному сигналам.

⁶ При моделировании ситуаций с одним отраженным сигналом наличие глобальных максимумов большой мощности в анализируемой на Шаге 4 разности установлено не было.

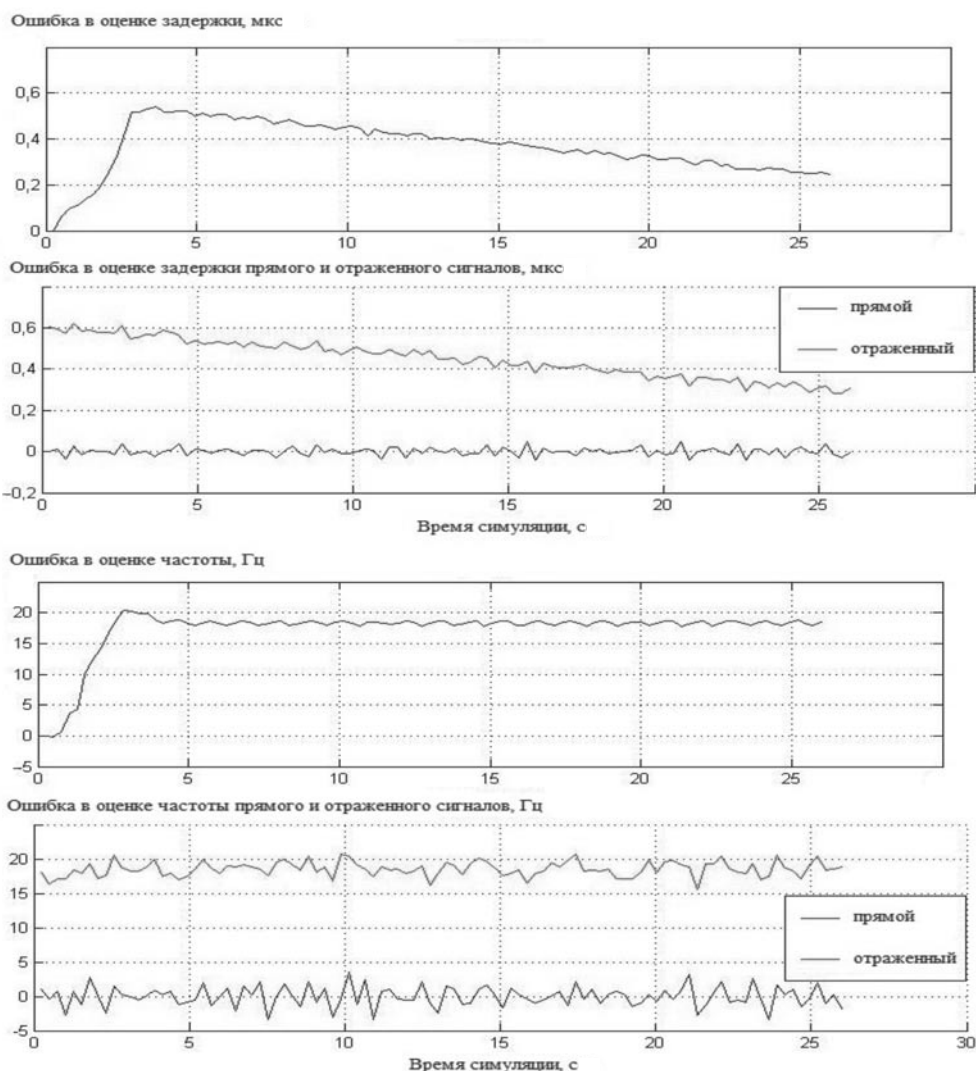


Рис. 3. Результаты тестирования на одном из сценариев

таем, что влияние отраженного сигнала на прямой является незначительным.

Пример работы метода

Работа предлагаемого метода протестирована с помощью аппаратного многоканального симулятора GPS сигнала GSS-6300 компании Spirent. БНН, сформированные приемником при проигрывании сценариев с различной конфигурацией параметров прямого и отраженного сигналов, обрабатывались с помощью предлагаемого метода с целью оценки его способности эффективно обнаруживать отраженный сигнал и формировать оценки РНП прямого.

На рис. 3 показаны результаты оценки параметров прямого и отраженного сигналов в одном из сценариев. В «нулевой» момент был добавлен отраженный сигнал на 3 дБ превышающий мощность прямого и отстоящий от него на 0,6 мкс по времени и на 18 Гц по частоте. Имитировался случай сближения приемника и отражающей поверхности при частичном ослаблении прямого сигнала⁷. Начальная мощность прямого сигнала составляла 32 дБГц.

Представленные результаты показыва-

⁷ Например, при прохождении сигнала через крону деревьев.

Таблица 2

Результаты тестирования метода

$f_{MP} \setminus C/N_0$, Гц	30 дБГц	35 дБГц	40 дБГц
15	Avg($\Delta\tau$) = 0,0171; $\sigma^2(\Delta\tau)$ = 0,0117; Avg(Δf) = 2,5348; $\sigma^2(\Delta f)$ = 0,7988	Avg($\Delta\tau$) = 0,0181; $\sigma^2(\Delta\tau)$ = 0,0079; Avg(Δf) = 2,7516; $\sigma^2(\Delta f)$ = 0,5009	Avg($\Delta\tau$) = 0,0188; $\sigma^2(\Delta\tau)$ = 0,0048; Avg(Δf) = 2,9110; $\sigma^2(\Delta f)$ = 0,3187
23	Avg($\Delta\tau$) = 0,0117; $\sigma^2(\Delta\tau)$ = 0,0140; Avg(Δf) = 3,2783; $\sigma^2(\Delta f)$ = 0,8985	Avg($\Delta\tau$) = 0,0136; $\sigma^2(\Delta\tau)$ = 0,0073; Avg(Δf) = 3,5354; $\sigma^2(\Delta f)$ = 0,5857	Avg($\Delta\tau$) = 0,0134; $\sigma^2(\Delta\tau)$ = 0,0056; Avg(Δf) = 3,6499; $\sigma^2(\Delta f)$ = 0,3971
31	Avg($\Delta\tau$) = 0,0106; $\sigma^2(\Delta\tau)$ = 0,0122; Avg(Δf) = 3,6730; $\sigma^2(\Delta f)$ = 3,8080	Avg($\Delta\tau$) = 0,0105; $\sigma^2(\Delta\tau)$ = 0,0087; Avg(Δf) = 3,4785; $\sigma^2(\Delta f)$ = 0,6551	Avg($\Delta\tau$) = 0,0103; $\sigma^2(\Delta\tau)$ = 0,0055; Avg(Δf) = 3,6796; $\sigma^2(\Delta f)$ = 0,4031
39	Avg($\Delta\tau$) = 0,0028; $\sigma^2(\Delta\tau)$ = 0,0144; Avg(Δf) = 2,1984; $\sigma^2(\Delta f)$ = 1,1169	Avg($\Delta\tau$) = 0,0055; $\sigma^2(\Delta\tau)$ = 0,0088; Avg(Δf) = 2,4486; $\sigma^2(\Delta f)$ = 0,6659	Avg($\Delta\tau$) = 0,0050; $\sigma^2(\Delta\tau)$ = 0,0062; Avg(Δf) = 2,5932; $\sigma^2(\Delta f)$ = 0,4314
47	Avg($\Delta\tau$) = -0,00003384; $\sigma^2(\Delta\tau)$ = 0,0156; Avg(Δf) = 1,3665; $\sigma^2(\Delta f)$ = 1,2209	Avg($\Delta\tau$) = 0,0016; $\sigma^2(\Delta\tau)$ = 0,0091; Avg(Δf) = 1,6059; $\sigma^2(\Delta f)$ = 0,6962	Avg($\Delta\tau$) = 0,0014; $\sigma^2(\Delta\tau)$ = 0,0054; Avg(Δf) = 1,7177; $\sigma^2(\Delta f)$ = 0,4660

ют неспособность в подобном случае классического EML-дискриминатора с узкой апертурой и классического частотного дискриминатора на основе функции atan давать правильные оценки РНП прямого сигнала. Оценки задержки и частоты постепенно сместились от истинных, равных нулю. В то же время, предложенный метод успешно детектирует отраженный сигнал, «разделяет» корреляционные пики прямого и отраженного сигналов и дает более точные оценки РНП прямого сигнала.

Тестирование метода

Результат работы предлагаемого метода зависит от мощности сигнала, а также от задержки и частотного сдвига отраженного сигнала относительно прямого. В табл. 2 представлены результаты статистической обработки оценок РНП, формируемых предлагаемым методом, при многократном (более 1000 раз) моделировании ситуации с

неблагоприятными условиями приема многолучевого сигнала: SMR = 3 дБ, задержка отраженного сигнала относительно прямого $\approx 0,1$ мкс.

В таблице следующие обозначения: f_{MP} – разность частот прямого и отраженного сигналов; $\Delta\tau$ и Δf – ошибка в оценке задержки (мкс) и частоты (Гц) прямого сигнала. Avg и σ^2 – их среднее значение и среднеквадратичное отклонение (СКО) соответственно. Как видно из представленных результатов, СКО ошибки снижается с ростом мощности сигнала. На низких отношениях C/N_0 возможно добиться меньшего СКО при повышении интервала некогерентного накопления $T_{МСОН}$, однако данный подход имеет обратную сторону: при движении пользователя возможно существенное изменение РНП сигнала за время $T_{МСОН}$ ⁸.

⁸ Данный вопрос требует дополнительного исследования.

В статье предложен и описан метод улучшения точности измерения РНП сигнала СРНС. Применение метода позволяет улучшить оценки РНП и мощности прямого сигнала при многолучевом распространении, что в свою очередь повышает качество позиционирования. Работа метода основана на применении частотно-временной модели корреляционного пика сигнала СРНС, формируемого и обрабатываемого системами слежения «без обратной связи». Преимуществами предложенного метода перед другими известными методами являются:

- возможность уменьшить ошибку слежения за частотой сигнала, что весьма важно для движущегося пользователя [10];
- применимость в случаях, когда мощность отраженного сигнала близка или превышает мощность прямого (в подобных случаях методы первой группы дают сильно смещенные оценки);
- метод не требует никаких аппаратных доработок архитектуры приемника, т. к. полностью реализуется программно;

- метод не требователен к полосе обработки сигнала и работает при двусторонней ширине полосы 6 МГц.

Основной областью применения данного метода являются случаи с движущимся пользователем при плохих условиях приема прямого сигнала. Подобные случаи часто встречаются при движении по местам с плотной городской застройкой, где большое количество стеклянных и железобетонных поверхностей является источником сильных отраженных сигналов.

Слабые стороны метода – его малая применимость для случая со стационарным пользователем (замирание сигнала не будет иметь ярко выраженной цикличности за время $T_{\text{МСОН}}$), а также сравнительная вычислительная сложность. В настоящее время проходят исследования по адаптации предложенного метода к случаю с неподвижным пользователем, а также по расширению метода до более общего случая с большим количеством предполагаемых отраженных сигналов.

СПИСОК ЛИТЕРАТУРЫ

1. McGraw, G.A. GNSS Multipath Mitigation Using Gated and High Resolution Correlator Concepts [Электронный ресурс] / G.A. McGraw, M.S. Braasch // Proc. of the 1999 National Technical Meeting of The Institute of Navigation. –San Diego, CA, Jan. 1999. –P. 333–342.
2. Garin, L. Strobe & Edge Correlator – Multipath Mitigation for Code [Электронный ресурс] / L. Garin, Dr. F. van Diggelen, J.M. Rousseau // Proc. of ION GPS-96. –Kansas City, S. –Part 1. –P. 657–664.
3. Jones, Jason. Theory and Performance of the Pulse Aperture Correlator [Электронный ресурс] / Jason Jones, Pat Fenton, Brian Smith // Technical Report. –Novatel, Sept. 2004.
4. Sahnoudi, M. Fast Iterative Maximum-Likelihood Algorithm (FIMLA) for Multipath Mitigation in Next Generation of GNSS Receivers [Электронный ресурс] / M. Sahnoudi, M.G. Amin // IEEE Trans. On Wireless Communication, Nov. 2008. – Vol. 7. –№ 11.
5. van Nee, Richard DJ. The Multipath Estimating Delay Lock Loop: Approaching Theoretical Accuracy Limits [Электронный ресурс] / Richard DJ. van Nee, Jaap Sierenveld, Patrick C Fenton [et al.] // Proc. of 1994 IEEE Position, Location and Navigation Symp. (PLANS'94). –P. 246–251.
6. Weill, L. Achieving Theoretical Accuracy Limits for Pseudorange in the Presence of Multipath [Электронный ресурс] / L. Weill // Proc. of the 8th Internat. Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1995). –1995. –P. 1521–1530.
7. Sahnoudi, M. Multipath mitigation techniques using maximum-likelihood principle [Text] / M. Sahnoudi, R. Landry Jr.// Inside GNSS. –2008. –Vol. 3. –№ 8. –P. 24–29.
8. Braasch, M. Comparison of Two Approaches for GNSS Receiver Algorithms: Batch Processing and Sequential Processing Considerations [Электронный ресурс] / M. Braasch // Proc. of the 18th Internat. Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005), Long Beach, CA, Sept. 2005. –P. 200–211.
9. Prata, Misra. Global Positioning System: Signal, Measurements and Performance [Text] / Prata Misra, Per Enge. – 2nd Ed. –Ganga-Jamuna Press, 2006.
10. Михайлов, Н.В. Использование сверхдлинных когерентных накоплений для устранения ошибок многолучевости в приемнике СРНС [Текст] / Н.В. Михайлов, С.С. Поспелов, М.В. Васильев, Д.Е. Юдакин, П.В. Глушков // Труды XVIII Санкт-Петербургской междунар. конф. по

интегрированным системам. –2011. –С. 330–335.

11. **Михайлов, Н.В.** Опыт использования метода SOFTFLEX в аппаратуре потребителей спутниковой навигации [Текст] / Н.В. Михайлов, В.В. Чистяков // Труды XIX Санкт-Петербургской междунар. конф. по интегрированным системам. –2012. –С. 324–330.

12. **Kunysz, W.** A Three Dimensional Choke Ring Ground Plane Antenna [Электронный

ресурс] / W. Kunysz // Proc. of the 16th Internat. Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS/GNSS 2003). –Portland, OR, Sept. 2003. –P. 1883–1888.

13. **Amin, M.G.** A Novel Interference Suppression Scheme for Global Navigation Satellite Systems Using Antenna Array [Электронный ресурс] / M.G. Amin, W. Sun // IEEE J. on Selected Areas in Communications. –2005. –№ 23 (5). –P. 999–1012.

REFERENCES

1. **McGraw G.A., Braasch M.S.** GNSS Multipath Mitigation Using Gated and High Resolution Correlator Concepts; Proc. of the 1999 National Technical Meeting of The Institute of Navigation; San Diego, CA, Jan. 1999. –P. 333–342.

2. **Garin L., van Diggelen Dr.F., Rousseau J.M.** Strobe & Edge Correlator Multipath Mitigation for Code; Proceedings of ION GPS-96; Kansas City, S, Part 1. –P. 657–664.

3. **Jones Jason, Fenton Pat, Smith Brian.** Theory and Performance of the Pulse Aperture Correlator; Technical Report; Novatel, Sept. 2004.

4. **Sahmoudi M., Amin M.G.** Fast Iterative Maximum-Likelihood Algorithm (FIMLA) for Multipath Mitigation in Next Generation of GNSS Receivers; IEEE Trans. On Wireless Communication, Nov. 2008. –Vol. 7. –№ 11.

5. **Richard DJ. van Nee, Jaap Siereveld, Patrick C Fenton, Bryan R Townsend.** The Multipath Estimating Delay Lock Loop: Approaching Theoretical Accuracy Limits; Proc. of 1994 IEEE Position, Location and Navigation Symp. (PLANS'94). –P. 246–251.

6. **Weill L.** Achieving Theoretical Accuracy Limits for Pseudorange in the Presence of Multipath; Proceedings of the 8th Internat. Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1995). –1995. –P. 1521–1530.

7. **Sahmoudi M., Landry Jr. R.** Multipath mitigation techniques using maximum-likelihood

principle; Inside GNSS. –2008. – Vol. 3. –№ 8. –P. 24–29.

8. **Braasch M.** Comparison of Two Approaches for GNSS Receiver Algorithms: Batch Processing and Sequential Processing Considerations; Proc. of the 18th Internat. Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005). – Long Beach, CA, Sept. 2005. –P. 200–211.

9. **Pratap Misra, Per Enge.** Global Positioning System: Signal, Measurements and Performance; 2 nd Ed. –Ganga-Jamuna Press, 2006.

10. **Mikhailov N.V.** Method of Extra-long Coherent Accumulation for Multipath Mitigation in GNSS Receiver; Proc. of the 18th St.-Petersburg Internat. Conf. on Integrated Navigation Systems, 2011. –P. 330–335. (rus)

11. **Mikhailov N.V.** Use of the SoftFlex Approach in the User Equipment for Satellite Navigation; Proc. of the 19th St.-Petersburg Internat. Conf. on Integrated Navigation Systems. –2012. –S. 324–330. (rus)

12. **Kunysz W.** A Three Dimensional Choke Ring Ground Plane Antenna; Proc. of the 16th Internat. Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS/GNSS 2003). –Portland, OR, Sept. 2003. –P. 1883–1888.

13. **Amin M.G., Sun W.** A Novel Interference Suppression Scheme for Global Navigation Satellite Systems Using Antenna Array; IEEE J. on Selected Areas in Communications. –2005. –№ 23 (5). –P. 999–1012.

ГЛУШКОВ Петр Валерьевич – инженер, компания *i-Free*.
197110, Санкт-Петербург, ул. Лодейнопольская, д. 5, лит. А.
E-mail: pglushkov@gmail.com

GLUSHKOV, Peter V. *i-Free company*.
197110, Lodeynopolskaya Str. 5 A, St.-Petersburg, Russia.
E-mail: pglushkov@gmail.com

МИХАЙЛОВ Николай Викторович – доцент кафедры бортовой радиоэлектронной аппаратуры Санкт-Петербургского государственного университета аэрокосмического приборостроения, кандидат технических наук.

190000, Санкт-Петербург, ул. Б. Морская, д. 67.
E-mail: nmikhailov@softnav.ru

MIKHAILOV, Nikolai V. *Saint-Petersburg State University of Aerospace Instrumentation.*
190000, B. Morskaya Str. 67, St.-Petersburg, Russia.
E-mail: nmikhailov@softnav.ru

ЮДАКИН Дмитрий Евгеньевич – аспирант кафедры медицинской радиоэлектроники Санкт-Петербургского государственного университета аэрокосмического приборостроения.

190000, Санкт-Петербург, ул. Б. Морская, д. 67.

E-mail: dyudakin@softnav.ru

YUDAKIN, Dmitry E. *Saint-Petersburg State University of Aerospace Instrumentation.*

190000, B. Morskaya Str. 67, St.-Petersburg, Russia.

E-mail: dyudakin@softnav.ru



УДК 004.8.032.26, 681.513.8

*Е.Н. Бендерская, К.В. Никитин***РЕКУРРЕНТНАЯ НЕЙРОННАЯ СЕТЬ КАК ДИНАМИЧЕСКАЯ СИСТЕМА
И ПОДХОДЫ К ЕЕ ОБУЧЕНИЮ***E.N. Benderskaya, K.V. Nikitin***RECURRENT NEURAL NETWORK AS DYNAMICAL SYSTEM
AND APPROACHES TO ITS TRAINING**

Представлены результаты аналитического исследования рекуррентных нейронных сетей (РНС) и их обобщающая классификация, выполненная с позиций динамических систем и с учетом нового класса РНС – резервуарных РНС. Систематизация знаний в данной предметной области позволила выделить основные динамические режимы работы РНС, а также определить наиболее перспективные направления в развитии методов обучения РНС с учетом выявленных достоинств и недостатков существующих подходов.

РЕКУРРЕНТНАЯ НЕЙРОННАЯ СЕТЬ. ДИНАМИЧЕСКАЯ СИСТЕМА. АЛГОРИТМЫ ОБУЧЕНИЯ. РЕЗЕРВУАРНЫЕ ВЫЧИСЛЕНИЯ. НЕУСТОЙЧИВАЯ ДИНАМИКА.

This paper presents the results of analytical study of recurrent neural networks (RNN) and their summary classification, done from the position of dynamical systems with regard to a new type of RNN – reservoir RNN. Knowledge systematization in the subject field allowed to distinguish main dynamical regimes of RNN and to describe the most perspective trends in the development of training algorithms in terms of the detected advantages and drawbacks of existing approaches.

RECURRENT NEURAL NETWORK. DYNAMICAL SYSTEM. TRAINING ALGORITHMS. RESEVOIR COMPUTING. UNSTABLE DYNAMICS.

Рекуррентные нейронные сети (РНС) – нейронные сети (НС) с обратными связями, благодаря которым они имеют собственную динамику, что отличает их от статических НС, например, нейронных сетей прямого распространения. Состояние РНС в текущий момент времени определяется не только входными сигналами, но и предыдущим состоянием. За счет этого РНС обладают бесконечной памятью, что отличает их от НС с задержками, которые запоминают информацию о входах только за ограниченный промежуток времени.

Ввиду того что РНС является динамической системой, целесообразно провести анализ возможности привлечения результатов исследования динамических систем в смежных областях науки, таких, как физика, теория нелинейных динамических систем, теория хаоса и других, для синте-

за РНС и решения вопросов их обучения [6, 9, 12, 18]. Необходимо отметить, что при этом термины «нейрон» и «НС» могут вообще не использоваться – вместо этого используются такие понятия, как связанная решетка, структура, среда, конечный автомат с памятью, клеточный автомат. Результатом такого анализа станет выявление перспективных направлений развития РНС в части их синтеза и обучения, что актуально ввиду больших потенциальных возможностей РНС, отмеченных во многих научных работах.

Классификация РНС

Систематизация существующих на сегодняшний день различных РНС позволила предложить обобщенную классификацию НС, выделив особенности, присущие РНС.

1. Структура:
 - 1) топология:
 - без обратных связей (прямого распространения, НС радиальных базисных функций);
 - с обратными связями (рекуррентные):
 - полносвязные (каждый с каждым);
 - с локальными связями (в небольшой окрестности);
 - с произвольными, случайными связями;
 - специфическая топология (самоорганизующиеся НС, сети теории адаптивного резонанса);
 - 2) дискретизация в пространстве:
 - дискретное пространство (решетка);
 - непрерывное пространство (среда);
 - 3) гомогенность связей:
 - гомогенные связи (одинаковые по типу и силе);
 - гетерогенные связи (различные по типу или силе);
 - 4) детерминированность:
 - полностью определенные связи;
 - случайные связи.
2. Тип элементов:
 - 1) модель нейрона (реалистичность):
 - искусственные нейроны (искусственные НС);
 - импульсные нейроны (импульсные НС);
 - 2) динамика нейрона:
 - статические нейроны (статические НС);
 - динамические нейроны (динамические НС):
 - нейроны с затухающей памятью;
 - осцилляторы (в т. ч. хаотические);
 - 3) функция активации нейрона:
 - аддитивная;
 - мультипликативная (шунтирующая);
 - расстояния (с нормализацией и без нормализации);
 - 4) передаточная функция нейрона:
 - пороговая (симметричная и несимметричная);
 - линейная с насыщением/без насыщения;
 - сигмоидальная (симметричная и несимметричная);
 - соревновательная (с жестким и мягким максимумами);
 - радиально-базисная;
 - В-сплайн.
3. Тип связей:
 - 1) постоянство:
 - фиксированные связи;
 - переменные связи (сила связи изменяется в течение работы НС);
 - 2) характер связей (для импульсных НС):
 - электрические – сильная связь (токовая);
 - химические – импульсно-связанные (pulse-coupled).
4. Динамика НС:
 - 1) наличие динамики:
 - динамика отсутствует (статические НС);
 - динамика есть (динамические НС)
 - 2) объем памяти:
 - конечная память (НС с задержками, КИХ-фильтры);
 - бесконечная память (РНС, БИХ-фильтры);
 - 3) установившийся режим:
 - аттрактор типа фиксированная точка;
 - предельный цикл (осцилляции, периодические колебания);
 - сложный аттрактор (тор, хаотический аттрактор);
 - 4) устойчивость динамики:
 - устойчивая;
 - на границе устойчивости (автоколебания);
 - неустойчивая;
 - 5) дискретизация по времени:
 - дискретное время (описывается разностными уравнениями);
 - непрерывное время (описывается дифференциальными уравнениями);
 - 6) синхронность динамики (для дискретного времени):
 - асинхронная (состояние изменяется сразу только у одного элемента);
 - синхронная (состояние изменяется сразу у всех элементов);
 - распределенная (состояние изменяется сразу у группы элементов);
5. Алгоритм обучения:
 - 1) с учителем:

- глобальное – на основе оптимизации: обратное распространение; Больцмана – случайное; генетические алгоритмы;
 - локальное – на основе правила Хебба;
- 2) без учителя – самоорганизация (правило Кохонена, синаптическая временная пластичность);
 - 3) с подкреплением (с критиком);
 - 4) без обучения.
6. Решаемая задача:
- 1) извлечение (генерация) признаков;
 - 2) классификация;
 - 3) ассоциативная память;
 - 4) аппроксимация;
 - 5) предсказание;
 - 6) оптимизация;
 - 7) управление, идентификация, диагностика;
 - 8) кластеризация, сегментация.

Структура РНС. Структура РНС является ее основной характеристикой, от которой напрямую зависят все ее свойства. Выделяют полносвязные РНС и РНС с локальными связями. В полносвязных РНС каждый нейрон связан с каждым. Это приводит к подчинению всех нейронов сети какому-то одному глобальному режиму. В РНС с локальными связями каждый нейрон связан лишь с нейронами в некоторой окрестности вокруг него. Такие связи формируются детерминированным или случайным образом. Локальная связность дает нейронам больше независимости, в разных областях РНС могут преобладать разные динамические режимы.

Существуют примеры РНС с изменяющейся структурой: в ходе работы нейронной сети у нее могут появляться (исчезать) нейроны и связи.

При небольшом количестве элементов структура РНС представляет собой *решетку* в одно-, двух- и максимум трехмерном *дискретном* пространстве. Состояние РНС характеризуется совокупностью значений выходов ее нейронов. Для математического описания таких РНС используются системы ДУ. Если количество элементов велико, то от дискретного пространства состояний переходят к *непрерывному*, и структура РНС представляет собой непрерывную *среду* с

распределенными параметрами. Динамика таких бесконечномерных в пространстве систем описывается специальными волновыми уравнениями – уравнениями в частных производных или интегрально-дифференциальными уравнениями.

Для анализа структуры РНС может применяться подход, основанный на теории графов. Он заключается в вычислении различных показателей, таких, как матрица достижимости, пути, циклы, маршруты, кластерный индекс и др. С помощью этих показателей можно приблизительно оценить взаимодействие элементов в РНС.

РНС является динамической системой, работающей в дискретном или непрерывном времени. В случае дискретного пространства значений выходов РНС изменение ее состояний может быть описано системой разностных или дифференциальных уравнений:

$$X[t+1] = g1(X[t], U[t])$$

$$dX/dt = g2(X, U),$$

где $X = (x_1, x_2, \dots, x_n)$ – состояние системы; n – размерность системы; x_i – значение i -го выхода РНС; U – входное воздействие; $g1$ и $g2$ – функция и отображение соответственно, описывающие зависимости состояния системы от предшествующих состояний и входного воздействия (определяют эволюцию системы во времени).

В зависимости от вида $g1$ и $g2$ РНС будет принадлежать к разному классу динамических систем, классификация которых частично совпадает с классификацией динамики для РНС, представленной выше. Одним из важнейших классификационных признаков является линейность или нелинейность $g1$ и $g2$, от чего зависит принадлежность РНС к классу линейных или нелинейных динамических систем.

Анализ системы ДУ может проводиться в зависимости от цели и особенностей динамической системы (линейная/нелинейная, непрерывность/дискретность по времени):

- 1) использование физических принципов работы динамической системы (анализ с позиций превращения энергии);
- 2) применение вариационных принципов

пов для системы ДУ;

3) применение формальных групп и теории инвариантов;

4) геометрический подход;

5) анализ существования и устойчивости периодических движений.

Нейродинамика — область знаний, в которой НС рассматриваются как динамические системы, но основной упор делается на проблему устойчивости [8].

Анализ динамики РНС

РНС с дискретным пространством состояний. *Устойчивость.* Одним из главных свойств РНС как динамической системы является устойчивость. РНС может быть устойчива (не устойчива) в малом и в большом. Устойчивость в большом гарантирует устойчивость во всем пространстве состояний, а устойчивость в малом — лишь в определенных точках. Теоремы Ляпунова дают аналитические критерии устойчивости нелинейных систем:

для устойчивости в большом необходимо подобрать специальную функцию, удовлетворяющую условиям второй теоремы Ляпунова;

для устойчивости в малом необходимо произвести линеаризацию системы в точке равновесия и проверить в ней устойчивость.

Как правило, устойчивость в большом удается выявить лишь для небольшого класса РНС (например, НС Хопфилда) [7]. А устойчивость в малом свидетельствует лишь о поведении системы в точках равновесия. Поэтому разработаны специальные численные характеристики, позволяющие судить об устойчивости системы по ее фазовой траектории.

Основные типы динамики. В зависимости от устойчивости различают три основных типа динамики РНС:

- устойчивая динамика, при которой за конечное время РНС сходится к устойчивому состоянию равновесия (в данном режиме при изменении входного сигнала возможны переходы системы из одного состояния в другое);

- колебательная динамика: состояние РНС описывает замкнутую циклическую

траекторию — предельный цикл; с одной стороны можно считать, что предельный цикл кодирует некоторую информацию, а с другой — что происходят колебания, при этом информация передается в фазовых соотношениях;

- неустойчивая динамика: траектории РНС при сколь угодно близких начальных состояниях расходятся со временем; если рост состояния системы никак не ограничен, то оно уходит в бесконечность, в противном случае устанавливается хаотическая динамика, при которой состояние системы спонтанно изменяется во времени внутри определенной области фазового пространства [2, 7].

Необходимо отметить, что существует еще хаотическая динамика, при которой изменения состояния РНС носят непредсказуемый характер, и со стороны случайного наблюдателя процесс является стохастическим. С точки зрения физики процесса, система «уходит в себя» и слабо реагирует на входные сигналы. В хаотическом режиме количество информации, содержащееся в состоянии РНС, большое, но извлечь эту информацию сложно. Во многих лабораториях исследуется поведение динамических систем в хаотических режимах и предлагаются варианты интерпретации этой динамики [10, 16]. Данный тип динамики является неустойчивым в малом, но устойчивым в большом и в некоторых случаях динамику можно рассматривать как колебательную.

Различные динамические режимы РНС могут рассматриваться как отдельные этапы решения конкретной задачи. Например, хаотический режим соответствует начальному поиску решений — выбору между различными гипотезами, колебательный режим отражает процессы переключения внимания с одного решения на другое, а устойчивый режим отвечает найденному решению [9].

Динамика РНС зависит как от ее структуры, так и от свойств составляющих ее элементов. При этом возможны случаи, когда РНС из устойчивых элементов при определенной структуре неустойчива и наоборот — когда РНС из неустойчивых, хаотических элементов является устойчи-

вой. Однако при исследовании устойчивых режимов НС, как правило, формируется из устойчивых элементов, при исследовании колебательных режимов – из осцилляторов-колебательных элементов и, наконец, при исследовании хаотических режимов – из хаотических осцилляторов [2, 7].

Теория бифуркаций, катастроф и другие разделы теории нелинейной динамики [5] позволяют аналитически исследовать динамику довольно простых систем (2-3 порядка). Это уровень нейрона (синапса). На уровне РНС эти теории не работают. Поэтому в общем случае для РНС произвольной структуры единственный возможный способ анализа динамики – численное моделирование и вычисление различных показателей динамики. Для РНС со сложными динамическими режимами можно вычислить специальные показатели, характеризующие эти режимы. Эти показатели определяются с использованием теории вероятности и математической статистики, теории случайных процессов, теории информации, теории детерминированного хаоса, теории фракталов и теории синхронизации.

Переходные и установившиеся режимы. Динамику РНС можно представить в виде суммы двух составляющих – установившейся и переходной. Переходная составляющая со временем затухает, после чего остается только постоянная составляющая. Для распознавания образов используются переходные и установившиеся составляющие (режимы) [1, 2, 6, 7]:

в переходном режиме после подачи входного образа его динамика накладывается на динамику РНС – по «снимку» результирующей динамики через некоторое время можно распознать входной образ; в данном случае используется свойство кратковременной памяти – учета предыстории изменения входного сигнала (стимула);

в установившемся режиме после подачи входного образа через какое-то время система выходит на аттрактор (фиксированное состояние, предельный цикл или хаотический аттрактор) – по этому аттрактору можно распознать входной образ.

РНС с устойчивыми режимами. РНС

данного класса являются устойчивыми в большом. После изменения их состояния или после подачи входного сигнала через некоторое время они сходятся к устойчивым точкам равновесия – фиксированным аттракторам.

Самым известным представителем таких РНС является НС Хопфилда (НСХ). Это полносвязная РНС с симметричной матрицей связей. Цель обучения НСХ – задать весовые коэффициенты таким образом, чтобы аттракторы совпадали с распознаваемыми образами.

На макроуровне НСХ описывается специальной функцией энергии (функцией Ляпунова). В ходе функционирования НСХ ее энергия уменьшается до тех пор, пока НСХ не попадет в состояние, соответствующее аттрактору. НСХ получила два основных применения – ассоциативная память и оптимизация.

НСХ породила целое направление в исследовании РНС – аттракторные НС [15]. Были предложены разные модификации, улучшающие ее основные недостатки (небольшой объем памяти, наличие ложных аттракторов – химер), а также скорость работы. В простейшем случае изменялся алгоритм обучения и, как следствие, – связность между нейронами. В некоторых подходах состояние каждого нейрона описывалось не одним числом, а вектором. Также использовались более сложные нелинейные модели нейронов и связей (терминальные аттракторы, синергетический компьютер Хакена). Во всех случаях принцип работы НС оставался тем же, что и у НСХ.

В целом, несмотря на большое количество моделей РНС с устойчивыми режимами, широкого распространения они не получили, а стали скорее удобными моделями для исследования (НСХ). В конечном итоге критике была подвергнута идея точечных аттракторов и предсказуемая динамика [17]:

в человеческом мозге основные вычисления осуществляются через хаотическую динамику, а сходимости к чему-то постоянному означает состояние покоя;

РНС с точечными аттракторами не могут обрабатывать динамические образы.

РНС с колебательными режимами или осцилляторные НС. Осцилляторными НС (ОНС) называются НС, состояние которых изменяется во времени по периодическому или близкому к периодическому закону. Иногда их еще называют НС с аттракторами типа «предельный цикл». С точки зрения биологии в основе ОНС лежит гипотеза о взаимосвязи феномена восприятия и колебательной активности нейронов.

Все модели ОНС различаются по следующим главным параметрам:

- тип элементов-осцилляторов;
- тип связей;
- структура.

В качестве осцилляторов могут использоваться детальные модели биологических нейронов (Ходжкина–Хаксли), упрощенные модели биологических нейронов (Хиндмарш–Роуз, Фитцху–Нагумо, Ижикевича), модели «интегрирование и возбуждение», модели популяций нейронов (Вилсона–Кована), а также модели физических осцилляторов, не имеющие никакого отношения к биологическому нейрону (генератор Ван-Дер-Поля, фазовый осциллятор) [2, 5, 17]. Главное требование к осциллятору заключается в том, чтобы он формировал на выходе колебательную активность. Одним из ключевых свойств элементов ОНС является их частота. В ОНС все элементы могут иметь одну частоту, либо быть разбиты на группы с разными частотами.

Тип связей зависит от модели осциллятора. Если это модель биологического нейрона, то связи осуществляются через химические синапсы, если нет, то через электрические. В зависимости от модели осциллятора электрическая связь может быть линейной или нелинейной. Связи могут обладать свойствами саморегуляции (внутренняя динамика) и пластичности (самообучение).

Наконец, наиболее важной характеристикой, отличающей одну модель осцилляторной НС от другой, является ее структура. Структура включает в себя следующие составляющие:

топология (связность) — полносвязная, с локальными связями, со слабыми связя-

ми и др.;

баланс возбуждающих и подавляющих связей (осцилляторов), наличие глобального ингибитора;

распределение силы связей, а также задержек (гомогенные и гетерогенные).

Все исследования ОНС можно разделить на две группы:

математическое исследование явлений синхронизации в ОНС;

применение ОНС для построения моделей обонятельной, зрительной коры и двигательной системы, а также памяти и внимания.

В рамках первой группы рассматривается проблема формирования нейронными структурами сложных пространственно-временных неоднородных структур (спиральных и плоских волн, солитонов). В рамках второй группы рассматривается применение ОНС для моделирования различных функций головного мозга. Исследованию ОНС посвящено большое количество научных работ. Часть из них затрагивает такой вопрос, как возникновение более сложных динамических режимов — хаотических.

РНС с хаотическими режимами. Хаос составляет базовую форму коллективной нейронной активности для всех процессов и функций восприятия. Он играет роль контролируемого источника шума с целью обеспечения непрерывного доступа к предыдущим запомненным образам и запоминанию новых. Хаос позволяет системе быть всегда активной, избавляя ее от необходимости пробуждаться или входить в устойчивое состояние каждый раз, когда изменяется входное воздействие.

Большинство исследователей сходятся во мнении, что наилучшим с точки зрения хранения и обработки информации является режим упорядоченного хаоса [5]. С одной стороны, в этом режиме есть все преимущества хаоса, с другой стороны, этим режимом можно управлять.

Множество состояний, через которое проходит траектория хаотической системы, называется *хаотическим аттрактором*. В зависимости от количества, силы связей и меры хаотичности элементов хаотическая

система может находиться в различных динамических режимах [16]. Наибольший интерес представляет упорядоченный режим работы решеток с формированием кластеров — хаотическая синхронизация. Этот режим находится между двумя крайними режимами — полным порядком и полным хаосом. Он позволяет использовать полезные свойства хаотических систем для решения задач кластеризации [1, 9].

РНС с непрерывным пространством состояний. При очень большом количестве нейронов РНС рассматривается не как дискретная в пространстве система, а как непрерывная среда. В этом случае для описания используются волновые уравнения [11]. Они определяют динамику нейронного поля $u(r, t)$, формируемого средой. Поле $u(r, t)$ характеризует активность РНС в момент времени t в точке с координатами r . Для задания связности используется специальная функция связи $w(r)$, определяющая зависимость силы связи от расстояния, например функция типа «мексиканская шляпа».

Динамику РНС-сред исследуют с помощью специального метода — анализа неустойчивости Тьюринга. Волновое уравнение среды имеет в качестве одного из состояний равновесия состояние покоя. Далее уравнение линеаризуется в окрестности этого состояния и вычисляются показатели, характеризующие тип динамики. При определенном подборе параметров (весовой функции, функции активации) динамика может быть неустойчивой. Такая неустойчивость носит название *неустойчивость Тьюринга* в честь ученого, впервые ее обнаружившего. Эта неустойчивость приводит к образованию стационарных неоднородных структур, названных *диссипативными структурами* или *паттернами Тьюринга* [10]. *Диссипативный* означает, что структура появляется за счет диссипативных процессов рассеивания энергии в открытой системе.

Вид и форма паттернов Тьюринга зависят в первую очередь от числа измерений среды. Для одномерной среды характерно появление глобальных периодических устойчивых структур, получивших назва-

ние «бугорковые» (bump) или «непрерывные» (continuous) аттракторы [14].

При определенных условиях в средах могут появляться более сложные динамические режимы — распространяющиеся волны. Они могут перемежаться с состояниями покоя, а могут существовать постоянно в форме солитонов. В двумерных средах были обнаружены спиральные волны. Есть гипотеза о том, что в реальных биологических системах из-за гетерогенности и нарушения симметрии волны не могут распространяться бесконечно и быстро затухают.

В целом процессы, наблюдаемые в средах, также могут быть обнаружены и в дискретных решетках, но для этого необходима соответствующая размерность для дискретного варианта.

Исследование полезных для решения задач обработки информации свойства таких РНС пока находится в начальной стадии. Поэтому в дальнейшем анализе проблем обучения РНС основное внимание сосредоточено на РНС с дискретным пространством состояний.

Обучение РНС

Классические подходы к обучению РНС. Одно из важнейших свойств НС — возможность обучения. Под обучением, согласно [8] понимается процесс, в котором параметры НС настраиваются за счет моделирования среды, в которую сеть помещена. Тип обучения зависит от способа подстройки этих параметров [8, 11]. Для обучения в НС из внешней среды должны поступать сигналы, изменяющие свободные параметры НС, что влечет за собой другую ответную реакцию НС на входные сигналы.

Обучение РНС в общем случае классифицируется как NP-полная задача даже для НС с одним скрытым слоем [12]. Для ряда случаев (определенная структура РНС и простые обучающие примеры) обучение имеет полиномиальную сложность.

К РНС применимо утверждение: «Чем сложнее устроена обучаемая машина, тем сложнее алгоритм обучения». Поскольку РНС сложнее обычных НС прямого рас-

пространения, то и алгоритмы их обучения сложнее алгоритмов обучения последних.

Для обучения РНС можно выделить несколько различных по своей сути групп подходов:

- использование известных соотношений между параметрами РНС и ее динамикой; пример – НСХ, НС Хакена и другие аттракторные РНС, в которых путем задания весовых коэффициентов можно закодировать желаемые аттракторы на базе правила Хэбба [8];

- использование алгоритмов обучения с учителем на базе методов оптимизации по типу алгоритма обратного распространения ошибки; пример – алгоритмы обратного распространения во времени (Back Propagation Through Time, ВРТТ), рекуррентное обучение в реальном времени (Real-Time Recurrent Learning, RTRL), рекуррентное обратное распространение (Recurrent Back Propagation, RBP) [8], алгоритмы, использующие Калмановскую фильтрацию; из-за того, что в этих алгоритмах функционал оптимизации формируется как сумма ошибок на некотором временном интервале, и при этом параметры на каждом шаге настраиваются одни и те же, существуют проблемы со сходимостью методов, временем работы и вычислительными затратами;

- использование алгоритма обучения с учителем при рассмотрении РНС как частично рекуррентной – сигналы с обратных связей рассматриваются как отдельные входные сигналы – контекстные нейроны, в результате задача обучения упрощается и сводится к обычным алгоритмам обучения НС прямого распространения; к данному типу обучения относятся НС Элмана, НС Джордана и другие НС;

- использование алгоритмов обучения без учителя (правило обучения Кохонена, синаптическая пластичность импульсных НС);

- отсутствие обучения в классическом смысле изменения параметров системы (неявное обучение) – настройка весовых коэффициентов случайными значениями; при этом функции обучения возложены

на специальное устройство – считыватель, который занимается классификацией динамики РНС; данный принцип лег в основу новой парадигмы НС – резервуарные вычисления [18].

Последний вариант обучения (без явного обучения через изменение весовых коэффициентов самой РНС) представляет собой новую парадигму обучения, в которой способ связи с окружающей средой представлен в изменении фазовых портретов поведения РНС и формировании определенной динамики в ответ на определенное воздействие (можно сопоставить с кратковременной памятью в НС прямого распространения при отклике на входное воздействие) [8]. Новая парадигма обучения связана с новым подходом в вычислениях – резервуарными вычислениями.

Резервуарные вычисления. Сложность и неэффективность существующих алгоритмов обучения РНС вынуждает искать новые подходы, стратегии для использования их вычислительных возможностей. Одним из таких подходов стало новое направление – резервуарные вычисления (РВ, Reservoir Computing, RC) [17, 18, 20].

Основная идея РВ заключается в использовании РНС как резервуара с богатой динамикой и мощными вычислительными возможностями. При этом резервуар формируется случайным образом, что исключает необходимость проводить его обучение. При подаче на вход резервуара непрерывного сигнала он попадает в определенный динамический режим или состояние. Это состояние зависит от входного сигнала. Резервуар формируется таким образом, чтобы для похожих входных сигналов это состояние было похожим, а для разных – разным. Выход резервуара соединен со специальными устройствами – считывателями, которые по состоянию резервуара решают поставленную задачу – производят классификацию, предсказание, кластеризацию [6]. В качестве считывателей могут использоваться статические обучающие машины с простыми алгоритмами обучения. Таким образом, резервуар интегрирует в своем состоянии динамику входного образа.

А считыватели по состоянию резервуара распознают входной образ.

Причины появления направления РВ связаны с тем, что статические модели НС в силу своих возможностей не могли решать задачу распознавания динамических образов, как и динамические НС с задержками, поэтому начали использоваться более сложные рекуррентные НС. Но проблема обучения РНС заставила искать новые подходы, и в результате появились резервуарные вычисления, сочетающие в себе богатую динамику РНС и мощные возможности по обучению статических НС. Было предложено несколько вариантов реализации резервуарных вычислений:

- сеть с эхо-состояниями (Echo State Network, ESN) [20], резервуар является обычной аналоговой РНС, функционирующей в дискретном времени;

- обратное распространение с декорреляцией (Backpropagation-Decorrelation, BPDC) [19]; модель похожа на ESN, используется алгоритм обучения BPDC, являющийся упрощением алгоритма RTRL; если в RTRL обучается вся РНС, то BPDC построен таким образом, чтобы обучались только считывающие модули;

- машина неустойчивых состояний (Liquid State Machine, LSM)[17], резервуар является импульсной РНС;

- эволюно (Evolino – EVolution of systems with Linear Output) [21] – резервуар составлен из специальных нейронов, моделирующих длительную кратковременную память, выходной слой – линейный;

- временная РНС (temporal RNN) – биологические резервуары [13] и представляющие собой настоящие кортикальные НС;

- прочие резервуары, представляющие собой произвольные динамические системы, как чисто аналитические, так и реальные оптические, биологические, физические, квантовые и другие системы.

Для формирования резервуаров существует три базовых способа [18]:

- использование общих рекомендаций для создания «хороших» резервуаров вне зависимости от решаемой задачи – выбор

топологии, связности, силы связей, задержек;

- адаптация резервуара – обучение без учителя с использованием примеров входных данных:

- на глобальном уровне – резервуар формируется так, чтобы при заданных входных данных он находился в нужном динамическом режиме и обладал необходимыми свойствами (например, разделения);

- на локальном уровне – самоорганизация резервуара с использованием правил синаптической пластичности при подаче входных данных;

- обучение резервуара с учителем с использованием примеров входных и соответствующих им выходных данных; в этом случае для конкретной задачи генерируется множество резервуаров с разными параметрами, оценивается качество распознавания для каждого резервуара и из них выбирается наилучший.

В базовом варианте резервуарные вычисления обладают универсальными вычислительными возможностями в смысле аппроксимации произвольных нелинейных динамических систем с затухающей памятью. Если добавить *обратные связи* со считывателя на резервуар, то появляется возможность аппроксимации систем с постоянной памятью, т. е. машин Тьюринга.

Считыватели являются простыми статическими обучающими машинами: взвешенная линейная регрессия, адалина (с обучением в реальном времени по методу наименьших квадратов), персептрон, k ближайших соседей, машины опорных векторов, статические НС.

Для представления о классах динамических систем на базе РВ и степени сложности систем на биологически подобных нейронах приведем пример: для описания МНС из 125 нейронов с наличием шумовой составляющей и динамическими синапсами необходимо 250 связанных нелинейных стохастических ДУ для нейронов и около 400 разностных уравнений третьего порядка для синапсов.

Использование существующих показателей динамических систем (например, эн-

тропии Колмогорова, фрактальной размерности, показателей синхронизации и др.) для определения режимов функционирования РНС даже для простых случаев хаотических нейронных сетей [1, 9] оказывается неэффективным [3, 4] в силу специфики использования динамических режимов для решения задач обработки информации. Это необходимость выявления особенностей именно совместной относительной динамики элементов в неоднородных нелинейных динамических системах большой размерности. Поэтому для анализа работы сложных РНС требуется использование методов ки-

бернетической физики и методов из новых разделов нелинейной динамики [4].

В данной статье проведен аналитический обзор существующих подходов к исследованию РНС и показаны предпосылки к появлению нового направления в области нейроинформатики – резервуарных вычислений. Проведенная систематизация разновидностей РНС и методов их обучения позволили предложить обобщенную классификацию РНС и показать место РНС с неустойчивой динамикой среди множества других классов РНС.

СПИСОК ЛИТЕРАТУРЫ

1. **Бендерская, Е.Н.** Осцилляторные нейронные сети с хаотической динамикой в задачах кластерного анализа [Текст] / Е.Н. Бендерская, С.В. Жукова // Нейрокомпьютеры: разработка, применение. – 2011. – № 7. – С. 74–86.
2. **Бендерская, Е.Н.** Моделирование нейронной активности мозга и бионспирированные вычисления [Текст] / Е.Н. Бендерская, К.В. Никитин // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. –СПб.: Изд-во Политехн. ун-та, 2011. –№ 6-2 (138). – С. 34–40.
3. **Бендерская, Е.Н.** Возможности применения некоторых характеристик синхронизации для выявления самоорганизующихся кластеров в осцилляторной нейронной сети с хаотической динамикой [Текст] / Е.Н. Бендерская // Нейрокомпьютеры: разработка, применение: научно-технический журнал. – М.: Научный центр нейрокомпьютеров, 2012. – № 11. – С. 69–73 .
4. **Магницкий, Н.А.** Новые методы хаотической динамики [Текст] / Н.А. Магницкий, С.В. Сидоров. – М.: Едиториал УРСС, 2004. – 320 с.
5. **Малинецкий, Г.Г.** Нелинейная динамика и хаос. Основные понятия: Учеб. пособие [Текст] / Г.Г. Малинецкий, А.Б. Потапов. – М.: Ком-Книга, 2006. – 240 с.
6. **Никитин, К.В.** Новый подход к применению рекуррентных нейронных сетей для решения задач распознавания образов [Текст] / К.В. Никитин, Е.Н. Бендерская // Научно-технические ведомости СПбГПУ. –СПб.: Изд-во Политехн. ун-та, 2007. – № 4-2 (52). – С. 85–92.
7. **Тюкин, И.Ю.** Адаптация в нелинейных динамических системах [Текст] / И.Ю. Тюкин, В.А. Терехов. –М.: Изд-во ЛКИ, 2008. – 384 с.
8. **Хайкин, С.** Нейронные сети: полн. курс [Текст] / С. Хайкин; Пер. с англ. – 2-е изд., испр. – М. –СПб. –Киев: ИД «Вильямс», 2006. – 1103 с.
9. **Benderskaya, E.N.** Nonlinear Trends in Modern Artificial Intelligence: A New Perspective [Text] / E.N. Benderskaya // Beyond AI: Interdisciplinary Aspects of Artificial Intelligence. Topics in Intelligent Engineering and Informatics. – Springer, 2013. – Vol. 4. – P. 113–124.
10. **Benderskaya, E.N.** Multidisciplinary Trends in Modern Artificial Intelligence: Turing’s Way [Text] / E.N. Benderskaya, S.V. Zhukova // AIECM – Turing 2013, Book Chapters: Artificial Intelligence, Evolutionary Computation and Metaheuristics. – Springer, 2013. – P. 320–343.
11. **Coombes, S.** Waves, bumps, and patterns in neural field theories [Text] / S. Coombes // Biological Cybernetics. – 2005. – Vol. 93. – № 2. – P. 91–108.
12. **Dasgupta, B.** On the Complexity of Training Neural Networks with Continuous Activation Functions [Text] / B. Dasgupta, H. Siegelmann, E.D. Sontag // IEEE Transactions on Neural Networks. – 1995. – Vol. 6. – № 6. – P. 1490–1504.
13. **Dominey, P.F.** Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning [Text] / P.F. Dominey // Biological Cybernetics. – 1995. – Vol. 73. – № 3. – P. 265–274.
14. **Dunin-Barkowski, W.L.** Hebb-Hopfield neural networks based on one-dimensional sets of neuron states [Text] / W.L. Dunin-Barkowski, N.B. Osovets // Neural Processing Letters. – 1995. – Vol. 2. – № 5. – P. 28–31.
15. **Feng, J.** Fixed-point attractor analysis for a class of neurodynamics [Text] / J. Feng, D. Brown // Neural Computation. – 1998. – Vol. 10. – P. 189–213.

16. **Kaneko, K.** Life: an introduction to complex systems biology [Text] / K. Kaneko. – Berlin: Springer-Verlag, 2006. – 369 p.

17. **Maass, W.** Real-time computing without stable states: a new framework for neural computations based on perturbations [Text] / W. Maass, T. Natschläger, H. Markram // Neural Computation. – 2002. – Vol. 11. – P. 2531–2560.

18. **Schrauwen, B.** An overview of reservoir computing theory, applications and implementations [Text] / B. Schrauwen, D. Verstraeten, J.V. Campenhout // Proc. of the 15th European Symp. on Artificial Neural Networks. – 2007. – P. 471–482.

19. **Steil, J.J.** Backpropagation-decorrelation: recurrent learning with $O(N)$ complexity [Text] / J.J. Steil // Proc. IJCNN. – 2004. – Vol. 1. – P. 843–848.

20. **Jaeger, H.** The «echo state» approach to analyzing and training recurrent neural networks: GMD report 148 [Text] / H. Jaeger. – German National Research Center for Information Technology. – St. Augustin, 2001. – 43 p.

21. **Schmidhuber, J.** Training Recurrent Networks by Evolino [Text] / J. Schmidhuber, D. Wierstra, M. Gagliolo, F. Gomez // Neural computation. – 2007. – Vol. 19. – № 3. – P. 757–779.

REFERENCES

1. **Benderskaia E.N., Zhukova S.V.** Ostsilliatornye neironnye seti s khaoticheskoi dinamikoi v zadachakh klaster'nogo analiza / Neirokomp'utery: razrabotka, primeneniye; Radiotekhnika. – 2011. – № 7. – S. 74–86. (rus)

2. **Benderskaia E.N., Nikitin K.V.** Modelirovanie neironnoi aktivnosti mozga i bionspirovannye vychisleniia / Nauchno-tekhnichestkie vedomosti SPbGPU. Informatika. Telecommunicacii. Upravlenie. – St.-Petersburg: Izd-vo Politehn. un-ta, 2011. – № 6-2(138). – S. 34–40. (rus)

3. **Benderskaia E.N.** Vozmozhnosti primeneniia nekotorykh kharakteristik sinkhronizatsii dlia vyavleniia samoorganizuiushchikhsia klasterov v ostsilliatornoj neironnoi seti s khaoticheskoi dinamikoi / Neirokomp'utery: razrabotka, primeneniye: nauchno-tekhnichestkii zhurnal. – Moscow: Nauchnyi tsentr neirokomp'uterov, 2012. – № 11. – S. 69–73. (rus)

4. **Magnitskii N.A., Sidorov S.V.** Novye metody khaoticheskoi dinamiki. – Moscow: Editorial URSS, 2004. – 320 s. (rus)

5. **Malinetskii G.G., Potapov A.B.** Nelineinaia dinamika i khaos. Osnovnye poniatia. – Moscow: KomKniga, 2006. – 240 s. (rus)

6. **Nikitin K.V., Benderskaia E.N.** Novyi podkhod k primeneniiu rekurrentnykh neironnykh setei dlia resheniia zadach raspoznavaniia obrazov / Nauchno-tekhnichestkie vedomosti SPbGPU. – St.-Petersburg: Izd-vo Politehn. un-ta, 2007. – № 4-2 (52). – S. 85–92. (rus)

7. **Tiukin I.Iu., Terekhov V.A.** Adaptatsiia v nelineinykh dinamicheskikh sistemakh. – Moscow: Izd-vo LKI, 2008. – 384 s. (rus)

8. Khaikin S. Neironnye seti: poln. Kurs; per. s angl; 2-e izd., ispr. – Moscow–St.-Petersburg–Kiev: ID “Vil'iams”, 2006. – 1103 s. (rus)

9. **Benderskaya E.N.** Nonlinear Trends in Modern Artificial Intelligence: A New Perspective / Beyond AI: Interdisciplinary Aspects of Artificial Intelligence.

Topics in Intelligent Engineering and Informatics. – Springer, 2013. – Vol. 4. – P. 113–124.

10. **Benderskaya E.N., Zhukova S.V.** Multidisciplinary Trends in Modern Artificial Intelligence: Turing's Way / AIECM – Turing 2013, Book Chapters: Artificial Intelligence, Evolutionary Computation and Metaheuristics. – Springer, 2013. – P. 320–343.

11. **Coomes S.** Waves, bumps, and patterns in neural field theories / Biological Cybernetics. – 2005. – Vol. 93. – № 2. – P. 91–108.

12. **Dasgupta B., Siegelmann H., Sontag E.D.** On the Complexity of Training Neural Networks with Continuous Activation Functions / IEEE Transactions on Neural Networks. – 1995. – Vol. 6. – № 6. – P. 1490–1504.

13. **Dominey P.F.** Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning / Biological Cybernetics. – 1995. – Vol. 73. – № 3. – P. 265–274.

14. **Dunin-Barkowski W.L., Osovets N.B.** Hebb-Hopfield neural networks based on one-dimensional sets of neuron states / Neural Processing Letters. – 1995. – Vol. 2. – № 5. – P. 28–31.

15. **Feng J., Brown D.** Fixed-point attractor analysis for a class of neurodynamics / Neural Computation. – 1998. – Vol. 10. – P. 189–213.

16. **Kaneko K.** Life: an introduction to complex systems biology. – Berlin: Springer-Verlag, 2006. – 369 p.

17. **Maass W., Natschläger T., Markram H.** Real-time computing without stable states: a new framework for neural computations based on perturbations / Neural Computation. – 2002. – Vol. 11. – P. 2531–2560.

18. **Schrauwen B., Verstraeten D., Campenhout J.V.** An overview of reservoir computing theory, applications and implementations / Proc. of the 15th European Symp. on Artificial Neural Networks. – 2007. – P. 471–482.

19. **Steil J.J.** Backpropagation-decorrelation: recurrent learning with $O(N)$ complexity / Proc. IJCNN. – 2004. – Vol. 1. – P. 843–848.
20. **Jaeger H.** The «echo state» approach to analyzing and training recurrent neural networks: GMD report 148. – St. Augustin: German National Research Center for Information Technology, 2001. – 43 p.
21. **Schmidhuber J., Wierstra D., Gagliolo M., Gomez F.** Training Recurrent Networks by Evolino / Neural computation. – 2007. – Vol. 19. – № 3. – P. 757–779.
-

БЕНДЕРСКАЯ Елена Николаевна – доцент кафедры компьютерных систем и программных технологий Санкт-Петербургского государственного политехнического университета, кандидат технических наук.

195251, Россия, Санкт-Петербург, Политехническая ул., д. 21.
Тел. +7(812)297-42-18, e-mail: helen.bend@gmail.com

BENDERSKAYA, Elena N. *St. Petersburg State Polytechnical University.*
195251, Politekhnikeskaya Str. 21, St.-Petersburg, Russia.
E-mail: helen.bend@gmail.com

НИКИТИН Кирилл Вячеславович – аспирант кафедры компьютерных систем и программных технологий Санкт-Петербургского государственного политехнического университета.

195251, Россия, Санкт-Петербург, Политехническая ул., д. 21.
Тел. +7(812)297-42-18, e-mail: execiter@mail.ru

NIKITIN, Kirill V. *St. Petersburg State Polytechnical University.*
195251, Politekhnikeskaya Str. 21, St.-Petersburg, Russia.
E-mail: execiter@mail.ru

УДК 621.396.663

И.А. Сухов, В.П. Акимов

**ПРИМЕНЕНИЕ АЛГОРИТМОВ «СВЕРХРАЗРЕШЕНИЯ»
К РАДИОПЕЛЕНГАТОРНОЙ АНТЕННОЙ РЕШЕТКЕ
ИЗ НАПРАВЛЕННЫХ ЭЛЕМЕНТОВ**

I.A. Sukhov, V.P. Akimov

**SUPER RESOLUTION TECHNIQUES FOR DIRECTION-FINDER
ANTENNA ARRAY WITH DIRECTIONAL ELEMENTS**

Проведена экспериментальная оценка точности пеленгации и разрешающей способности кольцевой антенной решетки из шунтовых вибраторов с применением алгоритмов «сверхразрешения». Показана возможность повышения разрешающей способности в результате применения алгоритма цифрового формирования лучей.

КЛАССИФИКАЦИЯ МНОЖЕСТВА СИГНАЛОВ. КОЛЬЦЕВАЯ АНТЕННАЯ РЕШЕТКА. РАДИОПЕЛЕНГАЦИЯ. АЛГОРИТМ ПЕЛЕНГАЦИИ. ФОРМИРОВАНИЕ ЛУЧЕЙ.

Super resolution technique is applied to the circular antenna array of shunt dipoles. Direction-finding accuracy and resolution ability are estimated experimentally. Method of the MUSIC algorithm improvement for the direction-finder is suggested.

MULTIPLE SIGNAL CLASSIFICATION. CIRCULAR ANTENNA ARRAY. DIRECTION-FINDING. DIRECTION-FINDING ALGORITHM. BEAMFORMING.

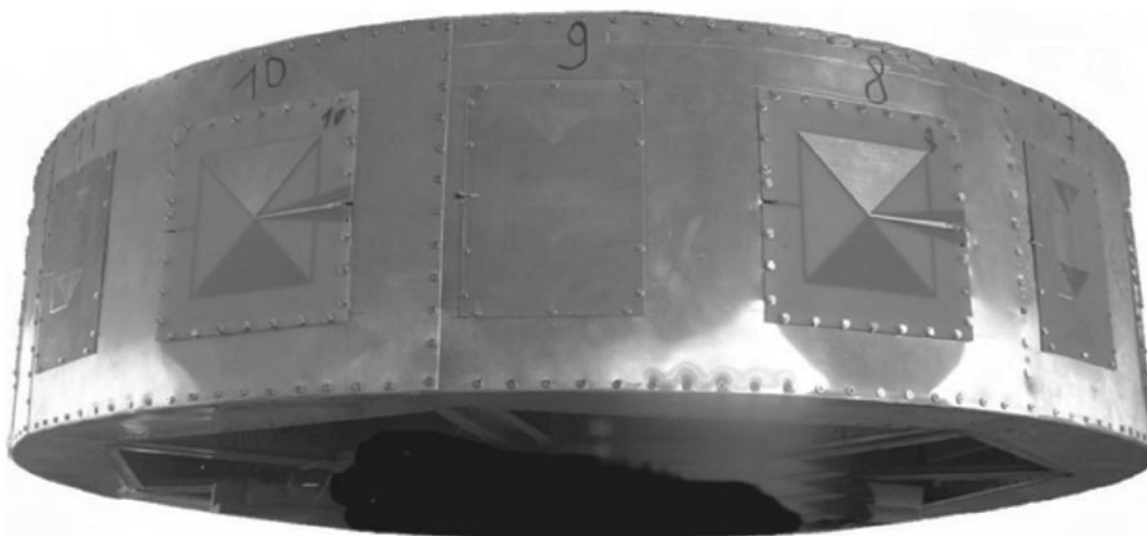
Задачи радиопеленгации осложняются наличием большого количества радиосредств, работающих в одной полосе частот, при этом источники радиоизлучения могут быть разнесены на малое расстояние по азимуту. Традиционные методы пеленгации оказываются не в состоянии разрешить близкорасположенные сигналы. В сложной электромагнитной обстановке предпочтительно использование методов сверхразрешения, т. к. они обеспечивают высокую точность оценки угла прихода одного сигнала и возможность определения пеленгов нескольких близкорасположенных сигналов на одной частоте.

В рамках данной статьи рассматривается вопрос применения сверхразрешающего алгоритма пеленгации MUSIC [1] к коль-

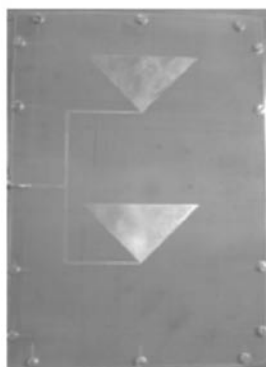
цевой антенной решетке из направленных элементов с использованием амплитуд и фаз сигналов, измеряемых на выходах элементов антенной решетки. Предлагается метод повышения разрешающей способности пеленгатора, обеспечивающий устойчивую работу радиопеленгатора, в случае воздействия нескольких сигналов на кольцевую антенную решетку.

Рассмотрим восьмиэлементную радиопеленгаторную антенную решетку диаметром 1,5 м (рис. 1 *a*). В данной антенной решетке расположены антенные элементы двух диапазонов частот, по восемь элементов на каждый диапазон. В настоящей статье рассматриваются антенные элементы верхнего диапазона частот. В качестве элементов антенной решетки верхнего диа-

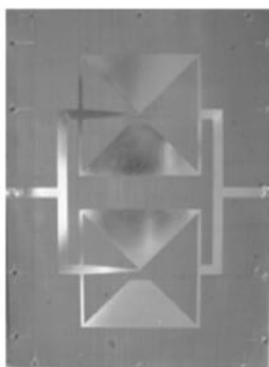
а)



б)



в)



г)

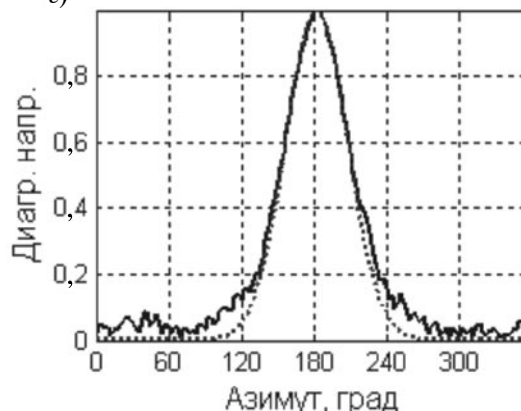


Рис. 1. Антенная решетка (а), шунтовой вибратор (б, в) и диаграмма направленности шунтового вибратора на частоте 1918 МГц (г)

пазона частот используются шунтовые печатные вибраторы, расположенные в резонаторах. Внешний вид антенных элементов представлен на рис. 1 б, вид антенного элемента без резонатора с обратной стороны – на рис. 1 в. Размер резонатора шунтовых вибраторов $200 \times 145 \times 45$ мм³. Резонатор предназначен для улучшения характеристик излучения шунтового вибратора. Вибраторные антенны выполнены на диэлектрике ФАФ-4Д толщиной 1 мм.

Проведено измерение комплексных диаграмм направленности антенных элементов во всем диапазоне углов в азимутальной плоскости для вертикальной поляризации с шагом $2,5^\circ$, с помощью прибора ZVL фирмы R&S. В качестве облучающей антенны

использована рупорная антенна П6-23М. На рис. 1 г представлена нормированная амплитудная диаграмма направленности шунтового вибратора по напряжению на частоте 1918 МГц. Ширина амплитудной диаграммы направленности по уровню 0,707 составляет 43° . Пунктирной линией представлена теоретическая диаграмма направленности, описываемая функцией $F_{\text{app}}(\varphi)$, аппроксимирующей измеренную диаграмму направленности. Выбранная для аппроксимации функция $F_{\text{app}}(\varphi)$ выглядит следующим образом:

$$F_{\text{app}}(\varphi) = \exp \left[- \left(\frac{3\pi}{8} \cdot \frac{\varphi}{\theta_{DN}} \right)^2 \right], \quad (1)$$

где θ_{DN} – ширина амплитудной диаграммы направленности в радианах.

Ширина диаграммы направленности θ_{DN} для функции $F_{\text{app}}(\varphi)$ выбирается в соответствии с шириной аппроксимируемой амплитудной диаграммы направленности, полученной в результате измерений. Теоретическая фазовая диаграмма рассчитывается исходя из предположения, что фазовый центр шунтового вибратора находится в центре антенного элемента. Теоретические диаграммы направленности вводятся для реализации программной модели радиопеленгатора. В процессе работы проводится проверка адекватности разрабатываемой программной модели радиопеленгатора.

На рассматриваемой частоте в случае прихода сигнала с направления, совпадающего с направлением максимума одной из диаграмм направленности антенной решетки, расстояние между соседними антенными элементами составляет порядка 0,7 длины волны, что потенциально может приводить к появлению неоднозначностей при определении пеленга.

Проведен натурный эксперимент для оценки точности пеленгации кольцевой антенной решеткой из шунтовых вибраторов с применением алгоритма пеленгации MUSIC (Multiple Signal Classification), использующего амплитуды и фазы сигналов (комплексные отсчеты) с выходов антенных элементов.

Алгоритм оценки пеленга MUSIC [1] включает в себя несколько этапов. На первом этапе измеряются амплитуды и фазы на выходах всех антенных элементов K раз и формируется матрица измеренных данных $\mathbf{X} = [x_1(t_k), x_2(t_k), \dots, x_N(t_k)]^T, k = 1, \dots, K$, размером $N \times K$, где N – количество антенных элементов. Рассчитывается корреляционная матрица $\mathbf{R} = \mathbf{X} \cdot \mathbf{X}^H$ и производится ее разложение по собственным векторам и собственным числам. По критерию Акайке [2] оценивается количество сигналов, воздействующих на антенную решетку M . Из собственных векторов-столбцов, соответствующих M наименьшим собственным числам, составляется шумовая матрица \mathbf{Q}_N , размером $N \times (N - M)$. На втором этапе строится пеленгационная характеристика

$P_{\text{MUSIC}}(\varphi)$ для всех возможных углов прихода сигнала φ в соответствии с формулой:

$$P_{\text{MUSIC}}(\varphi) = \frac{1}{\mathbf{a}(\varphi)^H \cdot \mathbf{Q}_N \cdot \mathbf{Q}_N^H \cdot \mathbf{a}(\varphi)}, \quad (2)$$

где $\mathbf{a}(\varphi)$ – матрица-столбец, определяющая амплитуды и фазы на всех антенных элементах в случае прихода сигнала с направления φ .

В данной работе матрица-столбец $\mathbf{a}(\varphi)$ измерена экспериментально при проведении калибровки антенной системы.

На третьем этапе необходимо найти M наибольших максимумов пеленгационной характеристики. Азимуты φ , на которых расположены найденные максимумы, соответствуют углам прихода сигнала.

При проведении эксперимента соотношение сигнал/шум определялось относительно уровня сигнала, принимаемого в направлении максимума диаграммы направленности антенного элемента. Для оценки пеленга использовалось $K = 10$ временных выборок значений амплитуд и фаз на всех антенных элементах. Максимумы амплитудных диаграмм направленности расположены на азимутах $0^\circ, 45^\circ, \dots, 315^\circ$. Источник сигнала последовательно располагался на азимутах от 0° до 45° с шагом 5° относительно антенной решетки. В результате проведения эксперимента получены графики зависимости среднеквадратического отклонения пеленга (СКО) от положения источника сигнала при разных соотношениях сигнал/шум (рис. 2).

Наибольшая точность пеленгации достигается в случае прихода сигналов в диапазоне углов между максимумами диаграмм направленности антенных элементов. Значения среднеквадратического отклонения пеленга малы, что связано с большой базой антенной решетки.

Теоретические зависимости построены на основе моделирования процесса пеленгации с использованием данных аппроксимации измеренных амплитудных и фазовых диаграмм. Наибольшее отличие теоретических и экспериментальных зависимостей среднеквадратического отклонения пеленга от азимута прихода сигнала наблюдается в области максимумов диаграмм направлен-

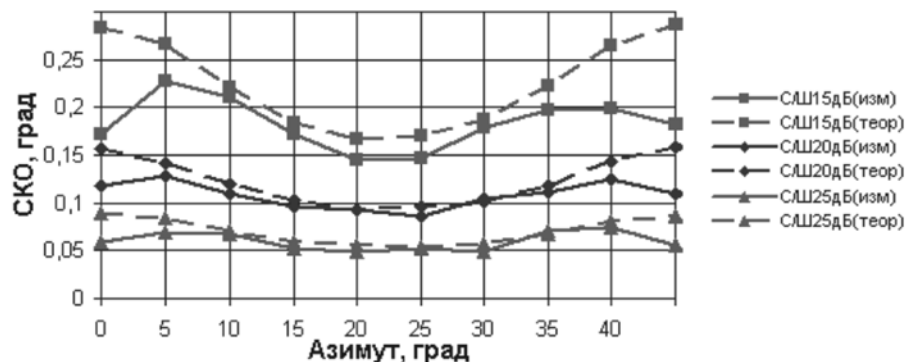


Рис. 2. График зависимости СКО от азимута прихода сигнала (—) экспериментальные данные; (- - -) теоретические данные

ности антенных элементов. Это связано с отличием теоретических и измеренных амплитудных диаграмм направленности. Наибольшее отличие диаграмм направленности наблюдается в области боковых лепестков.

Проведен натурный эксперимент для определения разрешающей способности пеленгатора с использованием рассматриваемой антенной решетки и применением алгоритма пеленгации MUSIC. При проведении эксперимента на антенную решетку воздействовало два сигнала с одинаковой мощностью на частоте 1918 МГц. Соотношение сигнал/шум 20 дБ относительно уровня сигнала в максимуме диаграммы направленности антенного элемента. Количество временных выборок для оценки пеленга 10, азимут одного из сигналов 70°, азимут второго сигнала изменялся от 64° до 69°. В результате проведения эксперимента получен график зависимости вероятности разрешения двух сигналов от расстояния между ними (рис. 3 а).

Два сигнала считаются разрешенными, если пеленгационная характеристика имеет два максимума в направлениях, близких к истинным пеленгам воздействующих сигналов.

Теоретическая зависимость вероятности разрешения сигналов в зависимости от расстояния между сигналами получена в результате использования данных аппроксимации измеренных амплитуд и фаз.

Из графика видно, что вероятность разрешения двух сигналов равна единице при расстоянии между ними в 6° и при сближе-

нии сигналов вероятность оценки быстро уменьшается до нуля. Теоретическая и экспериментальная зависимости практически совпадают, что говорит об адекватности разработанной модели для оценки точностных характеристик реальных пеленгаторов.

Для повышения разрешающей способности предлагается совместить алгоритм цифрового формирования лучей (Digital Beam Forming) и алгоритм пеленгации MUSIC. Полученный алгоритм DBF-MUSIC основан на последовательном применении к измеренным данным алгоритма DBF и алгоритма MUSIC. На первом этапе предлагаемого алгоритма необходимо провести предварительную приблизительную оценку направлений прихода сигналов любым алгоритмом пеленгации. Для этих целей целесообразно использовать алгоритм, не требующий значительных вычислительных затрат (например, неадаптивный формирователь луча [3]). Применение алгоритма цифрового диаграммообразования перед применением алгоритма пеленгации MUSIC позволяет ввести дополнительную пространственную фильтрацию сигналов, позволяющую выделить диапазон углов, требующий более высокой разрешающей способности. В качестве алгоритма цифрового диаграммообразования в рамках данной работы выбран классический алгоритм формирования луча [3]. Алгоритм DBF-MUSIC предполагает цифровое формирование диаграмм направленности с максимумами в направлениях $\Theta \in \{\theta_1, \theta_2, \dots, \theta_B\}$ (B – количество формируемых диаграмм

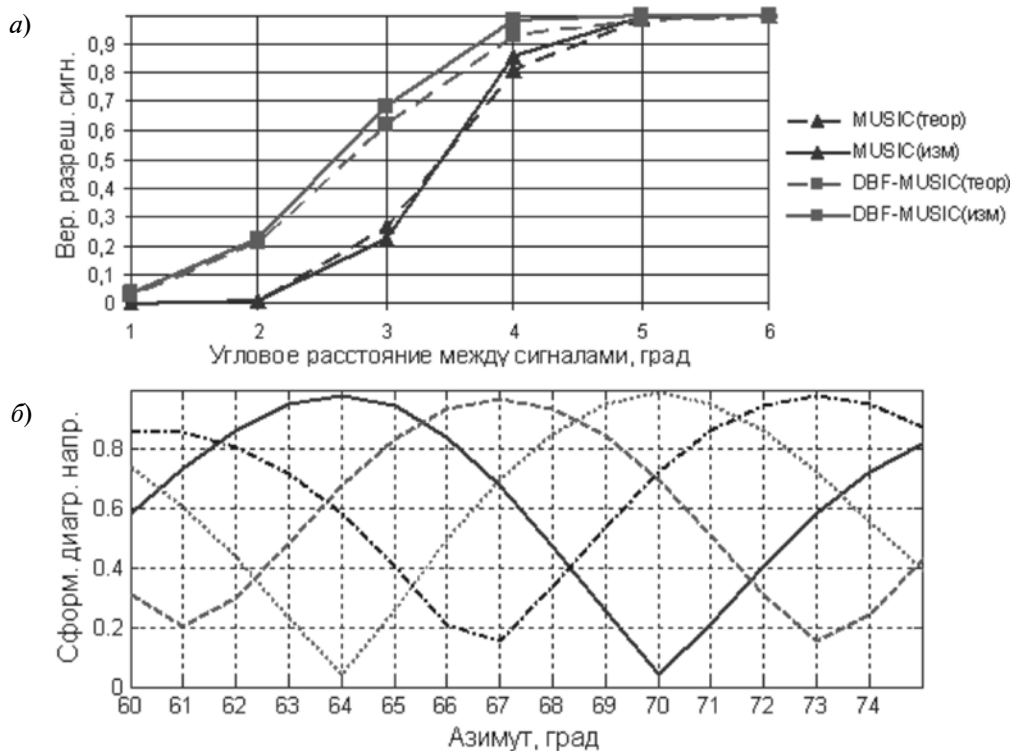


Рис. 3. График зависимости вероятности разрешения двух сигналов от расстояния между ними алгоритмом MUSIC и алгоритмом DBF-MUSIC (а); вид диаграмм направленности, сформированных по алгоритму DBF-MUSIC (б)

направленности), близких к предполагаемым направлениям прихода близкорасположенных сигналов. Для этого все отсчеты амплитуд и фаз сигналов, измеренные на выходах антенных элементов, умножаются на матрицу-столбец $\mathbf{a}(\theta_i), i = 1, \dots, B$ с получением выходных данных $y_i, i = 1, \dots, B$ по формуле:

$$y_i(t_k) = \mathbf{a}(\theta_i)^H \cdot \mathbf{X}, i = 1, \dots, B. \quad (3)$$

Данные, полученные в результате предварительной обработки, удобно записать в виде матрицы $\mathbf{Y} = [y_1(t_k), y_2(t_k), \dots, y_B(t_k)]^T, k = 1, \dots, K$, размером $B \times K$. На следующем этапе рассчитывается корреляционная матрица $\mathbf{R}_B = \mathbf{Y} \cdot \mathbf{Y}^H$, определяется количество воздействующих сигналов M по критерию Акайке. Из M собственных векторов-столбцов корреляционной матрицы \mathbf{R}_B , относящихся к наименьшим собственным числам, составляется матрица \mathbf{Q}_{NB} . Строится пеленгационная характеристика $P_{\text{DBF-MUSIC}}(\varphi)$ для диапазона углов, в направлении которых сформированы диа-

граммы направленности. Пеленгационная характеристика алгоритма DBF-MUSIC выглядит следующим образом:

$$P_{\text{DBF-MUSIC}}(\varphi) = \frac{1}{\mathbf{a}_S(\varphi)^H \cdot \mathbf{Q}_{NB} \cdot \mathbf{Q}_{NB}^H \cdot \mathbf{a}_S(\varphi)}, \quad (4)$$

где $\mathbf{a}_S(\varphi) = (\mathbf{a}(\theta_1)^H \cdot \mathbf{a}(\varphi) \quad \mathbf{a}(\theta_2)^H \cdot \mathbf{a}(\varphi) \quad \dots \quad \mathbf{a}(\theta_B)^H \cdot \mathbf{a}(\varphi))^T$; \mathbf{H} – символ транспонирования и комплексного сопряжения; \mathbf{T} – символ транспонирования.

Данные, полученные в результате воздействия двух сигналов на кольцевую антенную решетку из шунтовых вибраторов, обработаны согласно алгоритму DBF-MUSIC. В соответствии с предлагаемым алгоритмом сформированы четыре диаграммы направленности с максимумами в направлениях 64°, 67°, 70° и 73°. Вид сформированных диаграмм направленности представлен на рис. 3 б.

В результате обработки экспериментальных данных получены зависимости вероятности разрешения двух сигналов от расстояния между ними (рис. 3 а). Из графика видно, что предложенный алгоритм DBF-

MUSIC позволяет увеличить разрешающую способность пеленгатора по сравнению с классическим алгоритмом пеленгации MUSIC. Аналогично предлагаемый алгоритм применен для обработки теоретических амплитудных и фазовых диаграмм. На рис. 3 а представлена теоретическая зависимость вероятности разрешения двух сигналов от расстояния между ними, практически совпадающая с зависимостью, полученной экспериментально, что говорит об адекватности разработанной программной модели радиопеленгатора.

Проведено экспериментальное исследование применения алгоритма пеленга-

ции MUSIC и предложенного алгоритма повышения разрешающей способности DBF-MUSIC на примере антенной решетки из шунтовых вибраторов. Оценены точностные характеристики радиопеленгатора в случае пеленгации одного сигнала и разрешающая способность в случае воздействия двух сигналов на одной частоте. Показана возможность повышения разрешающей способности с помощью предложенного алгоритма DBF-MUSIC по сравнению с алгоритмом MUSIC. Разработана программная модель радиопеленгатора, позволяющая оценивать точностные характеристики и разрешающую способность пеленгатора.

СПИСОК ЛИТЕРАТУРЫ

1. Schmidt, R. Multiple emitter location and signal parameter estimation [Text] / R. Schmidt // Proc. RADC Spectrum Estimation Workshop. – 1979. – P. 243–258.
2. Akaike, H. A new look at statistical model identification [Text] / H. Akaike // IEEE Trans.

Automat. Control. – 1974. – № 19. – P. 716–723.

3. Krim, H. Two decades of array signal processing research: the parametric approach [Text] / H. Krim, M. Viberg // IEEE Signal Process. Mag. – 1999. – № 13 (4). – P. 67–94.

REFERENCES

1. Schmidt R. Multiple emitter location and signal parameter estimation / Proc. RADC Spectrum Estimation Workshop. – 1979. – P. 243–258.
2. Akaike H. A new look at statistical model identification / IEEE Trans. Automat. Control.

– 1974. – № 19. – P. 716–723.

3. Krim H., Viberg M. Two decades of array signal processing research: the parametric approach. / IEEE Signal Process. Mag. – 1999. – № 13 (4). – P. 67–94.

СУХОВ Игорь Александрович – аспирант Санкт-Петербургского государственного политехнического университета.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: sia23@mail.ru

SUKHOV, Igor A. – St. Petersburg State Polytechnical University.

195251, Politekhnikeskaya Str. 29, St.-Petersburg, Russia.
E-mail: sia23@mail.ru

АКИМОВ Валерий Петрович – профессор Санкт-Петербургского государственного политехнического университета, доктор физико-математических наук.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: valeri_akimov@mail.ru

AKIMOV, Valerij P. – St. Petersburg State Polytechnical University.

195251, Politekhnikeskaya Str. 29, St.-Petersburg, Russia.
E-mail: valeri_akimov@mail.ru

УДК 519.876.5 + 681.5.015.87

В.С. Кедрин, О.В. Кузьмин

ЧАСТОТНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ ПЕРИОДИЧЕСКИХ ФУНКЦИЙ С ПОМОЩЬЮ ОЦЕНКИ ЧИСЛЕННОГО РАНГА

V.S. Kedrin, O.V. Kuzmin

FREQUENCY ANALYSIS OF TIME SERIES PERIODIC FUNCTIONS BY MEANS ASSESSMENT OF THE NUMBER OF RANKS

Рассмотрены особенности применения метода сингулярного анализа для временных рядов, образованных аддитивной периодической функцией. Предложена методика исследования динамических свойств системы на основании оценки численного ранга.

ВРЕМЕННОЙ РЯД. ПЕРИОДИЧЕСКАЯ ФУНКЦИЯ. СИНГУЛЯРНОЕ РАЗЛОЖЕНИЕ. ЧИСЛЕННЫЙ РАНГ. ЧАСТОТНЫЙ АНАЛИЗ.

The article describes the features of the method of analysis for singular time series formed additive periodic. Proposed a method research of dynamic properties of the system based on the evaluation of the numerical rank.

TIME SERIES. PERIODIC FUNCTIONS. SINGULAR VALUE DECOMPOSITION. NUMERICAL RANK. FREQUENCY ANALYSIS.

Задача идентификации параметров функции, описывающей развитие сложного процесса, является существенно важной в рамках теории динамических систем. В случае регистрации этих параметров с постоянным шагом τ по времени в виде совокупности временных рядов

$$t_i = t_0 + (i - 1)\tau : y_i = y(t_i), \quad i = \overline{1, N},$$

указанная задача является сложной и многовариантной при отсутствии для их объяснения функциональной модели системы динамических уравнений. Данная проблема имеет место для многих сложных реальных систем, т. к. получение систем динамических уравнений в этом случае связано со значительными трудностями при большом количестве переменных (динамической размерности системы), либо

практически невозможно в силу их многосвязности и многосложности. Поэтому возникает необходимость развития современных методов обработки временных выборок с целью формирования приближенной оценки параметров идентификации динамики системы. Указанное требование особенно актуально в условиях развития современных систем мониторинга в режиме реального времени.

Постановка задачи

В работе [1] определены инвариантные (определяемые видом модели, а не ее числовыми параметрами) аналитические зависимости исследования временной (пространственной) выборки, порождаемой периодической функцией вида

$$f(x) = A \cos(\alpha x + \varphi) + B \sin(\alpha x + \varphi), \quad (1)$$

а также аддитивной комбинации данной функции

$$f(x) = \sum_{i=1}^I (A_i \cos(\alpha_i x + \varphi) + B_i \sin(\alpha_i x + \varphi)) \quad (2)$$

с помощью применения одного из современных дискретных методов – сингулярного анализа [2].

Данные зависимости получены на базе оценки $\text{rank}(\mathbf{A}, \varepsilon)$, т. н. численного ε -ранга матрицы \mathbf{A} , характеризующего такое количество сингулярных чисел s_j , которое с допустимой погрешностью ε определяет близость траекторной матрицы \mathbf{A} (см. [3, 4]) к матрице \mathbf{B} меньшего ранга $\text{rank}(\mathbf{B})$. Таким образом, определение численного (сингулярного) ранга r позволяет выявлять скрытые (латентные) внутренние взаимосвязи между элементами матрицы \mathbf{A} , что характеризует динамику исследуемой временной (пространственной) выборки процесса: число степеней свободы, порождающей его системы, или число переменных, определяющих фазовое состояние этой системы.

Выделенные в работе [1] ключевые зависимости численного ε -ранга $\text{rank}(\mathbf{A}, \varepsilon)$ от количества составляющих сложной гармонической функции вида (2), представляют интерес для разработки новой методологии оперативного анализа состояния динамической системы, что особенно актуально для мониторинга параметров управления в режиме реального времени сложных структурно-неустойчивых систем. В данной работе ставится задача создания методики анализа параметров исследуемой временной (пространственной) выборки процесса, имеющей ярко выраженный периодический характер, на базе оценки численного ε -ранга.

Методика определения частот периодических компонент на основании численного ранга

Выявленные в статье [1] свойства инвариантности численного ε -ранга $\text{rank}(\mathbf{A}, \varepsilon)$ относительно значений частот аддитивных составляющих $\forall a_i \neq \pi k, k \in \mathbb{N}$ и его пропорциональной зависимости от количества I периодических компонент при $\alpha_1 \neq \alpha_2 \neq \dots \neq \alpha_I$ для временных рядов, порождаемых пе-

риодической функцией вида (2), позволяют создать метод анализа частот (спектра) по исходной дискретной выборке временного ряда. Идея предлагаемого метода основана на принципе активной идентификации системы, когда используются специально выбранные тестовые сигналы и изучается влияние состояния системы на фиксированное воздействие. Применительно к контексту решаемой задачи авторы предлагают рассматривать изменение $\text{rank}(\mathbf{A}, \varepsilon)$ траекторной матрицы \mathbf{A} временного ряда, образованного сложением отсчетов исходного ряда и простейшей тестовой периодической компоненты заданной частоты:

$$f_T(x) = A_T \cos(\alpha_T x + \varphi) + B_T \sin(\alpha_T x + \varphi).$$

В случае «воздействия» на исходную временную выборку тестовой компоненты частоты α_T , не присутствующей в исходном временном ряду ($\forall a_i : a_i \neq \alpha_T$), будет наблюдаться увеличение $\text{rank}(\mathbf{A}, \varepsilon)$ в силу свойства его пропорциональной зависимости от количества I периодических компонент при $\alpha_1 \neq \alpha_2 \neq \dots \neq \alpha_I$.

При «воздействии» на исходную временную выборку тестовой компоненты частоты α_T , присутствующей в исходном временном ряду ($\forall a_i : a_i = \alpha_T$), в силу свойств инвариантности относительно значений частот аддитивных составляющих $\forall a_i \neq \pi k, k \in \mathbb{N}$ и дистрибутивности $\text{rank}(\mathbf{A}, \varepsilon)$ останется неизменным:

$$\begin{aligned} f(x) &= \{A_i \cos(\alpha_i x + \varphi) + B_i \sin(\alpha_i x + \varphi)\} + \\ &+ \{A_T \cos(\alpha_T x + \varphi) + B_T \sin(\alpha_T x + \varphi)\} = \\ &= (A_i + A_T) \cos(\alpha_i x + \varphi) + \\ &+ (B_i + B_T) \sin(\alpha_i x + \varphi), \forall a_i : a_i = \alpha_T. \end{aligned}$$

Сформулированная идея легла в основу создания итерационного алгоритма определения частот компонент периодических составляющих (спектра), присутствующих в исходном временном ряду, порождаемом периодической функцией вида (2) (рис. 1).

Исходными данными для алгоритма служат:

выборка исходного временного ряда y , порождаемого периодической функцией вида (2);

диапазон частот $(\alpha_{\text{start}}; \alpha_{\text{end}})$, в котором

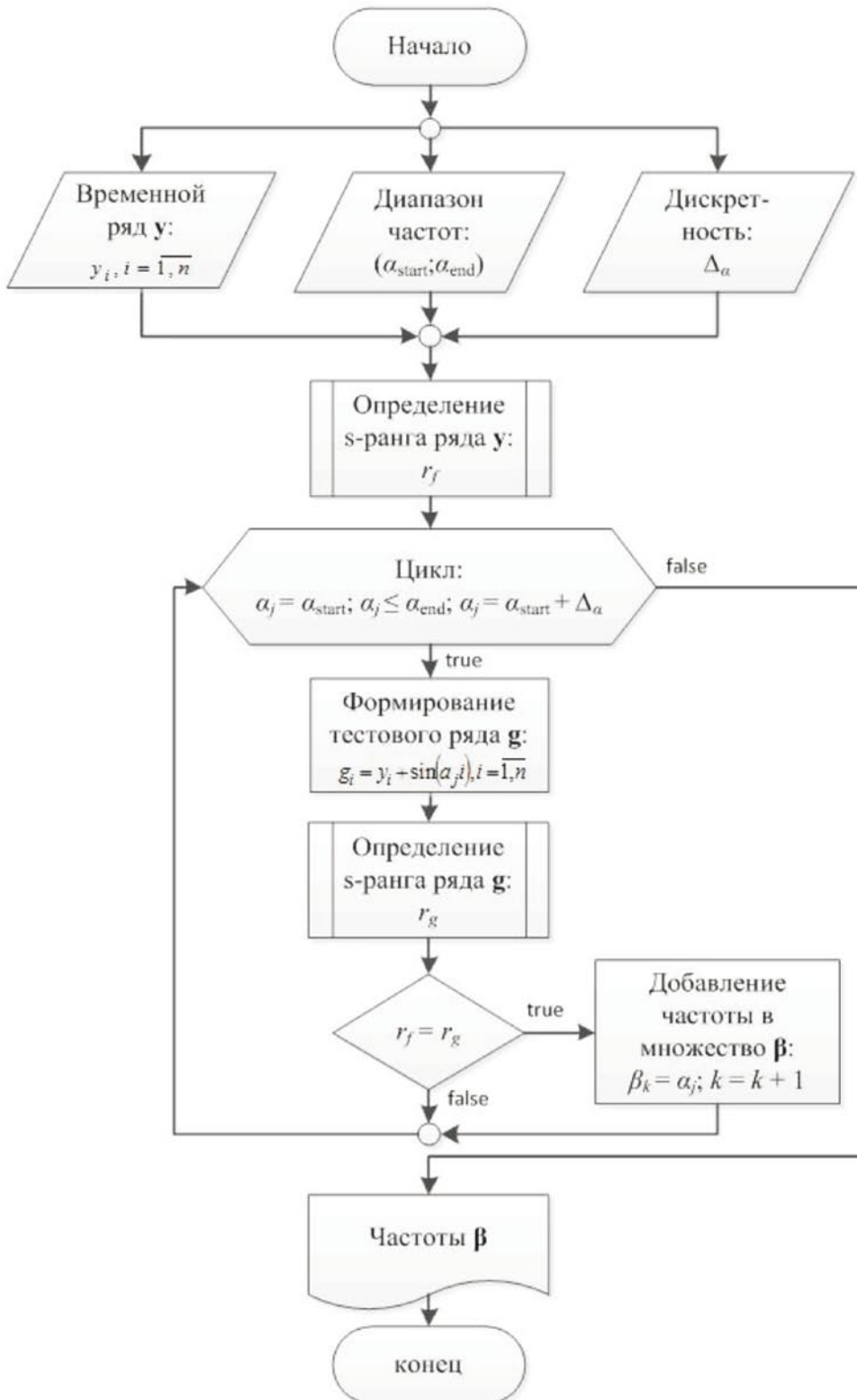


Рис. 1. Алгоритм определения частот периодических компонент

будет производиться поиск частот периодических составляющих, входящих в исходный ряд;

дискретность Δ_α , разбивающая диапазон $(\alpha_{\text{start}}; \alpha_{\text{end}})$ на l отсчетов:

$$\Delta_\alpha = \frac{\alpha_{\text{start}} - \alpha_{\text{end}}}{l}.$$

Особенности определения численного ε -ранга

В алгоритме, представленном в предыдущем разделе, центральное место занимает определение численного ε -ранга $\text{rank}(A, \varepsilon)$, основанное на выделении области значений $\text{range}(A)$ и нуль-пространства $\text{null}(A)$ матрицы A .

Представляет интерес определение критерия данного разделения. Интерпретация метода сингулярного разложения определяет, что разделение областей $\text{range}(A)$ и $\text{null}(A)$ может быть основано на исследовании динамики сингулярных чисел, последовательность которых расположена в порядке убывания модуля на главной диагонали матрицы S (см. [1, 2]).

В качестве иллюстрации сказанного выше приведем результаты анализа выборки сингулярных чисел, полученной в результате сингулярного разложения временного ряда, образованного функцией вида

$$f(i) = 0,1 \sin(i) + 0,2 \cos(2i), i = \overline{1; 51} \quad (3)$$

с параметром разложения $m = 21$ (табл. 1).

Из таблицы видно, что наиболее простым и очевидным критерием разделения областей $\text{range}(A)$ и $\text{null}(A)$ может служить величина абсолютной погрешности ε_s ,

выше которой сингулярные числа являются значимыми и попадают в область значений $\text{range}(A)$. В противном случае сингулярные числа являются незначимыми и характеризуют область $\text{null}(A)$. Так, при $\varepsilon_s = 0,01$ видно, что первые четыре сингулярных числа являются значимыми, следовательно, $\text{rank}(A, \varepsilon) = 4$.

В свою очередь сингулярные числа, определяющие $\text{null}(A)$, отличны от нуля и характеризуются определенной динамикой. Принятие малых ненулевых значений можно объяснить неравномерностью дискретных отсчетов, особенностями дискретного алгоритма сингулярного анализа, а также конечной точностью применяемых в ходе расчета переменных. Поэтому величина абсолютной погрешности численного ε -ранга ε_s не является достаточно гибким критерием разделения областей $\text{range}(A)$ и $\text{null}(A)$. Это легко показать, если изменить выборку исходного периодического ряда (3), прибавив еще одну компоненту близкой частоты к имеющейся в исходном ряду:

$$f(i) = 0,1 \sin(i) + 0,2 \cos(2i) + \sin(0,99i), i = \overline{1; 51}. \quad (4)$$

Выборка сингулярных чисел, полученных в результате сингулярного разложения ряда (4), представлена в табл. 2.

При выборе $\varepsilon_s = 0,01$ видно, что первые четыре сингулярных числа являются значимыми, следовательно, $\text{rank}(A, \varepsilon) = 4$, что не согласуется с представленной моделью функции (4), имеющей три периодических компоненты. Нетрудно заметить, что пятое

Таблица 1

Выборка сингулярных чисел

i	s_i	i	s_i	i	s_i
1	2,64515	8	3,13E-08	15	1,90E-08
2	2,46369	9	3,03E-08	16	1,64E-08
3	1,30862	10	2,92E-08	17	1,58E-08
4	1,22877	11	2,71E-08	18	1,42E-08
5	3,93E-08	12	2,18E-08	19	1,27E-08
6	3,84E-08	13	2,06E-08	20	1,25E-08
7	3,25E-08	14	2,02E-08	21	8,97E-09

Таблица 2

Выборка сингулярных чисел

i	s_i	i	s_i	i	s_i
1	14,5026	8	2,01E-07	15	8,79E-08
2	13,4774	9	1,62E-07	16	8,35E-08
3	2,63163	10	1,54E-07	17	7,43E-08
4	2,44769	11	1,40E-07	18	7,15E-08
5	0,0069	12	1,27E-07	19	6,06E-08
6	5,47E-03	13	1,09E-07	20	5,93E-08
7	2,69E-07	14	1,05E-07	21	4,80E-08

и шестое сингулярные числа хотя и принимают малые значения, но, тем не менее, отличаются на пять порядков от идущих далее сингулярных чисел. В связи с этим необходим выбор более гибкого критерия разделения областей $\text{range}(\mathbf{A})$ и $\text{null}(\mathbf{A})$. С этой целью предлагаем использование относительных моментных коэффициентов динамики – обратного цепного коэффициента роста:

$$Ks_i = \frac{s_{i+1}}{s_i}, \quad (5)$$

т. к. сингулярные числа s_i расположены в порядке убывания.

На основании коэффициента (5) можно определить точки разрыва в динамике выборки сингулярных чисел. Расчет моментных показателей обратного коэффициента роста для выборки сингулярных чисел, образованных в результате сингулярного разложения функции (4), представлен в табл. 3.

В случае графического представления изменения динамики сингулярных чисел

на основании показателя обратного коэффициента роста (рис. 3) целесообразно воспользоваться логарифмической шкалой:

$$\ln(Ks_i) = \ln\left(\frac{s_{i+1}}{s_i}\right),$$

т. к. точки перехода (разрыва) динамики могут принимать достаточно большие значения и существенно отличаются друг от друга.

Из рис. 2 видно, что динамика сингулярных чисел для ряда (4) имеет три точки разрыва, которые можно интерпретировать следующим образом:

первые две точки разрыва (между третьим – вторым и четвертым – пятым сингулярными числами соответственно) характеризуют отделение трех периодических компонент друг от друга;

третья точка разрыва (между шестым и седьмым сингулярными числами), имеющая наибольшее значение, определяет разделение областей $\text{range}(\mathbf{A})$ от $\text{null}(\mathbf{A})$.

Таким образом, на основании выбран-

Таблица 3

Обратные коэффициента роста, отражающие динамику сингулярных чисел

$(i+1)/i$	Ks_i	$(i+1)/i$	Ks_i	$(i+1)/i$	Ks_i
2/1	1,076	9/8	1,243	16/15	1,052
3/2	5,121	10/9	1,048	17/16	1,124
4/3	1,075	11/10	1,106	18/17	1,039
5/4	354,714	12/11	1,096	19/18	1,180
6/5	1,261	13/12	1,172	20/19	1,023
7/6	20311,930	14/13	1,037	21/20	1,236
8/7	1,339	15/14	1,192	22/21	–

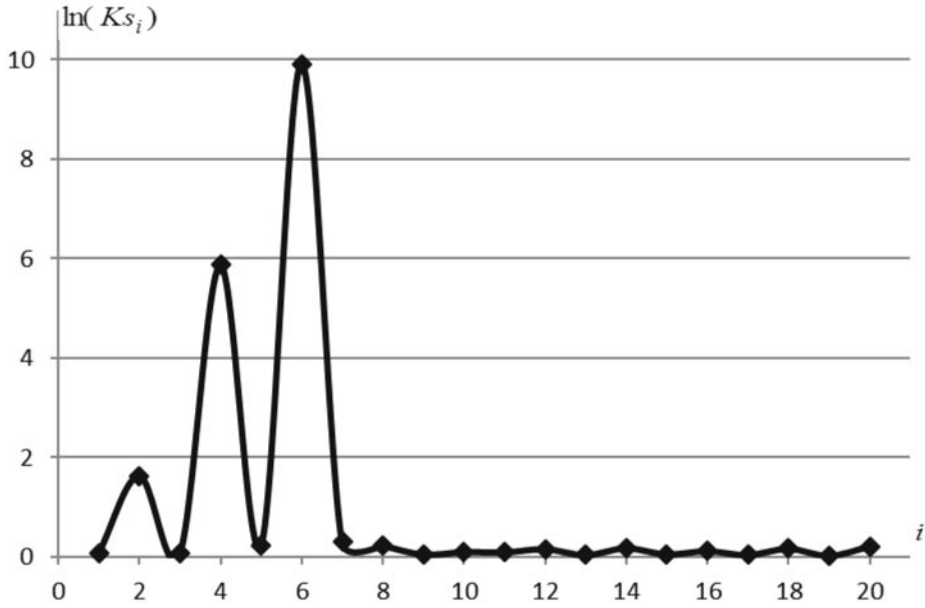


Рис. 2. Обратные коэффициента роста, отражающие динамику сингулярных чисел по логарифмической шкале

ного критерия разделения областей $\text{range}(A)$ и $\text{null}(A)$ имеем $\text{rank}(A, \epsilon) = 6$, что согласуется со сформированными в работе [1] выводами.

Численный эксперимент

Сформированный выше алгоритм использовался при проведении численного эксперимента по применению метода син-

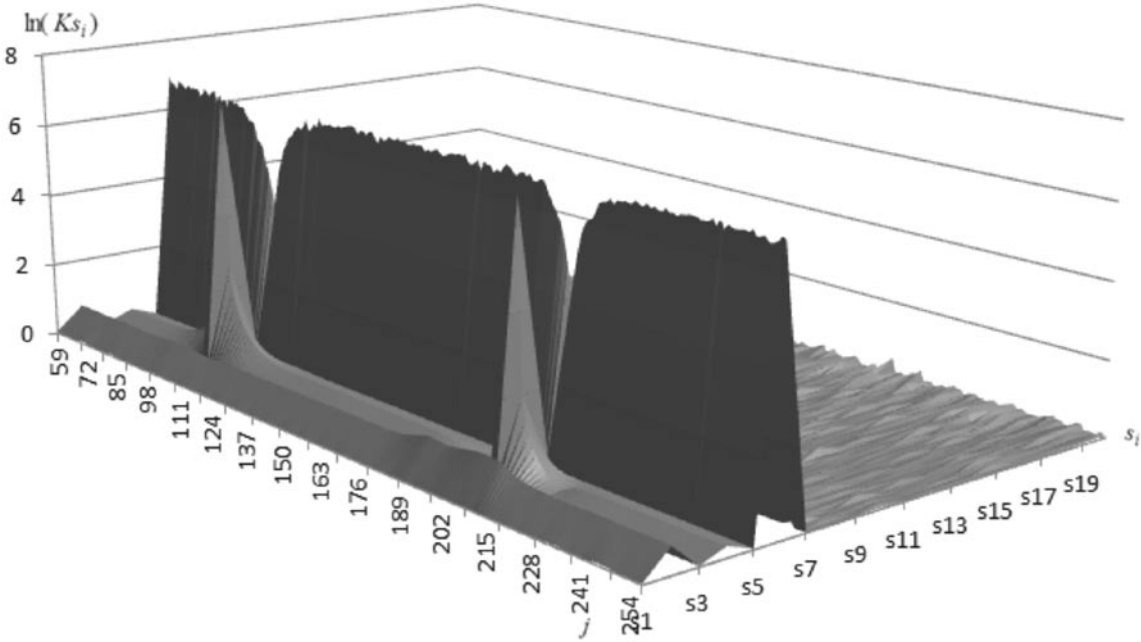


Рис. 3. Обратные коэффициента роста, отражающие динамику сингулярных чисел по логарифмической шкале для тестовых временных рядов, полученных сдвигом по частоте с заданной дискретностью



гулярного разложения к временной последовательности, образованной функцией вида (3).

Для определения присутствующих во временном ряде частот был задан диапазон тестовых частот (0; 3) с дискретностью 0,01.

Графическое изменение обратного коэффициента роста по логарифмической шкале $\ln(K_s)$ для множества рядов $\{g\}$, образованных сложением исходного ряда y и тестовых периодических компонент с определенной тестовой частотой α_T , представлено на рис. 3.

На основании эксперимента определено изменение численного ε -ранга траекторной

матрицы ряда A в точках с индексом 100 и 200. Следовательно, при заданной дискретности 0,01 частоты периодических компонент, присутствующих в исходной последовательности, равны 1 и 2 соответственно.

Таким образом, с помощью оценки численного ε -ранга удалось создать методику, позволяющую не только оценить динамическую размерность d (число степеней свободы) модели, но и идентифицировать параметры частот периодических компонент, присутствующих в исходном временном ряде, порождаемом периодической функцией вида (2).

СПИСОК ЛИТЕРАТУРЫ

1. Кузьмин, О.В. Анализ структуры гармонических рядов динамики на базе алгоритма сингулярного разложения [Текст] / О.В. Кузьмин, В.С. Кедрин // Проблемы управления. – 2013. – №1. – С. 26–31.

2. Воеводин, В.В. Матрицы и вычисления [Текст] / В.В. Воеводин, Ю.А. Кузнецов. – М.: Наука. Гл. ред. Физматлит, 1984. – 320 с.

3. Дойников, А.Н. Методика синтеза математических моделей рядов макроэкономических показателей на основе алгоритмов сингулярного разложения [Текст] / А.Н. Дойников, В.С. Кедрин, М.К. Сальникова // Вестник Иркутского государственного технического ун-та. – 2006. – № 2. – С. 138–142.

4. Дойников, А.Н. Моделирование нестационарных процессов с использованием алгоритмов их сингулярного разложения [Текст] / А.Н. Дойников, В.С. Кедрин, М.К. Сальникова // Научно-технические ведомости СПбГПУ. – СПб.: Изд-во Политехн. ун-та, 2006. – № 5 (47-1). – С. 143–147.

5. Голуб, Дж. Матричные вычисления: Пер. с англ. [Текст] / Дж. Голуб, Ч. Ван Лоун. – М.: Мир, 1999. – 548 с.

6. Golub, G.H. Singular Value Decomposition and least squares solutions [Text] / G.H. Golub, C. Reinsch // Handbook for Automatic Computation. Vol. 2. Linear Algebra; by J.H. Wilkinson, C. Reinsch (Eds.). – NY: Springer-Verlag, 1971. – P. 134–151.

REFERENCES

1. Kuz'min O.V., Kedrin V.S. Analiz struktury garmonicheskikh riadov dinamiki na baze algoritma singuliarnogo razlozheniia / Problemy upravleniia. – 2013. – №1. – P. 26-31. (rus)

2. Voevodin V.V., Kuznetsov Yu.A. Matritsy i vychisleniia. – Moscow: Nauka. Gl. redaksiia Fizmatlit, 1984 – 320 s. (rus)

3. Doinikov A.N., Kedrin V.S., Sal'nikova M.K. Metodika sinteza matematicheskikh modelei riadov makroekonomicheskikh pokazatelei na osnove algoritmov singuliarnogo razlozheniia / Vestnik Irkutskogo gosudarstvennogo tekhnicheskogo un-ta. – 2006. – № 2. – P. 138–142. (rus)

4. Doinikov A.N., Kedrin V.S., Sal'nikova

M.K. Modelirovanie nestatsionarnykh protsessov s ispol'zovaniem algoritmov ikh singuliarnogo razlozheniia / Nauchno-tekhnicheskie vedomosti SPbGPU. St.-Petersburg: Izd-vo Politehn. un-ta, 2006. – № 5. – P. 143–147. (rus)

5. Golub Dzh., Van Loun Ch. Matrichnye vychisleniia: Per. s angl. – Moscow: Mir, 1999. – 548 p. (rus)

6. Golub G.H., Reinsch C. Singular Value Decomposition and least squares solutions // Handbook for Automatic Computation. Vol. 2: Linear Algebra, by J.H. Wilkinson and C. Reinsch (Eds.). – NY: Springer-Verlag, 1971. – P. 134–151.

КЕДРИН Виктор Сергеевич – доцент кафедры информационных технологий филиала Иркутского государственного университета в г. Братске, кандидат технических наук.

665729, Россия, г. Братск, ул. Ленина, д. 34.

E-mail: kedrins@mail.ru

KEDRIN, Victor S. *Bratsk Branch of Irkutsk State University.*
665729, Lenin Str. 34, Bratsk, Russia.
E-mail: kedrinvs@mail.ru

КУЗЬМИН Олег Викторович – *заведующий кафедрой теории вероятностей и дискретной математики Института математики, экономики и информатики Иркутского государственного университета, доктор физико-математических наук, профессор.*

664003, Россия, г. Иркутск, б. Гагарина, д. 20.
E-mail: quzminov@mail.ru

KUZMIN, Oleg V. *Institute of Mathematics, Economics and Information Science of Irkutsk State University.*

664003, Gagarin Blvd. 20, Irkutsk, Russia.
E-mail: quzminov@mail.ru



УДК 681.513

*А.И. Писарев, И.И. Суляев***НЕЙРОЭМУЛЯТОР ТЕМПЕРАТУРНОГО РЕЖИМА ПЛАВКИ
МЕДНО-НИКЕЛЕВОГО СЫРЬЯ В ПЕЧИ ВАНЮКОВА***A.I. Pisarev, I.I. Sulyaev***NEURAL EMULATOR TEMPERATURE MODE OF COPPER-NICKEL RAW
MATERIALS SMELTING IN VANYUKOV FURNACE**

Одна из проблем плавки в печах Ванюкова – большие потери цветных металлов со шлаком, приводящие к снижению производительности процесса. На эффективное разделение продуктов плавки и, как следствие, на их химический состав, существенно влияет температура расплава. Описано построение модели температурного режима плавки медно-никелевого сырья в печи Ванюкова с помощью искусственных нейронных сетей. Разработанная модель позволит создать автоматическую систему управления, поддерживающую оптимальный температурный режим в печи.

ПЕЧЬ ВАНЮКОВА. НЕЙРОННАЯ СЕТЬ. НЕЙРОЭМУЛЯТОР. ТЕМПЕРАТУРНЫЙ РЕЖИМ. МЕДНО-НИКЕЛЕВОЕ СЫРЬЕ. ШТЕЙН. ПРОГНОЗ.

One of Vanyukov furnace smelting problems is large losses of non-ferrous metals with slag, which lead to lower process. Melt temperature significantly influences effective separation of melting products and consequently their chemical composition. The purpose of work is to create temperature smelting model of copper-nickel raw materials in Vanyukov furnace based on artificial neural networks and statistic data use. The elaborated model will create automatic control system supporting optimal temperature mode in furnace.

VANYUKOV FURNACE. NEURAL NETWORK. NEURAL NETWORK EMULATOR. TEMPERATURE MODE. COPPER-NICKEL RAW MATERIALS. MATTE, FORECAST.

Печь Ванюкова (ПВ) – автогенная плавильная печь шахтного типа, предназначенная для переработки медного никельсодержащего сырья, обеднения жидких конвертерных шлаков. Конечными продуктами плавки в ПВ являются: богатые по диоксиду серы отходящие газы (объемная доля до 70 %), бедные по содержанию цветных металлов жидкие отвальные шлаки (объемная доля менее 1 %) и жидкий штейн с высокой концентрацией меди (не менее 50 %) [1, с. 101]. Необходимые условия устойчивой работы печи: непрерывная загрузка шихты заданного состава; бесперебойная подача дутья фиксированного объема (воздух, кислород, природный газ); непрерывный выпуск продуктов плавки в сифоны; постоянство уровней шлаковой и штейновой ванны в плавильной зоне [1]. Объемная

модель печи Ванюкова приведена на рис. 1.

Для эффективного разделения продуктов плавки, уменьшения химических и механических потерь цветных металлов со шлаком необходимо поддерживать оптимальную температуру в шахте печи [2]. Температура штейна в ПВ должна поддерживаться в диапазоне 1050–1150 °С, шлака 1200–1250 °С [1, с. 29]. Стабилизация температуры расплава в ПВ возможна при неизменном расходе кислорода на тонну шихты. При этом эффективное регулирование температуры достигается согласованным изменением расходов воздуха и угля. Еще больший эффект дает управление расходом технологического кислорода и содержанием кислорода в дутье [1].

Температура шлака ниже 1200 °С приводит к низкой растворимости магнетита и его

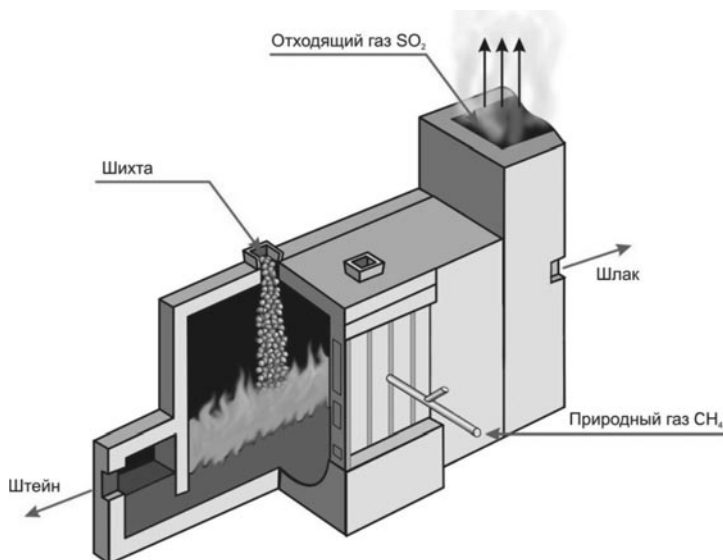


Рис. 1. Печь Ванюкова

выделению в самостоятельную фазу, происходит увеличение механических потерь цветных металлов со шлаком [1, с. 61].

Температура шлака выше 1250 °С вызывает ускоренный износ футеровки печи, перегрев отходящих газов. Стабилизация температуры приведет к снижению потерь цветных металлов со шлаком (рис. 2).

Таким образом, при окислительном процессе Ванюкова следует придерживаться оптимального диапазона температур, обеспечивающих нормальную производительность агрегата при наибольшем извлечении цветных металлов [2].

В статье разрабатывается нейроэмулятор теплового режима плавки медно-никелевого сырья в ПВ. Для этого используются статистические данные работы ПВ

Медного завода горно-металлургической компании «Норильский никель».

Исследованию и созданию температурной модели расплава в ПВ посвящены работы [3, 7], однако полученные математические модели не обладают свойством адаптации к внешним возмущениям, были учтены не все параметры процесса Ванюкова. Такие модели неэффективны в решении задачи оптимального управления печью.

Анализ входных факторов нейроэмулятора

Температура расплава в печи Ванюкова не измеряется в силу определенных причин (высокая температура, агрессивная среда и, тем самым, непродолжительный срок службы контрольно-измерительного прибора). Таким образом, единственным до-

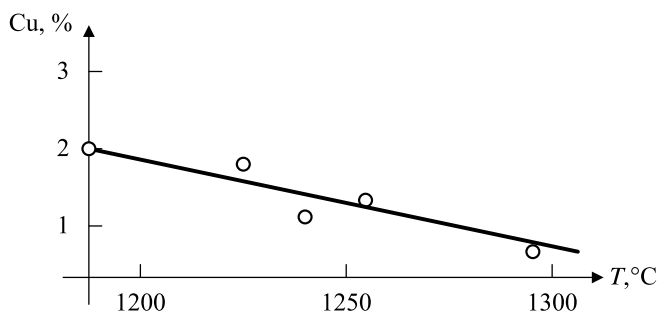


Рис. 2. График зависимости потерь меди со шлаком от температуры расплава в печи Ванюкова

стоверным параметром, косвенно характеризующим температурный режим, является температура воды на сливе с прикладных кессонов 1, 2 ряда горна печи, которая дает надежную информацию о тепловом режиме процесса, в частности, о температуре штейна [1].

В работе [1] при проведении испытаний на опытно-промышленном комплексе печи жидкой ванны было установлено, что перепад температуры в кессонах на 1 °С соответствует изменению температуры ванны расплава печи примерно на 20 °С.

В первом ряду кессонированного пояса печи расположено 22 водоохлаждаемых элемента, во втором – 5. Таким образом, с каждой стороны печи располагается 27

медных прикладных кессонов [8]. Охлаждающая вода подается отдельно на левую и правую сторону печи.

При разработке нейронной сети в первую очередь необходимо определить входные и выходные факторы [9]. В нашей модели только один выходной параметр – средняя температура воды на сливе с прикладных кессонов 1, 2 ряда правой стороны печи, косвенно характеризующая температуру штейна. Отметим, что давление и, как следствие, расход охлаждающей воды на каждую сторону печи разное, поэтому целесообразно выбрать только одну из сторон. Для определения степени влияния входных факторов на выходную величину проведены статистические исследования и выпол-

Таблица 1

Значения коэффициентов корреляции основных технологических параметров процесса Ванюкова

	Температура воды после охлаждения кессонов (правая сторона), °С
Расход КВС после смесителя, нм ³ /ч	0,57
Давление КВС после смесителя, кПа	0,63
Температура КВС после смесителя, °С	0,13
Расход кислорода к смесителю, нм ³ /ч	0,22
Давление кислорода к смесителю, кПа	0,1
Расход воздуха к смесителю, нм ³ /ч	0,3
Давление воздуха к смесителю, кПа	0,3
Содержание кислорода в КВС, %	-0,61
Расход природного газа к зоне плавления, нм ³ /ч	0,63
Давление природного газа к зоне плавления, кПа	0,44
Производительность конвейера 1, т/ч	-0,2
Производительность конвейера 2, т/ч	0,42
Скорость питателя 1, %	0,38
Скорость питателя 2, %	0,34
Скорость питателя 3, %	0,3
Скорость питателя 4, %	0,3
Скорость питателя 5, %	0,5
Скорость питателя 6, %	0,3
Скорость питателя 7, %	0,45
Скорость питателя 8, %	-0,4
Давление воды на охлаждение кессонов (правая сторона), кПа	0,65
Температура воды до охлаждения кессонов (правая сторона), °С	0,62

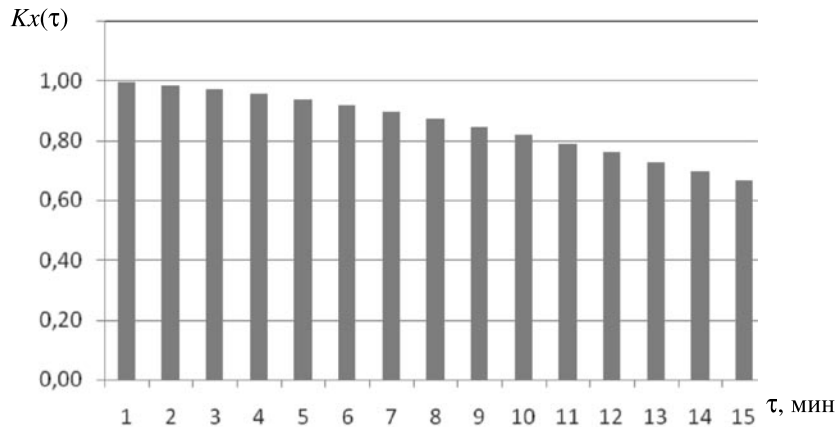


Рис. 3. График автокорреляционной функции средней температуры воды на сливе с прикладных кессонов 1, 2 ряда горна печи Ванюкова

нен корреляционный анализ (табл. 1).

Корреляционный анализ проводился на выборке объемом 200–300 минутных значений по каждому параметру процесса. Отметим, что при вычислении коэффициента парной корреляции между входным фактором и выходной величиной выбирался такой период работы печи, когда остальные параметры процесса были фиксированы или имели незначительные колебания. Такой подход позволяет увидеть реальное влияние отдельного фактора на температурный режим в шахте ПВ.

К примеру, коэффициент корреляции между температурой воды после охлаждения кессонов и расходом природного газа к зоне плавления составил больше 0,6, что отражает значимость фактора и высокую степень влияния на выходную переменную (см. табл. 1).

Из графика автокорреляционной функции $K_x(\tau)$ по температуре кессонов видно, что дискретность опроса температуры должна быть не более 13 мин (рис. 3). Следовательно, используемый в работе шаг квантования в 1 мин вполне приемлем.

С целью «отслеживания» динамики изменения температуры на вход модели подается средняя температура кессонов на текущем j и предыдущем временном шаге $j - 1$.

Таким образом, входными факторами модели являются:

x_{1j} – расход кислородо-воздушной смеси (КВС) после смесителя, $\text{нм}^3/\text{ч}$;

x_{2j} – давление КВС после смесителя, кПа;

x_{3j} – содержание кислорода в КВС, %;

x_{4j} – расход природного газа к зоне плавления, $\text{нм}^3/\text{ч}$;

x_{5j} – суммарный расход шихты, т/ч;

x_{6j} – температура воды до охлаждения кессонов печи, $^{\circ}\text{C}$;

x_{7j} – давление воды на охлаждение кессонов правой стороны печи, кПа;

$x_{8j} \dots x_{15j}$ – скорости питателей под расходными бункерами, косвенно характеризующие объемную производительность по каждому компоненту загружаемой шихты м/с;

$x_{16,j-1}$ – средняя температура воды на сливе с прикладных кессонов 1, 2 ряда горна печи Ванюкова на $j - 1$ шаге, $^{\circ}\text{C}$;

x_{17j} – средняя температура воды на сливе с прикладных кессонов 1, 2 ряда горна печи Ванюкова на j шаге, $^{\circ}\text{C}$.

Выходной параметр модели:

y – средняя температура воды на сливе с прикладных кессонов 1, 2 ряда горна печи Ванюкова на $j + 1$ шаге, $^{\circ}\text{C}$.

Преобработка обучающих данных

Как известно [6, 9, 10], для нейронных сетей весьма верно эмпирическое правило «мусор на входе – мусор на выходе», поэтому очень важна предварительная обработка (преобработка) исходных данных для обучения сети. Введем обозначение массива обучающих данных для нейронной сети:

$$Z = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{ij} \\ y_j \end{pmatrix}_{\substack{j=1+17 \\ j=1+10000}},$$

где i – номер входного параметра; j – момент времени.

В массиве Z каждому набору входных параметров x_{ij} в момент времени j соответствует выходное значение y_j . Таким образом, данные синхронизированы во времени с дискретностью одна минута. Размер обучающего набора соответствует одной неделе работы ПВ (10 000 значений по каждому параметру).

Из массива исключаем строки данных, когда ПВ простаивала или показания контрольно-измерительных приборов были недостоверными (об этом сигнализирует бит состояния параметра в контроллере).

Моделирование и прогнозирование временных рядов с помощью нейронной сети не предполагает пропусков, поэтому пропущенные значения в обучающей выборке заполняются методом линейной интерполяции. Неизвестное значение x_{ij} в момент времени t_j вычисляется следующим образом:

$$x_{ij} = \frac{t_j(x_{i,j-1} - x_{i,j+1}) + (t_{j-1}x_{i,j-1} - t_{j+1}x_{i,j+1})}{t_{j-1} - t_{j+1}}.$$

Преобразование данных осуществляется как по входным факторам x_{ij} , так и по выходному параметру y_j .

Для снижения «нагрузки» на сеть данные подвергаем нормированию, позволяющему предотвратить насыщение нейрона (значительные изменения входов нейрона приводят к незначительному изменению его выхода):

$$\hat{x}_{ij} = \frac{x_{ij} - x_{i \min}}{x_{i \max} - x_{i \min}}, \quad (1)$$

где \hat{x}_{ij} – нормированное значение i -го параметра; $x_{i \min} = \min\{x_{ij}\}$ – минимальное значение i -го параметра; $x_{i \max} = \max\{x_{ij}\}$ – максимальное значение i -го параметра.

Данные приводим к интервалу (0;1), соответствующему диапазону изменения

логистической функции активации каждого нейрона сети. Нормирование исходных данных снижает нагрузку на сеть, что позволяет улучшить соотношение «ошибка/время обучения» [9].

Далее каждый параметр подвергается фильтрации посредством соотношения:

$$\tilde{x}_{ij} = \tilde{x}_{i,j-1} \cdot K_n + \hat{x}_{ij} \cdot K_d, \quad (2)$$

где \tilde{x}_{ij} – отфильтрованное значение сигнала; $\tilde{x}_{i,j-1}$ – отфильтрованное значение сигнала на предыдущем шаге; \hat{x}_{ij} – фильтруемое значение сигнала; K_n, K_d – коэффициенты ($K_n = 0,97, K_d = 0,03$).

Задача фильтрации – очистка от шума, выделение «полезного» сигнала для обучения нейронной сети [10].

Данный метод фильтрации временного ряда позволяет улучшить качество нейросетевого прогноза [10].

Массив Z разбиваем на три множества: обучающее, контрольное и тестовое. Количество данных первого множества выбираем в пределах 50–70 % от исходного. На его основании происходит подстройка синаптических весов нейронной сети. Второе множество – контрольное, на котором при поиске оптимальной структуры выбирается наилучшая модель. Третье множество – тестовое. Итоговая модель проверяется на данных из этого множества, чтобы убедиться, что результаты, достигнутые на обучающем и контрольном множествах, реальны, а не являются артефактами процесса обучения [4].

Отметим, что универсального алгоритма выбора структуры нейросети нет, поэтому подбор количества слоев сети и нейронов в них осуществлялся экспериментально. Основным критерием выбора сети было оптимальное соотношение между временем и ошибкой обучения на обучающем множестве. После экспериментов с обучением различных нейронных сетей была выведена теоретическая кривая соотношения между временем и ошибкой обучения, из которого видно, что при дальнейшем увеличении числа итераций ($t > t_{\text{opt}}$), среднеквадратичная ошибка ε уменьшается незначительно, а вероятность возникновения явления переобучения сети возрастает [8].

В результате эвристического перебора множества моделей лучшее соотношение показала нейросеть со следующей структурой:

- тип нейронной сети – многослойный персептрон;
- количество слоев – 3;
- количество нейронов во входном слое – 17;
- количество нейронов в скрытом слое – 17;
- количество нейронов в выходном слое – 1;
- алгоритм обучения – метод Левенберга–Марквардта.

Выбранное количество нейронов в первом слое является логичным, т. к. количество входных факторов модели равно 17. Аналогично был выбран один нейрон в выходном слое, отвечающий за вывод нормированного значения средней температуры воды на сливе с прикладных кессонов 1,2 ряда горна печи на $j + 1$ шаге.

Активационная функция всех нейронов сети – логистическая (сигмоидальная), описываемая следующей формулой:

$$f(s) = \frac{1}{1 + e^{-\alpha \cdot s}}, \quad (3)$$

где α – параметр наклона логистической функции ($\alpha = 1$); s – взвешенная сумма входов нейрона.

Сигмоидальная функция является гладкой, а ее производная легко вычисляется,

что весьма существенно для алгоритма обучения нейронной сети.

В качестве алгоритма обучения сети выбран метод Левенберга–Марквардта (метод доверительных интервалов), суть которого заключается в минимизации среднеквадратичной ошибки на обучающем множестве посредством подбора весовых коэффициентов и смещений у нейронов [6]. Метод Левенберга–Марквардта – самый быстрый алгоритм обучения, но имеющий ряд важных ограничений:

- применяется только для сетей с одним выходным элементом;
- работает только со среднеквадратичной функцией ошибок;
- требует памяти порядка W^2 (где W – количество синаптических весов у сети, поэтому для обучения больших сетей он неэффективен) [6].

Накладываемые алгоритмом ограничения приемлемы в контексте поставленной задачи по прогнозированию температурного режима печи.

Формула вычисления среднеквадратичной ошибки имеет вид:

$$\varepsilon = \frac{1}{N} \sum_{i=1}^N (T_{эj} - T_j)^2 \rightarrow \min,$$

где $T_{эj}$ – экспериментальное значение температуры на j -м шаге; T_j – рассчитанное значение температуры на j -м шаге; N – объем обучающей выборки (10 000 значений).

Нейронная сеть реализует функцию ре-

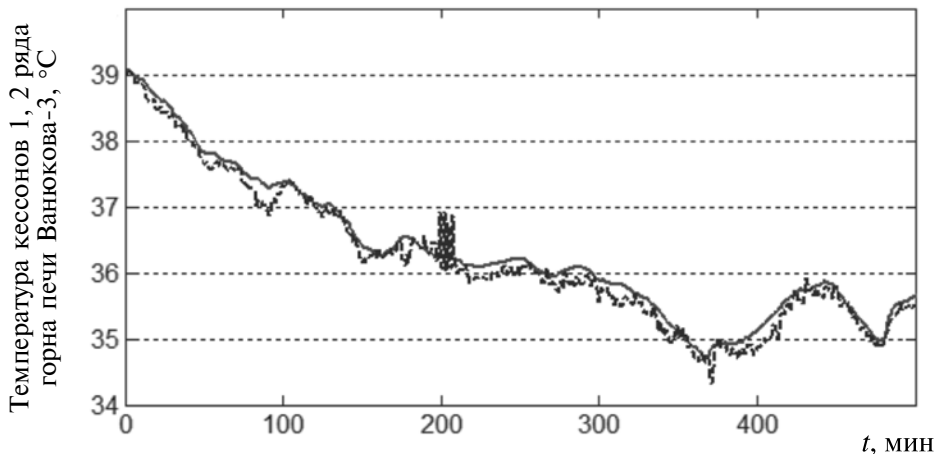


Рис. 4. Результат обучения персептрона структуры 17-17-1 (-----) теоретические данные; (—) экспериментальные данные

Таблица 2

Весовые коэффициенты и смещения нейронов сети

	Номер нейрона скрытого слоя сети																	Номер нейрона выходного слоя сети
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
w_1	1,26	2,6	2,15	-1,96	-0,3	2,06	-14,9	-0,56	1,4	-3,58	4,26	-2,64	0,27	0,95	-4,28	0,4	3,6	2,3
w_2	0,33	-2,08	0,94	1,57	-0,76	1,28	-6,74	0,58	1,21	-2,18	1,21	1,51	0,4	-1,56	0,72	3,57	5,39	-2,08
w_3	0,34	-2,26	-3,3	-3,21	0,44	-2,29	1,74	0,26	-1,94	1,97	-3,14	-1,19	-0,32	-2,37	-3,61	-2,8	-1,61	-2,22
w_4	0,64	0,4	0,36	-1,78	-1,06	2,31	7,51	-1,3	0,53	0,09	6,78	-7,3	4,35	0,17	-1,2	5,14	5,52	0,44
w_5	-1,78	-2,14	-1,29	0,74	-1,64	-2,97	-4,39	2,1	-0,51	-3,47	3,6	3,77	-1,9	-3,9	3,92	-0,31	1,14	-2,14
w_6	0,2	0,58	-1,03	0,69	0,39	0,44	12,31	-1,78	-0,49	-0,24	2,72	-0,26	-0,07	0,32	-0,12	5,96	4,99	0,58
w_7	-0,26	0,16	-0,79	0,6	0,03	0,47	9,5	-1,27	-0,24	-0,61	3,41	0,14	-0,32	0,11	0,2	0,63	2,69	0,1
w_8	-3,24	-4,19	2,74	-2,99	-0,78	-0,42	18,54	-2,24	-2,78	-1,59	10,96	-0,3	-5,54	1,17	0,66	0,67	1,56	-2,19
w_9	0,96	0,61	-0,89	0,65	1,16	-2,33	1,15	1,73	0,84	2,11	-10,17	0,37	-4,88	-0,64	-0,46	-3,87	3,01	0,61
w_{10}	-1,18	-2,02	0,56	0,82	-0,9	0,75	-5,13	1,07	0,98	-1,65	3,08	1,36	-0,82	-1,07	0,62	2,53	6,41	-2,02
w_{11}	0,28	-0,05	0,75	-0,68	-0,02	-0,5	-9,22	1,22	0,18	0,68	-3,43	-0,21	0,38	-0,04	-0,24	-5,08	5,38	-0,05
w_{12}	2,65	-2,58	3,5	-2,94	3,54	0,59	5,49	-0,75	-1,65	-1,27	-10,81	-2,57	0,44	4,18	-3,06	0,63	0,73	-2,58
w_{13}	-3,64	-1,15	1,94	3,22	1,34	-0,07	13,51	-0,09	-0,86	1,75	2,33	-2,7	-3,19	-1,07	-1,47	4,53	4,15	-1,17
w_{14}	-3,34	-2,04	0,43	0,27	-1,21	1,52	-2,7	1,52	1,54	-2,96	-2,79	4,24	1,66	-1,15	-1,36	1,61	-2,03	-2,04
w_{15}	-1,63	-4,31	1,04	-0,26	0,23	1,98	-17,76	-1,44	0,13	-1,98	-4,02	2,94	-2,16	-2,63	-1,56	3,15	-2,76	-4,31
w_{16}	0,53	0,73	1,11	-3,12	-1,37	0,63	-10,63	-3,47	0,86	-3,87	7,15	1,81	3,85	3,87	-2,06	2,98	-2,59	0,71
w_{17}	2,65	-0,88	-1,48	0,31	2,22	-2,03	-13,56	-1,19	0,82	0,78	7,84	1,3	4,46	-2,23	2,6	-1,43	3,08	-0,8
b	-6,4	-6,5	4,01	4,24	-7,2	-2	-3,02	-8,7	-0,4	-0,8	1,37	11,8	-5	-3,7	-3,5	2,35	0,21	-6,7

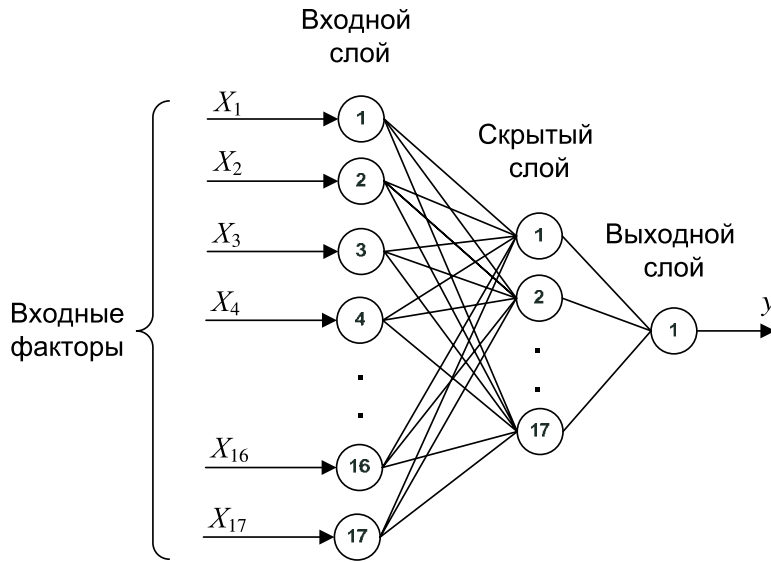


Рис. 5. Внутренняя структура нейроэмулятора

грессии от 17 переменных: $y_j = f(x_{1j}, x_{2j}, x_{3j}, x_{4j}, \dots, x_{17j})$.

Время обучения сети на обучающем множестве составило 20 с при 43 эпохах и среднеквадратичной ошибке 10^{-5} . Результаты обучения сети отображены на рис. 4. Отметим совпадение теоретических и экспериментальных данных с заданной ошибкой в широком диапазоне изменения температуры воды на сливе с прикладных кессонов 1, 2 ряда горна печи (от 34 до 40 °С).

В табл. 2 приведены синаптические веса

и смещения нейронов скрытого и выходного слоя сети, полученные в результате обучения. Отметим, что каждый нейрон имеет 17 весовых коэффициентов w_{km} (где k – номер нейрона в слое, m – порядковый номер весового коэффициента нейрона) и одно смещение b_k , отвечающее за уровень порога активации.

Проверка адекватности нейроэмулятора

Структура нейроэмулятора отображена на рис. 5. Математическая модель выпол-

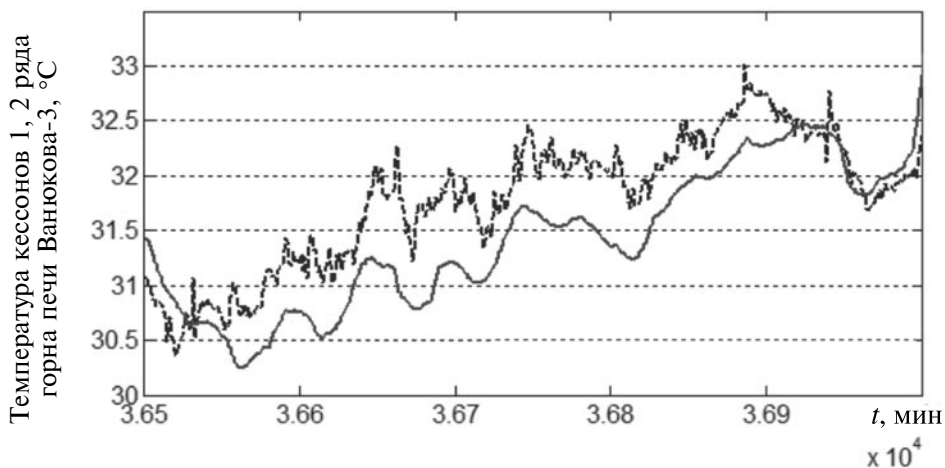


Рис. 6. Экспериментальные и теоретические данные средней температуры воды на сливе с прикладных кессонов 1, 2 ряда горна печи Ванюкова (-----) теоретические данные; (—) экспериментальные данные

нена на базе архитектуры многослойного перцептрона Розенблатта [4] с одним скрытым слоем, в котором количество нейронов равно количеству входных факторов. Нейроны входного слоя сети не участвуют в вычислениях, являясь лишь повторителями сигнала. Весовые коэффициенты таких нейронов равны единице, а смещения – нулю. Выходной слой содержит один нейрон, отвечающий за формирование отклика эмулятора. Нейроны последующего слоя сети связаны с нейронами предыдущего по принципу «каждый с каждым».

Для проверки степени адекватности модели на входы перцептрона подаем нормированный массив данных за один месяц 2010 г. работы ПВ. Полученный вектор значений с выхода модели переводим в реальный диапазон изменения температуры воды на сливе с прикладных кессонов 1, 2 ряда горна ПВ. Экспериментальные $T_{эi}$ и теоретические данные T_{pi} показаны на рис. 6. Рассчитанный множественный коэффициент корреляции

$$R^2 = \frac{\sum_{i=1}^N (T_{pi} - \bar{T}_э)^2}{\sum_{i=1}^N (T_{эi} - \bar{T}_э)^2} \approx 0,9$$

($N = 300$) показывает хорошее совпадение экспериментальных и теоретических данных [5, с. 65].

Однако для построения долгосрочного прогноза температуры (к примеру, на 30 мин вперед) необходимо учитывать динамику изменения каждого параметра в отдельности. При этом структура сети значительно усложняется, а время ее обучения увеличивается. Для уменьшения количества весовых коэффициентов сети возможно использование метода редукции, позволяющего исключить синаптические связи между нейронами, которые не влияют на качество прогноза. Это позволит уменьшить время обучения сети, не увеличивая ошибку прогноза.

В результате регрессионного анализа определена степень влияния входных параметров на температурный режим в шахте печи Ванюкова.

Разработан нейроэмулятор теплового режима плавки медно-никелевого сырья в печи Ванюкова.

Проведена экспериментальная проверка адекватности нейроэмулятора.

СПИСОК ЛИТЕРАТУРЫ

1. **Ванюков, А.В.** Плавка в жидкой ванне [Текст] / А.В. Ванюков, В.П. Быстров, А.Д. Васкевич. – М.: Metallurgia, 1988. – 208 с.
2. **Ванюков, А.В.** Шлаки и штейны цветной металлургии (свойства расплавов и пути снижения потерь металлов со шлаками) [Текст] / А.В. Ванюков, В.Я. Зайцев. – М.: Metallurgia, 1969. – 406 с.
3. **Вернигора, А.С.** Влияние технологических режимов плавки медного сульфидного сырья в печи Ванюкова на износ огневой поверхности кессонов: Дис. ... канд. техн. наук [Текст] / А.С. Вернигора. – М., 2009. – 157 с.
4. **Галушкин, А. И.** Нейронные сети: основы применения [Текст] / А.И. Галушкин. – М.: Горячая линия – Телеком, 2012. – 496 с.
5. **Дрейпер, Н.** Прикладной регрессионный анализ [Текст] / Н. Дрейпер, Г. Смит; В 2-х кн. Кн. 1; Пер. с англ. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 1986. – 366 с.
6. Нейронные сети: STATISTICA Neural Networks: Методология и технологии современного анализа данных [Текст] / Под ред.
- В.П. Боровикова. – 2-е изд., перераб. и доп. – М.: Горячая линия – Телеком, 2008. – 392 с.
7. **Семенова, И.Н.** Разработка системы управления температурным режимом плавки в печи Ванюкова [Текст] / И.Н. Семенова, И.А. Кирпичников // Цветные металлы. – 2009. – № 5.
8. **Суляев, И.И.** Нейросетевая модель температурного режима плавки медно-никелевого сырья в печи Ванюкова [Текст] / И.И. Суляев, А.И. Писарев // Науч. вестник Норильского индустриального ин-та. – 2013. – Вып. 13.
9. **Тадаусевич, Р.** Элементарное введение в технологию нейронных сетей с примерами программ [Текст] / Р. Тадаусевич, Б. Боровик, Т. Гончаж, Б. Леппер; пер. с польск. И.Д. Рудинского. – М.: Горячая линия – Телеком, 2011. – 418 с.
10. **Тимошенко, А.С.** Препроцессинг, построение и выбор оптимальных нейросетевых моделей в прогнозировании временных рядов [Текст] / А.С. Тимошенко, Л.Н. Ясницкий // Автоматизация и современные технологии.

– 2010. – №6. – С. 16–22.

11. Хайкин, С. Нейронные сети: полный

курс [Текст] / С. Хайкин. – 2-е изд; Пер. с англ.
– М.: ИД «Вильямс», 2006. – 1104 с.

REFERENCES

1. Vaniukov A.V. Plavka v zhidkoi vanne. – Moscow: Metallurgii, 1988. – 208 s. (rus)
2. Vaniukov A.V. Shlaki i shteyny tsvetnoi metallurgii (svoistva rasplavov i puti snizheniia poter' metallov so shlakami). – Moscow: Metallurgii, 1969. – 406 s. (rus)
3. Vernigora A.S. Vliianie tekhnologicheskikh rezhimov plavki mednogo sul'fidnogo syr'ia v pechi Vaniukova na iznos ognevoi poverkhnosti kessonov: Dis. ... kand. tekhn. nauk: 05.16.02. – Moscow, 2009. – 157 s. (rus)
4. Galushkin A.I. Neironnye seti: osnovy primeneniia. – Moscow: Goriachaia liniia – Telekom, 2012. – 496 s. (rus)
5. Dreiper N. Prikladnoi regressionnyi analiz. V 2-kh kn. Kn. 1; per. s angl. – 2-e izd., pererab. i dop. – Moscow: Finansy i statistika, 1986. – 366 s. (rus)
6. Neironnye seti: Statistika Neural Networks: Metodologii i tekhnologii sovremennogo analiza dannykh; pod red. V.P. Borovikova. – 2-e izd., pererab. i dop. – Moscow: Goriachaia liniia – Telekom, 2008. – 392 s. (rus)
7. Semenova I.N. Razrabotka sistemy upravleniia temperaturnym rezhimom plavki v pechi Vaniukova / Tsvetnye metally. – 2009. – № 5. (rus)
8. Suliaev I.I. Neurosetevaia model' temperaturnogo rezhima plavki medno-nikelevogo syr'ia v pechi Vaniukova / Nauchnyi vestnik Noril'skogo industrial'nogo in-ta. – 2013. – Vyp. 13. (rus)
9. Tadausevich R. Elementarnoe vvedenie v tekhnologiiu neironnykh setei s primerami programm; per. s pol'sk. I.D. Rudinskogo. – Moscow: Goriachaia liniia – Telekom, 2011. – 418 s. (rus)
10. Timoshenkov A.S. Preprotsessing, postroenie i vybor optimal'nykh neurosetevykh modelei v prognozirovanii vremennykh riadov / Avtomatizatsiia i sovremennye tekhnologii. – 2010. – № 6. – S. 16–22. (rus)
11. Khaikin S. Neironnye seti: polnyi kurs. – 2-e izd; per. s angl. – Moscow: ID «Vil'iams», 2006. – 1104 s. (rus)

ПИСАРЕВ Александр Иванович – *заведующий кафедрой электропривода и автоматизации технологических процессов и производств Норильского индустриального института, кандидат технических наук.*

663321, Россия, Норильск, ул. Озерная, д. 21.
Email: apissarev@yandex.ru

PISAREV, Alexander I. *Norilsk Industrial Institute.*
663321, Ozernaya Str. 21, Norilsk, Russia.
E-mail: apissarev@yandex.ru

СУЛЯЕВ Ильгиз Ильгизович – *аспирант кафедры электропривода и автоматизации технологических процессов и производств Норильского индустриального института.*

663305, Россия, Норильск, ул. Ветеранов, д. 28.
E-mail: ilgizfinland@mail.ru

SULYAEV, Ilgiz I. *Norilsk Industrial Institute.*
663305, Veteranov Str. 28, Norilsk, Russia.
E-mail: ilgizfinland@mail.ru



Вычислительные машины и программное обеспечение

УДК 004.432

Г.С. Петросян

ЯЗЫКОВЫЕ СРЕДСТВА ПОДДЕРЖКИ СИСТЕМАТИЧЕСКОЙ ОБРАБОТКИ ОШИБОК

G.S. Petrosyan

LANGUAGE SUPPORT FOR SYSTEMATIC ERROR HANDLING

Введено разделение ошибок времени выполнения на ожидаемые и неожиданные; предложены критерии, которым должны соответствовать языковые средства поддержки систематической обработки ошибок. Проанализированы популярные подходы к обработке ошибок с использованием механизма исключений и кодов возврата. Предложен подход, согласно которому обработка неожиданных ошибок должна состоять в автоматическом поддержании инвариантов состояния программы. На основе проведенного анализа предложен ряд языковых средств, соответствующий современным требованиям к механизму поддержки систематической обработки ошибок.

ЯЗЫКИ ПРОГРАММИРОВАНИЯ. ОБРАБОТКА ОШИБОК. ИСКЛЮЧЕНИЯ. КОДЫ ВОЗВРАТА. ИНВАРИАНТЫ.

Runtime errors are divided into expected and unexpected; requirements for language support for systematic error handling are proposed. Popular approaches to error handling — exceptions and return codes — are analyzed. Automatic treatment of unexpected errors by means of restoring program state invariants is proposed. Based on the conducted analysis, several mechanisms of language support for systematic error handling which satisfy stated requirements are proposed.

PROGRAMMING LANGUAGES. ERROR HANDLING. EXCEPTIONS. RETURN CODES. INVARIANTS.

Определимся с терминологией: *ошибкой* будем называть ситуацию, при которой часть программы не смогла успешно завершить действие, которое должна была выполнить. Например, в ходе расчета произошло деление на ноль, либо была предпринята попытка чтения из несуществующего файла. Причинами ошибок могут быть как внешние условия (такие, как отсутствие доступа в Интернет или неисправность подключенного оборудования), так и внутренние противоречия (такие, как попытка получить доступ к первому элементу пустого массива) — в данной статье не делается разделение ошибок на какие-либо классы.

Несмотря на то что начинающие про-

граммисты часто стремятся игнорировать обработку ошибок, их опытные коллеги знают, что систематическая обработка ошибок не менее важна, чем «основная» функциональность программы [1]. Объем кода, посвященного обработке ошибок, при этом может достигать 60 % [2]. Неудивительно, что многие языки программирования включают специальные средства, целью которых является поддержка систематической обработки ошибок.

Наиболее распространенное средство поддержки систематической обработки ошибок — механизм *исключений* (exceptions). Первым описанием механизма, похожего на используемый сегодня в «промышлен-

ных» языках (C++, Java, C#), можно считать [3]; формализован подход в [4]. Большое количество современных языков (в т. ч. Java, Scala, C#, F#, Python, PHP, Ruby и JavaScript) копируют вариант, впервые появившийся в C++ [5], и используют его как единственный механизм обработки ошибок. Однако, как показывает анализ реальных программ, такой подход не может считаться удовлетворительным [6].

Вместе с тем некоторые достаточно новые языки – Go [7] и Rust [8] – предполагают обработку ошибок, в т. ч. и без использования специальных языковых средств (тем не менее механизмы, похожие на исключения, присутствуют в обоих языках). Такой подход демонстрирует что исключения, в их общепринятом виде, не являются однозначно предпочтительным средством обработки ошибок.

В данной статье рассматриваются общие моменты обработки ошибок; анализируется обработка ошибок с помощью пространственного механизма исключений, а также без помощи специальных языковых средств (с использованием кода возврата). Выделяются сильные и слабые стороны каждого из подходов и предлагается новый, более предпочтительный в соответствии с критериями анализа.

Обработка ошибок

Какой должна быть реакция программы на возникновение ошибки в вызываемой подпрограмме? Для анализа разделим ошибки на ожидаемые и неожиданные. *Ожидаемыми* назовем ошибки, для которых существует предпочтительный механизм обработки, вытекающий из общей логики работы программы. Например, в случае, если веб-браузер не смог загрузить изображение, вместо него на странице должно быть использовано специальное, встроенное в браузер. Все другие ошибки назовем *неожиданными*.

Как следует из определений, специфическая реакция на ожидаемые ошибки – это часть общей логики работы программы. Определим такую реакцию как *альтернативу* – «если не получилось, то ...» в бизнес-логике. Напротив, обработка неожиданных ошибок не связана с бизнес-логикой и заключается в следующем:

освобождении ресурсов, используемых текущей операцией;

опционально: создание записи об ошибке журнале выполнения;

информировании вызывающего контекста (либо пользователя).

Отметим, что ошибки, вызванные нарушением контракта вызывающим контекстом (например, передача аргумента, для которого не выполняется предусловие), являются неожиданными, и сказанное выше относится к ним тоже.

Исходя из проведенного разделения, сформулируем требования к механизму обработки ошибок.

1. Обработка ожидаемых ошибок (использование *альтернативы*) должна выглядеть как составная часть бизнес-логики.

2. Неожиданные ошибки не должны игнорироваться по умолчанию.

3. Код обработки неожиданных ошибок не должен смешиваться с бизнес-логикой.

4. Код обработки неожиданных ошибок должен быть максимально кратким.

5. Информация об ошибке должна быть максимально подробной.

6. Информация об ошибке должна передаваться в вызывающий контекст автоматически.

7. Освобождение ресурсов при обработке ошибки должно происходить автоматически.

Рассмотрим обработку ошибок с помощью механизма исключений и с помощью кодов возврата, в т. ч. и в контексте указанных требований. Предполагается достаточное знакомство с языками C и C++.

Обработка ошибок с помощью механизма исключений. Будем рассматривать широко распространенный вариант механизма исключений, впервые введенный в C++ (большой сравнительный обзор других вариантов можно найти в [9]). Сильными сторонами обработки ошибок с помощью механизма исключений являются:

невозможность игнорировать ошибку, не обработанную явно;

автоматическое информирование вызывающего контекста об ошибке;

возможность передать подробную информацию об ошибке;

упрощение интерфейсов подпрограмм; отсутствие необходимости изменения ин-

терфейса в случае, если для подпрограммы, ранее не предполагающей ошибочного завершения, оно стало возможным («совместимость с будущим», forward compatibility).

Среди слабых сторон можно отметить:

громоздкий синтаксис обработки ожидаемых ошибок;

последовательную семантику исключений (только одно исключение может быть активно в некоторый момент времени; оно требует немедленной обработки), что не является оптимальным для параллельных программ;

более низкую скорость работы – либо в случае успешного, либо в случае ошибочного выполнения, в зависимости от реализации механизма исключений [10];

добавление неявных потоков управления.

Из последнего пункта следует возможность возникновения большого количества «ошибочных», не предусмотренных авторами, состояний программы в результате возникновения исключений и, в частности, несовместимость механизма с обработкой ошибок без использования исключений.

Проанализируем использование механизма исключений в соответствии с приведенными критериями 1–7.

Механизм исключений не делает различий между ожидаемыми и неожиданными ошибками. Как следствие, требование 1 не выполняется – использование альтернативы не выглядит частью бизнес-логики. Требование 3, напротив, выполняется.

Требование 2 выполняется – оно было одной из причин ввода механизма исключений в язык.

Требования 4, 6 и 7 выполняются, если используется (там, где это применимо) подход RAII [5]. Однако там, где он неприменим, освобождение ресурсов необходимо производить вручную, и код становится более громоздким и подверженным ошибкам.

Требование 5 выполняется только в случае, если программист будет самостоятельно перехватывать исключение по мере его продвижения по вызывающим контекстам и дополнять его необходимой информацией. Таким образом, требования 5 и 6 конфликтуют.

Обработка ошибок с помощью кодов возврата. Сильными сторонами обработки

ошибок с помощью использования кодов возврата являются:

отсутствие в языке дополнительных механизмов, которыми должен овладеть программист;

простота использования;

предсказуемая скорость выполнения кода;

совместимость с кодом, написанным на других языках (ABI-совместимость).

Среди слабых сторон можно выделить:

отсутствие обработки (игнорирование) ошибок по умолчанию;

громоздкий код обработки ошибок;

невозможность передать подробную информацию об ошибке;

отсутствие «совместимости с будущим» – потенциальная необходимость изменения интерфейса подпрограмм в случае появления новых ошибочных ситуаций;

сложность написания безопасного относительно исключений (exception-safe) кода [11].

Относительно критериев 1–7 можно отметить следующее: выполняется лишь важное требование 1; все прочие не выполняются.

Предлагаемый подход

Необходимость наличия единого механизма обработки ошибок. Несложно увидеть, что механизмы исключений и кодов возврата можно рассматривать как дополняющие друг друга: исключения хорошо работают для неожиданных ошибок, а коды возврата – для ожидаемых; исключения хорошо подходят для последовательного кода, а коды ошибок – для параллельного. Является ли правильным совмещение этих механизмов, подобно тому, как предлагают авторы Go и Rust?

Приведем простое рассуждение, которое показывает, что подобный подход неправилен. Что должен использовать автор подпрограммы при определении ошибочной ситуации? Так как он не может знать, каким образом будет использоваться его подпрограмма (например, будет ли эта ошибка ожидаемой или нет), то он не в состоянии обоснованно выбрать подходящий механизм.

Некоторые придерживаются противоположной точки зрения: автор подпрограммы может знать, например, какая ошибка будет

ожидаемой, а какая – нет. В случае подпрограммы открытия файла для чтения, ошибка «файл не существует» может рассматриваться как ожидаемая. Однако несложно представить пример, в котором это не так: допустим, если это служебный файл, до этого созданный самой программой в ходе выполнения. Очевидно, что в этом случае ошибка будет являться неожиданной.

Таким образом, в языке должен присутствовать единый механизм обработки ошибок, который предоставлял бы вызывающему (а не вызываемому) контексту выбирать семантику обработки.

Информирование об ошибке. Код, обнаруживший ошибку, должен сообщить о ней с использованием `fail`, аналога `throw` в C++ (здесь и далее для примеров используется модельный язык с синтаксисом и семантикой, близкими к C++):

```
double sqrt(double x) {
    if (x < 0) {
        fail InvalidArgumentError(x);
    }
}
```

Семантика `fail` аналогична `throw`, за исключением того, что в данные об ошибке автоматически добавляется информация о функции и строчке, в которой она произошла. Информирование вызывающего контекста об ошибке происходит автоматически, аналогично механизму исключений.

Обработка ожидаемой ошибки. Для обработки ожидаемой ошибки используется механизм `inline try`:

```
x = try sqrt(y);
switch x.kind() {
case error:
case result:
}
```

Механизм `inline-try` преобразует выражение, возвращающее данные типа `X`, в выражение, возвращающее данные типа `X|error` – объединение, наподобие `union` и `boost::variant<>` в терминах C++. Работа с результатом `inline-try` является явной частью бизнес-логики.

Обработка неожиданных ошибок. Желательно, чтобы обработка неожиданных ошибок не затрудняла чтение кода бизнес-логики. Кроме этого, неожиданные ошиб-

ки должны обрабатываться в обязательном порядке.

Предлагаемый механизм основан на наблюдении, что в обработке неожиданных ошибок главное – не собственно «обработка» ошибок, а поддержание инвариантов программы при условии возникновения ошибки. На основе этого наблюдения предлагается два механизма – `onerror` и `finally` (см. [12]):

```
x = action();
onerror rollback();
finally cleanup();
```

Код, ассоциированный с `onerror`, выполнится только при возникновении ошибки; при наличии нескольких подобных блоков в контексте их выполнение будет происходить в обратном порядке (аналогично выполнению деструкторов локальных объектов). Код, ассоциированный с `finally`, будет выполняться всегда при выходе из текущего контекста (опять же, аналогично коду деструкторов локальных объектов).

Такой подход позволяет группировать операции восстановления инвариантов и освобождения ресурсов с операциями бизнес-логики. Его преимуществом является хорошая масштабируемость на большое количество операций, а также легкость написания и чтения кода:

```
x = action1();
onerror rollback1();
finally cleanup1();

y = action2();
onerror rollback2();
finally cleanup2();

z = action3();
onerror rollback3();
finally cleanup3();
```

Соответствующий код на Java/C# (простая трансляция):

```
try {
    x = action1();
    try {
        y = action2();
        try {
            z = action3();
        } catch (Exception e) {
```

```

        rollback3();
        throw e;
    } finally {
        cleanup3();
    }
} catch (Exception e) {
    rollback2();
    throw e;
} finally {
    cleanup2();
}
} catch (Exception e) {
    rollback1();
    throw e;
} finally {
    cleanup1();
}

```

Механизм `onerror / finally` похож на механизм деструкторов локальных объектов в C++. В связи с этим предлагается уточнить и расширить понятие деструктора: деструктор заменяется специальными функциями `onerror()`, `onsuccess()` и `onexit()`, отвечающими за поддержание инвариантов в случае ошибки, успешного выполнения, и за освобождение ресурсов, соответственно.

Дополнение информации об ошибке. Нередко вызывающие контексты обладают полезной информацией о неожиданной ошибке. Эта информация, например, может пригодиться при отображении пользователю сообщения об ошибке. В широко используемом механизме исключений, однако, для добавления подобной информации к исключению код должен явно его перехватить, изменить и повторно возбудить, что не соответствует выдвинутому требованию автоматической обработки неожиданных ошибок.

Для решения указанной проблемы предлагается добавить в язык механизм аннотирования текущего действия информацией, потенциально полезной для пользователя:

```

f = fopen(filename, "rb");
do "reading from file", filename
read_from_file(f);

```

В случае возникновения неожиданной ошибки в `read_from_file`, генерируемое исключение будет автоматически дополнено именем файла — информацией, недоступ-

ной на момент возникновения ошибки.

Подобный механизм позволит автоматически получать после возникновения ошибки данные, что значительно улучшит информативность сообщения об ошибке для пользователя (по сравнению с существующими подходами, в т. ч. отображением трассировки стека выполнения).

В данной статье проанализировано использование популярных подходов к обработке ошибок — механизма исключений и кодов возврата. Введено разделение ошибок на ожидаемые и неожиданные и предложены критерии, которым должны соответствовать языковые средства поддержки систематической обработки ошибок. Предложена точка зрения, согласно которой обработка неожиданных ошибок должна быть автоматической и заключаться в поддержании инвариантов состояния программы.

На основании проведенного анализа и выдвинутых критериев предложен ряд языковых средств, который, будучи реализованным, позволит:

- использовать единый механизм для обработки всех возможных ошибок;

- объединить код обработки ожидаемых ошибок с бизнес-логикой (частью которой он должен, по определению, являться);

- обеспечить автоматическую обработку всех неожиданных ошибок, в т. ч. освобождение ресурсов и восстановление необходимых инвариантов состояния;

- сделать компактным, легким в сопровождении и явно выделенным из бизнес-логики код поддержания инвариантов и освобождения ресурсов;

- обеспечить автоматическое уведомление вызывающих контекстов об ошибке и автоматическое дополнение сообщения об ошибке контекстно-зависимой информацией.

Естественно, предложенный подход обладает и недостатками, большинство из которых являются следствием его схожести с механизмом обработки исключений. В связи с этим в дальнейшей работе планируется исследовать вопрос использования модели компиляции, автоматически (либо по запросу программиста) заменяющей использование семантики исключений семантикой кодов возврата.

СПИСОК ЛИТЕРАТУРЫ

1. **Shah, H.** Understanding exception handling: Viewpoints of novices and experts [Электронный ресурс] / H. Shah, C. Gorg, M. Harrold // Software Engineering, IEEE Transactions on. – 2010. – Vol. 36. – № 10. – P. 1–12.
2. **Gehani, N.** Exceptional C or C with Exceptions [Text] / N. Gehani // Software: Practice and Experience. – 1992. – Vol. 22. – № 10. – P. 827–848.
3. **Goodenough, J.B.** Exception handling: issues and a proposed notation [Text] / J.B. Goodenough // Communications of the ACM. – 1975. – Vol. 18. – № 12. – P. 683–696.
4. **Cristian, F.** Exception handling and software fault tolerance [Электронный ресурс] / F. Cristian // Computers, IEEE Transactions on. – 1982. – Vol. 100. – № 6. – P. 531–540.
5. **Страуструп, Б.** Дизайн и эволюция языка C++ [Текст] / Б. Страуструп. – СПб.: Питер, 2006. – 448 с.
6. **Cabral, B.** Exception handling: A field study in Java and .NET [Text] / B. Cabral, P. Marques // ECOOP 2007. – Springer Berlin Heidelberg, 2007. – P. 151–175.
7. Go Programming Language [Электронный ресурс] / Режим доступа: <http://golang.org> (Дата обращения 25.06.2013)
8. Rust Programming Language [Электронный ресурс] / Режим доступа: <http://rust-lang.org> (Дата обращения 25.06.2013)
9. **Garcia, A.** A Comparative Study of Exception Handling Mechanisms for Building Dependable Object-Oriented Software [Text] / A. Garcia, C. Rubira, A. Romanovsky, J. Xu // J. of systems and software. – 2001. – Vol. 59. – № 2. – P. 197–222.
10. **Goldthwaite, L.** Technical report on C++ performance [Электронный ресурс] / L. Goldthwaite // ISO/IEC PDTR. – 2006. – Vol. 18015.
11. **Stroustrup, B.** Exception Safety: Concepts and Techniques [Text] / B. Stroustrup // Advances in exception handling techniques. – Springer Berlin Heidelberg, 2001. – P. 60–76.
12. **Alexandrescu, A.** Generic: Change the Way You Write Exception-Safe Code Forever [Text] / A. Alexandrescu, P. Marginean // Dr. Dobb's Journal. – CMP Media LLC., 2003.

REFERENCES

1. **Shah H., Gorg C., Harrold M.** Understanding exception handling: Viewpoints of novices and experts / Software Engineering, IEEE Transactions on. – 2010. – Vol. 36. – № 10. – P. 1–12.
2. **Gehani N.** Exceptional C or C with Exceptions / Software: Practice and Experience. – 1992. – Vol. 22. – № 10. – P. 827–848.
3. **Goodenough J.B.** Exception handling: issues and a proposed notation / Communications of the ACM. – 1975. – Vol. 18. – № 12. – P. 683–696.
4. **Cristian F.** Exception handling and software fault tolerance / Computers, IEEE Transactions on. – 1982. – Vol. 100. – № 6. – P. 531–540.
5. **Straustrup B.** Dizain i evoliutsiia iazyka C++. – St.-Petersburg: Piter, 2006. – 448 s. (rus)
6. **Cabral B., Marques P.** Exception handling: A field study in Java and .NET / ECOOP 2007. – Springer Berlin Heidelberg, 2007. – P. 151–175.
7. Go Programming Language. Available at <http://golang.org> (Accessed 25.06.2013)
8. Rust Programming Language. Available at <http://rust-lang.org> (Accessed 25.06.2013)
9. **Garcia A., Rubira C., Romanovsky A., Xu J.** A Comparative Study of Exception Handling Mechanisms for Building Dependable Object-Oriented Software / Journal of systems and software. – 2001. – Vol. 59. – № 2. – P. 197–222.
10. **Goldthwaite L.** Technical report on C++ performance / ISO/IEC PDTR. – 2006. – Vol. 18015.
11. **Stroustrup B.** Exception Safety: Concepts and Techniques / Advances in exception handling techniques. – Springer Berlin Heidelberg, 2001. – P. 60–76.
12. **Alexandrescu A., Marginean P.** Generic: Change the Way You Write Exception-Safe Code Forever / Dr. Dobb's Journal. – CMP Media LLC., 2003.

ПЕТРОСЯН Григорий Сергеевич — аспирант кафедры прикладной математики Санкт-Петербургского государственного политехнического университета.


195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: gregory.petrosyan@gmail.com

PETROSYAN, Gregory S. *St. Petersburg State Polytechnical University.*

195251, Politeknicheskaya Str. 29, St.-Petersburg, Russia.

E-mail: gregory.petrosyan@gmail.com



Математическое моделирование: методы, алгоритмы, технологии

УДК 004.421

А.Ю. Потлов, С.Г. Проскурин

АЛГОРИТМ ЦВЕТОВОГО ДОПЛЕРОВСКОГО КАРТИРОВАНИЯ НАПРАВЛЕНИЙ ПОТОКОВ БИОЛОГИЧЕСКИХ ЖИДКОСТЕЙ В ОПТИЧЕСКОЙ КОГЕРЕНТНОЙ ТОМОГРАФИИ

A.Yu. Potlov, S.G. Proskurin

COLOR DOPPLER MAPPING ALGORITHM OF BIOLOGICAL LIQUIDS' DIRECTED FLOWS USING OPTICAL COHERENCE TOMOGRAPHY

Представлен и реализован новый алгоритм цветового доплеровского картирования одной выбранной скорости разнонаправленных потоков биологических жидкостей с использованием методов оптической когерентной томографии. Алгоритм отличается отсутствием 2π -неопределенности на получаемых картограммах, достигнутым за счет разделения исходных данных на части, соответствующие положительному и отрицательному сдвигу несущей частоты, с последующим независимым построением двумерного структурного и двумерного кодированного цветом изображения одной выбранной скорости и их комплексированием. Практическая реализация алгоритма осуществлена в среде LabVIEW с использованием потокового программирования. Тестирование и отладка полученного программного продукта произведены на базе гидродинамической модели кровеносного сосуда.

ЦВЕТОВОЕ ДОПЛЕРОВСКОЕ КАРТИРОВАНИЕ. ОДНА ВЫБРАННАЯ СКОРОСТЬ. ОПТИЧЕСКАЯ КОГЕРЕНТНАЯ ТОМОГРАФИЯ. КОМПЛЕКСИРОВАНИЕ ИЗОБРАЖЕНИЙ. ПОТОК БИОЛОГИЧЕСКОЙ ЖИДКОСТИ.

A new algorithm of color Doppler mapping of one specific velocity of biological liquids with shuttle flows using optical coherence tomography is presented. The algorithm does not have the disadvantage 2π -ambiguity of color Doppler images. It is realized using the separation of the raw data to two parts corresponding to positive and negative shifts of the carrier frequency. As a final result, the complexation of independently reconstructed structural image and 2D color-coded one specific velocity image is performed. The practical implementation of the algorithm was performed in the LabVIEW package using the stream programming. A blood vessel hydrodynamic phantom was used for testing and debugging the final software product.

COLOR DOPPLER MAPPING. ONE SPECIFIC VELOCITY. OPTICAL COHERENCE TOMOGRAPHY. IMAGES' COMPLEXATION. FLOW OF BIOLOGICAL FLUID.

Объединение принципов оптической когерентной томографии (ОКТ) и доплеровской диагностики позволяет создать многофункциональные системы для визуализации морфологических особенностей и структуры гемодинамики биообъектов [1]. В основе ОКТ гетеродинное детектирование и

обработка излучения отраженного от слоев различной оптической плотности. Оцифрованные интерферограммы (А-сканы) часто для последующей обработки записываются в файл. Название «А-сканы» дано по аналогии с ультразвуковыми исследованиями (УЗИ). Глубина когерентного зондирова-

ния плотных биологических тканей ОКТ системами при использовании длин волн $\lambda = 700\text{--}1500$ нм составляет 1–3 мм. Из-за сильного рассеяния оптического излучения в плотных биотканях ОКТ системы применяются, преимущественно, для исследования роговицы, стекловидного тела и сетчатки глаза. Их клиническая значимость обусловлена тем, что разрешение таких систем на один–два порядка превышает разрешение аналогичных УЗИ систем, которое составляет $\sim 1\text{--}0,1$ мм [2].

В ОКТ, так же, как и в УЗИ, обычно применяют два различных подхода к детектированию движений внутри биообъекта:

- 1) на основе регистрации сдвига фазы сигнала соседних интерферограмм;
- 2) на основе регистрации доплеровского сдвига отраженного и рассеянного назад сигнала.

В первом подходе при помощи кросс-, либо автокорреляционной функции оценивается среднее изменение фазы комплексного сигнала для нескольких соседних А-сканов, затем доплеровский сдвиг представляется как средний сдвиг фазы, из которого в свою очередь находят направление и скорость потока. Для современных алгоритмов детектирования движений, основанных на таком подходе, характерны быстрое действие, очень высокая чувствительность, но также и 2π -неопределенность. В случаях, когда сдвиг одного А-скана относительно другого достигает 2π , пропорциональность скорости сдвигу фазы теряется.

Второй подход (в связи с необходимостью регистрировать в ОКТ потоки преимущественно в капиллярах, артериолах и венах) является труднореализуемым. Низким скоростям соответствуют доплеровские сдвиги, попадающие в полосу сильных частотных шумов. Тем не менее если программно разделить доплеровский спектр относительно несущей частоты на положительную и отрицательную части и для каждой из них с помощью полосовых фильтров выделить интересующую пользователя «узкую» полосу частот, то можно реализовать визуализацию локализаций и направлений потоков биологических жид-

костей для ограниченного диапазона скоростей, который в свою очередь можно представить в виде одной заранее выбранной скорости (ОСВ) и точности ее определения. Цель данной работы – реализация такого алгоритма обработки интерференционного сигнала в ОКТ.

В работе применяется классическая схема визуализации скоростей, в которой положительный сдвиг частоты свидетельствует о наличии движения, направленного к датчику. Такое движение кодируется красным цветом. Отрицательный сдвиг свидетельствует о направлении движения от датчика и кодируется синим. Темные оттенки этих цветов соответствуют низким скоростям, светлые оттенки – высоким. Полученная картограмма с целью повышения информативности накладывается на стандартное 2D структурное изображение, выполненное в оттенках зеленого. Представленный алгоритм разработан с учетом ранее полученных результатов в построении структурных ОКТ изображений [3, 4] и отличается высоким быстродействием и качеством получаемых изображений.

Важно отметить, что визуализация ОСВ не снижает эффективности доплеровской диагностики в ОКТ [5]. Это связано с тем, что конкретным заболеваниям, при диагностике которых используется доплеровское картирование, как правило, соответствуют заранее известные направление и диапазон скоростей кровотока. Например, при незрелой осложненной катаракте максимальная линейная скорость кровотока в центральной артерии сетчатки составляет $63,3 \pm 23,9$ мм/с, при норме $150,9 \pm 25,5$ мм/с [6]. Пользователь может указать интересующий его диапазон (63 ± 24 мм/с) и получить анатомическую доплеровскую карту потоков биологических жидкостей. При этом скорости, попадающие в выбранный диапазон, будут визуализированы в виде чередующихся красных полос на структурном изображении сетчатки глаза.

Алгоритм цветового доплеровского картирования

Для получения структурных изображений в ОКТ, как правило, используют

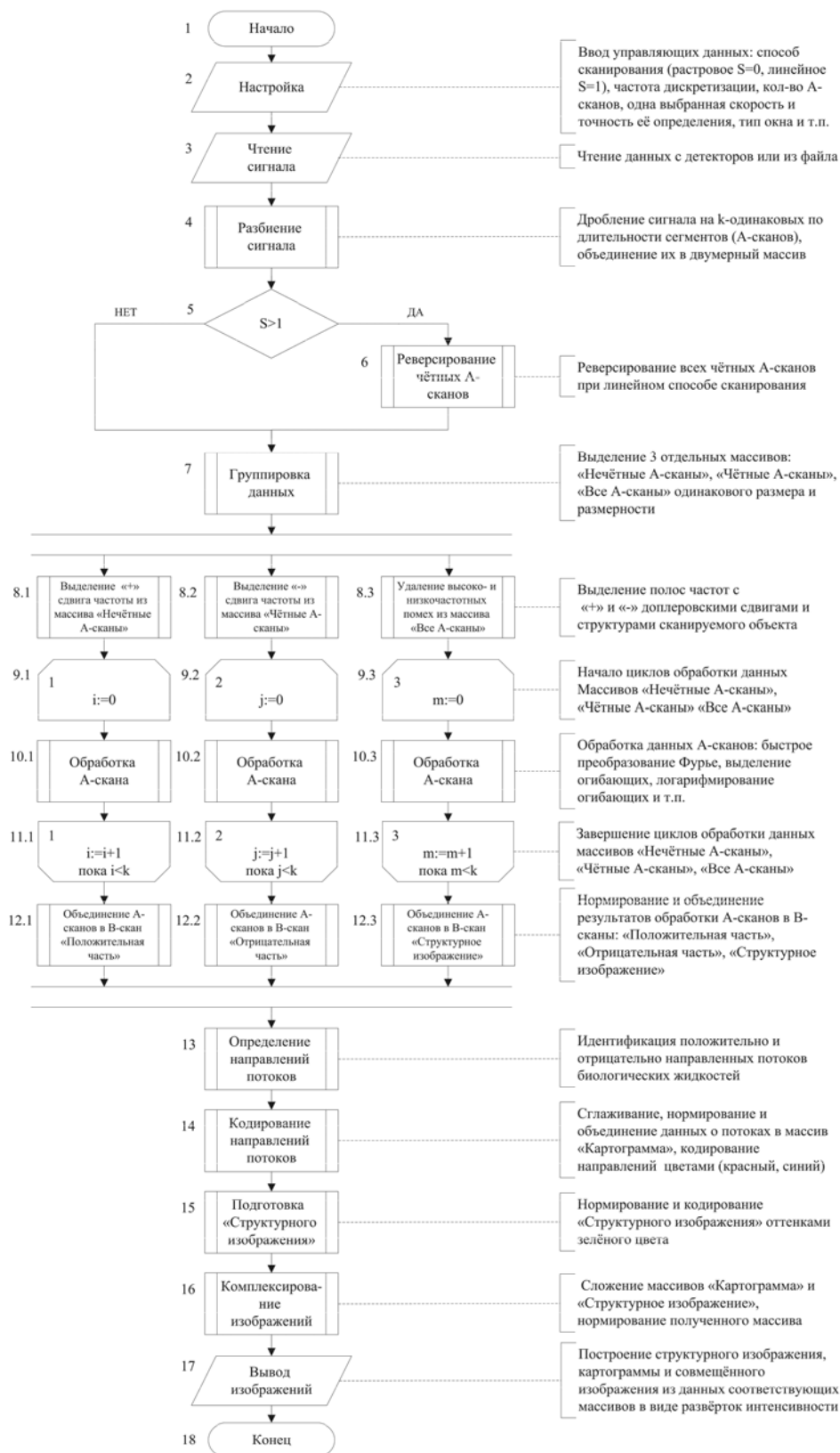


Рис. 1. Алгоритм цветного доплеровского картирования с комплексированием полученной ОСВ картограммы и структурного изображения

либо быстрый растровый (гальвано-сканер в Фурье области линзы), либо медленный линейный (с помощью ретрорефлектора) способ сканирования в плече образца интерферометра Майкельсона. Следует отметить, что оба способа приводят к сдвигу несущей частоты для всего интерференционного сигнала [1]. Причем на нечетных А-сканах сдвиг несущей частоты всегда положительный, а на четных – всегда отрицательный, но при этом одинаковый по модулю. В связи с этим нами предложено: разделение исходных данных на две части, соответствующие положительному и отрицательному сдвигу несущей частоты; независимое построение 2D структурного ОКТ изображения и 2D ОСВ изображения; последующее их комплексирование.

Подробный алгоритм получения цветовой картограммы в ОКТ приведен на рис. 1.

Исходные данные, полученные с оборадования или из файла (блок 3), в зависимости от настроек (блок 2) при работе программы (блок 1), в частности, от способа сканирования при их получении, подвергаются различной первичной обработке (блоки 4–7).

Изначально считанный сигнал представляет собой одномерный массив, поэтому в блоке 4 с учетом настроек программы (блок 2) осуществляется его дробление на отдельные А-сканы и формирование из них двумерного массива «Все А-сканы». Если (блок 5) сканирование осуществлялось с помощью ретрорефлектора, то четные А-сканы массива реверсируются (блок 6). Если (блок 5) использован растровый способ сканирования, то все А-сканы массива остаются без изменений.

Далее из массива «Все А-сканы» выделяются еще два массива (блок 7) «Нечетные А-сканы» и «Четные А-сканы». Чтобы не исказить размеры изображения, А-сканы в этих массивах чередуются с пустыми столбцами. В контексте потокового программирования эти массивы представляют собой параллельные данные – независимые секции одного набора данных, которые могут обрабатываться независимо друг от друга с последующим объединением результатов. В связи с этим далее в алгоритме органи-

зованные параллельные действия (блоки 8.1–12.1, 8.2–12.2, 8.3–12.3).

Для точной регистрации ОСВ внутри биообъекта цифровым способом выделяются: полоса частот из положительной части доплеровского спектра, представленной в массиве «Нечетные А-сканы» (блок 8.1); симметричная ей полоса частот из отрицательной части доплеровского спектра, т. е. из массива «Четные А-сканы» (блок 8.2); полезный сигнал из данных массива «Все А-сканы» (блок 8.3). Выделение сдвигов частоты (блоки 8.1, 8.2) производится с помощью полосовых фильтров Бесселя пятого порядка. Нижняя и верхняя частоты среза фильтров находятся из заданных пользователем скорости и точности ее определения с помощью эталонного доплеровского сдвига. Под ним понимается сдвиг частоты, соответствующий скорости 1 мм/с и вычисляемый по формуле доплеровского сдвига частоты. В проведенных экспериментах эталонный доплеровский сдвиг составлял $\approx 0,6$ кГц.

Групповая задержка фильтров Бесселя практически не изменяется по частотам полосы пропускания, вследствие чего форма фильтруемого сигнала на выходе такого фильтра в полосе пропускания сохраняется практически неизменной. Удаление высоко- и низкочастотных помех (блок 8.3) из сигнала массива «Все А-сканы» производится с помощью полосового фильтра Баттерворта третьего порядка. Заметим, что алгоритм ориентирован на обработку исходных экспериментальных данных, при получении которых уже проводилась аппаратная фильтрация низкочастотной компоненты спектров.

Затем отфильтрованный сигнал из массивов «Нечетные А-сканы», «Четные А-сканы», «Все А-сканы» подвергается параллельной циклической обработке по А-сканам (блоки 9.1–11.1, 9.2–11.2, 9.3–11.3). Блоки 10.1–10.3 отвечают за обработку отдельных А-сканов и являются одинаковыми (рис. 2). Каждый А-скан массивов «Нечетные А-сканы», «Четные А-сканы», «Все А-сканы» независимо подвергается дискретному оконному преобразованию Фурье (подблок А). На каждом сегменте

спектр представляет собой комплекснозначную функцию номера отсчета (момента времени). При этом амплитудный спектр — это модуль комплексного спектра, фазовый спектр — его аргумент. Для получения спектрограммы объединяют амплитудные спектры, вычисленные на коротких сегментах, в функцию двух переменных или матрицу. Полученная двумерная матрица с учетом размера окна преобразования и перекрытия нормируется под реальные размеры (подблок В). Затем с помощью преобразования Гильберта выделяются огибающие нормированного демодулированного сигнала (подблок С), и для устранения симметричности и сжатия динамического диапазона величин они логарифмируются (подблок D). В результате сигнал каждого А-скана вновь становится одномерным. В блоках 12.1–12.3 параллельно выполняется объединение результатов обработки интерференционного сигнала А-сканов в три двумерных массива: «Положительная часть», «Отрицательная часть», «Структурное изображение». На этом параллельные действия в алгоритме завершаются. Полученные массивы, по сути, представляют собой В-сканы, т. е. структурные изображения объекта, полученные из выделенной фильтром Бесселя полосы частот положительной части доплеровского спектра; симметричной ей полосы частот отрицательной части доплеровского спектра; очищенного от помех всего спектра сигнала.

Следующий шаг алгоритма — идентификация потоков биологических жидкостей (блок 13) — организован на базе поэлементного вычитания данных массивов «Положительная часть» и «Отрицательная часть» и последующего логического анализа полученных результатов. За неподвижную часть исследуемого объекта принимается то, что примерно ($\pm 5\%$ по абсолютной интенсивности сигнала) совпадает в этих двух массивах, а положительный и отрицательный потоки — то, что есть лишь в одном из них. Следует отметить, что в большинстве случаев неподвижная часть не будет представлена в этих двух массивах, но для полноты алгоритма мы учли вариант, при котором скорость сканирования в плече образца попала в заданный пользователем диапазон детектируемых скоростей потоков. По результатам работы блока 13 формируются массивы «Положительные потоки» и «Отрицательные потоки». Далее в блоке 14 выполняется сглаживание, нормирование, кодирование и объединение данных этих массивов. Для картограммы используется абсолютная цветовая шкала: положительный поток обозначается интенсивностью движения, равной единице, и кодируется красным цветом; отрицательный поток — интенсивностью движения, равной минус единице, и синим цветом; отсутствие движения — интенсивностью движения, равной нулю, и белым цветом.

В блоке 15 выполняется нормирование



Рис. 2. Универсальный блок «Обработка А-скана»

и кодирование данных массива «Структурное изображение». Сигнал массива нормируется таким образом, чтобы минимальное отличие от ОСВ сигнала составляло 20 раз. Для структурного изображения используется шкала оттенков зеленого: максимальная интенсивность светового поля для каждого отдельного сканирования кодируется зеленым цветом; нулевая интенсивность (нет отражения) – белым цветом; интенсивности в диапазоне от нуля до максимума – промежуточными цветами, вычисленными с помощью интерполяции.

Затем в блоке 16 осуществляется комплексирование структурного изображения и картограммы. Для этого одноименные массивы поэлементно складываются и нормируются под единую цветовую шкалу, включающую в себя все оттенки зеленого для визуализации внутренней структуры биологической ткани, красный и синий цвета для визуализации направления движения потоков жидкостей, а также белый цвет как начало отсчета (нет движения, нет отражения). В результате работы блока формируется массив «Совмещенное изображение».

Для получения структурного и доплеровского изображения по данным одноименных массивов строятся развертки интенсивности (блок 17), работающие с описанными выше шкалами. На этом алгоритм завершается (блок 18).

Результаты и обсуждение

Предлагаемый алгоритм удалось реализовать в виде специализированного про-

граммного продукта на графическом языке программирования G среды разработки и платформы для выполнения программ LabVIEW. Получено свидетельство о государственной регистрации этой программы для ЭВМ [7].

Для тестирования и отладки программного продукта использовалась простейшая модель кровеносного сосуда [5, 8]. Она представляет собой прозрачную трубку с внутренним диаметром 0,55 мм. Вместо крови использовался 1 % раствор интралипида в воде, по рассеивающим оптическим свойствам похожий на кровь. На рис. 3 приведены: структурное изображение, полученное при сканировании гидродинамической модели (а); картограмма, локализации и направления потока биологической жидкости (б); итоговое изображение (в). Использованное значение ОСВ 10 мм/с, точность ± 5 мм/с.

После отладки программного продукта проанализировали быстродействие предлагаемого алгоритма. Цветовое доплеровское ОСВ картирование для файла с оцифрованным интерференционным сигналом, состоящим из 60 А-сканов при числе дискретных значений в А-скане, равном 69 000, занимало $0,5 \pm 0,2$ с в зависимости от настроек программного продукта. Файл исходных данных имел текстовый формат и тип записанных данных П16 (целочисленные знаковые данные). При получении данных использовалось линейное сканирование. Оценка быстродействия производилась на ноутбуке с процессором Intel Core i3-2310M и оперативной памятью SO-DIMM DDR3

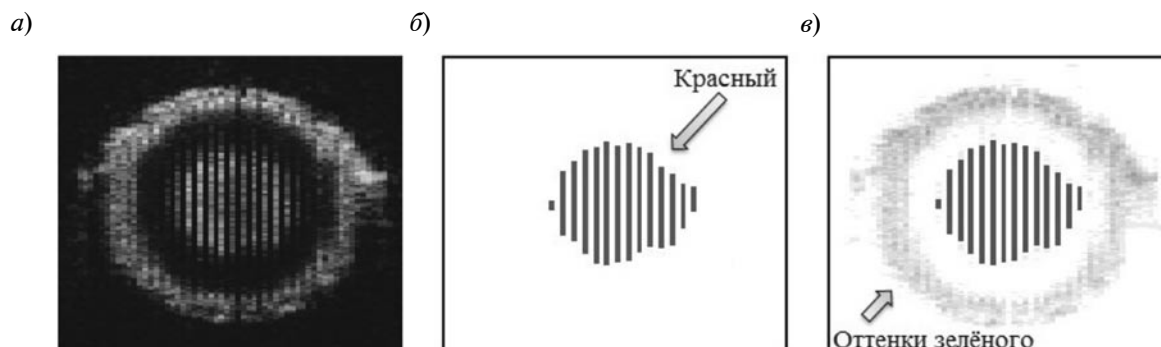


Рис. 3. ОКТ изображения прозрачной трубки с движущимся в ней раствором интралипида в воде: а – исходное структурное изображение; б – цветовая ОСВ картограмма; в – комплексированное изображение

объемом 3 Гб. Высокое быстродействие предлагаемого алгоритма обусловлено наличием в нем параллельных действий и обработкой не всех исходных данных (только частот, соответствующих выбранной скорости). Для сравнения, в тех же условиях работы алгоритму цветового картирования скорости по среднему сдвигу фазы, анализируемому автокорреляционной функцией, потребуется времени на 30 % больше.

Разработан алгоритм цветового доплеровского картирования заранее выбранных скоростей разнонаправленных потоков жидкостей, основанный на методах ОКТ и отличающийся высоким быстродействием и качеством получаемых картограмм (отсутствует 2π -неопределенность). Предложенное

цветовое ОСВ картирование позволяет получать анатомические доплеровские карты потоков биологических жидкостей, отражающие функциональное состояние биообъектов в норме и патологии. Алгоритм может применяться не только в ОКТ, но и в УЗИ.

Новизна предлагаемого алгоритма заключается в картировании выбранной пользователем скорости с заранее заданной точностью, на основе разделения спектра интерференционного сигнала на положительную и отрицательную части, выделения с помощью полосовых фильтров необходимых полос частот, построения для них ОСВ изображения, комплексирования этого изображения со структурным.

Работа выполнена в рамках контрактов № 9576 p/14219 и 10741p/16955 по программе У.М.Н.И.К. ФСРМФП в НТС.

СПИСОК ЛИТЕРАТУРЫ

1. **Зимняков, Д.А.** Оптическая томография тканей [Текст] / Д.А. Зимняков, В.В. Тучин // Квантовая электроника. —2002. —№ 10. —С. 849–867.
2. **Гундорова, Р.А.** Характеристика изменений органа зрения при ЧС [Текст] / Р.А. Гундорова, О.И. Кваша, Р.А. Нурмамедов // Сб. науч. трудов с междунар. участием: Российский общенациональный форум.—М.: Изд-во «Апрель», 2008. —С. 202–205.
3. **Проскурин, С.Г.** Построение структурного изображения биообъекта с использованием растрового усреднения в оптической когерентной томографии [Текст] / С.Г. Проскурин, А.Ю. Потлов, К.И.С. Галев, С.Н. Абдулкарим // Изв. Южного федерального университета. Технические науки. —2012. —№ 9. —С. 129–134.
4. **Фролов, С.В.** Построение структурных изображений в оптической когерентной томографии [Текст] / С.В. Фролов, А.Ю. Потлов, С.Г. Проскурин // Св-во о гос. рег. программы для ЭВМ в ФИПС № 2012615095. Зарег. 07.06.2012.
5. **Proskurin, S.G.** One specific velocity visualization in flows with complex geometry [Text]

/ S.G. Proskurin, R.K. Wang; V.V. Tuchin Ed. // Complex Dynamics and Fluctuations in Biomedical Photonics II. —Bellingham (USA), 2005. —Vol. 5696. —P. 129–135.

6. **Гаджиева-Азизова, С.А.** Изменение гемодинамических параметров кровотока в цилиарном теле при осложненной катаракте на фоне оперированной открытоугольной глаукомы [Текст] / С.А. Гаджиева-Азизова, А.Н. Мамедзаде, А.Т. Набиева, У.А. Аббасова // Oftalmoloqiya. —2011. —№ 1 (5). —С. 43–50.

7. **Фролов, С.В.** Цветовое картирование одной выбранной скорости в оптической когерентной томографии [Текст] / С.В. Фролов, А.Ю. Потлов, С.Г. Проскурин // Св-во о гос. рег. программы для ЭВМ в ФИПС № 2013614222. Зарег. 25.04.2013.

8. **Потлов, А.Ю.** Структурная и доплеровская визуализация поверхностных слоев биологических тканей в оптической когерентной томографии [Текст] // Новые информационные технологии. Тезисы докл. XXI Междунар. студенческой школы-семинара. —М.: Изд-во МИЭМ НИУ ВШЭ, 2013. —С. 236–237.

REFERENCES

1. **Zimniakov D.A., Tuchin V.V.** Opticheskaia tomografiia tkanei / Kvantovaia elektronika. —2002. —№ 10. —S. 849–867. (rus)
2. **Gundorova R.A., Kvascha O.I., Nurmamedov R.A.** Kharakteristika izmenenii organa zreniia pri ChS/ Sb. nach. tr. s mezhdunar. uchastiem Ros. obshchenatsional'nyi forum. —Moscow: Izd-vo «Aprel'», 2008. —S. 202–205. (rus)

3. **Proskurin S.G., Potlov A.Yu., Galeb K.I.S., Abdulkarim S.N.** Postroenie strukturnogo izobrazheniia bioob"ekta s ispol'zovaniem rastrovogo usredneniia v opticheskoi kogerentnoi tomografii / Izvestiia Iuzhnogo federal'nogo un-ta. Tekhnicheskie nauki. —2012. —№ 9. —S. 129–134. (rus)
4. **Frolov S.V., Potlov A.Yu., Proskurin S.G.** Postroenie strukturnykh izobrazhenii v opticheskoi

kogerentnoi tomografii; Sv-vo o gos. reg. programmy dlia EVM v FIPS № 2012615095. Zareg. 07.06.2012. (rus)

5. **Proskurin S.G., Wang R.K.** One specific velocity visualization in flows with complex geometry/ Complex Dynamics and Fluctuations in Biomedical Photonics II, Valery V. Tuchin Editor. –Bellingham (USA), 2005. –Vol. 5696. –P. 129–135.

6. **Gadzhieva-Azizova S.A., Mamedzade A.N., Nabieva A.T., Abbasova U.A.** Izmenenie gemodinamicheskikh parametrov krovotoka v tsiliarnom tele pri oslozhnennoi katarakte na fone operirovannoi otkrytougol'noi glaukomy / Oftalmologiya. –2011.

–№ 1 (5). – S. 43–50. (rus)

7. **Frolov S.V., Potlov A.Yu., Proskurin S.G.** Tsvetovoe kartirovanie odnoi vybrannoi skorosti v opticheskoi kogerentnoi tomografii / Sv-vo o gos. reg. programmy dlia EVM v FIPS № 2013614222. Zareg. 25.04.2013. (rus)

8. **Potlov A.Yu.** Strukturnaia i doplerovskaia vizualizatsiia poverkhnostnykh sloev biologicheskikh tkanei v opticheskoi kogerentnoi tomografii /Novye informatsionnye tekhnologii. Tezisy dokl. XXI Mezhdunar. studencheskoi shkoly-seminara.–Moscow: Izd-vo MIEM NIU VShE. –2013. –S. 236–237. (rus)

ПОТЛОВ Антон Юрьевич – аспирант кафедры биомедицинской техники Тамбовского государственного технического университета.

393251, Россия, Тамбовская обл., г. Рассказово, ул. Клубная, д. 22.

E-mail: zerner@yandex.ru

POTLOV, Anton Yu. Tambov State Technical University.

393251, Klubnaya Str. 22, Rasskazovo, Tambov region, Russia.

E-mail: zerner@yandex.ru

ПРОСКУРИН Сергей Геннадьевич – докторант кафедры биомедицинской техники Тамбовского государственного технического университета, доцент, кандидат физико-математических наук.

392027, Россия, г. Тамбов, ул. Рылеева, д. 79.

E-mail: spros@tamb.ru

PROSKURIN, Sergey G. Tambov State Technical University.

392027, Ryleeva Str. 79, Tambov, Russia.

E-mail: spros@tamb.ru



УДК 681.3.06

И.Г. Черноруцкий

ПРАКТИЧЕСКАЯ ОПТИМИЗАЦИЯ И НЕВЫПУКЛЫЕ ЗАДАЧИ

I.G. Chernorutskiy

PRACTICAL OPTIMIZATION AND NONCONVEX PROBLEMS

Рассмотрены основные особенности задач оптимизации при проведении реальных компьютерных вычислений. Обоснована необходимость учета таких характеристик прикладных экстремальных задач, как невыпуклость и жесткость целевых функционалов. Показано, что в указанных условиях классические оптимизационные методы ньютоновского и квазиньютоновского типов испытывают существенное влияние погрешностей и часто приводят к расходящимся процедурам.

МЕТОДЫ ОПТИМИЗАЦИИ. НЕВЫПУКЛЫЕ ЗАДАЧИ. НЬЮТОНОВСКИЕ МЕТОДЫ.

Main aspects of optimization problem in real computing are described. Requirement to take into account such parameters of applied extremal problem as rigidity and nonconvexity of criterion functional is founded. It is established that classic Newtonian and quasi-Newtonian optimization procedures are strongly influenced by inaccuracies in this case and nonconvergent procedures may appear.

OPTIMIZATION PROCEDURE. NONCONVEX OPTIMIZATION PROBLEM. NEWTONIAN PROCEDURE.

Рассматриваются практические аспекты применения методов однокритериальной конечномерной оптимизации (математического программирования) в задачах реального компьютерного моделирования при решении прикладных оптимизационных задач [1, 2]. Для определенности будем говорить о задачах поиска минимума. Основное внимание уделяется таким характерным чертам решаемых задач, как невыпуклость и плохая обусловленность (жесткость или овражность) минимизируемых функционалов. В теоретических работах по математическому программированию большой объем занимают исследования в области выпуклых задач оптимизации. Именно для таких задач могут быть получены наиболее законченные математические результаты, в частности, по доказательству сходимости алгоритмов минимизации (максимизации), установлению скорости сходимости и т. д. Однако, к сожалению, на практике выпуклые задачи встречаются редко. Утверждения, что в достаточно малой окрестности оптимума минимизируемая функция

(функционал) почти всегда становится выпуклой, лишены практического смысла, т. к. попадание в эту окрестность обычно и означает решение задачи. Следующий пример показывает, что выпуклости нет уже в простейших случаях.

Пример. Задача параметрической идентификации. Пусть некоторая реальная система описывается скалярным дифференциальным уравнением вида (задача Коши) $\frac{dx}{dt} = ax$, $x(0) = x_0$ с решением $x(t) = x_0 e^{at}$.

Начальное условие x_0 считаем известным.

Требуется по результатам измерения функции $x(t)$ получить оценку неизвестного параметра a . Пусть начальное значение $x_0 = 1$, а эталонное (искомое) значение равно $a = a^* = 1$. Тогда эталонная (измеряемая) функция имеет вид $x(t) = e^t$.

Найдем оценку неизвестного параметра a из сравнения эталонного решения с расчетным в соответствии с методом наименьших квадратов (задача параметрической идентификации).

Мы наблюдаем функцию e^t . Расчетное решение при произвольном параметре a и заданном $x_0 = 1$ имеет вид $x(t) = e^{at}$.

Найдем a^* из условия минимума:

$$J(a) = (e^{at} - e^t)^2 \rightarrow \min_a.$$

Сравнение расчетной и эталонной (измеренной) зависимости будем производить в одной точке для $t = 1$:

$$J_1(a) = (e^a - e)^2.$$

График этой зависимости представлен на рисунке.

Видно, что целевой функционал оказывается невыпуклым в значительной области изменения аргумента слева от оптимума. В более сложных случаях ситуация только усугубляется. В частности, легко проверить, что характер приведенной в примере зависимости сохраняется и при увеличении числа точек сравнения.

Таким образом, при решении практических задач оптимизации необходимо ориентироваться на методы, сохраняющие эффективность в невыпуклых ситуациях.

Ниже дан краткий обзор и характеристика некоторых популярных методов конечномерной оптимизации. Основное внимание уделяется анализу работоспособности алгоритмов при использовании невыпуклых целевых функционалов. Набор рассматриваемых методов по необходимости ограничен и включает в основном ньюто-

новские и квазиньютоновские процедуры, считающиеся в настоящее время наиболее эффективными при решении задач с гладкими функционалами. Одной из характерных особенностей решаемых на практике однокритериальных оптимизационных задач является также высокая степень жесткости целевого функционала [3]. Поэтому необходим анализ методов с позиций одновременного присутствия обеих характеристик – невыпуклости и жесткости целевого функционала.

Ньютоновские методы. Ньютоновские методы строят последовательность точек $\{x^k\}$ в соответствии с алгоритмом [4]:

$$x^{k+1} = x^k + h_k p_N^k, \quad h_k \in R^1, \quad (1)$$

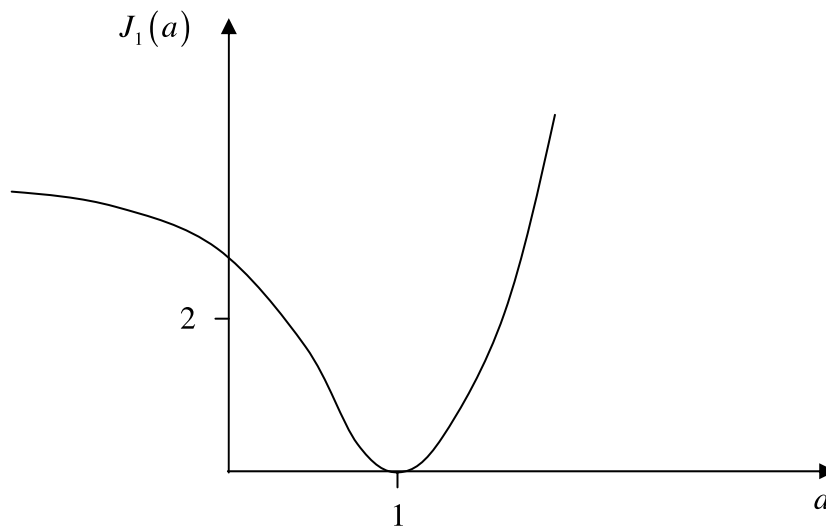
где $p_N^k = -[G_k]^{-1} g^k$ определяет т. н. *ньютоновское* направление спуска; h_k – длина шага;

$$G_k \triangleq J''(x^k); \quad g \triangleq J'(x^k).$$

Предполагается, что все матрицы G_k положительно определены ($G_k > 0$). Последнее условие гарантирует разрешимость задачи вычисления p_N^k исходя из системы уравнений

$$G_k p_N^k = -g^k. \quad (2)$$

Вектор x^{k+1} , построенный согласно (1), (2) при $h_k = 1$, является минимизатором аппроксимирующего квадратичного функционала, полученного как отрезок соответ-



Невыпуклая целевая функция

ствующего ряда Тейлора:

$$F_k(x) = J(x^k) + \langle g^k, x - x^k \rangle + \frac{1}{2} \langle G_k(x - x^k), x - x^k \rangle. \quad (3)$$

Ньютоновские методы (1) оказываются эффективными при решении задач безусловной минимизации выпуклых функционалов при невысоких степенях жесткости целевых функционалов. Кроме того, предполагается существование достаточно точных локальных квадратичных аппроксимаций основного функционала. Некоторые авторы утверждают, что это наиболее эффективные из всех применяемых ими для решения реальных задач методов при условии выпуклости целевого функционала. Однако это утверждение является слишком сильным. Последующий анализ и опыт реальных вычислений подтверждают это замечание.

Если матрица G_k не является положительно определенной, параболоид (3) может не иметь конечного минимизатора и, как правило, будет неограничен снизу. Поэтому для невыпуклых целевых функционалов процедура (2) непосредственно неприменима. Ньютоновское направление p_N^k при наличии отрицательных спектральных составляющих матрицы G_k может указывать в сторону возрастания функционала $J(x)$, что соответствует расходимости процесса.

Как отмечено в [5], «в настоящее время нет общепринятого определения «метода Ньютона» для расчета направления спуска при знаконеопределенной матрице G_k , поскольку среди специалистов нет согласия относительно того, как использовать локальную квадратичную аппроксимацию F_k в этом случае». Такая ситуация сохраняется и поныне.

Наиболее употребительны модификации метода Ньютона, в которых направление спуска p^k находится из решения линейной системы:

$$\bar{G}_k p^k = -g^k, \quad (4)$$

где \bar{G}_k — некоторая положительно определенная матрица, совпадающая с исходной матрицей Гессе G_k , если последняя положительно определена. Указанный метод

выбора направления p^k гарантирует глобальную сходимость процедуры независимо от характера выпуклости функционала $J(x)$ в окрестности текущей точки x^k .

Ниже рассмотрены наиболее известные процедуры построения матрицы \bar{G}_k на основе различных матричных разложений, позволяющих учитывать знаки собственных чисел G_k .

Методы, основанные на спектральном разложении. Спектральное разложение матрицы G_k имеет вид

$$G_k = U \Lambda U^T = \{g_{ij}\}, \quad (5)$$

$$g_{ij} = \sum_{m=1}^n \lambda_m u_i^m u_j^m,$$

где $U = [u^1, u^2, \dots, u^n]$; $\Lambda = \text{diag}(\lambda_i)$; $\{u^m, m \in [1:n]\}$ — система ортонормальных собственных векторов матрицы G_k ; $\{\lambda_m, m \in [1:n]\}$ — спектр матрицы G_k . Базовая схема, реализующая рассматриваемый подход, основана на выборе [5]:

$$\bar{G}_k = U \bar{\Lambda} U^T, \quad \bar{\Lambda} = \text{diag}[\max\{|\lambda_i|, \delta\}], \quad (6)$$

где $\delta > 0$ — параметр метода, определяющий границу «существенной положительности» любого из собственных чисел. Отмечаемый в литературе недостаток метода (4) с матрицей (6) связывается с трудоемкостью процедуры построения спектрального разложения (5), требующей до $4n^3$ арифметических операций. Кроме того, возникают известные из вычислительной линейной алгебры трудности в определении малых собственных чисел плохо обусловленной матрицы. Выбор параметра δ также до конца не алгоритмизирован.

Методы, основанные на факторизации Холецкого [5]. Разложение Холецкого для симметричной положительно определенной матрицы B имеет вид

$$B = LDL^T, \quad D = \text{diag}(d_i), \quad (7)$$

где L — нижняя треугольная матрица с единицами на диагонали; D — положительная диагональная матрица. Факторизация (7) непосредственно неприменима к знаконеопределенной симметричной матрице $B = G_k$. В [5] предложена модифицированная процедура, позволяющая равномерно ограничить рост элементов треугольного

фактора L на уровне $|r_{ik}| \leq \beta (i > k, r_{ik} \triangleq l_{ik} \sqrt{d_k})$ и гарантирующая «существенную» положительность $d_i > \delta > 0$ диагональных элементов матрицы D . В результате получаем

$$\bar{G}_k = LDL^T = G_k + C, \bar{G}_k > 0, \quad (8)$$

где C — неотрицательная диагональная матрица, зависящая от выбранного параметра β . Для сохранения численной устойчивости процедуры построения \bar{G}_k , а также для совпадения \bar{G}_k и G_k в случае положительно определенной G_k , целесообразно вычислять β из условия $\beta^2 = \max\{\gamma, \xi/\sqrt{n^2-1}, \varepsilon_M\}$, где ε_M — машинное эpsilon; ξ — максимальный модуль недиагонального элемента G_k ; γ — значение максимального из диагональных элементов G_k .

Основная цель построения разложения (8) заключается в сокращении вычислительных затрат по сравнению с (6) приблизительно до $\frac{1}{6}n^3$ арифметических операций. Существо дела при этом не затрагивается.

Аналогичные модификации метода Ньютона, основанные на других численно устойчивых процедурах факторизации знакоопределенных симметричных матриц, рассматриваются во многих работах. Большое число публикаций, посвященных рассматриваемым вопросам, с одной стороны, указывает на актуальность проблемы, а с другой — на отсутствие метода, полностью отвечающего предъявляемым практикой требованиям.

Методы доверительной окрестности. Методы разработаны для выпуклых задач оптимизации. Основная идея методов доверительной окрестности сводится к следующему [5].

Рассматриваются *квадратичные* методы, т. е. методы, основанные на последовательной квадратичной аппроксимации минимизируемого выпуклого функционала. Надежность прогноза в таких методах определяется областью справедливости (достаточной точности) локальной квадратичной модели. В методах Ньютона мы определяем направление в сторону минимума аппроксимирующей квадратичной зависи-

мости, а затем регулируем норму вектора продвижения, чтобы не уйти «слишком далеко». Здесь же предлагается поступать по иному: решается задача условной минимизации аппроксимирующей квадратичной функции при условии, что норма вектора продвижения ограничена сверху некоторым заданным параметром.

Таким образом, минимизирующая последовательность строится по правилу $x^{k+1} = x^k + p^k$, где вектор p^k на каждом шаге определяется как решение вспомогательной задачи вида

$$\begin{aligned} \frac{1}{2} \langle G_k p, p \rangle + \langle g^k, p \rangle &\rightarrow \min, \\ p &\in D \subset R^n, \\ D &= \{p \in R^n \mid \langle p, p \rangle^2 \leq \Delta\}. \end{aligned} \quad (9)$$

Величина Δ характеризует область (окрестность) «квадратичности» исходного функционала. Сформулированная задача может быть решена методом множителей Лагранжа.

Условие стационарности функции Лагранжа для задачи (9) приводит к методу определения p из системы линейных уравнений

$$[G_k + \beta E]p = -g^k, \beta > 0, \quad (10)$$

где β играет роль множителя Лагранжа. Алгоритм выбора β зависит от конкретной реализации МДО. Возможен непосредственный подбор оптимального β на основе многократного решения линейной системы (10). В ряде случаев вначале полагают $\beta = 0$. Если в процессе решения (10) при $\beta = 0$ выясняется, что G_k знакоопределена или $\|p\| > \Delta$, где Δ — установленное пороговое значение, то β определяется как решение нелинейного уравнения вида

$$\|[G_k + \beta E]^{-1}g^k\| = \Delta. \quad (11)$$

Далее по найденному β согласно (10) определяется искомым вектор p .

Основной недостаток всех методов рассматриваемого класса состоит в необходимости решения на каждом шаге итерационного процесса нелинейных алгебраических уравнений типа (11), что требует привлечения различных процедур регуляризации.

Кроме того, найденному оптимальному значению может отвечать очень плохо обусловленная матрица $G_k + \beta E$, что приводит к значительным вычислительным трудностям при решении системы (10). Аналогичные недостатки присущи и ньютоновским методам.

Указанные проблемы становятся практически неразрешимыми, если степень обусловленности матрицы G_k настолько высока, что информация о малых спектральных составляющих оказывается полностью утерянной на фоне больших собственных чисел за счет погрешностей задания G_k . Эти погрешности могут вызываться приближенным характером соотношений, применяемых для расчета производных, а также ограниченностью разрядной сетки компьютера. В указанных условиях, даже при использовании регуляризованных форм матричных разложений, гарантирующих, в частности, положительную определенность матрицы \bar{G}_k , генерируемые векторы p^k , оставаясь направлениями спуска, оказываются практически случайными, что резко замедляет сходимость соответствующей поисковой процедуры. Сказанное иллюстрируется следующим примером.

Численный пример. Вырождение метода Ньютона. Рассмотрим сильно выпуклый квадратичный функционал

$$f(x) = 1/2(Ax, x) - (b, x) + c, A > 0.$$

Тогда точка минимума определяется решением системы линейных алгебраических уравнений (необходимое условие экстремума $f'(x) = 0$):

$$Ax = b. \quad (12)$$

Спектральное разложение симметричной матрицы A имеет вид $A = TDT^T$, $T = (u^1, u^2, \dots, u^n)$, где столбцы u^i образуют ортонормальную систему собственных векторов матрицы A , а D — диагональная матрица собственных чисел λ_i матрицы A .

Получаем следующее выражение для минимизатора функционала $f(x)$:

$$\begin{aligned} x^* &= A^{-1}b = TD^{-1}T^T b = \\ &= T \left[\frac{1}{\lambda_1}(b, u^1), \dots, \frac{1}{\lambda_n}(b, u^n) \right]^T. \end{aligned} \quad (13)$$

Пусть в некотором модельном вычислительном устройстве возможно четырехзначное представление мантиссы числа в формате с плавающей точкой. И пусть точная матрица A имеет вид

$$A = 10^5 \begin{bmatrix} 0.5001 & 0.5 \\ 0.5 & 0.50001 \end{bmatrix},$$

$$\lambda_1 = 1, \lambda_2 = 10^5 + 1.$$

Видно, что значения элементов матрицы определяются «большим» собственным числом λ_2 .

После записи матрицы в память вычислительного устройства и округления значений элементов до четырех знаков мантиссы получим вырожденную матрицу:

$$A = 10^5 \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}, \lambda_1 = 0, \lambda_2 = 10^5.$$

В этих условиях классический метод Ньютона вообще неприменим, а его регуляризованные формы будут давать по существу случайные значения для «новых» малых собственных чисел и не будут содержать полезной информации об искомом векторе x^* . Плохая обусловленность системы (12) для жестких функционалов приводит к полной потере полезной информации.

Из рассмотренного примера наглядно видно, что для жестких даже сильно выпуклых функционалов резко возрастает влияние погрешностей, в т. ч. и погрешностей округления. В частности, изначально выпуклая задача может стать невыпуклой (матрица получает отрицательные собственные числа). Влияние высокой степени жесткости на точность получаемого решения определяется следующим.

Из аналитического представления (13) решения x^* следует, что значения собственных чисел входят в знаменатели соответствующих дробей и основное влияние на результат оказывают «малые» по модулю собственные числа. В то же время из спектрального разложения видно, что сами элементы матрицы представляют из себя некоторые линейные комбинации ее спектральных составляющих (собственных чисел) и численные значения элементов определяются в основном «большими» по

модулю собственными числами. В итоге небольшие относительные погрешности в представлении элементов матрицы Гессе приводят к большим относительным погрешностям для «малых» спектральных составляющих, а те, в свою очередь, полностью искажают окончательное решение. В этом, если угодно, и заключается «физический смысл» рассмотренного явления плохой обусловленности.

По этим же причинам любые «модификации» матрицы Гессе, в частности, выполняемые с целью аппроксимации невыпуклых задач выпуклыми, также приводят к потере полезной информации о «малых» собственных числах и не позволяют получить достаточно эффективные численные процедуры. Аналогично можно показать, что в невыпуклых квадратичных аппроксимациях жестких функционалов малые спектральные составляющие также несут основную полезную информацию о траектории спуска и основная цель состоит в построении алгоритмов, использующих в полной мере эту информацию.

Квазиньютоновские методы [5]. Квазиньютоновские методы, так же как и ньютоновские, основаны на использовании квадратичных моделей минимизируемых функционалов. Однако аппроксимация матрицы Гессе (либо обратной к ней) осуществляется последовательно, на основе наблюдений за изменением градиентов целевых функционалов в последовательных точках минимизирующей последовательности.

Часто указывается, что квазиньютоновские методы существенно более эффективны, чем ньютоновские, по причине более низкого порядка производных, определяющих схему метода. Однако реальная ситуация значительно сложнее. Пусть, например, минимизируется сильно выпуклый квадратичный функционал $J(x)$. Тогда для построения оптимизатора x^* квазиньютоновскому методу потребуется в общем случае вычислить n градиентов, где n – размерность пространства поиска. Последнее эквивалентно $2n^2$ вычислениям значений функционала $J(x)$ если используются двусторонние конечно-разностные аппроксимации первых производных. Указанного количества значений

функционала $J(x)$, очевидно, достаточно для построения аппроксимации $J''(x)$, необходимой для реализации любого из вариантов метода Ньютона. Следовательно, трудоемкости рассматриваемых процедур в указанных условиях приблизительно равны, если не учитывать дополнительные вычислительные затраты на процедуры одномерного поиска в квазиньютоновских методах. С другой стороны, доказано, что большинство вариантов квазиньютоновских методов (например, одна из наиболее эффективных схем Бройдена–Флетчера–Гольдфарба–Шенно) при минимизации сильно выпуклых квадратичных функционалов приводят к одной и той же траектории спуска, вырождаясь в хорошо изученные методы сопряженных градиентов.

Методы сопряженных градиентов. В то же время известна особенность методов сопряженных градиентов, существенно ограничивающая область их эффективного применения. Она заключается в понижении скорости сходимости для плохо обусловленных (жестких) задач оптимизации. Оценки, приведенные, например, в [6], показывают, что стандартные методы СГ сходятся по закону геометрической прогрессии со знаменателем q , близким к единице:

$$q \cong 1 - 2/\sqrt{\eta},$$

где η – степень жесткости минимизируемого функционала. Там же имеются указания на достаточно высокую скорость сходимости метода СГ «по функционалу», независимо от величины η :

$$J(x^k) - J(x^*) \leq \frac{L \|x^0 - x^*\|^2}{2(2k + 1)^2}, \quad L = \text{const}. \quad (14)$$

Однако оценки типа (14) получены в предположении $G(x) > 0$. Кроме этого, согласно (14) эффективно получаются значения функционала порядка $J(x^*)$, где x^* – минимизатор аппроксимирующего параболоида $f(x)$, что в общем случае не решает задачи. В соответствии с (14) реализуется указанная скорость сходимости по функционалу, но не по аргументу.

Предположение о невыпуклости вносит дополнительные трудности. Показано, что в этих условиях метод СГ по характери-

кам сходимости эквивалентен градиентному методу наискорейшего спуска со всеми вытекающими отсюда последствиями. Кроме отмеченных дефектов, общих для методов сопряженных градиентов и квазиньютоновских методов, последние имеют дополнительные недостатки, связанные с проблемой потери положительной определенности квазиньютоновских матриц из-за накопления вычислительных погрешностей в рекуррентных процедурах аппроксимации матриц Гессе [5].

В настоящее время на основе экспериментальных данных принята точка зрения, согласно которой ньютоновские методы с конечно-разностной аппроксимацией матрицы Гессе на основе аналитических выражений для градиентов надежнее, чем квазиньютоновские методы, и сходятся в более широком классе случаев, в частности, в задачах с очень плохой обусловленностью матрицы Гессе. С другой стороны, согласно тем же источникам, если аналитические выражения для градиентов отсутствуют и

применяются версии нулевого порядка соответствующих алгоритмов, в большинстве случаев более эффективными оказываются квазиньютоновские методы с конечно-разностной аппроксимацией градиентов. Данные замечания не снимают отмеченные выше недостатки указанных процедур, а отражают лишь некоторые сравнительные оценки, основанные на опыте проведения реальных вычислений.

Практические экстремальные задачи часто оказываются невыпуклыми и плохо обусловленными (жесткими). Основная полезная численная информация содержится в значениях «малых» по модулю собственных чисел матрицы Гессе. Эффективные алгоритмы минимизации жестких невыпуклых функционалов должны строиться не на ньютоновской или квазиньютоновской идеологии, приводящей к потере информации о малых спектральных составляющих, а на основных принципах теории жестких систем оптимизации [7–9].

СПИСОК ЛИТЕРАТУРЫ

1. **Черноруцкий, И.Г.** Методы параметрической оптимизации в задачах идентификации [Текст] / И.Г. Черноруцкий // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. –СПб.: Изд-во Политехнического ун-та, 2009. –№ 2 (76). –С. 150–155.
2. **Черноруцкий, И.Г.** Параметрические методы синтеза систем управления [Текст] / И.Г. Черноруцкий // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. –СПб.: Изд-во Политехнического ун-та, 2009. –№ 2 (76). –С. 111–115.
3. **Черноруцкий, И.Г.** Алгоритмические проблемы жесткой оптимизации [Текст] / И.Г. Черноруцкий // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. –СПб.: Изд-во Политехнического ун-та, 2012. –№ 6 (162). –С. 141–152.
4. **Ортега, Дж.** Итерационные методы решения нелинейных систем уравнений со многими неизвестными [Текст] Дж. Ортега, В. Рейнболдт.

- М.: Мир, 1975. – 560 с.
5. **Гилл, Ф.** Практическая оптимизация [Текст] / Ф. Гилл, У. Мюррей, М. Райт. –М.: Мир, 1985. – 509 с.
6. **Поляк, Б.Т.** Введение в оптимизацию [Текст] / Б.Т. Поляк. –М.: Наука, 1983. – 384 с.
7. **Ракитский, Ю.В.** Численные методы решения жестких систем [Текст] / Ю.В. Ракитский, С.М. Устинов, И.Г. Черноруцкий. –М.: Наука, 1979. – 208 с.
8. **Черноруцкий, И.Г.** Функции релаксации градиентных методов [Текст] / И.Г. Черноруцкий // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. –СПб.: Изд-во Политехнического ун-та, 2012. –№ 3 (150). –С. 66–72.
9. **Черноруцкий, И.Г.** Некоторые стандартные схемы параметрической оптимизации [Текст] / И.Г. Черноруцкий // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. –СПб.: Изд-во Политехнического ун-та, 2012. –№ 6 (162). –С. 128–133.

REFERENCES

1. **Chernorutskii I.G.** Metody parametricheskoi optimizatsii v zadachakh identifikatsii / Nauchno-tekhnicheskie vedomosti SPbGPU. Informatika.

- Telekommunikatsii. Upravlenie. –St.-Petersburg: Izd-vo Politehn. un-ta, 2009. –№ 2 (76). –S. 150–155. (rus)

2. **Chernorutskii I.G.** Parametricheskie metody sinteza sistem upravleniia /Nauchno-tekhnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie. –St.-Petersburg: Izd-vo Politehn. un-ta, 2009. –№ 2 (76). –S. 111–115. (rus)
3. **Chernorutskii I.G.** Algoritmicheskie problemy zhestkoi optimizatsii /Nauchno-tekhnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie. –St.-Petersburg: Izd-vo Politehn. un-ta, 2012. –№ 6 (162). –S. 141–152. (rus)
4. **Ortega Dzh.** Iteratsionnye metody resheniia nelineinykh sistem uravnenii so mnogimi neizvestnymi. –Moscow: Mir, 1975. – 560 s. (rus)
5. **Gill F.** Prakticheskaiia optimizatsiia. –Moscow: Mir, 1985. – 509 s. (rus)
6. **Poliak B.T.** Vvedenie v optimizatsiiu. –Moscow: Nauka, 1983. – 384 s. (rus)
7. **Rakitskii Iu.V.** Chislennye metody resheniia zhestkikh sistem. –Moscow: Nauka, 1979. – 208 s. (rus)
8. **Chernorutskii I.G.** Funktsii relaksatsii gradientnykh metodov / Nauchno-tekhnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie. –St.-Petersburg: Izd-vo Politehn. un-ta, 2012. –№ 3 (150). –S. 66–72. (rus)
9. **Chernorutskii I.G.** Nekotorye standartnye skhemy parametricheskoi optimizatsii / Nauchno-tekhnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie. –St.-Petersburg: Izd-vo Politehn. un-ta, 2012. –№ 6 (162). –S. 128–133. (rus)

ЧЕРНОРУЦКИЙ Игорь Георгиевич – директор Института информационных технологий и управления, заведующий кафедрой информационных и управляющих систем Санкт-Петербургского государственного политехнического университета, доктор технических наук, профессор.

195251, Россия, Санкт-Петербург, Политехническая ул., д. 21.

E-mail: director@icc.spbstu.ru

CHERNORUTSKIY, Igor G. St. Petersburg State Polytechnical University.

195251, Politekhnikeskaya Str. 21, St.-Petersburg, Russia.

E-mail: director@icc.spbstu.ru



УДК 681.3.016

Д.Е. Бортыяков, С.В. Мещеряков, Н.А. Солодилова

**ОБЕСПЕЧЕНИЕ КАЧЕСТВА АВТОМАТИЗИРОВАННОГО
ПРОЕКТИРОВАНИЯ МЕТАЛЛОКОНСТРУКЦИЙ ТЕХНОЛОГИЧЕСКИХ
СИСТЕМ НА ОСНОВЕ РАСПРЕДЕЛЕННОЙ БАЗЫ ДАННЫХ
ЭКСПЛУАТАЦИОННЫХ ДЕФЕКТОВ**

D.E. Bortyakov, S.V. Mescheryakov, N.A. Solodilova

**QUALITY ASSURANCE OF COMPUTER-AIDED DESIGN
FOR PRODUCTION METALWARE BASED ON DISTRIBUTED DATABASE
OF OPERATING DEFECTS**

Выполнен анализ базы данных эксплуатационных дефектов промышленных объектов на различных предприятиях Северо-Западного региона России. Выявлены наиболее характерные дефекты металлоконструкций и способы их устранения. Изменены традиционные математические модели расчета металлоконструкций с целью улучшения качества автоматизированного проектирования.

БАЗА ДАННЫХ. ЭКСПЛУАТАЦИОННЫЙ ДЕФЕКТ. МЕТАЛЛИЧЕСКИЕ КОНСТРУКЦИИ. АВТОМАТИЗИРОВАННОЕ ПРОЕКТИРОВАНИЕ. ОБЕСПЕЧЕНИЕ КАЧЕСТВА.

Production object database of operating defects on various enterprises in the North-West region of Russia is analyzed. The most problematic metalware and chronic defects are figured out as well as methods to resolve them. Traditional mathematic models of metalware estimation are modified for the purpose of quality assurance to computer-aided design.

DATABASE. OPERATING DEFECT. METALWARE. COMPUTER-AIDED DESIGN. QUALITY ASSURANCE.

Металлоконструкции являются неотъемлемой частью технологических и транспортно-перегрузочных систем и оказывают существенное влияние на эффективность функционирования всего производственного комплекса. При проектировании таких объектов, в особенности металлоемких конструкций большого размера, отсутствует возможность создания опытного образца, который можно было бы исследовать в различных условиях, вплоть до аварийных. Выполняемые проекторочные расчеты не учитывают различие в реальных производственных процессах и условиях эксплуатации.

Возникает необходимость корректировки математических моделей конструкций на основе экспериментальных результатов их практического применения при различных условиях работы и интенсивности на-

грузки. Полученные сведения представляют собой разрозненную и разнородную информацию, поскольку на разных предприятиях используются различные средства автоматизации и форматы данных. На многих из них технические паспорта, сведения об эксплуатации и другая учетная информация не структурированы в базе данных и хранятся в виде отдельных офисных документов. Все это затрудняет анализ статистики и делает его невозможным для проектировщиков и потребителей. Для полноценного анализа необходима интеграция разрозненных статистических данных в единую автоматизированную систему сбора и обработки информации.

При проектировании новых конструкций обычно стремятся по возможности использовать единообразные и сопоставимые методы анализа и синтеза, что находит от-

ражение в типовых инженерных методах расчетов. Создание каждой такой методики для конструкции в целом или отдельных ее элементов – сложный и длительный процесс, который приходится вести практически непрерывно на фоне модернизации и появления новых принципиальных схем. В этой ситуации внедрение единой базы данных эксплуатационных дефектов в современную практику автоматизированного проектирования представляется весьма актуальным и позволит специалистам гибко вносить изменения в имеющиеся типовые методики расчета, а также давать рекомендации инженерно-экспертным организациям по сбору дополнительных характеристик эксплуатации в целях выявления причин возникновения дефектов конструкций.

Анализ эксплуатационных дефектов технологических систем на основе распределенной базы данных

Интеграция распределенных данных в единую информационную систему выполнена с применением методов, предложенных в работах [1, 2]. База данных эксплуа-

тационных дефектов реализована на основе объектно-иерархических и реляционных моделей, представленных в работах [3, 4]. Экспериментальные данные для анализа собраны в базу данных с 469 действующих промышленных предприятий Северо-Западного региона России, на которых эксплуатируется более 1370 производственных объектов. Сведения о количестве обследованных объектов, сроках их эксплуатации и общем уровне выявленных дефектов приведены на рис. 1.

Средний возраст промышленных объектов составляет 37 лет. Тем не менее согласно статистике, прямой зависимости между их сроком службы и наличием дефектов не наблюдается. Например, после 50 лет эксплуатации количество дефектов соизмеримо с уровнем 20-летнего возраста. Это объясняется различием в интенсивности производства, графике работы, динамической нагрузке, погодных условиях – легких, средних, тяжелых.

Общее состояние обследованных промышленных объектов оценивалось экспертами по трехбалльной системе: исправное,

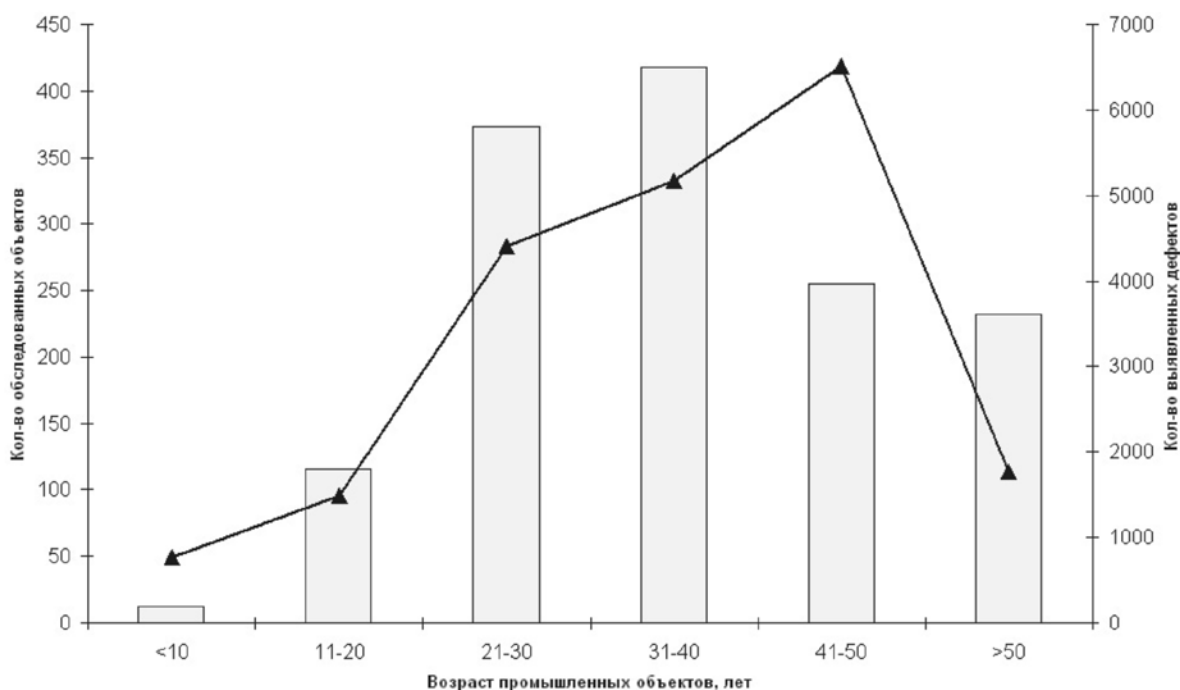


Рис. 1. Сведения о количестве и сроках эксплуатации промышленных объектов (□) кол-во обследованных объектов; (—▲) кол-во выявленных дефектов

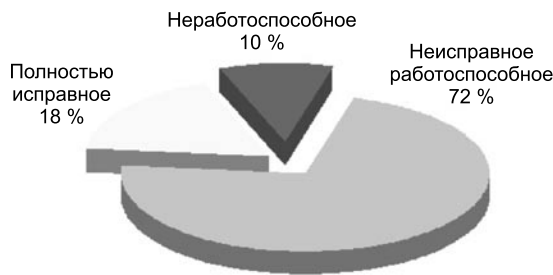


Рис. 2. Общее состояние обследованных промышленных объектов

неисправное работоспособное, неработоспособное (рис. 2). Для неработоспособного оборудования дальнейшая эксплуатация запрещена ввиду обнаруженных критических дефектов. Работоспособное оборудование, в отличие от исправного, предполагает наличие набора несущественных замечаний, которые разрешается устранить в течение определенного срока до очередного технического освидетельствования (от полугода до двух–трех лет в зависимости от условий

эксплуатации и характера нагрузки).

На всех промышленных объектах, погрузо-разгрузочных и транспортно-технологических комплексах в той или иной степени применяются следующие типы несущих металлоконструкций:

- балочные (более 10 %);
- балочно-ферменные клепаные (более 10 %);
- балочно-ферменные сварные (более 10 %);
- двухавровые (менее 10 %);
- коробчатые (более 10 %);
- рамные (менее 1 %);
- рельсовые (менее 10 %);
- ригельные (менее 1 %);
- сварные листовые (более 10 %);
- трубчатые сварные (более 10 %);
- уголковые (менее 1 %);
- ферменные клепаные (более 10 %);
- ферменные сварные (более 10 %);
- швеллерные (менее 10 %);
- шпренгельные (менее 10 %).

По результатам технических освидетель-

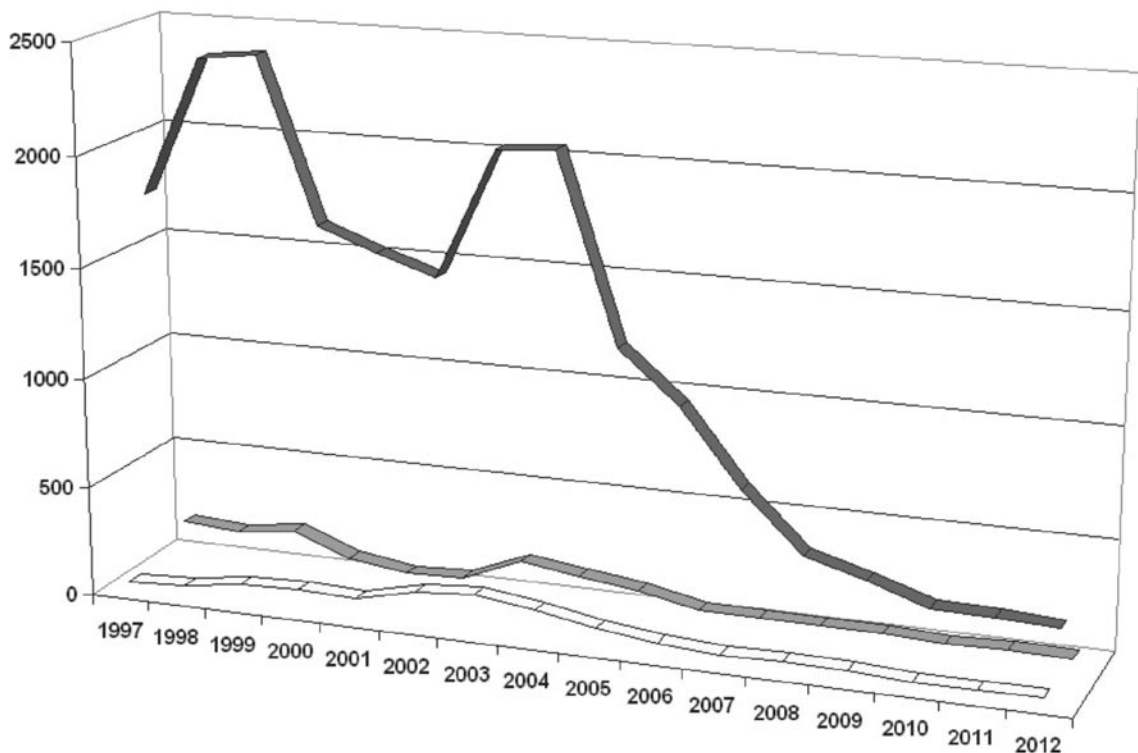


Рис. 3. Количество обследованных промышленных объектов и выявленных дефектов по годам (□) кол-во промышленных объектов; (■) общее кол-во дефектов; (▣) кол-во дефектов металлоконструкций

ствований за более чем 10 лет практически каждый обследованный промышленный объект имеет дефекты металлоконструкций при среднем уровне 15 дефектов на объект (рис. 3).

На обследованных объектах все металлоконструкции выдержали статические испытания, однако в результате динамического нагружения после снятия нагрузки у некоторых элементов металлоконструкций были обнаружены дефекты. Наиболее характерными из них, на основе анализа базы данных дефектов, являются трещины, остаточная деформация, прогиб, течи рабочей жидкости, ослабление заклепок или болтовых соединений.

В 35 % случаев по результатам обследования сделан либо капитальный ремонт с заменой отдельных элементов металлоконструкций, либо текущий ремонт местных деформаций с применением сварки.

Аналогичные исследования [5] подтверждают, что результаты замеров при статических и динамических испытаниях могут сильно отличаться друг от друга. На рис. 4 приведен пример результатов нивелировки высотного положения направляющих стального типа для порталного крана

по перемещению на складе завода железобетонных изделий. Во многих точках замеров (на участке от 3–5 до 8–10) наблюдается характерный прогиб металлического профиля под нагрузкой. При этом разность отметок высотного положения направляющих значительно превышает предельно допустимое отклонение 20 мм согласно действующим нормативам [6, 7].

Вопросы оценки динамических свойств металлоконструкций на стадии проектирования не затрагиваются или затрагиваются минимальным образом. Отсутствие анализа динамических процессов может привести к критическим ситуациям, например сильное отклонение груза из плоскости движения тележки может привести к поломке ног перегружателя. Расчетный прогноз потенциальной возможности появления таких динамических качеств необходимо осуществлять с использованием критериев эксплуатации и последующей корректировкой расчетных методик.

Расчетная математическая модель крановых металлоконструкций

При построении расчетной математической модели использована концепция

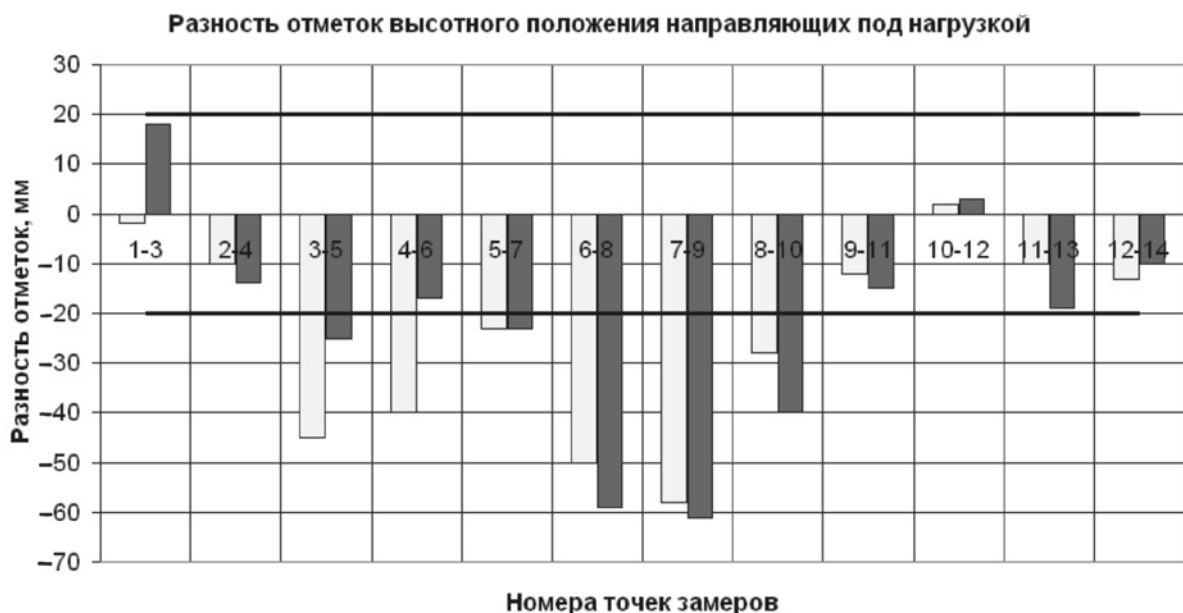


Рис. 4. Результаты замеров высотного положения направляющих под нагрузкой (□) ось А; (■) ось Б; (— —) предельное отклонение 20 мм

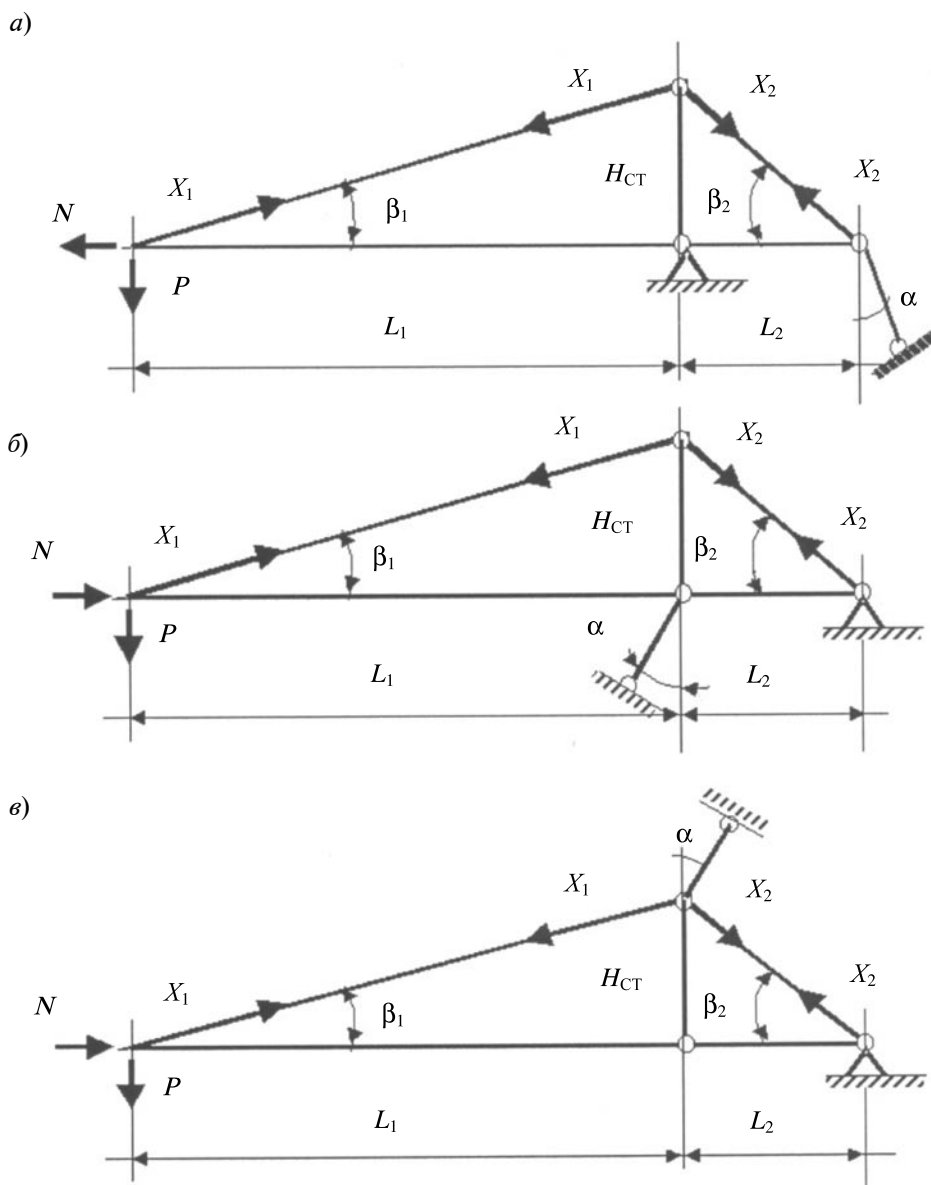


Рис. 5. Расчетные схемы шпренгельных конструкций:
 а – хобота; б, в – стрелы с креплением тягового звена соответственно к балке и к узлу шпренгелей

составной динамической многосвязной системы, которая базируется на типовых конструкциях, характеристики функционирования которых известны и хорошо изучены. Такое представление позволяет вскрыть конструктивные особенности и их потенциальные недостатки на ранней стадии и тем самым повысить продуктивность и качество процесса проектирования.

На основе сформулированной концепции в традиционные математические модели [8–11] внесены изменения с целью по-

лучения требуемых показателей. Помимо механизмов с жесткими звеньями, добавлен учет упругой составляющей и отклонений перемещения, вызванных податливостью звеньев.

На рис. 5 представлен пример расчетных схем для шпренгельных металлических конструкций в виде консольной балки на двух опорах со стойкой и раскосами, которая наиболее часто встречается в современной практике. За основу взят деформационный расчет конструкции, т. к. при прочностном

расчете для соблюдения условия равенства прочности элементов невозможно установить закономерность перераспределения нагрузок между элементами внутри статически неопределенной системы.

Усилия в раскосах принимаются в качестве неизвестных X_1 и X_2 . Система уравнений совместности перемещений одинакова для всех трех рассматриваемых расчетных случаев:

$$\delta_{11} X_1 + \delta_{12} X_2 + \Delta_{1P} = 0,$$

$$\delta_{21} X_1 + \delta_{22} X_2 + \Delta_{2P} = 0,$$

где δ_{ij} – перемещения j -й точки под действием единичной i -й силы; X_1 и X_2 – неизвестные внутренние усилия в раскосах; Δ_{iP} – перемещения i -й точки под действием внешней нагрузки P .

Перемещения всех точек системы под действием внутренних усилий X_1 и X_2 одинаковы для всех расчетных схем (рис. 5) и определяются известными методами [8, 9], но перемещения от внешней нагрузки будут разные. Поперечные силы не учитываются, т. к. для перемещений они не имеют существенного значения. Неизвестные параметры раскосов и стойки выделены отдельно:

$$\delta_{11} = C_1 + L_{p1}/F_{p1},$$

$$\delta_{22} = C_2 + L_{p2}/F_{p2},$$

где L_{p1} и F_{p1} , L_{p2} и F_{p2} – длины и площади поперечного сечения первого и второго раскосов соответственно; C_1 и C_2 – составляющие коэффициентов δ_{11} и δ_{22} , рассчитанные на основе параметров известных конструктивных элементов.

Тогда система уравнений имеет следующий вид:

$$(C_1 + L_{p1}/F_{p1})X_1 + \delta_{12}X_2 + \Delta_{1P} = 0,$$

$$(C_2 + L_{p2}/F_{p2})X_2 + \delta_{21}X_1 + \Delta_{2P} = 0.$$

Выражения для усилий X_1 и X_2 берутся исходя из вероятного распределения внешней нагрузки между балкой и раскосами. Если ввести коэффициент K_p , показывающий, какую часть внешней нагрузки берет на себя балка ($K_p < 1$), то выражения для усилий X_1 и X_2 при условии равновесия узлов будут иметь вид:

$$X_1 = P(1 - K_p)/\sin(\beta_1),$$

$$X_2 = X_1 \cos(\beta_1)/\cos(\beta_2) \text{ или}$$

$$X_2 = (X_1 \cos(\beta_1) - P \sin(\alpha))/\cos(\beta_2).$$

Решив данную систему, можно получить величины площадей раскосов F_{p1} и F_{p2} , при которых система удовлетворяет требованиям совместной деформации балочной и шпренгельной систем. Затем выполняется проверочный расчет. Если расчетные напряжения в раскосах получились больше допустимых, высота стойки последовательно уменьшается и расчет повторяется. Получив площади раскосов, удовлетворяющие прочностным и деформационным ограничениям, производится их проверка по условию общей устойчивости.

Реализация новой модели и дальнейшее развитие САПР позволяет оценить эффективность применения усовершенствованных элементов конструкции, в частности, систем гашения колебаний и демпфирующих устройств, на ранних стадиях проектирования.

Основная ценность распределенной базы данных эксплуатационных дефектов состоит в возможности анализировать сведения различных предприятий о техническом состоянии промышленных объектов. В процессе промышленной эксплуатации и по результатам периодического обследования выявляются наиболее слабые элементы конструкции и наиболее распространенные дефекты. Нередко эти дефекты вызваны особыми условиями эксплуатации и динамическими нагрузками, которые трудно рассчитать на этапе автоматизированного проектирования.

Описанная математическая модель расчета металлоконструкций и методика ее применения позволяют производить более точный расчет различных систем, улучшающих динамические свойства конструкций, и тем самым повысить качество автоматизированного проектирования. Кроме того, проведенные исследования помогли сформулировать практические рекомендации эксплуатирующим предприятиям и экспертным организациям по регулярному

контролю дополнительных параметров объектов с учетом фактической динамической нагрузки с целью их последующего анализа в базе данных.

Опыт автоматизированного проектирования и промышленной эксплуатации

металлоконструкций кранов подтверждает правильность такого подхода. По статистическим данным (см. рис. 3) на промышленных предприятиях Северо-Западного региона России количество выявленных дефектов в последние годы значительно снизилось.

СПИСОК ЛИТЕРАТУРЫ

1. **Мещеряков, С.В.** Эффективные технологии создания информационных систем [Текст]/ С.В. Мещеряков, В.М. Иванов.— СПб.: Изд-во «Политехника», 2005. — 312 с.

2. **Мещеряков, С.В.** Построение объектно-реляционных моделей баз данных с произвольным набором атрибутов [Текст]/ С.В. Мещеряков, В.М. Иванов // Системы управления и информационные технологии. — 2005. — № 4 (21). — С. 82–87.

3. **Mescheryakov, S.V.** A Successful Implementation of a Data Structure for Storing Multilevel Objects with Varying Attributes [Электронный ресурс]/ S.V. Mescheryakov // IBM, 2002. Режим доступа: <http://www.ibm.com/developerworks/data/zones/informix/library/techarticle/0212mescheryakov/0212mescheryakov.html>

4. **Мещеряков, С.В.** Реализация модели данных для описания иерархических объектов с произвольными атрибутами [Текст]/ С.В. Мещеряков, В.М. Иванов // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. — СПб.: Изд-во Политехн. ун-та, 2009. — № 1 (72). — С. 139–144.

5. **Мещеряков, С.В.** Анализ эксплуатационных дефектов грузоподъемного оборудования в системе автоматизированного проектирования [Текст] / С.В. Мещеряков, В.М. Иванов, Д.Е. Бортыков // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. — СПб.: Изд-во Политехн. ун-та, 2011. — № 6 (138). — Т. 2. — С. 95–102.

6. **РД 50:48:0075.03.05.** Рекомендации по устройству и безопасной эксплуатации надземных крановых путей: Руководящий нормативный документ. — М.: НПЦ «Путь К», 2005. — 213 с.

7. **ПБ 10-382-00.** Правила устройства и безопасной эксплуатации грузоподъемных кранов. — СПб.: ЦОТПБСП, 2000. — 129 с.

8. **Орлов, А.Н.** Обобщенная математическая модель стреловых кранов [Текст]/ А.Н. Орлов, Ф. Флюгель // Труды СПбГТУ. — СПб.: Изд-во СПбГТУ, 1995. — № 455. — С. 66–77.

9. **Бортыков, Д.Е.** Обеспечение качества системы изменения вылета порталных кранов на этапе проектировочного расчета [Текст]/ Д.Е. Бортыков, А.Н. Орлов // Повышение качества изделий в машиностроении: Межвуз. сб. научн. трудов. — Л.: Изд-во ЛПИ, 1990. — С. 101–104.

10. **Орлов, А.Н.** Нагрузки металлической конструкции крана при ее свободных колебаниях с грузом и управляемыми демпфирующими устройствами [Текст]/ А.Н. Орлов, В.Я. Головачев // Труды СПбГТУ. — СПб.: Изд-во СПбГТУ, 1993. — № 445. — С. 94–104.

11. **Манжула, К.П.** Оценка увеличения долговечности крановых металлоконструкций при включении в систему подвеса груза демпфирующих устройств [Текст]/ К.П. Манжула, А.Н. Орлов, В.Я. Головачев, А.М. Кабаков // Труды СПбГТУ. — СПб.: Изд-во СПбГТУ, 1993. — № 445. — С. 105–118.

REFERENCES

1. **Mescheryakov S.V., Ivanov V.M.** Effektivnyye tekhnologii sozdaniya informatsionnyh system. — St.-Petersburg: Izd-vo «Politekhnik», 2005. — 312 s. (rus)

2. **Mescheryakov S.V., Ivanov V.M.** Object-relational models construction database systems with varying attributes / Control Systems and Information Technology. — 2005. — № 4 (21). — S. 82–87. (rus)

3. **Mescheryakov S.V.** A Successful Implementation of a Data Structure for Storing Multilevel Objects with Varying Attributes. IBM, 2002. Available <http://www.ibm.com/developerworks/data/zones/informix/library/techarticle/0212mescheryakov/0212mescheryakov.html>

[informix/library/techarticle/0212mescheryakov/0212mescheryakov.html](http://www.ibm.com/developerworks/data/zones/informix/library/techarticle/0212mescheryakov/0212mescheryakov.html)

4. **Mescheryakov S.V., Ivanov V.M.** Realizatsiya modeli dannyih dlya opisaniya ierarhicheskikh ob'ektov s proizvolnyimi atributami (Implementation of a data model for defining hierarchical objects with varying attributes) / Nauchno-tekhnicheskie vedomosti SPbGPU. Informatika. Telekommunicacii. Upravlenie. — St.-Petersburg: Izd-vo Politehn. un-ta, 2009. — № 1 (72). — P. 139–144. (rus)

5. **Mescheryakov S.V., Ivanov V.M., Bortyakov D.E.** Analiz ekspluatatsionnyh defektov

gruzopod'emnogo oborudovaniya v sisteme avtomatizirovannogo proektirovaniya (Analysis of operating defects of lifting equipment in computer-aided design system) / Nauchno-tekhnicheskie vedomosti SPbGPU. Informatika. Telecommunicacii. Upravlenie. — St.-Petersburg: Izd-vo Politehn. un-ta, 2011. — № 6 (138). — Т. 2. — P. 95–102. (rus)

6. **RD 50:48:0075.03.05.** Rekomendatsii po ustroistvu i bezopasnoi ekspluatatsii nadzemnyh kranovyh putei: Rukovodyashchii normativniy document. — Moscow: NPC «Put' K», 2005. — 213 s. (rus)

7. **PB 10-382-00.** Pravila ustroistva i bezopasnoi ekspluatatsii gruzopod'emnyh kranov. — St.-Petersburg: COTPBSP, 2000. — 129 s. (rus)

8. **Orlov A.N., Flyugel' F.** Obobshchennaya matematicheskaya model' strelovyh kranov / Trudy SPbGTU. — St.-Petersburg: Izd-vo SPbGTU, 1995. — № 455. — S. 66–77. (rus)

9. **Bortyakov D.E., Orlov A.N.** Obespechenie kachestva sistemy izmeneniya vyleta portal'nyh kranov na etape proektirovochnogo rascheta / Povyshenie kachestva izdeliy v mashinostroenii: Mezhvuz. sb. nauchn. tr. — St.-Petersburg: Izd-vo LPI, 1990. — S. 101–104. (rus)

10. **Orlov A.N., Golovachev V.Ya.** Nagruzki metallicheskoj konstruksii kрана pri ee svobodnyh kolebaniyah s gruzom i upravlyaemymi dempфирuyushchimi ustroystvami / Trudy SPbGTU. — St.-Petersburg: Izd-vo SPbGTU, 1993. — № 445. — S. 94–104. (rus)

11. **Manzhula K.P., Orlov A.N., Golovachev V.Ya., Kabakov A.M.** Otsenka uvelicheniya dolgovechnosti kranovyh metallokonstruksiy pri vkluchenii v sistemu podvesa gruzа dempфирuyushchih ustroystv / Trudy SPbGTU. — St.-Petersburg: Izd-vo SPbGTU, 1993. — № 445. — S. 105–118. (rus)

БОРТЯКОВ Данил Евгеньевич — доцент кафедры транспортных и технологических систем Санкт-Петербургского государственного политехнического университета, кандидат технических наук.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: bortyakov@mail.ru

BORTYAKOV, Danil E. — *St. Petersburg State Polytechnic University.*

195251, Politekhnikeskaya Str. 29, St.-Petersburg, Russia.
E-mail: bortyakov@mail.ru

МЕЩЕРЯКОВ Сергей Владимирович — ведущий программист Санкт-Петербургского государственного политехнического университета, доктор технических наук, доцент.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: serg-phd@mail.ru

MESCHERYAKOV, Sergey V. — *St. Petersburg State Polytechnic University.*

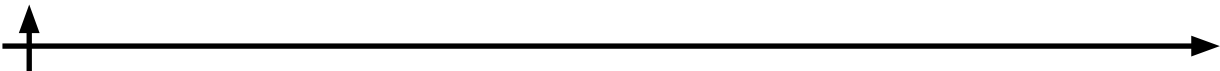
195251, Politekhnikeskaya Str. 29, St.-Petersburg, Russia.
E-mail: serg-phd@mail.ru

СОЛОДИЛОВА Наталья Алексеевна — старший преподаватель кафедры конструкторско-технологических инноваций Санкт-Петербургского государственного политехнического университета.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.
E-mail: solodna@yandex.ru

SOLODILOVA, Natalia A. — *St. Petersburg State Polytechnic University.*

195251, Politekhnikeskaya Str. 29, St.-Petersburg, Russia.
E-mail: solodna@yandex.ru



Электроника, технологии производства материалов электронной техники

УДК 621.376.9

М.М. Пилипко

ДЕЛЬТА-СИГМА МОДУЛЯТОР АНАЛОГО-ЦИФРОВОГО ПРЕОБРАЗОВАТЕЛЯ С ПРЕОБРАЗОВАНИЕМ ВО ВРЕМЕНИ

М.М. Piliipko

TIME-MODE DELTA-SIGMA MODULATOR FOR AN ANALOG-TO-DIGITAL CONVERTER

Рассмотрен принцип аналого-цифрового преобразования во временной области и устройства, реализующие данный принцип. Предложена дифференциальная структура дельта-сигма модулятора с преобразованием сигналов во времени. Работоспособность устройств подтверждена результатами схемотехнического моделирования.

ПРЕОБРАЗОВАТЕЛЬ НАПРЯЖЕНИЯ ВО ВРЕМЕННОЙ ИНТЕРВАЛ. ДЕЛЬТА-СИГМА МОДУЛЯТОР. ЦИФРОВЫЕ УСТРОЙСТВА.

Principle of time-mode analog-to-digital conversion is considered. Circuits utilizing this principle are presented. A differential structure of the time-mode delta-sigma modulator is proposed. Operability of the presented circuits has been confirmed by transistor-level simulation.

VOLTAGE-TO-TIME CONVERTER. DELTA-SIGMA MODULATOR. DIGITAL CIRCUITS.

Движущая сила развития МОП технологий – рост производства и повсеместное распространение цифровых устройств. В связи с этим оптимизация технологических процессов направлена, прежде всего, на улучшение характеристик цифровых схем: увеличение скорости переключения транзисторов, уменьшение напряжения питания и геометрических размеров элементов. Указанные изменения усложняют разработку аналоговых и аналого-цифровых устройств. В частности, следствием уменьшения толщины подзатворного диэлектрика является увеличение токов утечки через затвор. Снижение напряжения питания схем вынуждает использовать транзисторы в неоптимальных рабочих точках. Это ограничивает допустимый размах входного сигнала и вызывает проблемы с линейностью аналоговых схем.

Для обхода указанных проблем при разработке аналого-цифровых преобразователей (АЦП) ряд зарубежных авторов предлагает перейти от традиционного преобразования в области напряжений к преобразованию сигналов во временной области [1–3]. Принцип аналого-цифрового преобразования на основе обработки сигналов во временной области реализуется путем каскадного соединения нескольких блоков. Поскольку информационным сигналом АЦП, как правило, является входное напряжение, первый блок схемы представляет собой преобразователь напряжения во временной интервал. Следующий блок служит для необходимых операций с сигналом во временной области. Затем время-цифровой преобразователь формирует выходной цифровой код АЦП.

На основе данного подхода можно по-

строить АЦП различных типов. В мобильных системах телекоммуникаций нашли применение следующие АЦП с использованием дельта-сигма модуляции [4-5]. Тактовая частота дельта-сигма модулятора существенно превышает частоту Найквиста, что позволяет с высокой точностью отследить изменение амплитуды входного сигнала. Основу дельта-сигма модулятора составляют интеграторы, которые обычно имеют в своем составе транскондуктивный усилитель [6] или операционный усилитель [5, 7]. Интегратор накапливает разности между входным сигналом и сигналом обратной связи, а компаратор определяет знак сигнала на выходе интегратора на каждом такте. Во временной области аналогичным образом можно сравнивать моменты прихода фронта для входного и опорного сигналов.

В данной статье рассматривается принцип аналого-цифрового преобразования во временной области и устройства, реализующие данный принцип.

Элемент задержки, управляемый напряжением

Для обработки входного сигнала во временной области необходимо сопоставить уровню напряжения на входе определенный временной интервал – т. н. *время-разностную переменную* [3]. Для этой цели служит элемент задержки, управляемый напряжением (ЭЗУН). Фронт сигнала, поступающего на такой элемент, задерживается на время, пропорциональное напряжению на управляющем входе.

На рис. 1 а показано условное обозначение элемента задержки, а на рис. 1 б – временная диаграмма. Фронт сигнала Φ_I поступает на вход элемента в момент времени t_I и задерживается до момента t_O , определяемого входным напряжением V_{IN} . Выражение для время-разностной переменной имеет вид $\Delta T_{O,I} = G_\Phi V_{IN} + b_\Phi$, где G_Φ – крутизна характеристики преобразования, а b_Φ – постоянная аддитивная задержка.

Наиболее широко известна схема ЭЗУН на основе инвертора с ограничением тока (рис. 1 в) [3, 8]. Принцип работы данной схемы следующий. При низком уровне входного сигнала Φ_I транзистор М1 открыт, а конденсатор С заряжен до напряжения питания. На выходе Φ_O сигнал низкого уровня. По фронту входного сигнала закрывается транзистор М1 и открывается транзистор М2, конденсатор разряжается через транзисторы М3 и М4, причем скорость разряда определяется напряжением V_{IN} . Выходной инвертор (М5-М6) играет роль компаратора и переключается в тот момент, когда напряжение на конденсаторе становится меньше порогового напряжения инвертора. Таким образом, момент появления фронта сигнала на выходе Φ_O линейно зависит от напряжения V_{IN} . По спаду входного сигнала транзистор М2 закрывается, а транзистор М1 открывается. Скорость заряда конденсатора через транзистор М1 не зависит от напряжения V_{IN} .

На рис. 1 г приведен вид передаточной характеристики ЭЗУН, построенной по результатам моделирования схемы на основе

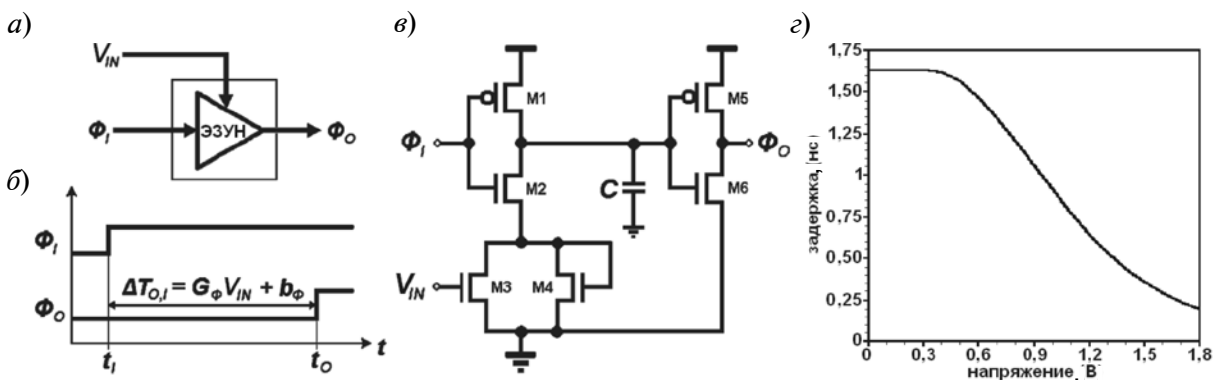


Рис. 1. Элемент задержки, управляемый напряжением:

а – обозначение; б – временная диаграмма; в – схема; г – передаточная характеристика

стандартной 0,18 мкм МОП технологии с напряжением питания 1,8 В. В диапазоне входных напряжений 0,7–1,1 В нелинейность характеристики не превышает 0,5 %. Крутизна характеристики составляет $G_{\Phi} = 710$ пс/В. Аддитивная часть времени задержки соответствует задержке ЭЗУН при максимальном управляющем напряжении (в данном случае 1,1 В) и равна $b_{\Phi} = 740$ пс. Временные характеристики и потребляемая мощность элемента задержки зависят от величины конденсатора C и геометрических размеров транзистора МЗ. Если конденсатор C не использовать, работоспособность ЭЗУН сохранится благодаря наличию паразитных емкостей транзисторов.

Преобразователь напряжения и интеграторы на основе элемента задержки

В обычном аналого-цифровом преобразователе происходит сравнение входного напряжения с некоторым опорным уровнем (уровнем аналоговой земли). Аналогичное сравнение возможно и во временной области. Для этого импульсный сигнал Φ_I подается параллельно на два ЭЗУН. На управляющий вход одного из них поступает входное напряжение V_{IN} , а на управляющий вход другого – опорное напряжение V_{REF} (рис. 2 а). Полученная схема является преобразователем напряжения во временной интервал. Временная диаграмма работы данной схемы приведена на рис. 2 б. Моменты времени появления фронтов сигналов на выходах Φ_O и Φ_{REF} определяются следующими выражениями [3]:

$$t_O = t_I + G_{\Phi} V_{IN} + b_{\Phi},$$

$$t_{REF} = t_I + G_{\Phi} V_{REF} + b_{\Phi},$$

а интервал времени между ними равен $\Delta T_O = t_O - t_{REF} = G_{\Phi} (V_{IN} - V_{REF})$.

Если выразить переменную часть входного напряжения как $v_{in} = V_{IN} - V_{REF}$, то получим $\Delta T_O = G_{\Phi} v_{in}$. Таким образом, отклонение время-разностной переменной от опорного значения прямо пропорционально отклонению входного напряжения от опорного уровня. Отметим, что различие параметров элементов задержки может приводить к таким ошибкам преобразования,

как сдвиг относительно нуля и изменение крутизны характеристики.

Подключив к выходу ЭЗУН инвертор, а выход инвертора – ко входу ЭЗУН, можно получить генератор, управляемый напряжением [9]. Сравнение моментов прихода фронта в сигнальном и опорном генераторах позволяет реализовать операцию интегрирования во временной области (рис. 2 в). Поскольку аддитивная задержка b_{Φ} , а также время задержки ЭЗУН по спаду и время задержки инвертора постоянны, то изменение время-разностной переменной для n -го такта можно записать как

$$\Delta T_O(n) = \Delta T_O(n-1) + G_{\Phi} v_{in}(n-1).$$

Временная диаграмма для первых трех тактов интегратора приведена на рис. 2 г. Начальное значение $\Delta T_O(0)$ считается равным нулю. Последующие значения вычисляются по приведенной рекуррентной формуле. Следует заметить, что в данной структуре отсутствует внешний тактовый генератор. Частота опорного генератора определяется размерами транзисторов и уровнем опорного напряжения V_{REF} .

Интегратор с двумя или несколькими входами может быть построен путем включения соответствующего числа ЭЗУН в кольцевые генераторы. На рис. 2 д показан интегратор с двумя входными сигналами $V_{IN,1}(n)$ и $V_{IN,2}(n)$. При этом в опорном генераторе также используются два элемента задержки, управляемые сигналом V_{REF} . Разностное уравнение для интегратора в этом случае имеет вид:

$$\Delta T_O(n) = \Delta T_O(n-1) + G_{\Phi} [v_{in,1}(n-1) + v_{in,2}(n-1)].$$

Применительно к аналого-цифровому преобразованию на основе дельта-сигма модуляции желательным является каскадное соединение интеграторов. Однако в данном случае это представляет определенную трудность, поскольку выходным сигналом интегратора является временной интервал, а входным – напряжение. Поэтому необходим промежуточный блок для обратного преобразования интервала времени в уровень напряжения.

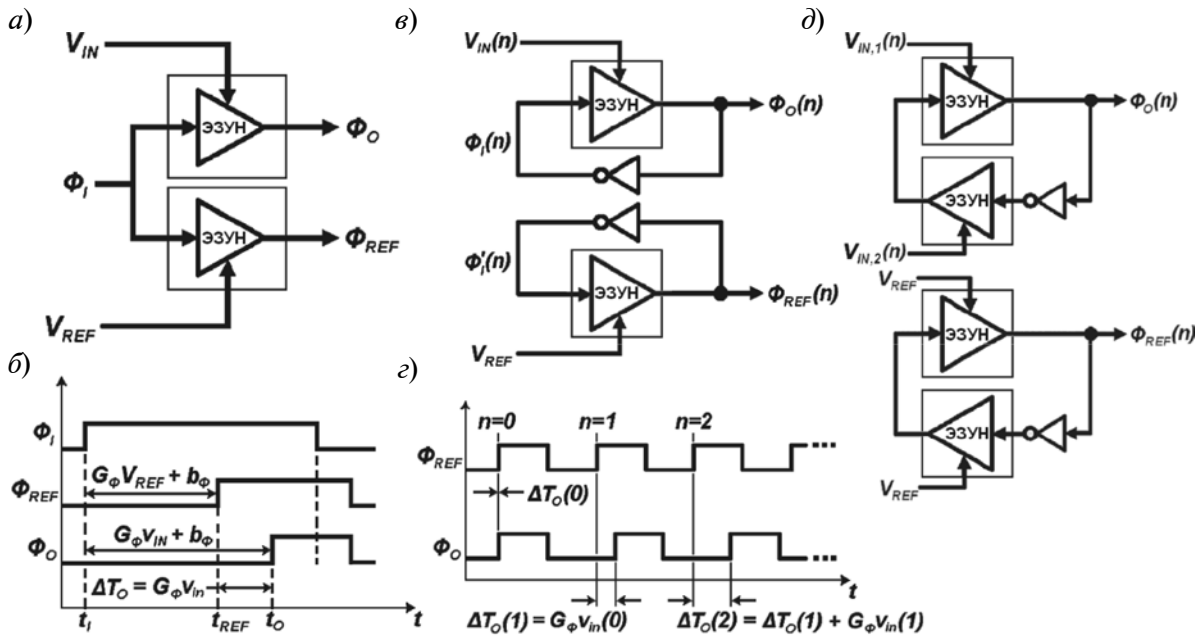


Рис. 2. Преобразователь напряжения во временной интервал и интеграторы время-разностных переменных

Последним этапом аналого-цифрового преобразования является перевод время-разностной переменной в цифровой вид. Знак переменной можно определить при помощи триггера задержки (D-триггера). Если фронт сигнала Φ_O опережает фронт тактового сигнала Φ_{REF} , то на выходе триггера будет сформирован сигнал логической единицы, в противном случае на выходе будет логический ноль. Устройство квантования с количеством уровней сравнения два и более [10] может быть организовано с помощью соответствующего количества элементов задержки и триггеров.

Дельта-сигма модулятор с преобразованием сигналов во времени

Дельта-сигма модулятор первого порядка состоит из интегратора, компаратора и цепи обратной связи с цифро-аналоговым преобразователем (ЦАП). В предыдущем разделе рассмотрены интегратор с двумя входами и компаратор сигналов во временной области. Если элементы задержки в интеграторе могут работать в полном диапазоне напряжений, то ЦАП в обратной связи не нужен, поскольку может использоваться непосредственно выходной сигнал

D-триггера. В противном случае следует применить одноразрядный ЦАП на основе двух аналоговых ключей. Как видно, все каскады дельта-сигма модулятора состоят из цифровых компонентов, следовательно, эта система обеспечивает все преимущества цифровой КМОП технологии. Аналого-цифровой преобразователь будет работать с высокой тактовой частотой, потреблять малую мощность и занимать небольшую площадь на кристалле. Более того, следует ожидать сохранения этих преимуществ с развитием цифровой КМОП технологии.

На рис. 3 а показана структура дельта-сигма модулятора с преобразованием во времени, основанная на двух кольцевых генераторах и D-триггере [3]. На элементы задержки одного генератора подаются входной аналоговый сигнал и инверсный выходной сигнал, на элементы задержки другого – опорное напряжение. Данный генератор формирует опорную частоту интегратора и тактирует D-триггер. Цель работы системы состоит в подстройке частоты сигнального генератора к частоте опорного генератора при помощи импульсов выходного кода. Разностное уравнение, описывающее работу интегратора, имеет вид

$$\Delta T_o(n) = \Delta T_o(n-1) + G_\Phi [v_{in}(n-1) - v_o(n-1)],$$

т. е. приращение время-разностной переменной на текущем такте равно разности входного и выходного сигналов на предыдущем такте, умноженной на крутизну характеристики ЭЗУН [3]. Сигнал ошибки квантования, порождаемый компаратором в виде D-триггера, определяется отличием время-разностной переменной от выходного сигнала, умноженного на крутизну ЭЗУН:

$$\Delta T_E(n) = G_\Phi v_o(n) - \Delta T_o(n).$$

Подставив данную модель сигнала ошибки в выражение для интегратора, можно получить разностное уравнение для дельта-сигма модулятора первого порядка:

$$v_o(n) = v_{in}(n-1) + \frac{1}{G_\Phi} [\Delta T_E(n) - \Delta T_E(n-1)].$$

Для обеспечения устойчивости к синфазным помехам аналого-цифровой преобразователь должен иметь дифференциальную структуру. В работе [3] предложено подавать на элементы задержки опорного кольцевого генератора (рис. 3 а) инверсное входное напряжение и прямой выход-

ной сигнал. Сравнение сигналов при помощи D-триггера в этом случае позволяет устранить воздействие синфазных помех и четные гармоники, обусловленные нелинейностью элементов задержки. Однако недостатком данного подхода является отсутствие синхронизации: частота опорного генератора изменяется, поэтому импульсы, поступающие на выход и в цепь обратной связи, непостоянны по длительности. Выходной сигнал такого модулятора перед подачей на устройства цифровой обработки требует преобразования в синхронный код, что приводит к увеличению уровня шума и сужению динамического диапазона.

Предлагается дифференциальная структура (рис. 3 б), лишенная указанного недостатка. В структуре используются два кольцевых генератора и два D-триггера. На управляющие входы каждого генератора подается один из входных сигналов (прямой V_{IN+} или инверсный V_{IN-}) и инверсный выходной сигнал соответствующего D-триггера. Сигнал синхронизации CLK поступает на оба D-триггера от внешнего генератора либо от третьего кольцевого генератора, на управляющие входы которого подано постоянное напряжение V_{REF}

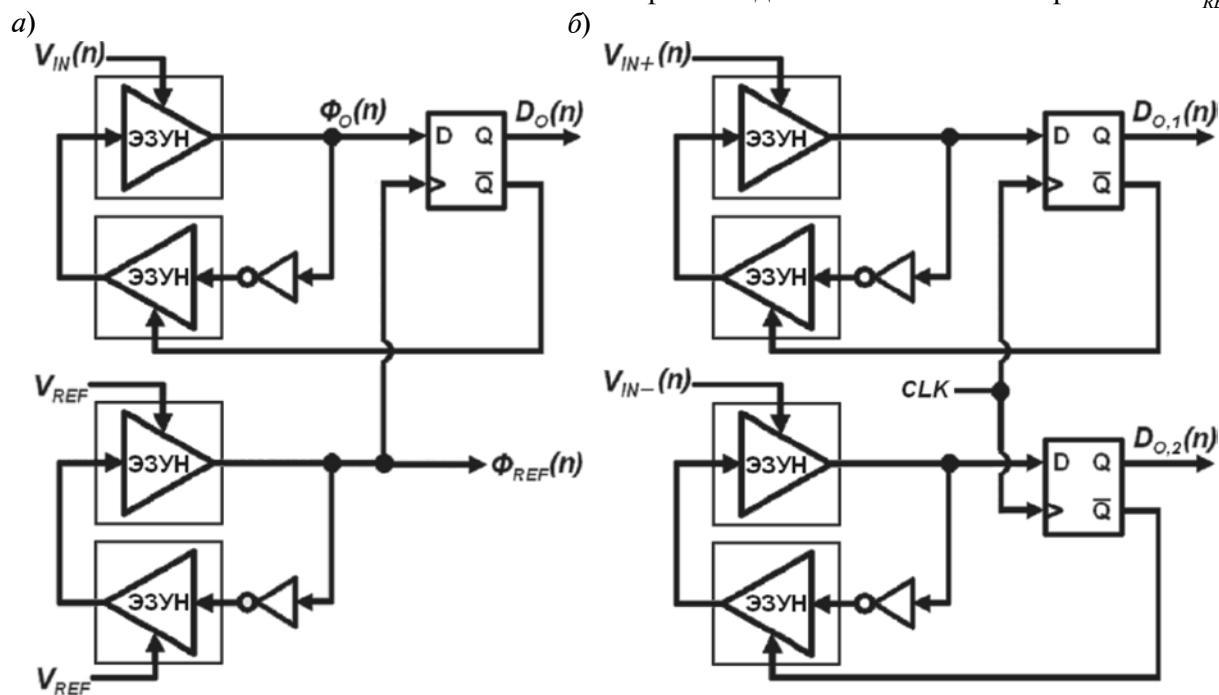


Рис. 3. Структуры дельта-сигма модулятора

Разность выходных кодов триггеров дает двухразрядный код, в котором отсутствуют входные синфазные составляющие и четные гармоники полезного сигнала.

Для проверки работоспособности предложенной структуры собрана схема с использованием стандартной 0,18 мкм МОП технологии с напряжением питания 1,8 В. Проведено моделирование дельта-сигма модулятора в среде Cadence Virtuoso. При тактовой частоте 280 МГц и коэффициенте передискретизации 128 отношение сигнал/шум составило 58 дБ, средняя потребляемая мощность не превышает 1,1 мВт.

В статье изложен принцип действия АЦП с преобразованием сигналов во временной

области. Рассмотрены основные устройства, необходимые для построения такого АЦП: элемент задержки, управляемый напряжением; интегратор время-разностных сигналов; компаратор. Приведена структура дельта-сигма модулятора первого порядка с преобразованием во времени. Показан недостаток известной дифференциальной структуры модулятора – асинхронность, предложена синхронная структура. Моделирование схемы для стандартной 0,18 мкм МОП технологии показало, что дифференциальный дельта-сигма модулятор первого порядка обеспечивает 9–10 разрядов разрешающей способности для сигналов с частотами до 1,1 МГц при средней потребляемой мощности 1,1 мВт.

СПИСОК ЛИТЕРАТУРЫ

1. **Yang, H.Y.** A time-based energy-efficient analog-to-digital converter [Text] / H.Y. Yang, R. Sarpeshkar // IEEE J. Solid-State Circuits. – 2005. – Vol. 40. – № 8. – P. 1590–1601.
2. **Wismar, U.** 0.2 V 7.5 uW 20 kHz sigma-delta modulator with 69 dB SNR in 90 nm CMOS [Text] / U. Wismar, D. Wisland, P. Andreani / Proc. Eur. Solid-State Circuits Conf. – 2007. – P. 206–209.
3. **Taillefer, C.S.** Delta-Sigma A/D Conversion Via Time-Mode Signal Processing [Text] / C.S. Taillefer, G.W. Roberts // IEEE Trans. Circuits and Systems I. – 2009. – Vol. 56. – № 9. – P. 1908–1920.
4. **Коротков, А.С.** Аналого-цифровые преобразователи на основе дельта-сигма модуляторов [Текст] / А.С. Коротков, М.В. Теленков // Зарубежная радиоэлектроника. – 2002. – № 12. – С. 53–72.
5. **Коротков, А.С.** Дельта-сигма модулятор с частотой дискретизации 50 МГц на основе 0,18 мкм КМОП технологии [Текст] / А.С. Коротков, М.М. Пилипко, Д.В. Морозов, Й. Хауэр // Микроэлектроника. – 2010. – Т. 39. – № 3. – С. 230–240.
6. **Коротков, А.С.** Схемотехника современных интегральных усилителей [Текст] / А.С. Коротков, Д.В. Морозов // Зарубежная радио-

электроника. – 1998. – № 6. – С. 41–75.

7. **Морозов, Д.В.** Дельта-сигма модулятор аналого-цифрового преобразователя с троичным кодированием данных [Текст] / Д.В. Морозов, М.М. Пилипко, А.С. Коротков // Микроэлектроника. – 2011. – Т. 40. – № 1. – С. 64–75.

8. **Ljuslin, C.** An integrated 16-channel CMOS time to digital converter [Text] / C. Ljuslin, J. Christiansen, A. Marchioro [et al.] // IEEE Trans. Nuclear Science. – 1994. – Vol. 41. – № 4. – Part 1–2. – P. 1104–1108.

9. **Балашов, Е.В.** Микро- и наноэлектронные аналоговые и цифровые устройства систем связи [Текст] / Е.В. Балашов, А.С. Коротков, Д.В. Морозов, М.М. Пилипко, И.М. Пятак, Р.И. Романов, И.А. Румянцев // Всерос. науч.-техн. конф. Проблемы разработки перспективных микро- и наноэлектронных систем (МЭС). Сб. трудов. – 2012. – № 1. – С. 9–20.

10. **Пятак, И.М.** Схемное решение блока компараторов двухразрядной ячейки конвейерного АЦП [Текст] / И.М. Пятак, Д.В. Морозов, М.М. Пилипко // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. – СПб.: Изд-во Политехн. ун-та, 2011. – № 5 (133). – С. 49–52.

REFERENCES

1. **Yang H.Y., Sarpeshkar R.** A time-based energy-efficient analog-to-digital converter / IEEE J. Solid-State Circuits. – 2005. – Vol. 40. – № 8. – P. 1590–1601.
2. **Wismar U., Wisland D., Andreani P.** 0.2 V 7.5 uW 20 kHz sigma-delta modulator with 69 dB SNR

in 90 nm CMOS / Proc. Eur. Solid-State Circuits Conf. – 2007. – P. 206–209.

3. **Taillefer C.S., Roberts G.W.** Delta-Sigma A/D Conversion Via Time-Mode Signal Processing / IEEE Trans. Circuits and Systems I. – 2009. – Vol. 56. – № 9. – P. 1908–1920.

4. **Korotkov A.S., Telenkov M.V.** Analogo-tsifrovye preobrazovateli na osnove del'ta-sigma moduliatorov / Zarubezhnaia radioelektronika. – 2002. – № 12. – S. 53–72. (rus)

5. **Korotkov A.S., Pilipko M.M., Morozov D.V., Khauer I.** Del'ta-sigma moduliator s chastotoi diskretizatsii 50 MGts na osnove 0,18 mkm KMOP tekhnologii / Mikroelektronika. – 2010. – T. 39. – № 3. – S. 230–240. (rus)

6. **Korotkov A.S., Morozov D.V.** Skhemotekhnika sovremennykh integral'nykh usilitelei / Zarubezhnaia radioelektronika. – 1998. – № 6. – S. 41–75. (rus)

7. **Morozov D.V., Pilipko M.M., Korotkov A.S.** Del'ta-sigma moduliator analogo-tsifrovogo preobrazovatelia s troichnym kodirovaniem dannykh / Mikroelektronika. – 2011. – T. 40. – № 1. – S. 64–75. (rus)

8. **Ljuslin C., Christiansen J., Marchioro A.,**

Klingsheim O. An integrated 16-channel CMOS time to digital converter / IEEE Trans. Nuclear Science. – 1994. – Vol. 41. – № 4. – Part 1–2. – P. 1104–1108.

9. **Balashov E.V., Korotkov A.S., Morozov D.V., Pilipko M.M., Piatak I.M., Romanov R.I., Rumiantsev I.A.** Mikro- i nanoelektronnye analogovye i tsifrovye ustroistva sistem svyazi / Vseros. nauch.-tekhn. konf. Problemy razrabotki perspektivnykh mikro- i nanoelektronnykh sistem (MES). Sb. tr. – 2012. – № 1. – S. 9–20. (rus)

10. **Piatak I.M., Morozov D.V., Pilipko M.M.** Skhemnoe reshenie bloka komparatorov dvukhradnoi iacheiki konveiernogo ATsP / Nauchno-tekhnicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Upravlenie. – St.-Petersburg: Izd-vo Politehn. un-ta, 2011. – № 5 (133). – S. 49–52. (rus)

ПИЛИПКО Михаил Михайлович – доцент кафедры интегральной электроники Санкт-Петербургского государственного политехнического университета, кандидат технических наук.

195251, Россия, Санкт-Петербург, ул. Политехническая, д. 29.

E-mail: m_m_pilipko@rambler.ru

PILIPKO, Mikhail M. St. Petersburg State Polytechnical University.

195251, Politekhnikeskaya Str. 29, St.-Petersburg, Russia.

E-mail: m_m_pilipko@rambler.ru

ПРИГЛАШАЕМ НА КОНФЕРЕНЦИЮ



КОСТРОМА, КГТУ, 10-12 ОКТЯБРЯ 2013 ИНСТРУМЕНТЫ И МЕТОДЫ АНАЛИЗА ПРОГРАММ

Приглашаем принять участие в Международной научно-практической конференции «Tools & Methods of Program Analysis» («Инструменты и методы анализа программ, ТМРА-2013»), посвящённой одному из наиболее актуальных и важных направлений программной инженерии - анализу качества программного обеспечения.

Конференция пройдёт 10-12 октября 2013 года в Костромском государственном технологическом университете (КГТУ) и соберёт студентов, аспирантов и молодых учёных, специализирующихся в области разработки программного обеспечения и заинтересованных в независимой экспертизе своих исследований, а также признанных учёных, готовых поделиться своими инновационными разработками. Организаторы ожидают интересных практических докладов и от ведущих фирм-разработчиков программного обеспечения.

К участию в программе конференции принимаются заявки в форме конкурсных докладов для отбора на рецензирование независимыми экспертами конференции «ТМРА-2013», в качестве которых выступают российские и зарубежные учёные в области разработки и анализа ПО.

Приглашенные докладчики

- Карпов Юрий Глебович, СПбГПУ
- Петренко Александр Константинович, ИСП РАН
- Захаров Владимир Анатольевич, МПКБ МГУ
- Ицкxон Владимир Михайлович, СПбГПУ

Организаторы мероприятия:

- Факультет автоматизированных систем и технологий Костромского государственного технологического университета (ФАСТ КГТУ)
- Институт проблем информатики Российской академии наук (ИПИ РАН)
- Кафедра компьютерных систем и программных технологий института информационных технологий и управления Санкт-Петербургского государственного политехнического университета (КСПТ ИИТУ СПбГПУ)
- Компания «Инновационные Трейдинговые Системы» (ООО «ИТС»)

Спонсор конференции компания «Exactpro Systems LLC»

Регистрация на конференцию уже открыта! Количество мест ограничено.

По вопросам участия обращаться по телефону +7 495 640 24 60 или по электронной почте:


tmpa-org@exactprosystems.com

Подробности на официальном сайте конференции www.tmpaconf.org и в социальных сетях:



Организаторы и партнеры:





Управление в социальных и экономических системах

УДК 004.4'22

О.Х. Бостонов, Г.Н. Зверев

СЕМИОТИЧЕСКАЯ ИНТЕГРАЦИЯ И УПРАВЛЕНИЕ БИЗНЕС-ПРОЦЕССАМИ ОРГАНИЗАЦИЙ НЕФТЯНОЙ ОТРАСЛИ

O.H. Bostonov, G.N. Zverev

SEMIOTIC INTEGRATION AND CONTROL OF BUSINESS PROCESSES IN PETROLEUM ORGANIZATION

Проведен анализ современного состояния проблемы интеграции информационных систем. Проблему интеграции и повышения эффективности контроля и управления бизнес-процессами предложено решать путем построения адекватной иерархической модели предприятия и его подразделений в виде сети операционных производств, оценки статистических параметров информационных потоков и решаемых задач.

ИНТЕГРАЦИЯ. ИНФОРМАЦИОННЫЕ СИСТЕМЫ. СЕМИОТИКА. ЦЕЛОСТНОСТЬ.

The paper provides analysis of the current state of the issue regarding integration of information systems. The issue of integration and efficiency of the control and management in business processes is proposed to be solved by constructing an adequate hierarchical model of the company and its subsidiaries in the form of a network operating productions, estimation of statistical parameters of information flows and tasks.

INTEGRATION. INFORMATION SYSTEMS. SEMIOTICS. CONSISTENCY.

Типичный научно-исследовательский проектный институт (НИПИ) нефтяной компании оснащен многими информационными системами различного назначения, сотнями приложений унаследованных, коммерческих, собственных разработок, десятками web-сайтов различных подразделений института. Так, в отделе создания проектов разработки нефтяных месторождений функционирует не менее трех-четырех информационных систем, взаимозависимых по информации и управлению, решающих задачи геологии и гидродинамики залежей, вскрытия пластов, закачки флюидов, экономических показателей добычи и т. д.

При одновременной работе специалистов разного профиля над проектом разработки залежи нефти неизбежно возникают

формальные и содержательные несоответствия при реализации функций источников и потребителей информации, ее форматов и семантики (смыслов данных и действий). Имеющиеся средства выявления и устранения синтаксических и семантических коллизий не позволяют своевременно корректировать бизнес-процессы, что серьезно снижает эффективность работы НИПИ.

Выходом из данной ситуации является разработка средств интеграции разнородных систем, обеспечивающих достаточную автономность, надежность, эффективность их функционирования, единство информационного пространства проектов и бизнес-процессов, целостность используемых данных. Под целостностью данных понимается их готовность к использованию в динамике

процесса проектирования разработки месторождения, непротиворечивость, актуальность и достоверность информации (достаточная адекватность реалиям).

Цель данной статьи – анализ текущего состояния проблемы интеграции информационных технологий в нефтяной отрасли, разработка моделей управления бизнес-процессами и методов создания программного обеспечения средств интеграции организационно-технических систем, обеспечивающих глобальную целостность данных.

Анализ состояния проблемы интеграции

Решению разнообразных задач интеграции существующих и создаваемых информационных систем и технологий посвящено большое количество книг и статей, из которых значительную долю занимают описания паттернов (шаблонов) – формализованного опыта создания проектных решений в архитектуре программных приложений, объектно-ориентированном проектировании, в интеграционных решениях и технологии обмена сообщениями (см. например [1, 2]).

Для согласованного и успешного функционирования интегрированных систем необходимо, во-первых, наладить эффективные связи между объединяемыми компонентами источников и/или приемников информации и, во-вторых, создать распределенное управление информационными процессами, сочетающее вполне очерченную автономность, независимость действий для асинхронного взаимодействия и подчиненность общим задачам, целям и критериям. Подходящие модели управления будут рассмотрены далее, а здесь остановимся на связях интегрируемых компонентов объединенной системы.

В общем случае различают следующие виды интеграции:

- по форматам данных;
- по функциям и процессам;
- по управлению бизнес-процессами;
- по каналам связи;
- по языкам и семантике сообщений.

В процессах создания проекта разработки залежи нефти участвуют специалисты А,

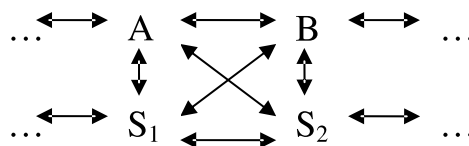


Рис. 1. Типы каналов связей по информации и управлению

В, ... одной, а чаще нескольких предметных областей, а также интегрируемые информационные системы S_1, S_2, \dots , использующие в общем случае различные терминосистемы, языки программирования, платформы, форматы данных, языки общения.

В подобной структуре возникают три типа интерфейсов, каналов связей по информации и управлению (см. рис. 1): 1) $A \leftrightarrow B$ – общение специалистов А, В, ... на естественном языке и языках предметных областей в диалоге, в пересылаемых сообщениях и т. п.; 2) $A \leftrightarrow S$ – человеко-машинный интерфейс сообщений на искусственном информационном языке; 3) $S_1 \leftrightarrow S_2$ – обмен сообщениями, командами на согласованном языке программно-аппаратных коммуникаций.

При взаимодействии систем в источниках и приемниках информации, в каналах связи выполняется трансляция синтаксиса и семантики сообщений на языке общения (общий для них язык). Например, при общении специалистов $A \leftrightarrow B$ смыслы сообщения с внутреннего языка понимания конкретного субъекта переводятся и кодируются в смыслы языка общения. Модели таких преобразований относятся к представлениям понятийных знаний в базисах теоретической информатики, семиотики и лингвистики [3] и служат основой семиотической (синтаксической и семантической) интеграции, согласования смыслов терминов, форматов данных, команд, приложений, в которых управление семантикой сообщений осуществляет естественный интеллект специалиста. В работе [2] описаны наиболее распространенные способы синтаксической интеграции: передача файла, общая база данных, удаленный вызов процедуры, обмен данными. Интеграция систем и бизнес-

процессов существенно повышает эффективность функционирования предприятия, однако серьезно усложняет процессы синхронизации, контроля и управления бизнес-процессами предприятия.

Проектирование и принятие интеграционных решений

Основная цель интеграции – обеспечение связей по информации и управлению между автономными либо зависимыми информационными системами и программными приложениями различного происхождения, реализованными на различных платформах, языках, географически распределенных, не имеющих самостоятельных средств интеграции и административно независимых от предприятия – в нашем случае, НИПИ.

Проектирование интеграции начинается с формирования класса бизнес-процессов, выполняемых системой, ее состава, структуры, регламента функционирования. В рассматриваемом в рамках статьи отделе проектирования разработки нефтяных месторождений существуют бизнес-процессы, связанные с созданием геологической и гидродинамической модели месторождения, согласованием и оформлением проекта разработки месторождения.

Далее осуществляется формализация системных соглашений, применяемых стандартов и спецификаций данных и функций, строятся канонические модели синтаксиса и семантики бизнес-процессов: глоссарий предметной области, онтологические модели и т. д. Проектные решения общих, перечисленных выше, и локальных задач интеграции принимаются с учетом факторов, влияющих на функционирование создаваемой системы в заданном классе бизнес-процессов, и критериев принятия интеграционных решений [4, 5]. Перечислим основные влияющие факторы на процесс интеграции и бизнес-процессы: несоответствия между программными приложениями: языки программирования, платформы, форматы данных, обозначения и их семантика, рассогласование бизнес-процессов во времени, необходимость синхронизации и асинхронного взаимодействия, ненадеж-

ность сети и низкая скорость передачи данных, ограниченный контроль бизнес-процессов.

Различают частные и общие (сводные, интегральные) критерии принятия интеграционных решений. Частные критерии решений: обеспечение надежной прямой и обратной связи между интегрируемыми приложениями, источником и приемником сообщений, требуемая оперативность, своевременность доставки данных и принятия решений, достаточный контроль достоверности передачи сообщений по каналам связи, повторная передача при сбоях, информирование приемника и/или источника о состоянии сообщений в канале связи, минимальное дублирование данных и функций в подсистемах и приложениях.

Эти и другие частные критерии предъявляют противоречивые требования. Компромиссные решения достигаются по сводным (интегральным) критериям: минимальные затраты ресурсов на интеграцию подсистем, внесение изменений в приложения, установление связей, создание сопрягающих компонентов, минимальные затраты на функционирование интегрированной системы, поддержка гибкости и управляемости при неизбежных изменениях ее состава, структуры и выполняемых функций, оптимальный компромисс между «свободой действий» автономных подсистем в асинхронном режиме и «подчинением» функций в синхронном бизнес-процессе и достижении общей цели.

Модель распределенного управления интегрированной системой

Структура НИПИ, функции его подразделений и отдельных сотрудников в основном соответствуют постоянно меняющимся условиям и задачам проектирования разработки месторождений и реализации бизнес-процессов. Каждое из подразделений оснащено необходимыми информационными системами и программными приложениями, которые должны обеспечивать успешное выполнение текущих задач. Поэтому управление интегрированной информационной системой при реализации бизнес-процессов целиком подчиняется управляю-

щим командам руководства НИПИ и его служб.

Достаточно адекватной и результативной моделью иерархического распределенного управления организационно-техническими системами и бизнес-процессами в них является сеть операционных продукций [3, 6]. Формализм операционных продукций позволяет в наиболее общей форме описать иерархическую структуру распределенного управления процессами интегрированных систем, статику и динамику, алгоритмы и мультиалгоритмы их функционирования в конструктивно-процедурной и дескриптивно-декларативной семантике, свободно изменять аспекты и степени подробности описаний человеко-машинных систем.

Термины «продукция» и «система продукций» введены Эмилем Постом (1943 г.) в связи с построением логических и символьных исчислений, уточнением понятия математического алгоритма, машины Тьюринга, машины Поста, при этом под системой продукций Пост понимал множество допустимых подстановок $\{p \rightarrow t\}$ букв в цепочки символов в среде формальной грамматики. Сами подстановки или правила символьных (синтаксических) преобразований — *логические* или *синтаксические продукции* — интерпретируются как шаги генерации синтаксически правильного текста либо шаги логического вывода.

В информатике продукционные модели получили распространение и дальнейшее развитие в интеллектуальных и экспертных системах. В них знания эксперта представлены множеством подстановок, продукционных правил — *экспертных продукций* с дескриптивной (логической) либо конструктивной (процедурной) семантикой. Логическое правило продукции $a \rightarrow b$ описывает логическую зависимость знаний: «если a истинно, то b тоже истинно, если b ложно, то a также ложно». Конструктивное знание — продукционное правило $a \rightarrow f$ определяет императивную связь знаний и умений, возможность или необходимость действия, выполнения функции f при истинности условия a . Экспертные продукции собираются в упорядоченные или неупорядоченные классы, которые экспертная система после наступления очередного события последовательно просматривает и выполняет продукционное правило, если его условие a истинно.

В работе [6] Д.А. Поспелов существенно расширил понятие продукции и ввел модели внешнего управления системой и внутреннего управления продукцией. Последующие обобщения продукционных моделей, сохраняя структуру продукции, предложенную в этой работе, были выполнены в статье [7] применительно к моделированию организационно-технических систем и бизнес-процессов и затем в книге [3].

Каждому уровню иерархии организационно-технической системы и ее управляющему компоненту сопоставляется операционная продукция, а системе в целом и ее внешнему окружению — *продукционная сеть* — совокупность взаимосвязанных операционных продукций, описывающих реализацию основных внутренних и внешних функций системы и *распределенное иерархическое* управление внешними и внутренними бизнес-процессами.

В продукционной сети и вмещающей ее продуктивной среде происходит обмен материальными и информационными объектами, ресурсами, заданиями, данными, исполняемыми функциями (компетенциями) для достижения локальных и глобальных целей с наилучшими показателями и критерийными оценками. Таким образом, структура предприятия и его управляющие функции модельно отображаются в продукционную сеть.

Операционная продукция PO (альтернативные термины: продукционный объект, продукционный субъект, продукционный агент) в соответствии с работой [6] есть пятерка объектов: $PO = (NP, SP, P_{in}, F, P_{out})$, связи между которыми определяют структуру операционной продукции и ее функциональную семантику. Первый элемент продукции есть уникальное *имя* продукции NP , ее идентификатор или порядковый номер в продукционной системе, обеспечивающий внешнее обращение к продукции, ее активацию, передачу данных и управляющих команд.

Сфера деятельности SP, применимости продукции к объектам, состояниям и событиям есть описатель области действий, компетенций, степеней свободы принятия решений и их исполнений, прав и обязанностей производственного субъекта. В сети продукции сферы их деятельности разграничивают полномочия и образуют иерархии разбиений и покрытий, древесные и плексные классификации действий и проблемных ситуаций [3].

Основные функции производственного объекта *PO* выполняет *ядро продукции F*, которое осуществляет последовательные или параллельные действия, материально-информационные операции, может иметь произвольную иерархическую структуру и содержать в своем составе внутренние продукции. Ядро продукции начинает работу после его активации *входным предпроцессором* P_{in} , главная функция которого – вычислить *предусловие* активации ядра по заданному предикату и если оно принимает значение «истина», то P_{in} передает управление ядру *F*, если же *предусловие* не выполняется, то ядро остается в пассивном состоянии.

Дополнительные функции предпроцессора включают обеспечение информационной связи с управляющими продуктами с учетом субординаций, а также сбор информации в сфере деятельности продукции *SP*, необходимой для вычисления *предусловия*, оценка ее качества, например, заключение договора о проведении работ ядром *F* и т. п. Активированная внешним субъектом либо своим предпроцессором продукция становится производственным агентом, отслеживающим информацию в своей сфере деятельности и выполняющим необходимые действия в производственной среде.

По завершению работы активированное ядро *F* передает управление *выходному постпроцессору* P_{out} продукции, который вычисляет признаки *постусловия* успешности/неуспешности работы ядра, оценивает качество его действий и результатов, если они получены, анализирует причины неудач, отправляет сигналы о невыполнении внешним управляющим продукциям либо повторно активирует ядро.

Регламент функционирования взаимосвязанных операционных продукций определяется общесистемными требованиями к производственной сети. Так, предпроцессоры и постпроцессоры продукций могут выполнять функции внутреннего контроля входных и выходных объектов продукции либо эти функции выполняют внешние продукции, при этом контроль может осуществляться только по входу, только по выходу, либо по входу и выходу продукций и их процессоров. Другое общесистемное требование относится к режиму активации продукции. Исходное состояние продукции – пассивное, для перевода в активное состояние необходимо определить режим работы предпроцессора: непрерывное отслеживание сферы *SP* либо по расписанию, случайной выборке ее состояний во времени и т. д.

Еще одно системное требование относится к согласованию форм и семантики параметров и моделей, которыми обмениваются продукции. Прежде всего это относится к логическим формам *предусловий* и *постусловий* продукций, заданных в шкалах классической логики {да, нет}, {истина, ложь}, в шкалах трилогики и тетралогики для неопределенных и противоречивых ситуаций {да, нет, не знаю, абсурд}, в шкале частотной логики или субъективных вероятностей [3]. Продукции в сети представляются полным составом, либо некоторые компоненты могут отсутствовать, выноситься при совпадении для всего класса однотипных продукций в описатель класса, скажем, если сферы действий у них одни и те же. Если продукция выполняет чисто контрольные функции пред- и постпроцессором, то ядро *F* в продукции отсутствует. Если контроль осуществляется только по входу, то постпроцессор может оказаться излишним и т. д.

При построении производственных сетей существенную роль играет принятая в системных требованиях классификация операционных продукций в соответствии с декомпозицией целей, выполняемых функций, допустимых решений возникающих внешних и внутренних задач производственной сети, функционирующей в априори определенной продуктивной среде. Прежде

всего выделяют *головные* (корневые) продукции, обеспечивающие целеполагание и высшие уровни стратегии управления, ниже которых находятся *управляющие* и *управляемые*, исполнительные продукции, которые также могут осуществлять функции управления нижестоящими продуктами, продукции *планирования, проектирования, прогнозирования* будущих состояний, событий, действий и их результатов, *информаторы* других продукции о текущем состоянии, происшедших и ожидаемых значимых событиях в системе и вне ее, *диспетчеры*, распределяющие материальные и информационные транзакты (потокные объекты) и выполняемые функции, обеспечивающие синхронизацию, согласование, координацию, коллаборацию целей и действий, *адресаторы, маршрутизаторы* входных, внутренних, выходных транзактов системы, *контролирующие* продукции, которые оценивают качество выполнения работ, коррекцию результатов и выполняемых функций, защиту системы от критических ситуаций.

Особый класс в продукционных сетях составляют *метапродукции* — это продукция, которые порождают, удаляют, изменяют другие продукции, а также связи в сети продукции. Их предназначение — реорганизация продукционной сети, порождение новых, недостающих продукции для достижения поставленных целей, исключение неэффективных объектов либо их модификацию. В абстрактном представлении метапродукции выполняют вполне определенные операции над продуктами, их генерацию и аннуляцию, разбиение на более мелкие функции при специализации либо укрупнение и объединение функций, делегирование полномочий другим продукциям, изменение их сфер действий SP , ядер F , пред- и постпроцессоров P_{in} , P_{out} , их связей с другими продуктами.

Метапродукции есть средство описания процессов в системах с переменным составом, структурой, функциями, изменяемыми горизонтальными и вертикальными, свободными и обязательными связями по информации, по ресурсам, по управлению. Метапродукции в качестве внутренних модулей могут содержать продукции проекти-

рования, планирования, диагноза, прогноза и упреждения нежелательных последствий. В этот класс включают *рекурсивные* метапродукции, порождающие себе подобные продукции и однородные иерархии управляющих и исполнительных структур.

Построив сеть операционных функций, описывающую законы и правила функционирования предприятия при реализации заданного класса бизнес-процессов, и оценив фактические либо прогнозные частотные (вероятностные) распределения их параметров, можно получить количественные и качественные характеристики интеграционных решений, выработать мероприятия по их оптимизации.

Метод полисемической декомпозиции. Программная реализация

Синтаксическая интеграция, рассмотренная выше, в реальных процессах должна дополняться семантической интеграцией. С этой целью в работах [5, 8] предложен метод полисемической декомпозиции, который включает четыре этапа проектирования интеграционного решения:

- понятия предметной области формализуются на метаязыке проектирования;
- понятия группируются во множества и образуют прикладные языки проектной области;
- семантика этих языков формализуется в виде правил соответствия понятий, их проверки и преобразования;
- при изменении предложений на прикладных языках семантические процессоры автоматически выполняют их проверку, согласование и преобразование понятий.

Каждая из объединяемых информационных систем формализуется в виде модуля. К каждому из модулей относится набор понятий определенной предметной области, семантика которых представлена подходящими онтологиями [4], указываются связи между понятиями каждого из модулей с понятиями из других модулей; определяются правила автоматического преобразования и согласования понятий.

Реализованное интеграционное решение основано на обмене сообщениями, поскольку данный способ интеграции позволяет обеспечить быструю разработку

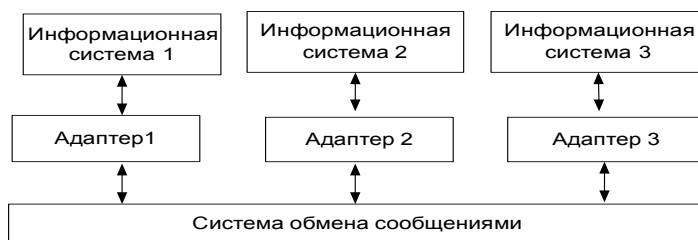


Рис. 2. Архитектура промежуточного программного обеспечения

программного обеспечения и необходимую гибкость при изменениях бизнес-процессов предприятия при наличии необходимого набора адаптеров (рис. 2). Адаптер представляет собой посредник между данной информационной системой и остальными информационными системами. В его обязанности входит реагирование на события, связанные с изменениями, происходящими с выделенными сущностями, и передача сообщения о случившихся изменениях остальным адаптерам, подписанным на данный вид сообщений.

Преобразование данных, основанное на заранее определенных правилах, на уровне системы S2–S2 происходит внутри адаптера. В частности, для систем геологической информационной системы и системы хранения данных добычи используются соответственно адаптер для текстового источника и ORACLE-адаптер.

Для настройки интеграционного решения используется система автоматизации проектирования, позволяющая определить модули, входящие в их состав понятия и связи между ними, а также правила преобразования понятий. Для облегчения создания интеграционного решения предусмотрена специальная утилита «Генератор программного кода», позволяющая сгенерировать на основе полученной из системы автоматизации проектирования конфигурационной схемы часть программного кода адаптеров.

При необходимости внесения изменений в текущее интеграционное решение, например, добавления новой информации

онной системы, можно воспользоваться сохраненной схемой взаимосвязей, добавить новый объект типа «модуль», определить понятия и связи для новой системы и вновь воспользоваться утилитой «Генератор программного кода».

Проведен анализ современного состояния проблемы интеграции информационных систем, рассмотрены различные виды интеграции, возможности использования паттернов (шаблонов проектирования) для реализации оптимального или приемлемого интеграционного решения. Проблему интеграции и повышения эффективности контроля и управления бизнес-процессами предложено решать путем построения адекватной иерархической модели предприятия и его подразделений в виде сети операционных продукций, в которой каждому исполнителю, группе специалистов, подразделениям, организации в целом соответствуют исполнительные, контролирующие и управляющие производственные объекты, которые реализуют заданные функции, получают оценки статистических параметров информационных потоков и решаемых задач.

Согласование смыслов при обмене сообщениями и семиотическая интеграция информационных систем выполнена на основе метода полисемической декомпозиции. Реализованное интеграционное решение основано на обмене сообщениями, оно обеспечивает быструю разработку и необходимую гибкость программных приложений при изменениях бизнес-процессов предприятия.

СПИСОК ЛИТЕРАТУРЫ

1. Фаулер, М. Архитектура корпоративных программных приложений [Текст] / М. Фаулер. – М.: ИД «Вильямс», 2004. – 554 с.
2. Хоп, Г. Шаблоны интеграции корпоратив-

ных приложений [Текст] / Г. Хоп, Б. Вульф. – М.: ИД «Вильямс», 2007. – 672 с.

3. Зверев, Г.Н. Теоретическая информатика и ее основания [Текст] / Г.Н. Зверев; В 2-х т.

– М.: Физматлит, 2007. – Т. 1. – 592 с.; 2009. – Т. 2. – 576 с.

4. **Галямов, А.Ф.** Интеграция и управление организационными системами с использованием онтологий [Текст] / А.Ф. Галямов, О.Х. Бостонов // Вестник ВГТУ. Проблемно-ориентированные системы управления. – 2012. – № 2. – С. 9–12.

5. **Бостонов, О.Х.** Средства автоматизации проектирования для интеграции информационных систем [Текст] / О.Х. Бостонов // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. – СПб.: Изд-во Политехн. ун-та, 2011. – № 1 (115). – С. 154–159.

6. **Поспелов, Д.А.** Продукционные модели

[Текст] / Д.А. Поспелов // Искусственный интеллект; В 3-х кн., Кн. 2. Модели и методы: Справочник. – М.: Радио и связь, 1990. – С. 49–56.

7. **Zverev, G.N.** The System of Operational Production for Modelling Technical-Organizational Systems and Business-process [Электронный ресурс] / G.N. Zverev, A.G. Turganov, R.U. Kulyev // Workshop on Computer Science and Information Technologies, CSIT-2005, Ufa, 2005. – P. 61–65.

8. **Тюрганов, А.Г.** Формализованные понятийные модели для ИТ-проектирования организационно-технических систем [Текст] / А.Г. Тюрганов // Сб. науч. трудов X национ. науч.-техн. конф. РАИИ с междунар. участием (КИИ-2006). – М.: Физматлит, 2006. – Т. 1. – С. 183–188.

REFERENCES

1. **Fauler M.** Arkhitektura korporativnykh programmnykh prilozhenii. – Moscow: ID «Vil'iams», 2004. – 554 s. (rus)

2. **Khop G., Vul'f B.** Shablony integratsii korporativnykh prilozhenii. – Moscow: ID «Vil'iams», 2007. – 672 s. (rus)

3. **Zverev G.N.** Teoreticheskaia informatika i ee osnovaniia; 2-kh t. – Moscow: Fizmatlit, 2007. – Т. 1 – 592 s.; 2009. Т.2. – 576 s. (rus)

4. **Galiyomov A.F., Bostonov O. Kh.** Integratsiia i upravlenie organizatsionnymi sistemami s ispol'zovaniem ontologii / Vestnik VGTU. Problemno-orientirovannye sistemy upravleniia. – 2012. – № 2. – С. 9–12. (rus)

5. **Bostonov O.Kh.** Sredstva avtomatizatsii proektirovaniia dlia integratsii informatsionnykh sistem / Nauchno-tekhicheskie vedomosti SPbGPU. Informatika. Telekommunikatsii. Uprav-

lenie. – St.-Petersburg: Izd-vo Politehn. un-ta, 2011. – № 1. – S. 154–159. (rus)

6. **Pospelov D.A.** Produktсионnye modeli / Iskustvennyi intellect; V 3-kh kn.; Kn. 2. Modeli i metody: Spravochnik. – Moscow: Radio i sviaz', 1990. – S. 49–56. (rus)

7. **Zverev G.N., Turganov A.G., Kulyev R.U.** The System of Operational Production for Modelling Technical-Organizational Systems and Business-process; Workshop on Computer Science and Information Technologies; CSIT-2005, Russia, Ufa. – P. 61–65. (rus)

8. **Tiurganov A.G.** Formalizovannye poniatiiyne modeli dlia IT-proektirovaniia organizatsionno-tekhicheskikh sistem // Sb. nauch. trudov natsional'noi nauch.-tekh. konf. RAIИ s mezhd. uchastiem KII-2006. – Moscow: Fizmatlit, 2006. – Т. 1. – S. 183–188. (rus)

БОСТОНОВ Оскар Хамзович – аспирант кафедры компьютерной математики Уфимского государственного авиационного университета.

450000, Россия, г. Уфа, ул. К. Маркса, д. 12.

E-mail: bostonov@yandex.ru

BOSTONOV, Oskar H. Ufa State Aviation Technical University.

450000, K. Marksa Str. 12, Ufa, Russia.

E-mail: bostonov@yandex.ru

ЗВЕРЕВ Геннадий Никифорович – профессор кафедры компьютерной математики Уфимского государственного авиационного университета, доктор технических наук.

450000, Россия, г. Уфа, ул. К. Маркса, д. 12.

ZVEREV, Gennady N. Ufa State Aviation Technical University.

450000, K. Marksa Str. 12, Ufa, Russia.



Высокопроизводительные вычисления

Для решения широкого круга задач современных фундаментальных и прикладных исследований в таких областях, как инфокоммуникационные системы, облачные вычисления, беспроводные сенсорные сети и многих других, требуются высокопроизводительные и надежные вычисления цифровой обработки сигналов, криптографические преобразования в реальном масштабе времени и с высокой степенью достоверности.

Все статьи, публикуемые в данном разделе, освещают различные подходы к использованию параллельных вычислений. Большинство статей написано по материалам докладов участников всероссийских конференций «Параллельная компьютерная алгебра» и «Проблемы математики и радиофизики в области информационной безопасности», поддержанных Российским фондом фундаментальных исследований, проектами 12-07-06055-г и ФЦП «Научно-педагогические кадры инновационной России на 2005–2013 гг.», государственный контракт № 02.741.12.22.13.

Представленные в разделе статьи будут интересны как специалистам в области высокопроизводительных вычислений, так и научным сотрудникам, занимающимся практическим применением высокопроизводительных структур.

*Заведующий кафедрой прикладной математики
и математического моделирования Института математики
и естественных наук Северо-Кавказского федерального университета,
Заслуженный деятель науки и техники РФ, доктор технических наук,
профессор Н.И. Червяков*

УДК 514.743.2

А.Н. Макоха

ИССЛЕДОВАНИЕ КАНОНИЧЕСКИХ ТРИВЕКТОРОВ ВОСЬМОГО РАНГА С ПОМОЩЬЮ ТЕОРИИ ГРАФОВ И ГРУПП ПОДСТАНОВОК

A.N. Makoha

RESEARCH OF CANONICAL TRIVECTORS OF EIGHTH GRADE BY MEANS OF THEORY OF THE GRAPHS AND GROUPS OF SUBSTITUTIONS

С помощью теории графов исследованы все типы канонических тривекторов восьмого ранга на предмет их однозначного представления. В случае их неоднозначности найдены группы подстановок, переводящие один тривектор в другой.

РАНГ ТРИВЕКТОРА. КАНОНИЧЕСКИЙ ВИД ТРИВЕКТОРА. ГИПЕРПЛОСКОСТЬ. ЦИКЛОМАТИЧЕСКОЕ ЧИСЛО ГРАФА. ГРУППА ПОДСТАНОВОК.

In the article using the graph theory investigated all types of canonical trivectors eighth grade for their unambiguous representation. In the event of ambiguities found substitutions groups trivectors transform one into the other.

GRADE TRIVECTORS. CANONICAL FORM TRIVECTORS. HYPERPLANE. CYCLOMATIC NUMBER OF A GRAPH. SUBSTITUTION GROUPS.

Классификацию тривекторов (кососимметрических тензоров третьей валентности) восьмого ранга независимо друг от друга провели Гуревич [1] и Лонго [15].

Классификация, проведенная Гуревичем, основана на системе арифметических инвариантов $(\rho_0\rho_1\rho_2; \sigma_1\sigma_2\sigma_3)$, где числа $\rho_0, \rho_1, \rho_2, \sigma_1, \sigma_2, \sigma_3$ означают последовательно ранг тривектора и ранги по одному из индексов специально введенных тензоров (комитантов).

Лонго провел классификацию соответствующих тривекторам линейных комплексов плоскостей, основываясь на понятиях особых точек, особых подпространств, ассоциированных с особыми точками, следами комплекса плоскостей и др. Классификация линейных комплексов плоскостей, проведенная Лонго, как оказалось, эквивалентна классификации тривекторов, проведенной Гуревичем. Согласно обеим классификациям имеется всего 23 попарно неэквивалентных типа тривекторов, ранг которых не превышает восьми, при этом 13 типов имеют ранг, равный восьми.

При построении геометрической конструкции линейных комплексов плоскостей, ассоциированных с каждым типом тривекторов, желательно иметь простые координатные симплексы, которые отличались бы друг от друга лишь транспозициями координатных гиперплоскостей. Предлагаемый нами метод позволяет найти такие преобразования координатных гиперплоскостей и, что более важно, найти все варианты канонических тривекторов, принадлежащих одному и тому же типу, и тем самым ликвидировать неоднозначность при их исследовании.

Постановка и описание задачи

Канонические виды всех 23 возможных типов тривекторов, ранг которых не превышает восьми, согласно [3] можно объединить в следующей формуле:

$$\begin{aligned}
 W = & \alpha e^1 \wedge e^2 \wedge e^3 + \\
 & + \beta e^5 \wedge e^6 \wedge e^7 + \gamma e^1 \wedge e^4 \wedge e^5 + \\
 & + \delta e^2 \wedge e^4 \wedge e^6 + \varepsilon e^3 \wedge e^4 \wedge e^7 +
 \end{aligned} \tag{1}$$

$$+\lambda e^2 \wedge e^7 \wedge e^8 + \mu e^3 \wedge e^6 \wedge e^8, \quad (1)$$

где \wedge – символ внешнего умножения в алгебре Грассмана; $e^i (i = \overline{1,8})$ – линейно независимые гиперплоскости аффинного или проективного пространства над алгебраически замкнутым полем характеристики 0; коэффициенты $\alpha, \beta, \gamma, \delta, \varepsilon, \lambda$ и μ принимают значения из множества $\{0,1\}$. Десять типов канонических тривекторов, ранг которых меньше восьми, получаются из равенства (1) при $\lambda = \mu = 0$. Остальные 13 типов тривекторов имеют ранг, равный восьми. Для них коэффициент $\mu = 1$. Для тривекторов категории B (для них арифметический инвариант $\sigma_3 = 0$, коэффициент $\lambda = 0$, а для тривекторов категории A (для них $\sigma_3 \neq 0$) $\lambda = 1$.

Разложение (1) обладает тем свойством, что каждая пара слагаемых в его правой части имеет не более чем одну общую гиперплоскость данного базиса.

К исследованию канонических тривекторов с помощью теории графов нас натолкнула работа [2], в которой находятся и исследуются условия применимости к каноническим тривекторам т. н. *основного правила* для разыскания весовых коэффициентов различных сочетаний индексов для тривекторов любого ранга. Следуя [2], каждое отличное от нуля слагаемое в правой части равенства (1) назовем *листом* соответствующего канонического тривектора, и для краткости будем обозначать его через W_k , где k принимает одно из значений $\alpha, \beta, \gamma, \delta, \varepsilon, \lambda, \mu$. Например, листы $e^1 \wedge e^4 \wedge e^5$ и $e^3 \wedge e^6 \wedge e^8$, входящие в каждый из 13 канонических тривекторов восьмого ранга, будут обозначаться соответственно через W_γ и W_μ , при этом $W_\gamma \cap W_\mu = \emptyset$. В общем случае каждому листу W_k отвечает сочетание T_k трех попарно различных индексов из множества $\{1, 2, \dots, r\}$, где r – ранг тривектора. При перестановке какой-либо пары индексов знак листа в силу косой симметрии меняется на противоположный. Для канонических тривекторов восьмого ранга минимальное число листов равно трем, максимальное – семи (см. [3, с. 71, 72]).

Поставим в соответствие каждому ка-

ноническому тривектору два неориентированных взаимно дополняющих друг друга графа G и \bar{G} следующим образом. Листам тривектора будут отвечать вершины графов G и \bar{G} . Пары листов, имеющих общую гиперплоскость, в основном графе G будет отвечать ребро, соединяющее соответствующие листам вершины. В дополнительном графе \bar{G} , наоборот, пары листов, не имеющих общей гиперплоскости, отвечает ребро, соединяющее вершины, соответствующие этим листам. Максимальное число ребер графа G $m = C_n^2 - 1$, где n – число его вершин. В силу того, что $W_\gamma \cap W_\mu = \emptyset$, случай $m = 0$ исключается. Полный граф канонического тривектора получается объединением $G \cup \bar{G}$ основного и дополнительного графов. Такое соответствие в общем случае не является функцией: нескольким каноническим тривекторам может отвечать один и тот же основной (а, следовательно, и дополнительный) граф с фиксированным числом вершин и ребер с точностью до изоморфизма, и существуют графы с заданным числом вершин и ребер, которым не соответствует ни один канонический тривектор [14]. Однако, как мы покажем ниже, все тривекторы, по крайней мере, восьмого ранга, интерпретируемые изоморфными графами, эквивалентны между собой, т. е. принадлежат одному и тому же типу. Этот факт означает, что существуют подстановки индексов, переводящие один канонический тривектор в другой.

Ранее в работах [4–13] и др., опираясь на инвариантные геометрические образы, найденные в работе [3], для всех тривекторов восьмого ранга нами были построены группы Ли линейных преобразований пространства тривектора, сохраняющих данный тривектор, его канонический вид и, как следствие, ассоциированный с ним линейный комплекс плоскостей. Эти группы мы назвали основными группами автоморфизмов соответствующих тривекторов.

Основной задачей настоящей работы является *нахождение всех вариантов канонических видов тривекторов восьмого ранга с помощью аппарата теории графов и групп подстановок.*

Изложение основных результатов

Обозначим для дальнейшего через n – число листов (вершин) дополнительного графа \bar{G} ; через m – число его ребер; x, y, z – индексы из множества $\{\alpha, \beta, \delta, \varepsilon, \lambda\}$. Цикломатическое число графа \bar{G} будем вычислять по формуле $v(\bar{G}) = m(\bar{G}) - n(\bar{G}) + p(\bar{G})$, где p – число его связных компонент. Через U обозначим группу всех подстановок, переводящих канонические тривекторы друг в друга; через e – тождественную подстановку.

Зафиксировав вершины, отвечающие листам W_γ и W_μ , и перебирая подходящие наборы оставшихся листов с различными реберными связями между ними, будем последовательно рассматривать всевозможные попарно неизоморфные дополнительные графы с 3, 4, 5, 6 и 7 вершинами.

1. $n = 3$. В этом случае возможные значения для числа m : $m = 0$, $m = 1$, $m = 2$ и $m = 3$. При $m > 3$ неизбежно будут повторяться ребра (т. е. в графе будут присутствовать параллельные ребра), что исключается в записи канонических тривекторов.

1.1 Конфигурация графа при $m = 0$ невозможна, поскольку листы W_γ и W_μ не имеют общих гиперплоскостей, и тогда добавленный лист W_x тоже не должен иметь общих гиперплоскостей с листами W_γ и W_μ . Из этого следует, что для канонического тривектора потребовалось бы девять линейно независимых гиперплоскостей, что невозможно (их максимальное число равно рангу тривектора и равно восьми).

1.2 Конфигурация графа при $m = 3$ также невозможна, т. к. в противном случае листы W_γ и W_μ имели бы общую гиперплоскость.

1.3 При $m = 1$ возможны два варианта: 1) добавляемый лист W_x имеет общую гиперплоскость с листом W_γ ; 2) добавляемый лист W_x имеет общую гиперплоскость с листом W_μ . При этом дополнительный граф имеет два ребра.

Конфигурация графа в первом варианте невозможна. Действительно, согласно этой конфигурации в качестве листа W_x должен быть взят один из оставшихся листов

$W_\alpha, W_\beta, W_\delta, W_\varepsilon, W_\lambda$, который с листом W_γ имеет общую гиперплоскость, а с листом W_μ не имеет общей гиперплоскости. Перебирая поочередно оставшиеся листы, мы видим, что

$$W_\alpha \cap W_\gamma = \{e^1\}, \quad W_\alpha \cap W_\mu = \{e^3\} \neq \emptyset;$$

$$W_\beta \cap W_\gamma = \{e^5\}, \quad W_\beta \cap W_\mu = \{e^6\} \neq \emptyset;$$

$$W_\delta \cap W_\gamma = \{e^4\}, \quad W_\delta \cap W_\mu = \{e^6\} \neq \emptyset;$$

$$W_\varepsilon \cap W_\gamma = \{e^4\}, \quad W_\varepsilon \cap W_\mu = \{e^3\} \neq \emptyset;$$

$$W_\lambda \cap W_\gamma \neq \emptyset, \quad W_\lambda \cap W_\mu = \{e^8\} \neq \emptyset.$$

Остается рассмотреть второй вариант. Поскольку в листах W_γ и W_μ участвуют гиперплоскости e^1, e^4, e^5, e^3, e^6 и e^7 , а в каноническом тривекторе должны участвовать все восемь гиперплоскостей, то в качестве листа W_x должен быть взят лист, содержащий гиперплоскости e^2 и e^7 . Среди таких листов имеется лишь один – это лист $W_\lambda = e^2 \wedge e^7 \wedge e^8$, который удовлетворяет конфигурации при $m = 1$ во втором варианте, т. к. $W_\gamma \cap W_\mu = \emptyset$, $W_\gamma \cap W_\lambda = \emptyset$, $W_\mu \cap W_\lambda = \emptyset$, $W_\lambda \cap W_\mu = \{e^8\} \neq \emptyset$. Таким образом, при $m = 1$ (или в дополнительном графе при $m = 2$) мы имеем единственный канонический тривектор $W = W_\lambda + W_\gamma + W_\mu$, представляющий тривектор типа (888; 411). Кстати, это единственный тривектор восьмого ранга, граф которого состоит из двух компонент связности (т. е. этот граф несвязный). Он также характеризуется еще и тем, что это единственный канонический тривектор, граф которого имеет цикломатическое число, равное нулю (как для основного, так и для дополнительного графа), т. е. основной граф является ациклическим.

2. $n = 4$. Возможное количество ребер соответствующего полного графа равно $m = C_4^2 = 6$. Случай полного графа невозможен (иначе все листы имели бы общую гиперплоскость; кроме того, как отмечалось выше, $W_\gamma \cap W_\mu = \emptyset$). Невозможен также случай всех изолированных вершин графа. Поэтому подлежат рассмотрению лишь значения $m = 1, m = 2, m = 3, m = 4$ и $m = 5$. Мы теперь должны найти два листа W_x и W_y так, чтобы выпол-



нялись соответствующие связи на графах. Покажем, что случай $m = 3$ невозможен. Поскольку лист W_μ связан ребрами со всеми остальными листьями, кроме листа W_γ , то для основных графов возможны лишь три конфигурации:

1) лист W_γ не связан с листом W_x , но связан с листом W_y . Тогда в качестве листа W_x можно взять лишь лист W_λ , а на место листа W_y остается четыре претендента: W_α , W_β , W_δ и W_ε . Но лист W_λ связан с каждым из этих листов, в то время как этих связей быть не должно;

2) лист W_γ не связан с листьями W_x и W_y . В этом случае на роль листа W_x снова претендует только лист W_λ . Но тогда лист W_γ окажется связанным с каждым из оставшихся четырех листов, которые претендуют на роль листа W_y , что также невозможно;

3) лист W_γ не связан ребром с листом W_y , но связан с листом W_x . Тогда на место листа W_y претендует лишь лист W_λ , который связан со всеми оставшимися листьями, что невозможно.

Остальные три конфигурации (их всего шесть) являются соответственно дополнительными графами к рассмотренным выше основным графам, поэтому на них переносятся те же рассуждения.

Перейдем к рассмотрению случаев $m = 1$ и $m = 2$. Двойственными к ним являются соответственно случаи $m = 5$ и $m = 4$. Поэтому достаточно рассмотреть первые два случая.

Пусть $m = 1$. На роль листов W_x и W_y не может претендовать лист W_λ , т. к. он не связан с листом W_γ . Тогда на роль листов W_x и W_y остается четыре претендента: W_α , W_β , W_δ и W_ε . При этом эти листы должны быть попарно связаны между собой. Но $W_\alpha \cap W_\beta = \emptyset$, поэтому пара W_α и W_β отпадает. Поскольку листы W_x и W_y входят в основной граф симметрично, то следует рассмотреть пары (W_α, W_δ) , $(W_\alpha, W_\varepsilon)$, (W_β, W_δ) , (W_β, W_ε) , $(W_\delta, W_\varepsilon)$.

Пара (W_α, W_δ) (невозможна, т. к. в ней отсутствует гиперплоскость e^7). Невозможна также пара (W_β, W_δ) , т. к. в ней отсутствует гиперплоскость e^2 . Таким образом, мы имеем три возможности:

$$W^{(1)} = W_\gamma + W_\mu + W_\alpha + W_\varepsilon,$$

$$W^{(2)} = W_\gamma + W_\mu + W_\beta + W_\delta,$$

$$W^{(3)} = W_\gamma + W_\mu + W_\delta + W_\varepsilon,$$

или в бескоординатной форме

$$W^{(1)} = e^1 \wedge e^4 \wedge e^5 + e^3 \wedge e^6 \wedge e^8 + e^1 \wedge e^2 \wedge e^3 + e^3 \wedge e^4 \wedge e^7,$$

$$W^{(2)} = e^1 \wedge e^4 \wedge e^5 + e^3 \wedge e^6 \wedge e^8 + e^5 \wedge e^6 \wedge e^7 + e^2 \wedge e^4 \wedge e^6,$$

$$W^{(3)} = e^1 \wedge e^4 \wedge e^5 + e^3 \wedge e^6 \wedge e^8 + e^2 \wedge e^4 \wedge e^6 + e^3 \wedge e^4 \wedge e^7.$$

Покажем, что графы тривекторов $W^{(1)}$, $W^{(2)}$ и $W^{(3)}$ изоморфны и, следовательно, полученные канонические тривекторы эквивалентны. Последнее утверждение означает, что любой из них может быть переведен в другой с помощью некоторого невырожденного преобразования координатных гиперплоскостей e^1, e^2, \dots, e^8 в семимерном проективном пространстве. Действительно, подстановка $\varphi = (134658)(27)$, где числами 1, 2, ..., 8 обозначены номера гиперплоскостей, отображает $W^{(1)} \rightarrow W^{(3)}$, $W^{(3)} \rightarrow W^{(2)}$, а подстановка φ^2 отображает $W^{(1)} \rightarrow W^{(2)}$. Обратная подстановка производит обратные преобразования. При необходимости у некоторых гиперплоскостей следует поменять знак. Все три тривектора $W^{(1)}$, $W^{(2)}$ и $W^{(3)}$ принадлежат типу $(863;100)$. Цикломатическое число соответствующих графов равно двум, что свидетельствует о наличии двух базисных циклов $(W_\gamma W_x W_y)$ и $(W_\mu W_x W_y)$ на этих графах.

Остается рассмотреть последний, третий случай при $n = 4$, $m = 2$. Здесь необходимо рассмотреть две конфигурации: 1) листы W_x и W_y связаны друг с другом, но не связаны с листьями W_γ и W_μ ; 2) листы W_x и W_y связаны друг с другом, связаны оба с листом W_μ , а лист W_y связан с листом W_γ (лист W_x не связан с листом W_γ).

1) По аналогии со случаем $m = 1$, найдем листы W_x и W_y , перебирая листы $W_\alpha, W_\beta, W_\delta$ и W_ε . Ясно, что на роль листов W_x и W_y могут претендовать только листы W_α и W_β , т. к. кроме листов W_γ и W_μ только эта пара

листов не имеет общих гиперплоскостей. Поэтому получаем единственный канонический тривектор $W = W_\alpha + W_\beta + W_\gamma + W_\mu$, принадлежащий типу (884;400). Цикломатическое число соответствующего графа равно единице, т. е. мы имеем один цикл, проходящий по одному разу через все вершины и все ребра. Получили *гамильтонов* граф, степень каждой вершины которого равна двум.

2) На роль листов W_x и W_y претендуют листы W_α , W_β , W_δ , W_ϵ и W_λ . Так как лист W_γ не должен иметь общей гиперплоскости с листом W_x , а с листом W_y должен иметь общую гиперплоскость и, кроме того, общую гиперплоскость должны иметь листы W_x и W_y , то в качестве листа W_x можно взять один из листов W_α , W_δ , и W_ϵ , а в качестве листа W_y — листы W_β и W_λ , при этом пару W_α и W_β следует исключить, как и пару W_β и W_ϵ . Таким образом, рассматриваемой конфигурации удовлетворяют четыре тривектора:

$$\begin{aligned} W^{(1)} &= W_\gamma + W_\mu + W_\alpha + W_\lambda, \\ W^{(2)} &= W_\gamma + W_\mu + W_\delta + W_\lambda, \\ W^{(3)} &= W_\gamma + W_\mu + W_\epsilon + W_\lambda, \\ W^{(4)} &= W_\gamma + W_\mu + W_\beta + W_\lambda. \end{aligned}$$

Покажем, что графы этих тривекторов изоморфны и, следовательно, соответствующие тривекторы эквивалентны. Для этого рассмотрим подстановки

$$\begin{aligned} \varphi_1 &= (1)(27)(36)(4)(5)(8), \\ \varphi_2 &= (14)(2)(36)(5)(7)(8), \\ \varphi_3 &= (14)(27)(3)(5)(6)(8), \\ \varphi_4 &= (1)(2)(36)(45)(7)(8), \\ \varphi_5 &= \varphi_1 \circ \varphi_4 = (1)(27)(3)(45)(6)(8). \end{aligned}$$

Тогда $\varphi_1^2 = \varphi_2^2 = \varphi_3^2 = \varphi_4^2 = \varphi_5^2 = e$ и $\varphi_1 : W^{(3)} \leftrightarrow W^{(4)}$; $\varphi_2 : W^{(1)} \leftrightarrow W^{(3)}$; $\varphi_3 : W^{(1)} \leftrightarrow W^{(4)}$; $\varphi_4 : W^{(2)} \leftrightarrow W^{(4)}$; $\varphi_5 : W^{(2)} \leftrightarrow W^{(3)}$.

Полагая



Рис. 1. Условное изображение дополнительного графа

$$\varphi_6 = \varphi_3 \circ \varphi_4 = (154)(27)(36)(8),$$

$$\varphi_7 = \varphi_4 \circ \varphi_3 = (145)(27)(36)(8),$$

получим $\varphi_6 : W^{(1)} \rightarrow W^{(2)}$, $\varphi_7 : W^{(2)} \rightarrow W^{(1)}$. Введем подстановки

$$\begin{aligned} \varphi_8 &= \varphi_1 \circ \varphi_6 = \varphi_2 \circ \varphi_4 = \\ &= \varphi_3 \circ \varphi_5 = (154)(2)(3)(6)(8)(7); \end{aligned}$$

$$\begin{aligned} \varphi_1 \circ \varphi_7 &= \varphi_4 \circ \varphi_2 = \varphi_5 \circ \varphi_3 = \\ &= \varphi_7 \circ \varphi_1 = \varphi_6^2 = \varphi_8^2 = (145)(2)(3)(6)(7)(8); \end{aligned}$$

$$\begin{aligned} \varphi_{10} &= \varphi_2 \circ \varphi_9 = \varphi_3 \circ \varphi_7 = \\ &= \varphi_4 \circ \varphi_8 = \varphi_5 \circ \varphi_6 = \varphi_6 \circ \varphi_3 = \\ &= \varphi_7 \circ \varphi_5 = \varphi_8 \circ \varphi_2 = (15)(2)(36)(4)(7)(8); \end{aligned}$$

$$\begin{aligned} \varphi_{11} &= \varphi_1 \circ \varphi_{10} = \varphi_2 \circ \varphi_7 = \varphi_3 \circ \varphi_9 = \\ &= \varphi_4 \circ \varphi_6 = \varphi_5 \circ \varphi_8 = \varphi_6 \circ \varphi_2 = \\ &= \varphi_7 \circ \varphi_4 = \varphi_8 \circ \varphi_3 = \varphi_9 \circ \varphi_5 = \\ &= \varphi_{10} \circ \varphi_1 = (15)(27)(3)(4)(6)(8). \end{aligned}$$

Простые вычисления показывают, что $\varphi_6 \circ \varphi_7 = \varphi_7 \circ \varphi_6 = \varphi_8 \circ \varphi_9 = \varphi_9 \circ \varphi_8 = \varphi_{10}^2 = \varphi_{11}^2 = e$. Легко проверить, что полученное множество из 12 подстановок с приведенным выше правилом умножения образует группу, которая является группой автоморфизмов рассматриваемой совокупности тривекторов. При этом подмножество H , состоящее из четырех подстановок $e, \varphi_1, \varphi_2, \varphi_3$, образует циклическую подгруппу четвертого порядка. Цикломатическое число графа рассматриваемого тривектора равно единице, поэтому этот граф имеет только один фундаментальный цикл $W_\delta W_\mu W_\lambda$ и одну висячую вершину W_γ . Все четыре канонических тривектора, как было показано выше, попарно эквивалентны между собой и принадлежат типу (888; 621).

3. $n = 5$. Максимально возможное число ребер графа $m = C_5^2 = 10$.

При этом, как и ранее, следует исключить случаи полного графа и всех изолированных вершин. Поэтому подлежат рассмотрению лишь значения $m = 1, 2, 3, 4, 5$, т. к. значения $m = 9, 8, 7$ являются двой-

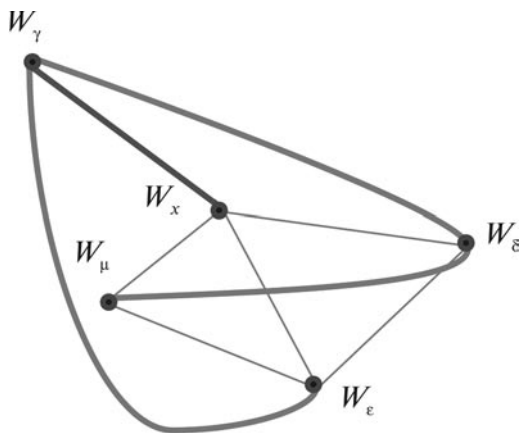


Рис. 2. Одно из остовных деревьев графа G

ственными соответственно значениям $m = 1, 2, 3$, а значение $m = 5$ является двойственным самому себе. Теперь мы должны найти три листа W_x, W_y и W_z так, чтобы выполнялись соответствующие связи на графах. На рис. 1 приведено условное изображение искомого дополнительного графа \bar{G} .

3.1 $m = 1$. Поскольку в каждый канонический тривектор должны входить листы W_γ и W_μ , то только эти листы могут быть соединены единственным ребром. Следовательно, на роль листов W_x, W_y и W_z могут претендовать лишь тройки $W_\alpha, W_\delta, W_\epsilon$ и $W_\delta, W_\epsilon, W_\beta$, так как из $C_5^3 = 10$ троек только две указанные тройки имеют по одной общей гиперплоскости как между собой, так и с листьями W_γ и W_μ . В остальных тройках найдется пара листов, не имеющих общей гиперплоскости. Таким образом, имеем два канонических тривектора:

$$W^{(1)} = W_\gamma + W_\mu + W_\alpha + W_\delta + W_\epsilon,$$

$$W^{(2)} = W_\gamma + W_\mu + W_\delta + W_\epsilon + W_\beta,$$

принадлежащие типу (874; 200). Остается найти подстановку φ , переводящую тривекторы $W^{(1)}$ и $W^{(2)}$ друг в друга. Такой подстановкой является подстановка $\varphi = (15)(27)(36)(4)(8)$, при этом $\varphi^2 = e$. Получаем циклическую группу (e, φ) второго порядка. Цикломатическое число соответствующих графов равно пяти, что свидетельствует о наличии пяти базисных циклов в основном графе G . Для того чтобы

убедиться в этом, построим одно из остовных деревьев в графе G (на рис. 2 остовное дерево выделено жирными ребрами). В качестве листа W_x можно взять лист W_α или лист W_β .

Добавляя к остовному дереву по одному из оставшихся ребер, получим один из наборов базисных циклов: $(W_\gamma W_x W_\delta)$, $(W_\gamma W_x W_\mu W_\delta)$, $(W_\gamma W_x W_\epsilon)$, $(W_\gamma W_\delta W_\epsilon)$.

3.2 $m = 2$ (двойственное значение $m = 8$). Как видно из рис. 3, в этом случае возможны лишь две конфигурации с двумя ребрами a и b . Рассмотрим каждую из них.

3.2 a – На роль листов W_x и W_y могут претендовать из оставшихся четырех листов из шести возможных пар только пять пар: (W_α, W_δ) , (W_α, W_ϵ) , (W_β, W_δ) , (W_β, W_ϵ) , (W_δ, W_ϵ) . Пара (W_α, W_β) не подходит, так как $W_\alpha \cap W_\beta = \emptyset$, и мы имели бы три ребра, что невозможно по условию выбора числа m . Таким образом, для рассматриваемого случая мы имеем пять канонических тривекторов:

$$W^{(1)} = W_\gamma + W_\mu + W_\lambda + W_\alpha + W_\delta,$$

$$W^{(2)} = W_\gamma + W_\mu + W_\lambda + W_\alpha + W_\epsilon,$$

$$W^{(3)} = W_\gamma + W_\mu + W_\lambda + W_\beta + W_\delta,$$

$$W^{(4)} = W_\gamma + W_\mu + W_\lambda + W_\beta + W_\epsilon,$$

$$W^{(5)} = W_\gamma + W_\mu + W_\lambda + W_\delta + W_\epsilon.$$

Рассмотрим подстановку $\varphi_1 = (1)(23)(4)(5)(67)(8)$. Очевидно, что $\varphi_1^{-1} = \varphi_1$. Эта подстановка отображает $W^{(1)} \leftrightarrow W^{(2)}$, $W^{(3)} \leftrightarrow W^{(4)}$, $W^{(5)} \leftrightarrow W^{(5)}$. В первых двух случаях надо поменять знаки у гиперплоскостей e^4 и e^8 на

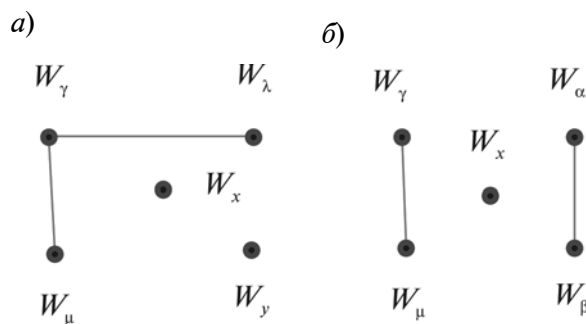


Рис. 3. Дополнительные графы для $n = 5, m = 2$

противоположный знак. Подстановка $\varphi_2 = (15)(27)(36)(4)(8)$ также совпадает со своей обратной и отображает $W^{(1)} \leftrightarrow W^{(4)}$, $W^{(2)} \leftrightarrow W^{(3)}$, $W^{(5)} \leftrightarrow W^{(5)}$. Подстановка $\varphi_3 = \varphi_1 \circ \varphi_2 = \varphi_2 \circ \varphi_1 = (15)(26)(37)(4)(8)$ отображает $W^{(1)} \leftrightarrow W^{(3)}$, $W^{(2)} \leftrightarrow W^{(4)}$, $W^{(5)} \leftrightarrow W^{(5)}$ ($\varphi_3^{-1} = \varphi_3$). Наконец, $\varphi_1 \circ \varphi_3 = \varphi_3 \circ \varphi_1 = \varphi_2$, $\varphi_2 \circ \varphi_3 = \varphi_3 \circ \varphi_2 = \varphi_1$. Таким образом, множество подстановок $\{e, \varphi_1, \varphi_2, \varphi_3\}$ образует коммутативную инволютивную циклическую группу четвертого порядка, которая отображает тривекторы $W^{(1)}$, $W^{(2)}$, $W^{(3)}$ и $W^{(4)}$ друг в друга, а каждая из подстановок этой группы отображает тривектор $W^{(5)}$ в себя, т. е. оставляет его неподвижным. Остается найти хотя бы одну подстановку, переводящую тривектор $W^{(5)}$ хотя бы в один из тривекторов $W^{(1)}$, $W^{(2)}$, $W^{(3)}$, $W^{(4)}$. Тем самым мы доказали бы изоморфизм соответствующих графов и эквивалентность всех пяти тривекторов, принадлежащих типу (888; 741). Сложность нахождения такой подстановки вызвана следующим обстоятельством. Запишем в виде таблицы ранг каждой гиперплоскости, входящей в тот или иной канонический тривектор.

Из табл. 1 видно, что гиперплоскость e^4 входит три раза только в тривектор $W^{(5)}$. В этом и состоит трудность нахождения подстановки, переводящей тривектор в какой-либо другой тривектор. Из этого факта можно заключить, что тривектор $W^{(5)}$ не эквивалентен ни одному из тривекторов $W^{(1)}$, $W^{(2)}$, $W^{(3)}$, $W^{(4)}$. Следовательно, граф тривектора $W^{(5)}$ не изоморфен ни одному из графов, соответствующих тривекторам $W^{(1)}$, $W^{(2)}$, $W^{(3)}$, $W^{(4)}$, хотя все графы имеют одно и то же цикломатическое число, равное четырем. Построив остов-

ное дерево для графа, изображенного на рис. 3 а, получим четыре фундаментальных цикла: $(W_\gamma W_x W_y)$, $(W_\gamma W_x W_\lambda W_y)$, $(W_\gamma W_y)$ и $(W_\gamma W_x W_\lambda W_\mu W_y)$.

Вычисляя арифметические инварианты σ_1 , σ_2 и σ_3 для тривектора $W^{(5)}$, убеждаемся, что этот тривектор принадлежит типу (888; 822).

3.2 б – Очевидно, что на роль листа W_x могут претендовать лишь листы W_δ и W_ε (лист W_λ не может претендовать на роль листа W_x , т. к. тогда появилось бы еще одно ребро). Поэтому здесь имеем два канонических тривектора:

$$W^{(1)} = W_\gamma + W_\mu + W_\alpha + W_\beta + W_\delta,$$

$$W^{(2)} = W_\gamma + W_\mu + W_\beta + W_\alpha + W_\varepsilon,$$

принадлежащие типу (886; 410).

Инволютивная подстановка $\varphi = (15)(27)(36)(4)(8)$ переводит тривекторы $W^{(1)}$ и $W^{(2)}$ друг в друга. Цикломатическое число для соответствующего графа $\nu(G) = 4$, что свидетельствует о наличии четырех фундаментальных циклов в этом графе: $(W_\alpha W_\mu W_\beta W_\gamma)$, $(W_\alpha W_\gamma W_\beta W_x)$, $(W_\alpha W_\gamma W_x)$ и $(W_\alpha W_x W_\mu)$.

3.3 $m = 3$ (двойственное значение $m = 7$). Здесь имеется единственная возможность для канонического тривектора

$$W = W_\gamma + W_\mu + W_\alpha + W_\beta + W_\lambda,$$

принадлежащего типу (888; 873). Цикломатическое число соответствующего основного графа $\nu(G) = 3$, что свидетельствует о наличии трех фундаментальных циклов в этом графе: $(W_\alpha W_\lambda W_\mu)$, $(W_\alpha W_\gamma W_\beta W_\lambda)$, $(W_\alpha W_\gamma W_\beta W_\mu)$.

3.3 Значения $m = 4$ и $m = 5$ невозможны, поскольку максимально возможное

Таблица 1

Ранги гиперплоскостей, входящих в канонические тривекторы

	e^1	e^2	e^3	e^4	e^5	e^6	e^7	e^8
$W^{(1)}$	1	3	2	2	1	2	1	2
$W^{(2)}$	2	2	3	2	1	1	2	2
$W^{(3)}$	1	2	1	2	2	3	2	2
$W^{(4)}$	1	1	2	2	2	2	3	2
$W^{(5)}$	1	2	2	3	1	2	2	2



количество ребер в дополнительном графе $m = 3$.

4. $n = 6$. Сразу отбрасываем значения $m = 0$ (полный граф), $m = 4$, $m = 5$ и их дополнения $m = 10$, $m = 6$, $m = 5$. Остается рассмотреть случаи $m = 1$ (дополнение $m = 9$), $m = 2$ (дополнение $m = 8$) и $m = 3$ (дополнение $m = 7$).

4.1 $m = 1$. Этот случай отпадает. Действительно, листья W_γ и W_μ должны присутствовать обязательно. На оставшиеся четыре места претендует $C_5^4 = C_5^1 = 5$ четверок: $(W_\alpha W_\beta W_\lambda W_\delta)$, $(W_\alpha W_\beta W_\lambda W_\epsilon)$, $(W_\alpha W_\beta W_\delta W_\epsilon)$, $(W_\alpha W_\lambda W_\delta W_\epsilon)$, $(W_\beta W_\lambda W_\delta W_\epsilon)$. Ни одна из этих четверок не подходит, т. к. образуется еще одно ребро: либо (W_α, W_β) (первые три четверки), либо (W_γ, W_λ) (оставшиеся две четверки).

4.2 $m = 2$. Здесь возможны два варианта, показанные на рис. 4 а и б.

4.2 а — На роль листьев W_x , W_y , W_z претендуют $C_4^3 = C_4^1 = 4$ тройки листьев: $(W_\alpha W_\beta W_\delta)$, $(W_\alpha W_\beta W_\epsilon)$, $(W_\alpha W_\delta W_\epsilon)$, $(W_\beta W_\delta W_\epsilon)$. Первые две тройки отпадают, т. к. $W_\alpha \cap W_\beta = \emptyset$ и мы имели бы три ребра на графе. Остаются две последние возможности, которые дают два канонических тривектора:

$$W^{(1)} = W_\gamma + W_\lambda + W_\mu + W_\alpha + W_\delta + W_\epsilon,$$

$$W^{(2)} = W_\gamma + W_\lambda + W_\mu + W_\beta + W_\delta + W_\epsilon.$$

Подстановка $\varphi = (15)(27)(36)(4)(8)$ отображает эти тривекторы друг в друга (ср. случай 3.2 б). Имеем инволютивную циклическую группу второго порядка. Тривекторы $W^{(1)}$ и $W^{(2)}$ принадлежат типу (888; 852). Цикломатическое число соответству-

ющего основного графа $v(G) = 8$, что свидетельствует о наличии в этом графе восьми фундаментальных циклов: $(W_\gamma W_\epsilon W_\lambda W_\delta)$, $(W_\mu W_\delta W_\lambda)$, $(W_\gamma W_x W_\lambda W_\epsilon)$, $(W_\lambda W_\delta W_\epsilon)$, $(W_x W_\lambda W_\delta)$, $(W_x W_\lambda W_\epsilon)$, $(W_x W_\lambda W_\mu)$, $(W_\mu W_\lambda W_\epsilon)$.

4.2 б — На роль листьев W_x и W_y могут претендовать только листья W_δ и W_ϵ (лист W_λ не может быть задействован, т. к. $W_\lambda \cap W_\gamma = \emptyset$ и мы имели бы на графе не два, а три ребра). Таким образом, в этом случае получается лишь один канонический тривектор $W = W_\gamma + W_\mu + W_\alpha + W_\beta + W_\delta + W_\epsilon$, принадлежащий типу (887; 520). Здесь, как и в предыдущем случае, в соответствующем основном графе имеется восемь фундаментальных циклов ($v(G) = 8$): $(W_\gamma W_\alpha W_\delta)$, $(W_\gamma W_\alpha W_\epsilon)$, $(W_\alpha W_\delta W_\mu)$, $(W_\alpha W_\delta W_\epsilon)$, $(W_\mu W_\alpha W_\epsilon)$, $(W_\gamma W_\alpha W_\epsilon W_\beta)$, $(W_\gamma W_\alpha W_\mu W_\beta)$, $(W_\gamma W_\alpha W_\delta W_\beta)$.

4.3 $m = 3$. На роль листа W_x претендуют листья W_δ и W_ϵ . В соответствии с этим имеем два канонических тривектора

$$W^{(1)} = W_\gamma + W_\mu + W_\lambda + W_\alpha + W_\beta + W_\epsilon,$$

$$W^{(2)} = W_\gamma + W_\mu + W_\lambda + W_\alpha + W_\beta + W_\delta.$$

Тригекторы $W^{(1)}$ и $W^{(2)}$ переводятся друг в друга инволютивной подстановкой $\varphi = (1)(23)(4)(5)(67)(8)$ и принадлежат типу (888; 885). Цикломатическое число соответствующего основного графа $v(G) = 7$ и граф имеет семь фундаментальных циклов: $(W_\mu W_\alpha W_\lambda)$, $(W_\alpha W_\lambda W_x)$, $(W_\mu W_\alpha W_x)$, $(W_\mu W_\alpha W_x W_\beta)$, $(W_\gamma W_\alpha W_x W_\beta)$, $(W_\gamma W_\alpha W_x)$, $(W_\mu W_\alpha W_\gamma)$.

5. $n = 7$. Здесь $m = 3$, $v(G) = 12$ и мы имеем единственный канонический тривектор

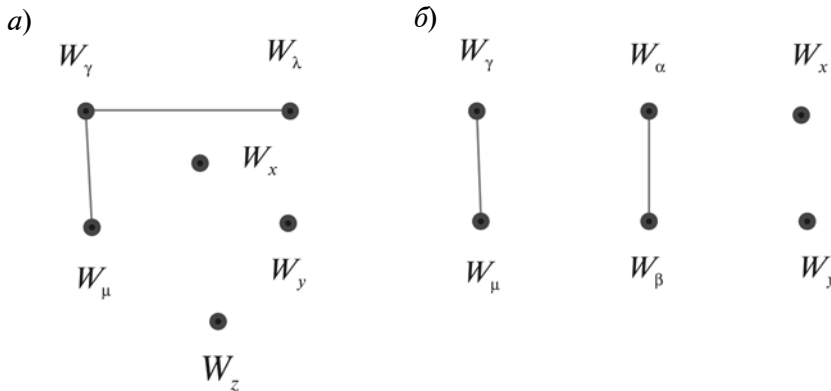


Рис. 4. Дополнительные графы для $n = 6$, $m = 2$

Таблица 2

Сводная таблица результатов исследований

Тип тривектора	$v(G)$	Количество канонических тривекторов	Количество листов	Порядок группы подстановок
(863; 100)	2	3	4	2
(874; 200)	5	2	5	2
(884; 400)	1	1	4	-
(886; 410)	4	2	5	2
(887; 520)	8	1	6	-
(888; 411)	0	1	3	-
(888; 621)	1	4	4	4
(888; 741)	4	4	5	4
(888; 822)	4	1	5	-
(888; 852)	8	2	6	2
(888; 873)	3	1	5	-
(888; 885)	7	2	6	2
(888; 888)	12	1	7	-

$W = W_\alpha + W_\beta + W_\gamma + W_\delta + W_\epsilon + W_\lambda + W_\mu$, принадлежащий общему типу (888; 888).

Случаем $n = 7$ завершается исследование всех тринадцати канонических тривекторов восьмого ранга.

Результаты исследований можно представить в виде сводной таблицы.

Анализируя данные табл. 2, можно сделать следующие выводы:

из тринадцати канонических типов тривекторов восьмого ранга шесть типов пред-

ставляются однозначно с точностью до групп Ли линейных преобразований;

в случаях, когда имеется несколько канонических тривекторов для одной и той же конфигурации вершин и ребер графа, найдены группы подстановок индексов, отображающие попарно один тривектор на другой, тем самым устанавливая изоморфизм соответствующих графов;

найжены фундаментальные циклы всех графов, при этом оказалось, что лишь один граф ацикличесен.

СПИСОК ЛИТЕРАТУРЫ

1. Гуревич, Г.Б. Алгебра тривектора. Ч. II [Текст] / Г.Б. Гуревич // Труды семинара по вект. и тенз. анализу. —М.: МГУ, 1948. — Вып. 6. —С. 28–124.
2. Гуревич, Г.Б. О некоторых линейных операторах, связанных с тривектором [Текст] / Г.Б. Гуревич // Труды семинара по вект. и тенз. анализу. —М.: МГУ, 1978. — Вып. XVIII. —С. 176–196.
3. Макоха, А.Н. Особые точки тривекторов восьмого ранга в P_7 [Текст] / А.Н. Макоха //Сб. Геометрия погруженных многообразий. —М., 1972. —С. 69–97.

4. Макоха, А.Н. Геометрия тривектора типа (863; 100) [Текст] / А.Н. Макоха //Сб. Геометрия однородных пространств. —М., 1976. —С. 34–45.
5. Макоха, А.Н. Линейные преобразования, сохраняющие тривектор типа (874; 200) [Текст] / А.Н. Макоха //Сб. Геометрия погруженных многообразий. —М., 1979. —С. 62–66.
6. Макоха, А.Н. Линейный комплекс плоскостей, ассоциированный с тривектором типа (886; 410) [Текст] / А.Н. Макоха //Сб. Современная геометрия. —Л., 1980. —С. 44–63.

7. **Макоха, А.Н.** Линейные операторы, связанные с тривектором типа (887; 520), и основная группа автоморфизмов этого тривектора [Текст] / А.Н. Макоха // Изв. вузов. Матем. —1981. —№ 7. —С. 46–63.

8. **Макоха, А.Н.** Группа автоморфизмов комплекса плоскостей типа A_6 и его геометрическая конструкция [Текст] / А.Н. Макоха // Изв. вузов. Матем.; рук. деп. в ВИНТИ; № 2119-92 ДЕП. — Казань, 1982. —16 с.

9. **Макоха, А.Н.** Группа автоморфизмов линейного комплекса плоскостей типа A_3 и особые прямые этого комплекса [Текст] / А.Н. Макоха // Изв. вузов. Матем. —1986. —№ 8. —С. 40–46.

10. **Макоха, А.Н.** Группа автоморфизмов линейного комплекса плоскостей типа A_4 [Текст] / А.Н. Макоха // Изв. вузов. Матем.; рук. деп. в ВИНТИ; № 8966-В87. —1987. —12 с.

11. **Макоха, А.Н.** Группа линейных преобразований, сохраняющих тривектор типа

(888; 852) [Текст] / А.Н. Макоха // Изв. вузов. Матем. —1988. —№ 2. —С. 45–49.

12. **Макоха, А.Н.** Линейный комплекс плоскостей общего типа категории A [Текст] / А.Н. Макоха // Вестник СГПУ. Естественные науки. —Ставрополь: Изд-во СГПУ, 1995. — Вып. 2. —С. 91–94.

13. **Макоха, А.Н.** Группы автоморфизмов тривекторов восьмого ранга [Текст] / А.Н. Макоха // Международная геометрическая школа-семинар памяти Н.В. Ефимова. Тезисы докладов. —Абрау-Дюрсо: Изд-во РГУ, 1996. —С. 50–51.

14. **Макоха, А.Н.** Интерпретация канонических тривекторов на графах [Текст] / А.Н. Макоха // Труды участников Междунар. школы-семинара по геометрии и анализу памяти Н.В. Ефимова. —Абрау-Дюрсо: Изд-во РГУ, 2002. — С. 47–49.

15. **Longo, C.** Sui complessi lineari di piani [Text] / C. Longo // Ann. mat. pura ed appl. —1954. —№ 37 (4). —P. 61–138.

REFERENCES

1. **Gurevich G.B.** Algebra trivektora. Ch. II / Trudy seminarov po vekt. i tenz. analizu. —Moscow: MGU, 1948. — Vyp. 6. —S. 28–124. (rus)

2. **Gurevich G.B.** O nekotorykh lineinykh operatorakh, svyazannykh s trivektorom / Trudy seminarov po vekt. i tenz. analizu. Vyp. XVIII. —Moscow: MGU, 1978. —S. 176–196. (rus)

3. **Makokha A.N.** Osobyie tochki trivektorov vos'mogo ranga v P_7 / Sb. Geometriia pogruzhennykh mnogoobrazii. —Moscow, 1972. —S. 69–97. (rus)

4. **Makokha A.N.** Geometriia trivektora tipa (863; 100) / Sb. Geometriia odnorodnykh prostranstv. —Moscow, 1976. —S. 34–45. (rus)

5. **Makokha A.N.** Lineinye preobrazovaniia, sokhraniayushchie trivektor tipa (874; 200) / Sb. Geometriia pogruzhennykh mnogoobrazii. —Moscow, 1979. —S. 62–66. (rus)

6. **Makokha A.N.** Lineinyi kompleks ploskosti, assotsirovannyi s trivektorom tipa (886; 410) / Sb. Sovremennaya geometriia. —Leningrad, 1980. —S. 44–63. (rus)

7. **Makokha A.N.** Lineinye operatory, svyazannyye s trivektorom tipa (887; 520), i osnovnaya gruppy avtomorfizmov etogo trivektora / Izv. vuzov. Matem. —1981. —№ 7. —S. 46–63. (rus)

8. **Makokha A.N.** Gruppy avtomorfizmov kompleksa ploskosti tipa A_6 i ego geometricheskaya konstruktsiya / Izv. vuzov. Matem. ruk. dep. v VINITI;

№ 2119-92 DEP. — Kazan', 1982. —16 s. (rus)

9. **Makokha A.N.** Gruppy avtomorfizmov lineinogo kompleksa ploskosti tipa A_3 i osobyie priamyye etogo kompleksa / Izv. vuzov. Matem. —1986. —№ 8. —S. 40–46. (rus)

10. **Makokha A.N.** Gruppy avtomorfizmov lineinogo kompleksa ploskosti tipa A_4 / Izv. vuzov. Matem.; ruk. dep. v VINITI; № 8966-В87; 1987. —12 s. (rus)

11. **Makokha A.N.** Gruppy lineinykh preobrazovaniy, sokhraniayushchikh trivektor tipa (888; 852) / Izv. vuzov. Matem. —1988. —№ 2. —S. 45–49. (rus)

12. **Makokha A.N.** Lineinyi kompleks ploskosti obshchego tipa kategorii A / Vestnik SGPU. Estestvennyye nauki. —Stavropol': Izd-vo SGPU. —1995. —Vyp. 2. —S. 91–94. (rus)

13. **Makokha A.N.** Gruppy avtomorfizmov trivektorov vos'mogo ranga / Mezhdunarodnaya geometricheskaya shkola-seminar pamiati N.V. Efimova. Tezisy dokladov. —Абрау-Дюрсо: Изд-во РГУ, 1996. —S. 50–51. (rus)

14. **Makokha A.N.** Interpretatsiya kanonicheskikh trivektorov na grafakh / Trudy uchastnikov Mezhdunarodnoi shkoly-seminara po geometrii i analizu pamiati N.V. Efimova. — Абрау-Дюрсо: Изд-во РГУ, 2002. — S. 47–49. (rus)

15. **Longo C.** Sui complessi lineari di piani / Ann. mat. pura ed appl. —1954. —№ 37 (4). —P. 61–138.

МАКОХА Анатолий Николаевич — доцент кафедры прикладной математики и компьютерных технологий Северо-Кавказского федерального университета, кандидат физико-математических наук.

355029, Россия, г. Ставрополь, пр. Кулакова, д. 2.

E-mail: anmakoha@mail.ru

МАКОНА, Anatoliy N. *North-Caucasian Federal University.*

355029, prosp. Kulakova 2, Stavropol, Russia.

E-mail: anmakoha@mail.ru



УДК 528.854

В.Б. Мнухин

ИНТЕГРАЛЬНОЕ ПРЕОБРАЗОВАНИЕ ДЛЯ РАСПОЗНАВАНИЯ СИММЕТРИИ ИЗОБРАЖЕНИЙ

V.B. Mnukhin

INTEGRAL TRANSFORM FOR SYMMETRY RECOGNITION OF GRAY-LEVEL IMAGES

Предложено интегральное преобразование для распознавания симметрии полутоновых изображений. Вводимое преобразование оказывается инвариантным относительно сдвигов и вращений, но зависит от масштабирования. Свойства преобразования строго доказаны; на их основе предложен ряд тестов, позволяющих эффективно распознавать группу симметрий. Рассмотрены вопросы применимости предложенного метода к анализу симметрии цифровых изображений.

РАСПОЗНАВАНИЕ СИММЕТРИИ. ПОЛУТОНОВЫЕ ИЗОБРАЖЕНИЯ. ГРУППА СИММЕТРИЙ. ЧАСТОТНАЯ ОБЛАСТЬ. ИНТЕГРАЛЬНОЕ ПРЕОБРАЗОВАНИЕ. ПРЕОБРАЗОВАНИЕ ФУРЬЕ–МЕЛЛИНА.

A transform for symmetry recognition in 2D gray-level images is considered. The transform is invariant under rotations and translations, but is not scale-invariant. The rigorous mathematical description of its properties is provided and a series of tests for symmetry recognition is produced. Some details of the discrete case realization are given and illustrated with experiment results.

SYMMETRY RECOGNITION. GRAY-LEVEL IMAGES. SYMMETRY GROUP. FREQUENCY DOMAIN. INTEGRAL TRANSFORM. FOURIER–MELLIN TRANSFORM.

Изучение симметрии изображений в настоящее время — одно из активно развиваемых направлений теоретической информатики. Работы в этой области активно стимулируются, в частности тем, что группы симметрий изображений не зависят от их размеров, поворотов, яркости, плотности и центрирования, являясь тем самым сильными дескрипторами изображенных объектов [1–3]. Данные о симметрии объекта позволяют:

производить сжатие и компактное хранение видеоинформации;

осуществлять как обычный, так и ассоциативный поиск видеоданных, необходимость в котором возникает при создании современных интеллектуальных систем;

делать анализ графической информации, направленный на повышение качества и достоверности работы систем распознавания образов и др.

В настоящей статье рассматривается задача распознавания группы симметрий

изображения с точностью до изоморфизма. Для ее решения предлагается новое двумерное интегральное преобразование, являющееся непрерывным по первой переменной и дискретным по второй. Идея его построения основана [4] на рассмотрении Фурье-образа изображения в полярной системе координат, что позволяет трансформировать вращения исходного изображения в сдвиги его образа.

Отметим, что вводимое преобразование до некоторой степени аналогично т. н. *преобразованию Фурье–Меллина* [5], инвариантному относительно как сдвигов и вращений, так и масштабирования. Заметим, что последнее важно для эффективного решения задач совмещения изображений [6], однако избыточно при распознавании симметрии. В то же время использование преобразованием Фурье–Меллина полярно-логарифмической системы координат приводит к значительным вычислительным сложностям. В частности, основанный на

нем метод [7] распознавания симметрии требует априорного знания координат центра симметрии.

Предлагаемый в статье подход свободен от указанных недостатков. Простота аналитического выражения вводимого преобразования позволяет сформулировать и строго доказать его свойства в окончательной форме. На их основе устанавливается ряд следствий, позволяющих эффективно распознавать группу симметрий. Их применение к анализу симметрий цифровых изображений иллюстрируется результатами экспериментов.

Группа симметрий изображения

Будем считать, что на действительной плоскости \mathbb{R}^2 введена декартова система координат xOy . Под *изображением* (точнее, под *плоским непрерывным полутоновым финитным изображением*) будем понимать действительную неотрицательную ограниченную функцию $f(x,y)$, определенную всюду на \mathbb{R}^2 , но отличную от нуля только внутри некоторой области $I \subset \mathbb{R}^2$ конечного диаметра d . Рассмотрим следующие элементарные преобразования изображений:

1) *сдвиг* T_a на вектор $a = (x_0, y_0)$:
 $T_a[f] = f(x - x_0, y - y_0)$;

2) *вращение* R_α на угол α вокруг начала координат O :

$$R_\alpha[f] = f(x \cos \alpha - y \sin \alpha, x \sin \alpha + y \cos \alpha);$$

3) *отражение* M относительно оси Ox :
 $M[f] = f(x, -y)$.

Более сложные преобразования получаются как композиции элементарных. В частности, преобразование M_L *отражения относительно прямой L* , проходящей через точку a и наклоненной к оси Ox на угол α , записывается как $M_L = T_a R_{2\alpha} M T_a^{-1}$.

(Действительно, если L проходит через начало координат, то равенство $M_L = R_{2\alpha} M$ следует из элементарных геометрических соображений; в общем случае остается заметить, что точку a можно сначала сместить в начало координат, а после выполнения преобразования $R_{2\alpha} M$ вернуть в прежнее положение с помощью соответствующих сдвигов.)

Определение 1. Изображение f обладает *отражательной симметрией* относительно прямой L (или *симметрично относительно L*), если

$$M_L[f(x, y)] = f(x, y) \text{ для всех } (x, y) \in \mathbb{R}^2.$$

В этом случае L называется *осью симметрии*.

Заметим, что центром вращения может служить произвольная точка плоскости. Легко видеть, что соответствующее преобразование $R_{a,\alpha}$ *вращения на угол α вокруг точки a* можно записать как

$$R_{a,\alpha} = T_a R_\alpha T_a^{-1}.$$

Определение 2. Будем говорить, что изображение f в точке a обладает *вращательной симметрией бесконечного порядка*, если $R_{a,\alpha}[f] = f$ для всех углов α . Если же найдется угол β , ($0 < \beta < 2\pi$), такой, что и $R_{a,\beta}[f] = f$, и $R_{a,\alpha}[f] \neq f$ для всех меньших углов α , $0 < \alpha < \beta$, то в точке a изображение имеет *вращательную симметрию порядка $k = 2\pi/\beta > 1$* , где, как нетрудно показать, число k является целым. Точка a называется *центром* вращательной симметрии.

Всевозможные композиции элементарных преобразований T_a , R_α и M образуют бесконечную *группу перемещений* Γ , являющуюся подгруппой группы всех аффинных преобразований плоскости [8]. Таким образом, Γ естественно действует на множестве изображений.

Определение 3. *Группой симметрий* изображения f называется его стабилизатор $\Gamma(f)$ в группе перемещений. Другими словами, это множество всех тех перемещений плоскости, которые не изменяют данное изображение.

Классы изоморфизма групп симметрий финитных изображений хорошо известны [8]. Это

1) *единичная группа I* , соответствующая изображениям без нетривиальных симметрий;

2) *бесконечная ортогональная группа $O(2)$* , порождаемая вращением бесконечного порядка и являющаяся группой симметрий окружности;

3) *конечные циклические группы Z_k* порядка $k > 2$, порождаемые вращением во-

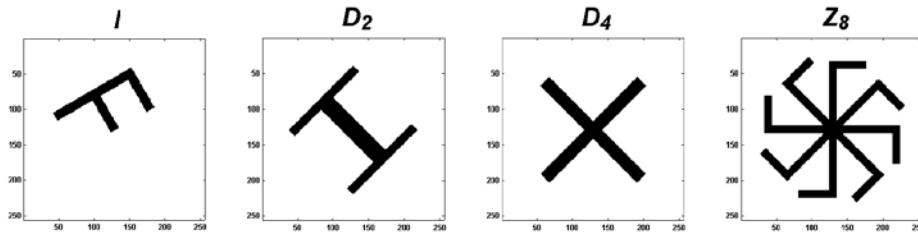


Рис. 1. Примеры изображений с небольшими группами симметрий

круг некоторой точки на угол $2\pi/k$;

4) конечные *диэдральные группы* D_k порядка $2k \geq 2$, порождаемые отражением и/или вращением порядка k с центром на оси отражательной симметрии. (Несмотря на изоморфизм $D_1 \cong Z_2$, мы отличаем группу D_1 , порождаемую единственным *отражением*, от группы Z_2 , порождаемой *вращением* на 180° , поскольку соответствующие этим группам симметрии изображений различны). На рис. 1 приведены примеры изображений с небольшими группами симметрий.

***H*-преобразование и его свойства**

Введем интегральное преобразование, играющее ключевую роль в нашем подходе к распознаванию симметрии. Оно основано на двумерном преобразовании Фурье, определяемом [9] следующим образом:

$$F[f] = F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi i(xu + yv)} dx dy.$$

Заметим, что сходимость интеграла немедленно вытекает из наложенных на функцию $f(x, y)$ ограничений. Фурье-образ $F(u, v)$ определен на плоскости uOv , называемой *частотной областью*. Введем в ней полярную систему координат $rO\varphi$ с тем же самым началом O и с полярной осью $\varphi = 0$, совпадающей с положительной полуосью Ou . Рассмотрим Фурье-образ $F(u, v)$ в системе координат $rO\varphi$, полагая $\Phi(r, \varphi) = F(r \cos \varphi, r \sin \varphi)$.

Наложим на функцию f дополнительное ограничение. Поскольку реальные изображения являются гладкими, условимся считать $f(x, y)$ бесконечно дифференцируемой в каждой точке по любому направлению. Тогда для всякого фиксированного полярного угла $\varphi_0 \in [0, \pi]$ модуль функции $\Phi(r, \varphi_0)$ быстро убывает с ростом радиуса r ,

что гарантирует корректность следующего определения.

Определение 4. *H*-преобразованием $H[f]$ функции $f(x, y)$ будем называть следующую функцию двух переменных:

$$H(w, n) = \int_0^{+\infty} \int_0^\pi |\Phi(r, \varphi)| e^{-i(wr + 2n\varphi)} dr d\varphi,$$

где *непрерывная* переменная w принимает всевозможные действительные значения, а *дискретная* переменная n пробегает множество всех целых чисел \mathbb{Z} . Удобно рассматривать $H[f]$ как бесконечную (в обе стороны) последовательность функций

$$\dots, H(w, -2), H(w, -1), H(w, 0), H(w, 1), H(w, 2), \dots$$

Функцию $H(w, n)$ условимся называть *H-образом* для f .

Рассмотрим свойства введенного преобразования. Отметим, что поскольку $H[f]$ зависит только от энергетического спектра $|F[f]|$ изображения, *H*-преобразование не является ни обратимым, ни линейным, однако для произвольной константы $c \in \mathbb{R}$ справедливо тождество $H[cf] = |c| \cdot H[f]$. Покажем, как меняется *H*-образ изображения при его сдвигах и вращениях.

Утверждение 1. *Преобразование $H[f]$ инвариантно относительно сдвигов: $H[T_a f] = H[f]$.*

Доказательство. Воспользуемся известным свойством преобразования Фурье [9]: $|F[T_a f]| = |e^{-2\pi i(x_0 u + y_0 v)} F(u, v)| = |F(u, v)|$, где $a = (x_0, y_0)$. Из него вытекает, что функция $|\Phi(r, \varphi)|$ не меняется при сдвигах, а значит, не меняется при сдвигах и $H[f]$. ■

Утверждение 2. *При вращении изображения f на угол α вокруг начала координат, его *H*-образ $H(w, n)$ умножается на фазовый множитель:*

$$H[R_\alpha f] = e^{2in\alpha} H(w, n).$$

Доказательство. Как известно [9], непрерывное преобразование Фурье коммутирует с вращениями, то есть $FR_\alpha = R_\alpha F$. Поэтому, обозначая Фурье-образ в полярной системе координат как $F[f] = \Phi(r, j)$, получим $F[R_\alpha f] = R_\alpha[F[f]] = \Phi(r, \varphi + \alpha)$, и

$$H[R_\alpha f] = \int_0^{+\infty} \int_0^\pi |\Phi(r, \varphi + \alpha)| e^{-i(wr+2n\theta)} dr d\varphi.$$

Выполняя замену переменной $\theta = \varphi + \alpha$, получаем

$$H[R_\alpha f] = e^{2i\alpha} \int_0^{+\infty} \int_\alpha^{\pi+\alpha} |\Phi(r, \theta)| e^{-i(wr+2n\theta)} dr d\theta.$$

Поскольку f принимает только действительные значения, $F(u, v) = F^*(-u, -v)$, и, следовательно, функция $|\Phi(r, \theta)|$ является периодической по θ с периодом π . Так как n – целое, периодическим будет и все подынтегральное выражение. Следовательно,

$$\int_0^{+\infty} \int_\alpha^{\pi+\alpha} |\Phi(r, \theta)| e^{-i(wr+2n\theta)} dr d\theta = H[f],$$

откуда и вытекает требуемое тождество. ■

Рассмотрим поведение H -преобразования при вращениях $R_{a,\alpha}$ вокруг произвольной точки a . Пользуясь предыдущими результатами и тождеством $R_{a,\alpha} = T_a R_\alpha T_a^{-1} = T_a R_\alpha T_{-a}$, немедленно получаем

Следствие 1. Выбор центра вращения изображения f не влияет на его H -преобразование, $H[R_{a,\alpha} f] = H[R_\alpha f]$, а модуль H -преобразования инвариантен относительно вращений, $|H[R_{a,\alpha} f]| = |H[f]|$. В частности,

$$H[R_{\alpha,\pi/4} f] = i^n H(w, n),$$

$$H[R_{\alpha,\pi/2} f] = (-1)^n H(w, n), \quad H[R_{\alpha,\pi} f] = H[f]. \quad \blacksquare$$

Утверждение 3. Отражение изображения f от любой из координатных осей Ox или Oy меняет знак переменной n в функции $H(w, n) = H[f]$:

$$\begin{aligned} H[Mf] &= H[f(x, -y)] = \\ &= H(w, -n) = H[f(-x, y)]. \end{aligned}$$

Доказательство. Пусть $F[f] = F(r \cos \varphi, r \sin \varphi) = \Phi(r, \varphi)$, тогда $F[Mf] = F[f(x, -y)] = \Phi(r, -\varphi)$. Подставляя это выражение в определение 4 и выполняя в интеграле замену переменных $\theta = -\varphi$, получим, с учетом периодичности подынтегрального выражения, что $H[Mf] = H(w, -n)$. Последнее равенство вытекает из соотно-

шения $f(-x, y) = R_\pi M[f]$. ■

Случай отражения M_L относительно произвольной оси L чуть более сложен.

Утверждение 4. Если прямая L образует с осью Ox угол α , а H -образ исходного изображения f равен $H(w, n) = H[f]$, то $H[M_L f] = e^{4i\alpha} H(w, -n)$.

Доказательство. Ранее было замечено, что $M_L = T_a R_{2\alpha} M T_{-a}$, где a – произвольная точка на L . Учитывая предыдущие результаты, получаем

$$\begin{aligned} H[M_L f] &= H[T_a R_{2\alpha} M T_{-a} [f]] = \\ &= H[R_{2\alpha} M T_{-a} [f]] = \\ &= e^{4i\alpha} H[M[T_{-a} f]] = e^{4i\alpha} H(w, -n), \end{aligned}$$

что завершает доказательство. ■

Распознавание симметрии с помощью H -преобразования

Результаты предыдущего раздела приводят к ряду тестов на наличие у изображений того или иного типа симметрии. Вспоминая, что отражательная симметричность относительно оси L равносильно условию $M_L[f] = f$, немедленно получаем из утверждения 4 следующий результат.

Утверждение 5. Если изображение f симметрично относительно некоторой прямой, образующей с осью Ox угол α , то H -образ этого изображения удовлетворяет тождеству $H(w, n) \equiv e^{4i\alpha} H(w, -n)$.

В частности, если для некоторых значений переменных w и n справедливо неравенство $|H(w, n)| \neq |H(w, -n)|$, то изображение не обладает отражательной симметрией.

Рассмотрим тесты на вращательную симметрию.

Утверждение 6. Если изображение f допускает вращательную симметрию бесконечного порядка (и значит, имеет группу симметрий $O(2)$), то $H(w, n) \equiv 0$ для всех $n \neq 0$.

Доказательство. Как уже отмечалось, непрерывное преобразование Фурье коммутирует с вращениями, а значит, сохраняет вращательные симметрии исходного изображения. Следовательно, для изображения с группой $O(2)$ функция $\Phi(r, \varphi)$ не зависит от полярного угла φ . Пусть $\Phi(r, \varphi) = \Psi(r)$, тогда при $n \neq 0$ имеем

$$H(w, n) = \int_0^{+\infty} \int_0^\pi |\Psi(r)| e^{-i(wr+2n\varphi)} dr d\varphi = \\ = \int_0^{+\infty} |\Psi(r)| e^{-iwr} dr \int_0^\pi e^{-2in\varphi} d\varphi = 0. \blacksquare$$

Для формулировки аналогичного результата об изображениях, допускающих вращательные симметрии только конечных порядков, введем следующую функцию на множестве \mathbb{Z} целых чисел:

$$h(k) = \begin{cases} k, & \text{если } k \text{ нечетно,} \\ k/2, & \text{если } k \text{ четно.} \end{cases}$$

Утверждение 7. Если изображение f допускает вращение порядка k , то $H(w, n) \equiv 0$ для всех $n \neq 0 \pmod{h(k)}$.

Другими словами, периодическое появление в последовательности

$$\dots, H(w, -2), H(w, -1), H(w, 0), \\ H(w, 1), H(w, 2), \dots$$

нулевых функций указывает на вращательную симметрию изображения, причем «ширина нулевого интервала» связана с порядком симметрии.

Доказательство. Действительно, в этом случае $R_{a, 2\pi/k}[f] = f$. Переходя к H -образам, получаем $H[R_{a, 2\pi/k}f] = H[f]$, или $(1 - e^{4\pi in/k})H(w, n) \equiv 0$. Значит, если $e^{4\pi in/k} \neq 1$, то $H(w, n) = 0$ для всех $n \in \mathbb{Z}$. Остается заметить, что при нечетном k равенство $e^{4\pi in/k} = 1$ означает, что n делится на k ; а при четном k — что n делится на $k/2$. \blacksquare

Из утверждений 5 и 7 вытекает тест на диэдральность.

Следствие 2. Если изображение f имеет диэдральную группу симметрий D_k порядка $2k$, то

$$H(w, n) = \begin{cases} e^{4\pi in\alpha} H(w, -n), & n \equiv 0 \pmod{h(k)}, \\ 0, & \text{иначе,} \end{cases}$$

где α является углом наклона произвольной оси отражательной симметрии данного изображения к Ox .

(Корректность предыдущего утверждения следует из того, что разность углов наклона двух различных осей симметрии всегда кратна π/k .)

Отметим, что предыдущие результаты дают только *необходимые условия* наличия у

изображения того или иного типа симметрии, или же *достаточные условия* ее отсутствия. Кроме того, из утверждения 7 вытекает невозможность различить вращения порядков $2k + 1$ и $2(2k + 1)$ с помощью *только* H -преобразования. Это связано с тем, что Фурье-спектр $|F[f]|$ *любого* изображения обладает центральной симметрией. Как следствие, H -преобразования изображений с группами Z_{2k+1} и $Z_{2(2k+1)}$ устроены аналогично. Вместе с тем нетрудно заметить, что изображение с группой $Z_{2(2k+1)}$ всегда будет центрально-симметричным, а с группой Z_{2k+1} — нет. Несложный тест на центральную симметричность изображений предложен в [4], где могут быть также найдены дальнейшие детали.

Распознавание симметрии цифровых изображений

Рассмотрим некоторые детали применения результатов предыдущего раздела для практически значимого случая цифровых изображений.

Назовем *цифровым изображением* размера $N \times N$ произвольную действительную неотрицательную ограниченную функцию $f(x, y)$ двух дискретных аргументов, определенную на множестве $[0, N - 1] \times [0, N - 1] \subset Z^2$. Под *диаметром* цифрового изображения будем понимать размер d наименьшей квадратной области, вне которой функция $f(x, y)$ принимает только нулевые значения.

Чтобы ввести дискретный аналог H -преобразования, воспользуемся дискретным преобразованием Фурье (ДПФ):

$$F(u, v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \exp \left\{ -\frac{2\pi i}{N} (ux + vy) \right\}.$$

Кроме того, необходимо задать на $[0, N - 1] \times [0, N - 1]$ некоторый аналог полярной системы координат, или, что равносильно, определить оператор *дискретного вращения* \mathfrak{R}_α . Отметим, что это является нетривиальной задачей, допускающей различные решения, а вполне адекватное определение полярной системы координат на дискретной плоскости представляется невозможным [10]. Обычно \mathfrak{R}_α вводится как $\mathfrak{R}_\alpha[f] = f([X] \bmod N, [Y] \bmod N)$, где $X = x \cos \alpha - y \sin \alpha$, $Y = x \sin \alpha + y \cos \alpha$,

а $[X]$ и $[Y]$ означают ближайшие целые к X и Y соответственно. Это позволяет ввести дискретное H -преобразование по аналогии с определением 4. Отметим, что H -преобразование цифрового $(N \times N)$ -изображения является комплексной $(N \times N)$ -матрицей.

К сожалению, ДПФ не коммутирует с дискретными вращениями \mathfrak{R}_α , что препятствует немедленному переносу результатов предыдущего раздела на дискретный случай. Вместе с тем можно показать, что для цифровых изображений фиксированного диаметра d имеет место сходимость $\|(\mathfrak{R}_\alpha F - F\mathfrak{R}_\alpha)[f]\| \xrightarrow{N \rightarrow \infty} 0$, что позволяет использовать полученные для непрерывного случая результаты в дискретном случае при выполнении условий $N \gg 0$ и $d \ll N$.

Рассматривая симметрию цифровых изображений, следует учитывать, что поня-

тие симметрии непрерывных изображений ранее трактовалось как двузначное: функция $f(x,y)$ считалась либо обладающей некоторым типом симметрии, либо нет. При этом для реальных объектов симметрия никогда не бывает идеальной и говорить можно только о той или иной *мере симметричности* объекта. Более того, следует учитывать, что даже абсолютно симметричное относительно некоторой оси цифровое изображение может потерять эту симметричность при поворотах. Поэтому анализ H -преобразований цифровых изображений следует проводить статистическими методами.

В качестве примера рассмотрим рис. 2, показывающий визуализации матриц $\|\ln(1 + |H(w, n)|)\|$ для четырех изображений рис. 1. На всех визуализациях ось On направлена горизонтально, а ось Om —

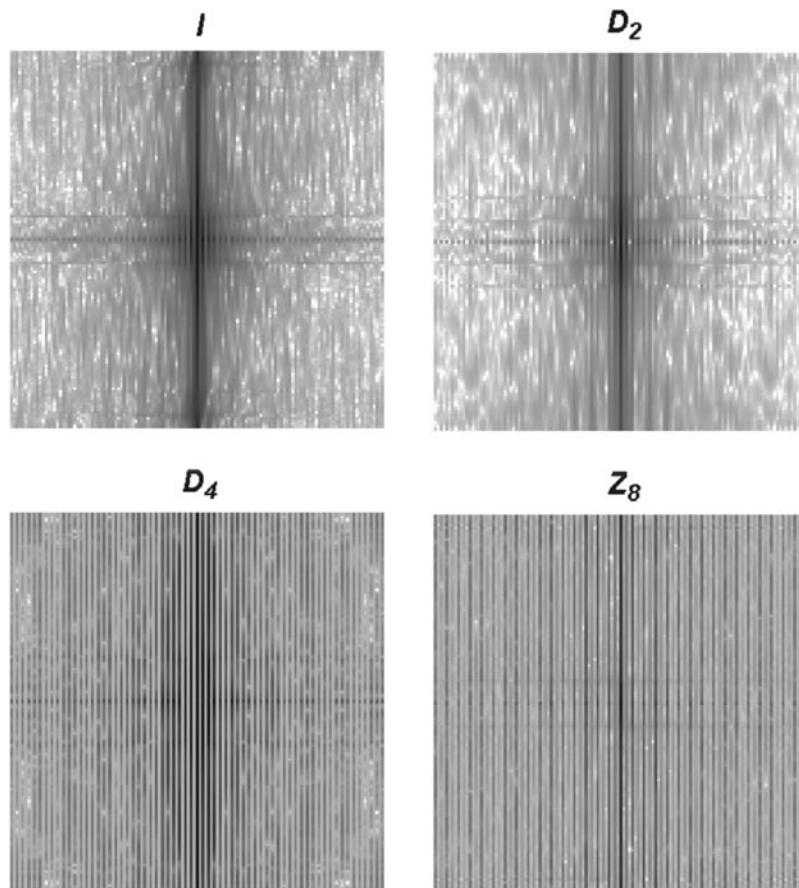


Рис. 2. Примеры визуализации H -преобразований изображений



вертикально; точка O совпадает с геометрическим центром; темные оттенки соответствуют большим значениям величин, а светлые — меньшим. Правый верхний рисунок соответствует изображению буквы **H** с диэдральной группой D_2 порядка 4. Хорошо заметна характерная симметричность рисунка относительно Ow , вытекающая из следствия 2. Статистическая проверка выполнения условия $H(w, n) = e^{4ina} H(w, -n)$ для всех w и n позволяет уточнить тип симметрии и оценить наклон оси симметрии исходного изображения. В частности, так можно отбросить гипотезу о нетривиальности симметрии у изображения буквы **F**, соответствующего левому верхнему рисунку.

Левый нижний рисунок соответствует изображению Андреевского креста с группой D_4 порядка 8. Симметричность относительно Ow указывает на наличие отражательной симметрии, а вытекающая из утверждения 7 характерная «полосатость» — на наличие вращения четвертого порядка. Правый нижний

рисунок соответствует 8-лучевой свастике с группой Z_8 . Можно заметить, что визуализация H -преобразования утратила Ow -симметричность, что говорит об отсутствии у свастики отражательной симметрии. В то же время ширина «темных полос» увеличилась до трех, что указывает на наличие вращательной симметрии восьмого порядка.

Таким образом, в работе предложен метод распознавания групп симметрий плоских полутоновых изображений, основанный на использовании нового двумерного интегрального преобразования, непрерывного по одной переменной и дискретного по другой. Сформулированы и строго доказаны свойства этого преобразования. На их основе предложен ряд следствий, позволяющих эффективно распознавать симметрии как непрерывных, так и цифровых изображений

Работа выполнена при поддержке РФФИ, проекты № 11-07-00591 и №13-07-00327.

СПИСОК ЛИТЕРАТУРЫ

1. **Chertok, M.** Spectral Symmetry Analysis [Text] / M. Chertok, Y. Keller // IEEE Trans. on Pattern Analysis and Machine Intelligence. — 2010. — Vol. 32. — P. 1227–1238.
2. **Каркищенко, А.Н.** Классификация изображений периодических структур на основе непрерывного преобразования симметрии [Текст] / А.Н. Каркищенко, В.Б. Мнухин // Матер. VIII Междунар. конф. Интеллектуализация обработки информации. — М.: МАКС Пресс, 2010. — С. 359–362.
3. **Каркищенко, А.Н.** Преобразование симметрии периодических структур в частотной области [Текст] / А.Н. Каркищенко, В.Б. Мнухин // Матер. XV Всерос. конф. Математические методы распознавания образов. — М.: МАКС Пресс, 2011. — С. 386–389.
4. **Каркищенко, А.Н.** Распознавание симметрии изображений в частотной области [Текст] / А.Н. Каркищенко, В.Б. Мнухин // Матер. IX Междунар. конф. Интеллектуализация обработки информации. — М.: Торус Пресс, 2012. — С. 426–429.
5. **Reddy, S.** A FFT-based technique for transla-

tion, rotation, and scale invariant image registration [Text] / S. Reddy, B. Chatterji // IEEE Trans. on Image Processing. — 1996. — Vol. 5. — P. 1266–1271.

6. **Derrode, S.** Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description [Text] / S. Derrode, F. Ghorbel // Computer Vision and Image Understanding. — 2001. — Vol. 83. — P. 57–78.

7. **Derrode, S.** Shape analysis and symmetry detection in gray-level objects using the analytical Fourier–Mellin representation [Text] / S. Derrode, F. Ghorbel. // Signal Processing. — 2004. — Vol. 84. — P. 25–39.

8. **Никулин, В.В.** Геометрии и группы [Текст] / В.В. Никулин, И.П. Шафаревич. — М.: Наука, 1983. — 239 с.

9. **Poularikas, A.D.** Transforms and Applications Handbook [Text] / A.D. Poularikas. — CRC Press, 2010. — 1336 p.

10. **Beylkin, G.** On the fast Fourier transform of functions with singularities [Text] / G. Beylkin // Appl. Comput. Harmon. Anal. — 1995. — Vol. 2. — P. 363–381.

REFERENCES

1. **Chertok M., Keller Y.** Spectral Symmetry Analysis / IEEE Trans. on Pattern Analysis and Machine Intelligence. — 2010. — Vol. 32 — P. 1227–1238.

2. **Karkishchenko A.N., Mnukhin V.B.** Klassifikatsiia izobrazhenii periodicheskikh struktur na osnove nepreryvnogo preobrazovaniia simmetrii

/ Mater. VIII Mezhdunar. konf. Intellektualizatsiia obrabotki informatsii. – Moscow: MAKS Press, 2010. – S. 359–362. (rus)

3. **Karkishchenko A.N., Mnukhin V.B.** Preobrazovanie simmetrii periodicheskikh struktur v chastotnoi oblasti / Mater. XV Vseros. konf. Matematicheskie metody raspoznavaniia obrazov. – Moscow: MAKS Press, 2011. – S. 386–389. (rus)

4. **Karkishchenko A.N., Mnukhin V.B.** Raspoznavanie simmetrii izobrazhenii v chastotnoi oblasti / Mater. IX Mezhdunar. konf. Intellektualizatsiia obrabotki informatsii. – Moscow: Torus Press, 2012. – S. 426–429. (rus)

5. **Reddy S., Chatterji B.** A FFT-based technique for translation, rotation, and scale invariant image registration / IEEE Trans. on Image Processing.

–1996. –Vol. 5. – P. 1266–1271.

6. **Derrode S., Ghorbel F.** Robust and efficient Fourier–Mellin transform approximations for gray-level image reconstruction and complete invariant description / Computer Vision and Image Understanding. – 2001. – Vol. 83. – P. 57–78.

7. **Derrode S., Ghorbel F.** Shape analysis and symmetry detection in gray-level objects using the analytical Fourier–Mellin representation / Signal Processing. – 2004. –Vol. 84. – P. 25–39.

8. **Nikulin V.V., Shafarevich I.R.** Geometrii i gruppy. – Moscow: Nauka, 1983. –239 s. (rus)

9. **Poularikas A.D.** Transforms and Applications Handbook. – CRC Press, 2010. – 1336 p.

10. **Beylkin G.** On the fast Fourier transform of functions with singularities / Appl. Comput. Harmon. Anal. – 1995. – Vol. 2. – P. 363–381.

МНУХИН Валерий Борисович – доцент кафедры высшей математики Таганрогского технологического института Южного федерального университета, кандидат физико-математических наук.

347916, Россия, г. Таганрог, ул. Циолковского, д. 55.

E-mail: mnukhin.valeriy@mail.ru

MNUKHIN, Valery B. *Taganrog Technological Institute of Southern Federal University.*

347916, Tsiolkovsky Str. 55, Taganrog, Russia

E-mail: mnukhin.valeriy@mail.ru



УДК 621.391

Н.И. Червяков, М.Г. Бабенко, П.А. Ляхов, И.Н. Лавриненко

**ПРИБЛИЖЕННЫЙ МЕТОД
ОПРЕДЕЛЕНИЯ ЗНАКА ЧИСЛА В СИСТЕМЕ ОСТАТОЧНЫХ КЛАССОВ
И ЕГО ТЕХНИЧЕСКАЯ РЕАЛИЗАЦИЯ**

N.I. Chervyakov, M.G. Babenko, P.A. Lyakhov, I.N. Lavrinenko

**APPROXIMATE METHOD FOR DETERMINING THE NUMBER SIGN
IN RESIDUE NUMBER SYSTEM AND IT'S TECHNICAL SALES**

Предложен приближенный метод определения знака числа в системе остаточных классов, основанный на быстром приближенном вычислении относительной величины числа к диапазону системы. Описана техническая реализация метода, отличающаяся простотой реализации. Показано, что применение данного метода позволяет существенно сократить время выполнения основных проблемных операций в системе остаточных классов.

СИСТЕМА ОСТАТОЧНЫХ КЛАССОВ. МОДУЛЯРНАЯ АРИФМЕТИКА. ПОЗИЦИОННАЯ ХАРАКТЕРИСТИКА. ПРИБЛИЖЕННЫЙ МЕТОД. ЗНАК ЧИСЛА.

This paper proposes an approximate method for determining the sign of the number in the residue number systems. This method is based on fast approximate calculation of the relative magnitude of the number to the range of the system. Technical implementation of the proposed method, which is easy to implement. It is shown that the application of the proposed method significantly reduces the time to perform basic operations in the problem of residue number systems.

RESIDUE NUMBER SYSTEM. MODULAR ARITHMETIC. POSITIONAL CHARACTERISTICS. APPROXIMATE METHOD. NUMBER SIGN.

Современное состояние развития инфокоммуникационных технологий в области обработки и передачи данных характеризуется интенсивным внедрением новых принципов и подходов к обработке информации. Результаты теоретических и практических разработок отечественных и зарубежных специалистов со всей определенностью указывают на то, что одним из перспективных многообещающих путей решения задач сокращения времени обработки данных и повышения надежности вычислительных средств является применение различных форм параллельной обработки данных, в т. ч. и на основе числовых систем с параллельной структурой. Одно из магистральных направлений среди современных подходов к созданию отказоустойчивых высокопроизводительных средств обработки данных — использование системы остаточных классов (СОК) [1].

Если фиксированный ряд положительных чисел p_1, p_2, \dots, p_n назвать основаниями (модулями) СОК, то под системой остаточных классов понимается такая непозиционная система счисления, в которой любое целое положительное число A представляется в виде набора остатков (вычетов) от деления представляемого числа на выбранные основания системы $A = (a_1, a_2, \dots, a_n)$, где a_i — наименьшие неотрицательные вычеты (остатки) числа по модулям p_1, p_2, \dots, p_n . Цифры a_i данного представления по выбранным модулям образуются следующим образом

$$\begin{aligned} a_i &= \text{res} + A \pmod{p_i} = \\ &= A - \left[\frac{A}{p_i} \right] p_i, (\forall i \in [1, n]), \end{aligned} \quad (1)$$

где $\left[\frac{A}{p_i} \right]$ — целочисленное частное, p_i —

основания (модули) — взаимно-простые числа. В теории чисел доказано, что если $\forall i \neq j (p_i, p_j) = 1$, то представление (1) является единственным, при условии $0 \leq A \leq P$, где $P = p_1 p_2 \dots p_n = \prod_{i=1}^n p_i$ — диапазон представления чисел, т. е. существует число A , для которого:

$$\begin{aligned} A &\equiv \alpha_1 \pmod{p_1}; \\ A &\equiv \alpha_2 \pmod{p_2}; \\ &\dots \\ A &\equiv \alpha_n \pmod{p_n}. \end{aligned}$$

Основным преимуществом такого представления является тот факт, что выполнение операций сложения, вычитания и умножения реализуется очень просто по формуле

$$\begin{aligned} A * B &= (\alpha_1, \alpha_2, \dots, \alpha_n) * (\beta_1, \beta_2, \dots, \beta_n) = \\ &= ((\alpha_1 * \beta_1) \pmod{p_1}, (\alpha_2 * \beta_2) \pmod{p_2}, \\ &\dots, (\alpha_n * \beta_n) \pmod{p_n}), \end{aligned}$$

где * обозначает одну из операций: сложение, вычитание или умножение. Эти операции носят название *модульных*, т. к. для их выполнения в СОК достаточно одного такта обработки численных значений, причем эта обработка происходит параллельно и величина числа в каждом разряде не зависит от других разрядов.

К достоинствам такого представления и обработки чисел относится также мало-разрядность остатков, что позволяет эффективно применять табличные методы обработки. Вычислительные системы, построенные на основе СОК, обладают высокой производительностью и надежностью [2]. Однако возникают серьезные трудности при реализации непозиционных процедур, к которым относятся: нахождение вычета (остатка) числа; определение знака числа (в СОК знак числа представлен в неявном виде); сравнение модулярных чисел; определение переполнения; операции деления, масштабирования, расширения, исправления ошибок и др. [3]. Выполнение этих операций является довольно проблематичным. Большинство приложений СОК не требуют использования этих операций. Фундаментальной операцией здесь является

определение знака модулярного числа, которое может использоваться при обнаружении переполнения динамического диапазона, сравнении величин чисел, исправлении ошибок и других операций, время выполнения которых может быть уменьшено до времени выполнения модульного деления вместе со сложением, вычитанием и умножением, а также масштабированием вместе с расширением [4].

Необходимо отметить, что даже в тех случаях, когда СОК ограничена приложениями, в которых преобладающими операциями являются сложение и вычитание, нет возможности полностью исключить проблематичные операции. Так, в вычислениях особенно важно масштабирование, т. к. во многих приложениях, для которых СОК особенно хорош, могут встретиться операции, приводящие к росту чисел, которые, в свою очередь, могут привести к переполнению [5]. Поэтому чтобы гарантировать, что все результаты лежат в пределах допустимого диапазона, необходимо проводить контроль в процессе производимых вычислений. В связи с этим возникает необходимость быстрого выполнения указанных выше операций.

Приближенный метод выполнения основных проблемных операций в системе остаточных классов

В настоящее время известны следующие методы определения позиционных характеристик (ПХ) модулярного представления чисел [2, 6]:

- ортогональных базисов;
- интервальных оценок;
- с использованием коэффициентов обобщенной позиционной системы счисления (ОПСС) и др.

Суть метода ортогональных базисов состоит в переходе к позиционному представлению через модуль всей системы $P = p_1 p_2 \dots p_n$, что разрушает идею модулярной арифметики. Недостаток метода ортогональных базисов заключается в том, что приходится иметь дело с большими числами ортогональных базисов и, кроме того, действия сложения и умножения надо выполнять в позиционной системе счисления, а полученный результат необходимо вводить в

диапазон вычитанием величины, кратной P , которая определяется рангом числа. Нахождение ранга числа связано с вычислительной сложностью. Метод интервальных оценок сокращает модуль до величины $\frac{P}{p_i}$, и только метод ОПСС позволяет выполнять преобразования по модулю p_i . Метод ОПСС универсальный. К недостаткам этого метода можно отнести его сложность и избыточность ПХ при вычислении некоторых немодульных процедур.

Метод ортогональных базисов и метод ОПСС – точные методы определения ПХ, а метод интервальных оценок – приближенный метод, к характеристике которого относится ПХ номера интервала числа, суть которого состоит в том, что числовой диапазон P разбивается на p_i интервалов

$$\left[j \frac{P}{p_i}, (j+1) \frac{P}{p_i} \right], \text{ при } j = 1, 2, \dots, p_i - 1.$$

Определение номера интервала, в котором расположено число, позволяет получить оценку немодульного числа по его величине с точностью до величины интервала, что ограничивает область его применения.

Процесс определения знака числа сводится к операции выявления принадлежности интервала, в котором находится число, представленное в СОК, к группе положительных или отрицательных интервалов по заданному p_i , на которые разбит диапазон P . Число интервалов определяется величиной $\frac{P}{p_i}$. В случае если диапазон разбивается на нечетное число интервалов по выбранному модулю p_i , т. е. все основания нечетные, то имеется критический интервал, который является границей между положительными и отрицательными числами. В этом случае критический интервал делится на две части, а процесс определения знака числа при этом сводится к сравнению остатка по модулю p_i , что резко усложняет процесс определения знака числа.

С целью повышения эффективности вычисления ПХ предлагается новый приближенный метод определения ПХ, позволяющий реализовать практически все немодульные процедуры модулярного кода.

Исторически так сложилось, что поиск некоторого компромисса в удовлетворении требований, предъявляемых к ПХ, привел исследователей к введению таких характеристик модулярной алгебры, как ранг, след, нормированный ранг, неточный ранг, ядро числа и др. [1, 5, 7]. Анализ этих ПХ показал, что значение модулярной величины по ним определяется сложно и не всегда однозначно. Кроме того, при выполнении некоторых немодульных операций нет необходимости в точном их определении, а достаточно знать значения в пределах каких-то интервалов, т. е. при определении этих характеристик появляется избыточная информация, которая не используется. Эта идея и подтолкнула к поиску такой ПХ, которая бы не содержала избыточной информации, на нахождение которой требуются дополнительные вычислительные ресурсы.

С другой стороны широко используются методы выполнения проблемных операций, основанные на выборе модулей СОК специального вида [8–10]. Однако такой подход накладывает существенные ограничения на используемую СОК, что сильно затрудняет его использование в некоторых приложениях, особенно в тех, где необходимо динамическое изменение структуры СОК: диапазона и модулей. К таким приложениям относятся, например, криптографическая защита информации и проектирование отказоустойчивых вычислительных систем, функционирующих в СОК.

Анализ немодульных операций показал, что их можно представить точно или приближенно, поэтому методы вычисления ПХ можно разделить на две группы:

методы точного вычисления ПХ;

методы приближенного вычисления ПХ.

Подробное описание методов точного вычисления ПХ можно найти, например, в [1–3, 6]. В данной статье исследуются приближенные методы вычисления ПХ, позволяющие существенно сократить аппаратные и временные затраты, обусловленные операциями, выполняемыми над позиционными кодами уменьшенной разрядности. В связи с этим возникает задача использования приближенного метода при вычислении определенного ряда немодуль-

ных процедур: определения интервалов чисел, знака числа, сравнения, насколько одно число больше или меньше другого, в том случае, когда не требуется знания точного значения.

С целью упрощения процесса сравнения модулярных чисел рассмотрим приближенный метод, позволяющий реализовать основные классы процедур принятия решений: анализ наличия определенного значения в конкретном разряде; проверку равенства (неравенства) двух значений; сравнение двух значений (больше, меньше), обеспечивающих решение основного круга задач, возникающих при аппаратной или программной реализации реальных процессов.

Суть приближенного метода сравнения модулярных чисел основана на использовании относительных величин анализируемых чисел к полному диапазону, определяемому Китайской теоремой об остатках [1], которая связывает позиционное число A с его представлением в остатках $(\alpha_1, \alpha_2, \dots, \alpha_n)$, где α_i – наименьшие неотрицательные вычеты числа, относительно модулей системы остаточных классов p_1, p_2, \dots, p_n следующим выражением:

$$A = \left\lfloor \sum_{i=1}^n \frac{P}{p_i} |P_i^{-1}|_{p_i} \alpha_i \right\rfloor_P, \quad (2)$$

где p_i – модули СОК; $P = \prod_{i=1}^n p_i$ – диапазон СОК; $P_i = \frac{P}{p_i} = p_1 p_2 \dots p_{i-1} \dots p_{i+1} p_n$ – произведение всех модулей СОК, кроме i -го; $|P_i^{-1}|$ – мультипликативная инверсия P_i относительно p_i .

Если левую и правую части выражения (2) разделить на величину P , соответствующую диапазону чисел, то получим приближенное значение

$$\left\lfloor \frac{A}{P} \right\rfloor = \left\lfloor \sum_{i=1}^n \frac{|P_i^{-1}|_{p_i}}{p_i} \alpha_i \right\rfloor \approx \left\lfloor \sum_{i=1}^n k_i \alpha_i \right\rfloor, \quad (3)$$

где $k_i = \frac{|P_i^{-1}|_{p_i}}{p_i}$ – константы выбранной системы; α_i – разряды числа, представленного в СОК, при этом значение каждой суммы будет в интервале $[0,1)$. Конечный резуль-

тат суммы определяется после суммирования и отбрасывания целой части числа с сохранением дробной части суммы. Дробная часть $F(A) = \left\lfloor \frac{A}{P} \right\rfloor \in [0,1)$ содержит информацию как о величине числа, так и о его знаке. Если $\left\lfloor \frac{A}{P} \right\rfloor \in \left[0, \frac{1}{2}\right)$, то число A положительное и $F(A)$ равна величине A , разделенной на P . В противном случае A – отрицательное число и $1 - F(A)$ показывает относительную величину числа A по отношению к P . Исследования показали, что функция $F(A)$ может использоваться при разработке алгоритмов вычисления основных проблемных операций в системе остаточных классов [11]. Целая часть числа представляет собой ранг числа, т. е. такую непозиционную характеристику, которая показывает, сколько раз диапазон системы P был превзойден при переходе от представления чисел в системе остаточных классов к его позиционному представлению. При необходимости определение ранга числа может производиться непосредственно в процессе выполнения операции суммирования констант k_i . Дробная часть может быть записана так же как $A \bmod 1$, потому что $A = \lfloor A \rfloor + A \bmod 1$ [4]. Количество разрядов дробной части числа определяется максимально возможной разностью между соседними числами. При необходимости точного сравнения необходимо вычислить значение (3), которое является эквивалентом преобразования из СОК в позиционную систему счисления. Для решения задач основных процедур принятия решения достаточно знать приблизительно значения чисел $\left\lfloor \frac{A}{P} \right\rfloor$ и $\left\lfloor \frac{B}{P} \right\rfloor$ по отношению к динамическому диапазону $[0,1)$, которое выполняется достаточно просто, но при этом верно определяется соотношениями $A = B$, $A > B$ или $A < B$.

Итак, приближенный метод вычисления ПХ может описываться следующей последовательностью действий [4]:

1. Вычисление констант СОК $k_i = \frac{|P_i^{-1}|_{p_i}}{p_i}$ с требуемой точностью.



2. Вычисление приближенных значений $\alpha_i k_i$, где k_i – константы, найденные в п. 1, $1 \leq \alpha_i \leq p_i - 1$.

3. Вычисление приближенного значения ПХ $\left| \sum_{i=1}^n k_i \alpha_i \right|$ в интервале $[0, 1)$.

Рассмотрим использование информации, содержащейся в $F(A)$, для вычисления проблемных операций в СОК.

Конструируются некоторые правила Ψ_i , $i = 1, \dots, 4$, согласно которым вычисляется i -я немодульная операция (определение знака числа, сравнение чисел, обнаружение ошибки и переполнения, а также локализация ошибочного разряда).

Правило Ψ_1 . Определение знака числа в случае, если $p_1 = 2$:

если $\left| \frac{A}{P_1} \right| < \frac{1}{2}$, то число положительное;

если $\left| \frac{A}{P_1} \right| > \frac{1}{2}$, то число отрицательное.

Правило Ψ_2 . Сравнение модулярных чисел A и B :

если $\left| \frac{A}{P_1} \right| - \left| \frac{B}{P_1} \right| = 0$, то $A = B$;

если $\left| \frac{A}{P_1} \right| - \left| \frac{B}{P_1} \right| > 0$, то $A > B$;

если $\left| \frac{A}{P_1} \right| - \left| \frac{B}{P_1} \right| < 0$, то $A < B$.

Правило Ψ_3 . Обнаружение ошибки и переполнения динамического диапазона:

если $\left| \frac{\bar{A}}{P_{\text{изб}}} \right| < \left| \frac{M}{P_{\text{изб}}} \right|$, то ошибки нет, где \bar{A} – искаженное число; $P_{\text{изб}} = p_{n+1} p_{n+2} P$ – избыточный диапазон при двух избыточных модулях p_{n+1} и p_{n+2} ; $M = P = \prod_{i=1}^n p_i$ – рабочий диапазон;

если $\left| \frac{\bar{A}}{P_{\text{изб}}} \right| \geq \left| \frac{M}{P_{\text{изб}}} \right|$, то ошибка есть и установлено переполнение динамического диапазона.

Правило Ψ_4 . Локализация неисправного канала:

если $\left| \frac{\bar{A}_i}{P_i} \right| < \left| \frac{M_i}{P_i} \right|$, то в разряде i нет ошибки;

если $\left| \frac{\bar{A}_i}{P_i} \right| \geq \left| \frac{M_i}{P_i} \right|$, то в разряде i есть

ошибка, где $\bar{A}_i = (\alpha_1, \alpha_2, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_n, \alpha_{n+1}, \alpha_{n+2})$ – проекция искаженного числа \bar{A} , $M_i = (m_1, m_2, \dots, m_{i-1}, m_{i+1}, \dots, m_n, m_{n+1}, m_{n+2})$ – проекция рабочего диапазона.

Алгоритм определения знака модулярного числа и его техническая реализация

Основой для выполнения немодульных операций, указанных в правилах Ψ_2 – Ψ_4 , является немодульная операция определения знака числа (правило Ψ_1). Пусть для вычислений используются константы $\bar{k}_i^{(N)}$, представленные в виде двоичных дробей, содержащих N двоичных знаков после запятой, полученные отбрасыванием всех остальных разрядов, начиная с $N+1$ -го. Тогда точные значения констант K_i будут лежать в диапазоне

$$\bar{k}_i^{(N)} \leq k_i \leq \bar{k}_i^{(N)} + 2^{-N}. \quad (4)$$

Для определения ПХ числа в СОК необходимо вычислять величину $\left| \sum_{i=1}^n k_i \alpha_i \right|$ из формулы (3).

Утверждение. Наименьшее значение N_{min} , при котором восстановление позиционной величины числа из представления в СОК по формуле (3) будет корректным, равно

$$N_{\text{min}} = \lceil \log_2 (P\rho) \rceil, \quad (5)$$

где $\rho = -n + \sum_{i=1}^n p_i$.

Доказательство. Из (4) следует, что точное значение величины $\sum_{i=1}^n k_i \alpha_i$ из формулы (3) будет лежать в диапазоне

$$\begin{aligned} \sum_{i=1}^n \bar{k}_i^{(N)} \alpha_i &\leq \sum_{i=1}^n k_i \alpha_i \leq \\ &\leq \sum_{i=1}^n \bar{k}_i^{(N)} \alpha_i + 2^{-N} \sum_{i=1}^n \alpha_i. \end{aligned} \quad (6)$$

Так как величина $\left| \sum_{i=1}^n k_i \alpha_i \right|$ соответствует точному местоположению числа A на числовой оси, то для однозначного (точного) определения величины числа A не-

обходимо таким образом подобрать параметр N , чтобы в произвольный интервал $\left[\sum_{i=1}^n \bar{k}_i^{(N)} \alpha_i, \sum_{i=1}^n \bar{k}_i^{(N)} \alpha_i + 2^{-N} \sum_{i=1}^n \alpha_i \right]$ попадало лишь одно возможное число из диапазона СОК. Это требование равносильно условию $2^{-N} \sum_{i=1}^n \alpha_i < \frac{1}{P}$.

Если обозначить $\rho = -n + \sum_{i=1}^n p_i$, то наименьшее значение N_{\min} , при котором возможно точное восстановление позиционной формы числа с использованием формулы (3), определяется формулой $N_{\min} = \lceil \log_2(P\rho) \rceil$. Утверждение доказано.

При снижении точности вычислений (уменьшении величины N) диапазон (6) будет увеличиваться так, что в него будут попадать несколько чисел. Однако при решении задачи определения знака числа с использованием правила Ψ_1 возможно использование таких расширенных диапазонов, если учитывать появление зон ошибочного определения знака числа.

Пусть в процессе вычислений используются константы, округленные до N двоичных знаков после запятой, $1 \leq N < N_{\min}$. Тогда в интервал $\left[\sum_{i=1}^n \bar{k}_i^{(N)} \alpha_i, \sum_{i=1}^n \bar{k}_i^{(N)} \alpha_i + 2^{-N} \sum_{i=1}^n \alpha_i \right]$

попадет не более $2^{-N} P \sum_{i=1}^n \alpha_i$ чисел СОК. Так как $0 \leq \alpha_i \leq p_{i-1}$, то

$$\max \left\{ 2^{-N} P \sum_{i=1}^n \alpha_i \right\} = 2^{-N} P \left(-n + \sum_{i=1}^n p_i \right) = 2^{-N} P \rho.$$

Ввиду увеличения интервала будет происходить наложение его границ на ту область числовой оси, где определенная по формуле $\left\{ \sum_{i=1}^n \bar{k}_i^{(N)} \alpha_i \right\}$ и реальный знак чис-

ла в СОК из формулы $\left\{ \sum_{i=1}^n k_i \alpha_i \right\}$ могут не совпадать. На рис. 1 а показаны возможные направления появления зон неопределенности при округлении констант [12]. На рис. 1 б показано появление зон неопределенности при отбрасывании разрядов вместо округления. В последнем случае формируются только две зоны неопределенности вместо четырех. Это позволяет в два раза уменьшить количество проверяемых условий в процессе выполнения алгоритма, что позволяет существенно упростить его формулировку и реализацию. Стрелками отмечены направления роста зон ошибочного определения знака числа при снижении точности. Зонам, изображенным на рис. 1 б, соответствуют следующие диапазоны:

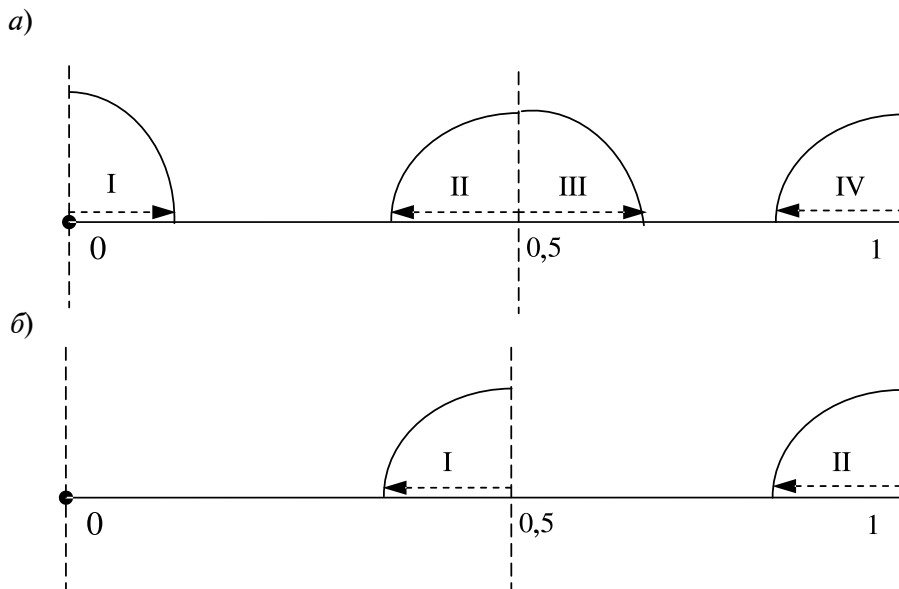


Рис. 1. Появление зон ошибочного определения знака числа в СОК: а – при округлении констант; б – при отбрасывании разрядов

зона ошибочного определения знака I: $[0,5 - 2^{-N\rho}, 0,5)$;

зона ошибочного определения знака II: $[1 - 2^{-N\rho}, 1)$.

Для определения знака числа необходимо выполнить проверку двух условий:

если $0 < \frac{A}{P} < 0,5 - 2^{-N\rho}$, то число A – положительное;

если $0,5 < \frac{A}{P} < 1 - 2^{-N\rho}$, то число A – отрицательное.

В том случае, если не выполняется ни одно из указанных условий, число попадает в одну из зон неопределенности и требуется дополнительная итерация для получения точного результата с использованием N_{\min} знаков после запятой (5).

На рис. 2 приведена схема для определения знака модулярного числа, представленного по четырем модулям p_i , $i = 1, 2, 3, 4$. Схема содержит входные регистры RG_i , $\forall i = [1...4]$ для временного хранения остатков чисел по соответствующим модулям, просмотрные таблицы LUT_i , $\forall i = [1...4]$ для хранения произведений $\left\lfloor \frac{P_i^{-1}}{p_i} \right\rfloor \cdot \alpha_i$ и параллельный сумматор.

Схема работает следующим образом. Код числа A , для которого необходимо определить интервал, что равносильно определению знака числа, поступает на входные регистры RG_i в двоичном коде (каждый разряд СОК кодируется двоичным кодом). Сигналы с выходов регистров

поступают на входы просмотрных таблиц LUT. В просмотрных таблицах хранятся произведения констант k_i и остатков α_i , то

есть $\left\lfloor \frac{P_i^{-1}}{p_i} \right\rfloor \alpha_i$, представленных в естественной форме двоичной дроби в дополнительном коде. Количество элементов памяти (N) просмотрных таблиц определяется выражением $N = \sum_{i=1}^n p_i$.

Выходные сигналы просмотрных таблиц в дополнительном двоичном коде поступают на вход сумматора, в котором уже записана константа $0,5 - 2^{-N\rho}$, во время начальной установки. (Дополнительный код используется для того чтобы операцию вычитания заменить операцией сложения). Знак результата сложения определяет интервал (первый или второй), что соответственно определяет знак числа.

Моделирование алгоритма

Для моделирования разработанного алгоритма определения знака числа с использованием ПХ на основе приближенного метода были выбраны следующие СОК:

1. СОК₁ {2, 3, 5, 7}.
2. СОК₂ {7, 17, 19, 29}.
3. СОК₃ {2, 3, 5, 11, 13, 19, 23, 29, 79}.

Данный выбор объясняется тем, что СОК₁ наиболее просто и наглядно отображает свойства вычислений в остаточных классах и используется многими авторами (например [6]); СОК₂ и СОК₃ позволяют

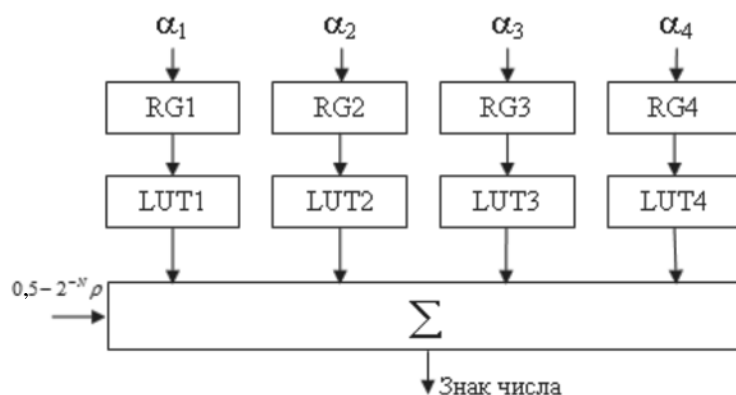


Рис. 2. Схема определения знака числа

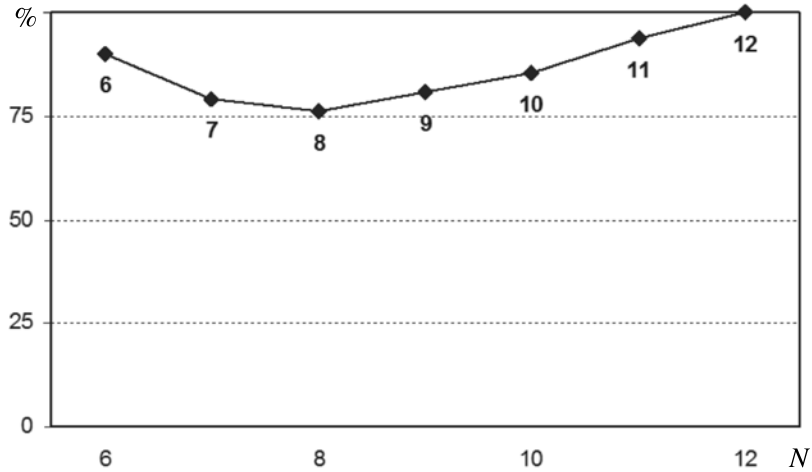


Рис. 3. График зависимости времени работы алгоритма (% от максимального) от точности вычислений для СОК₁

представить числовые диапазоны в 16 бит и 32 бит соответственно [2].

При моделировании замерялось время выполнения операции определения знака числа при постепенном уменьшении величины $N \leq N_{\min}$. Для СОК₁ имеем следующие характеристики: $P = 210$, $\rho = 16$, $N_{\min} = 12$. Результаты моделирования разработанного алгоритма для СОК₁ показаны на рис. 3. На рисунке изображено, насколько (в процентах) снижается время работы алгоритма при выборе различных значений N . При снижении величины N от $N_{\min} = 12$ до $N = 8$ происходит уменьшение времени работы алгоритма за счет уменьшения разрядности обрабатываемых чисел. Однако при дальнейшем уменьшении N начинается увеличение времени

выполнения алгоритма. Этот факт объясняется тем, что при $N < 8$ зоны неопределенности при определении знака числа увеличиваются настолько, что уже значительная доля чисел из диапазона СОК требует дополнительной (уточняющей) итерации алгоритма с максимальной точностью вычислений.

Наилучший показатель времени работы алгоритма достигается при $N = 8$, в этом случае время работы алгоритма составляет $\approx 76,2\%$ от максимального времени работы алгоритма при $N_{\min} = 12$. Таким образом, использование предложенного алгоритма для СОК₁ позволяет сократить время определения знака числа примерно в 1,31 раз по сравнению с точными методами.

Для СОК₂ имеем: $P = 65569$, $\rho = 68$,

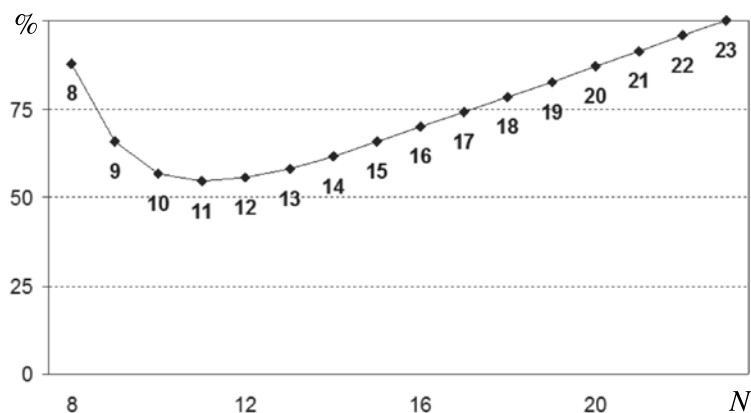


Рис. 4. График зависимости времени работы алгоритма (% от максимального) от точности вычислений для СОК₂

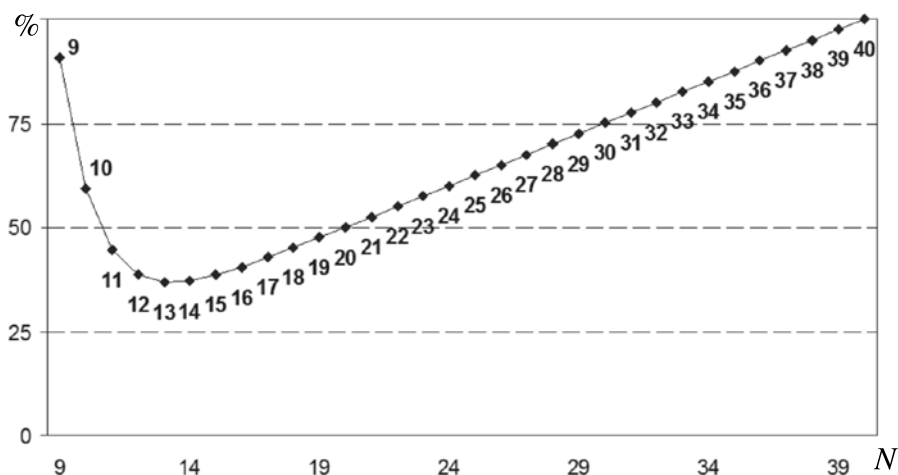


Рис. 5. График зависимости времени работы алгоритма (% от максимального) от точности вычислений для СОК₃

$N_{\min} = 23$. Результаты моделирования разработанного алгоритма для СОК₂ показаны на рис. 4. Поведение кривой скорости алгоритма аналогично случаю СОК₁. Снижение величины N от $N_{\min} = 23$ до $N = 11$ приводит к уменьшению времени работы алгоритма. При дальнейшем уменьшении N начинается увеличение времени выполнения алгоритма.

Самая высокая скорость работы алгоритма наблюдается при $N = 11$, в этом случае время работы алгоритма составляет $\approx 54,5\%$ от максимального времени работы алгоритма при $N_{\min} = 74$. Таким образом, использование предложенного алгоритма для СОК₂ позволяет сократить время определения знака числа примерно в 1,84 раз по сравнению с точными методами.

Наконец, СОК₃ имеет параметры: $P = 4295006430$, $\rho = 175$, $N_{\min} = 40$. Результаты моделирования разработанного алгоритма для СОК₃ показаны на рис. 5. Снова наблюдается снижение времени работы алгоритма при снижении точности до $N = 13$. При дальнейшем снижении время работы алгоритма увеличивается.

Самая высокая скорость работы алгоритма наблюдается при $N = 13$, в этом случае время работы алгоритма составляет $\approx 36,8\%$ от максимального времени работы алгоритма при $N_{\min} = 40$. Таким образом, использование предложенного алгоритма для СОК₂ позволяет сократить время опре-

деления знака числа примерно в 2,72 раз по сравнению с точными методами.

Противоречие между вычислительной сложностью определения основных проблемных процедур в СОК и их быстродействием разрешено путем замены абсолютных величин их относительными значениями и простотой их вычисления, которая сохраняет адекватную связь числовых значений модулярных величин с их представлениям в СОК и позволяет существенно повысить скорость выполнения немодульных операций. Благодаря этому применение СОК может дать значительные преимущества не только в тех приложениях, в которых основная доля вычислений приходится на точное умножение, возведение в степень больших чисел в сочетании со сложением и вычитанием, но и в которых часто появляется необходимость в делении либо сравнении и определении знака числа, а также при проверке не «выходят» ли результаты за пределы допустимых значений и др.

Решена фундаментальная проблема реализации основных проблемных операций в СОК, которые ранее определяли наибольший вклад в алгоритмическую сложность и сдерживали широкое применение СОК при разработке новых классов вычислительных систем. Предложенная техническая реализация приближенного метода определения знака числа проста для применения

на практике, поэтому внедрение полученных результатов позволит расширить область применения модулярной арифметики. Полученные новые результаты эффективного выполнения немодульных процедур являются развитием теории математических основ разработки и проектирования высокопроизводительных и надежных вычислительных систем, функционирующих

в системе остаточных классов. Применение предложенного алгоритма позволяет сократить время выполнения проблемной операции в 1,31–2,72 раза (в зависимости от диапазона СОК) по сравнению с известными методами.

Работа выполнена при финансовой поддержке РФФИ (проекты № 13-07-00478-а) и ФЦП № 14.В37.21.1128.

СПИСОК ЛИТЕРАТУРЫ

1. **Акушский, И.Я.** Машинная арифметика в остаточных классах [Текст] / И.Я. Акушский, Д.И. Юдицкий. – М.: Сов. радио, 1968. – 440 с.
2. **Червяков, Н.И.** Модулярные параллельные вычислительные структуры нейропроцессорных систем [Текст] / Н.И. Червяков, П.А. Сахнюк, А.В. Шапошников, С.А. Ряднов. – М.: Физматлит, 2003. – 288 с.
3. **Червяков, Н.И.** Нейрокомпьютеры в остаточных классах [Текст] / Н.И. Червяков, П.А. Сахнюк, А.В. Шапошников, А.Н. Макоха; под ред. А.И. Галушкина. – М.: Радиотехника, 2003. – 272 с.
4. **Червяков, Н.И.** Методы, алгоритмы и техническая реализация основных проблемных операций, выполняемых в системе остаточных классов [Текст] / Н.И. Червяков // Инфокоммуникационные технологии. – 2011. – № 4. – С. 4–12.
5. **Burgess, N.** Scaling an RNS Number Using the Core Function [Text] / N. Burgess // XVI IEEE Symp. Computer Arithmetic, Santiago de Compostella. – 2003. – P. 262–271.
6. **Omondi, A.** Residue Number Systems: Theory and Implementation [Text] / A. Omondi, B. Premkumar. – Imperial College Press, 2007. – 296 p.
7. **Амербаев, В.М.** Теоретические основы ма-

шинной арифметики [Текст] / В.М. Амербаев. – Алма-Ата: Наука, 1976. – 324 с.

8. **Gallagher, D.** The digit parallel method for fast RNS to weighted number system conversion for specify moduli $(2^k-1, 2^k, 2^{k+1})$ [Text] / D. Gallagher, F.E. Petry, P. Sirmivasan // IEEE Trans. on Circuits and System II: Analog and Digital Signal Processing. – 1997. – Vol. 44. – № 1. – P. 53–57.

9. **Tomczak, T.** Fast sign detection for rns $(2^n-1, 2^n, 2^{n+1})$ [Text] / T. Tomczak // IEEE Trans. on Circuits and Systems-I: Regular papers. – 2008. – Vol. 55. – № 6. – P. 1502–1511.

10. **Wang, Y.** Adder based residue to binary numbers converters for $(2^n-1, 2^n, 2^{n+1})$ [Text] / Y. Wang, M. Aboulhamid, H. Shen // IEEE Trans. Signal Processing. – 2002. – Vol. 50. – № 7. – P. 1772–1779.

11. **Hung, C.Y.** An Approximate Sign Detection Method for Residue Numbers and its Application to RNS Division [Text] / C.Y. Hung, B. Parhami // Computers Math. Applic. – 1994. – Vol. 27. – № 4. – P. 23–35.

12. **Червяков, Н.И.** Метод определения знака числа в системе остаточных классов на основе приближенных вычислений [Текст] / Н.И. Червяков, П.А. Ляхов. – Нейрокомпьютеры: разработка, применение. – 2012. – № 12. – С. 56–64.

REFERENCES

1. **Akushskii I.Ia., Iuditskii D.I.** Mashinnaiia arifmetika v ostatochnykh klassakh. – Moscow: Sov. radio, 1968. – 440 s. (rus).
2. **Cherviakov N.I., Sakhniuk P.A., Shaposhnikov A.V., Riadnov S.A.** Moduliarnye parallel'nye vychislitel'nye struktury neuroprotsessornykh sistem. – Moscow: Fizmatlit, 2003. – 288 s. (rus)
3. **Cherviakov N.I., Sakhniuk P.A., Shaposhnikov A.V., Makokha A.N.** Neirokomp'iutery v ostatochnykh klassakh; pod red. A.I. Galushkina. – Moscow: Radiotekhnika, 2003. – 272 s. (rus)
4. **Cherviakov N.I.** Metody, algoritmy i tekhnicheskaiia realizatsiia osnovnykh problemnykh operatsii, vypolniaemykh v sisteme ostatochnykh

klassov / Infokommunikatsionnye tekhnologii. – 2011. – № 4. – S. 4–12. (rus)

5. **Burgess N.** Scaling an RNS Number Using the Core Function / XVI IEEE Symp. Computer Arithmetic, Santiago de Compostella. – 2003. – P. 262–271.

6. **Omondi A., Premkumar B.** Residue Number Systems: Theory and Implementation. – Imperial College Press, 2007. – 296 p.

7. **Amerbaev V.M.** Teoreticheskie osnovy mashinnoi arifmetiki. – Alma-Ata: Nauka, 1976. – 324 s. (rus)

8. **Gallagher D., Petry F.E., Sirmivasan P.** The digit parallel method for fast RNS to weighted



number system conversion for specify moduli $(2k-1, 2k, 2k+1)$ / IEEE Trans. on Circuits and System II: Analog and Digital Signal Processing. – 1997. – Vol. 44. – № 1. – P. 53–57.

9. **Tomczak T.** Fast sign detection for rns $(2n-1, 2n, 2n+1)$ / IEEE Trans. on Circuits and Systems-I: Regular papers. – 2008. – Vol.55. – № 6. – P. 1502–1511.

10. **Wang Y., Aboulhamid M., Shen H.** Adder based residue to binary numbers converters for $(2n-1, 2n, 2n+1)$ / IEEE Trans. Signal Processing.

– 2002. – Vol.50. – № 7. – P. 1772–1779.

11. **Hung C.Y., Parhami B.** An Approximate Sign Detection Method for Residue Numbers and its Application to RNS Division / Computers Math. Applic. – 1994. – Vol. 27. – № 4. – P. 23–35.

12. **Cherviakov N.I., Liakhov P.A.** Metod opredeleniia znaka chisla v sisteme ostatochnykh klassov na osnove priblizhennykh vychislenii / Neurokomp'iutery: razrabotka, primenenie. – 2012. – № 12. – S. 56–64. (rus)

ЧЕРВЯКОВ Николай Иванович – *заведующий кафедрой прикладной математики и математического моделирования Института математики и естественных наук Северо-Кавказского федерального университета, доктор технических наук, профессор.*

355000, Россия, г. Ставрополь, пр. Кулакова, д. 2.

E-mail: k-fmf-primath@stavsru

CHERVYAKOV, Nikolay I. *North-Caucasian Federal University.*

355000, prosp. Kulakova 2, Stavropol, Russia.

E-mail: k-fmf-primath@stavsru

БАБЕНКО Михаил Григорьевич – *доцент кафедры прикладной математики и математического моделирования Института математики и естественных наук Северо-Кавказского федерального университета, кандидат физико-математических наук.*

355000, Россия, г. Ставрополь, пр. Кулакова, д. 2.

E-mail: whbear@yandex.ru

BABENKO, Michael G. *North-Caucasian Federal University.*

355000, prosp. Kulakova 2, Stavropol, Russia.

E-mail: whbear@yandex.ru

ЛЯХОВ Павел Алексеевич – *доцент кафедры прикладной математики и математического моделирования Института математики и естественных наук Северо-Кавказского федерального университета, кандидат физико-математических наук.*

355000, Россия, г. Ставрополь, пр. Кулакова, д. 2.

E-mail: ljahov@mail.ru

LYAKHOV, Pavel A. *North-Caucasian Federal University.*

355000, prosp. Kulakova 2, Stavropol, Russia.

E-mail: ljahov@mail.ru

ЛАВРИНЕНКО Ирина Николаевна – *доцент кафедры высшей алгебры и геометрии Института математики и естественных наук Северо-Кавказского федерального университета, кандидат физико-математических наук.*

355017, Россия, г. Ставрополь, пр. Кулакова, д. 2.

LAVRINENKO, Irina N. *North-Caucasian Federal University.*

355000, prosp. Kulakova 2, Stavropol, Russia.

УДК 519.7

С.А. Диченко, Н.И. Елисеев, О.А. Финько

КОНТРОЛЬ ОШИБОК ФУНКЦИОНИРОВАНИЯ ГЕНЕРАТОРОВ ДВОИЧНЫХ ПСП, РЕАЛИЗОВАННЫХ НА АРИФМЕТИЧЕСКИХ ПОЛИНОМАХ

S.A. Dichenko, N.I. Eliseev, O.A. Finko

ERROR FUNCTION GENERATOR BINARY PRS CONTROL IMPLEMENTED ON ARITHMETIC POLYNOMIALS

Предложена методика повышения безопасности функционирования средств криптографической защиты информации (СКЗИ), в частности, узлов формирования двоичных псевдослучайных последовательностей (ПСП), действующих в условиях помех, генерируемых злоумышленником. Системы булевых характеристических уравнений реализуются линейными арифметическими полиномами, позволяющими распараллелить процесс вычисления элементов ПСП. «Арифметизация» логического счета, в свою очередь, позволила применить аппарат избыточных модулярных кодов для контроля ошибок функционирования узлов генерации ПСП и обеспечить тем самым, необходимую безопасность их функционирования в составе СКЗИ.

ДВОИЧНАЯ ПСЕВДОСЛУЧАЙНАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ. ПАРАЛЛЕЛЬНЫЕ ЛОГИЧЕСКИЕ ВЫЧИСЛЕНИЯ ПОСРЕДСТВОМ АРИФМЕТИЧЕСКИХ ПОЛИНОМОВ. МОДУЛЯРНАЯ АРИФМЕТИКА. КОНТРОЛЬ ОШИБОК ФУНКЦИОНИРОВАНИЯ СРЕДСТВ КРИПТОГРАФИЧЕСКОЙ ЗАЩИТЫ ИНФОРМАЦИИ. ГЕНЕРАЦИЯ АППАРАТНЫХ ОШИБОК. КРИПТОГРАФИЯ. ШИФРУЮЩАЯ ГАММА.

A method of improving safety of the cryptographic protection of information (CPS), in particular the formation of binary nodes pseudorandom sequence (PRS), operating in a noise generated by an attacker. System of Boolean equations realize linear characteristic polynomial arithmetic, allowing parallelize the process of calculating the elements of the PRS. «Arithmetization» logical accounts, in turn, allowed the use of redundant modular device codes for error control operation of generating units and to provide bandwidth, thus the necessary security of their operation in the CPS.

BINARY PSEUDORANDOM SEQUENCE. PARALLEL LOGICAL CALCULATIONS BY POLYNOMIALS ARITHMETIC. MODULAR ARITHMETIC. THE ERROR CONTROL OPERATION OF THE CRYPTOGRAPHIC PROTECTION OF INFORMATION. GENERATION OF HARDWARE ERRORS. CRYPTOGRAPHY. CIPHER SCHEME.

Из перечня известных атак на СКЗИ важным является новый малоизученный вид атак, основанный на генерации аппаратных ошибок функционирования узлов СКЗИ [1]. Выработка мер защиты от данного вида атак необходима для решения задач обеспечения безопасности функционирования СКЗИ. Безопасность функционирования СКЗИ обеспечивается в т. ч. и за счет повышения достоверности их функционирования. В настоящее время необходимый уровень достоверности функционирования СКЗИ достигается и с помощью привле-

чения избыточного оборудования (резервирования), и с привлечением временной избыточности за счет различного рода повтора вычислений (реализации прямых и обратных преобразований с последующим сравнением результатов) [2].

Известно, что хорошие результаты для повышения достоверности функционирования цифровых устройств дают различные методы избыточного кодирования. Однако для логических типов данных, подверженных криптопреобразованиям, обеспечение кодового контроля вызывает множество за-

труднений [3]. В то же время известно, что контроль ошибок арифметических вычислений может эффективно обеспечиваться за счет использования методов избыточного модулярного кодирования, применение которых для осуществления контроля логических типов данных стало возможным, благодаря полученной в [4, 5] возможности представления логических операций арифметическими выражениями, в частности, арифметическими полиномами.

Цель статьи – повышение безопасности функционирования узлов СКЗИ методами модулярной арифметики.

Алгоритм генерации двоичных ПСП, реализованный на арифметических полиномах

Одним из основных узлов СКЗИ, как известно [1], наиболее подверженных атакам, основанных на генерации аппаратных ошибок, являются генераторы двоичных ПСП, т. к. от качества их функционирования напрямую зависит качество функционирования СКЗИ.

Генератор ПСП имеет важнейшее значение для различных криптоалгоритмов и систем генерации ключевого материала [6–8]. Наиболее распространенными и проверенными практикой являются алгоритмы генерации ПСП, основанные на использовании рекуррентных логических выражений и неприводимых полиномов [6–8].

В частности, наиболее простым по структуре является рекуррентный регистр сдвига с обратной связью, реализуемой некоторой функцией f (см. рисунок).

Из [9–14] известно, что большинство криптографических функций можно реализовать посредством арифметических полиномов. В частности, в [9, 13, 14] рассмотрены параллельные генераторы ПСП, основанные

на линейных числовых полиномах (ЛЧП), где w -й блок участка двоичной ПСП можно представить посредством одного ЛЧП. Благодаря этому методу на выходе генератора может быть получен не один, а блок новых элементов ПСП необходимой длины.

Суть метода состоит в следующем. Пусть имеется характеристическое уравнение:

$$x_q = x_{q+\varphi-\tau} \oplus x_{q-\tau},$$

где $x_q, x_{q+\varphi-\tau}, x_{q-\tau} \in \{0, 1\}$; $q \geq t$; $q \in N$, полученное на основе тринома (частный случай):

$$D(\chi) = \chi^\tau + \chi^\varphi + 1,$$

где τ – степень тринома, $\tau \in N$, $1 < \varphi < \tau - 1$, $\varphi \in N$.

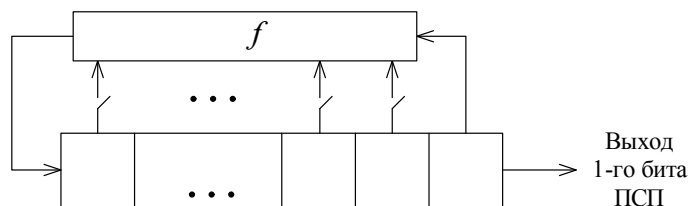
В соответствии с [13, 14] получим систему характеристических уравнений для участка ПСП длины τ :

$$\begin{cases} x_q = x_{q+\varphi-\tau} \oplus x_{q-\tau}, \\ x_{q+1} = x_{q+\varphi-\tau+1} \oplus x_{q-\tau+1}, \\ \dots \\ x_{q+\tau-1} = x_{q+\varphi-1} \oplus x_{q-1}, \end{cases}$$

где $[x_{q-\tau} \ x_{q-\tau+1} \ \dots \ x_{q-1}]$ – вектор начальных условий; $[x_q \ x_{q+1} \ \dots \ x_{q+\tau-1}]$ – вектор участка ПСП; $x_\varphi \in \{0, 1\}$; $\varphi = q - \tau + 1, \dots, q + \tau - 1$.

Систему характеристических уравнений представим как систему булевых функций (БФ), которую в свою очередь, в соответствии с правилами, приведенными в [4, 5, 15], преобразуем в систему ЛЧП:

$$\begin{cases} L_q(x_{q-\tau}, x_{q-\tau+1}, \dots, x_{q-1}) = \sum_{i=q-\tau}^{q-1} g_{q,i} x_i, \\ L_{q+1}(x_{q-\tau}, x_{q-\tau+1}, \dots, x_{q-1}) = \sum_{i=q-\tau}^{q-1} g_{q+1,i} x_i, \\ \dots \\ L_{q+\tau-1}(x_{q-\tau}, x_{q-\tau+1}, \dots, x_{q-1}) = \sum_{i=q-\tau}^{q-1} g_{q+\tau-1,i} x_i, \end{cases}$$



Общий вид рекуррентного регистра сдвига с обратной связью

где $g_{\varepsilon,i}$ (здесь и далее) принимает значение ноль или единица в зависимости от вхождения в ε -й ЛЧП x_i ; $\varepsilon = q, q + 1, \dots, q + \tau - 1$. Результат вычисления ε -го ЛЧП системы представим двоичным машинным словом длины $l_\varepsilon = \left\lceil \log \left(\sum_{i=q-\tau}^{q-1} g_{\varepsilon,i} \right) \right\rceil + 1$.

Полученную систему ЛЧП представим посредством одного ЛЧП:

$$\begin{aligned} U &= L(x_{q-\tau}, x_{q-\tau+1}, \dots, x_{q-1}) = \\ &= L_q(x_{q-\tau}, x_{q-\tau+1}, \dots, x_{q-1}) + \\ &+ \sum_{j=q+1}^{q+\tau-1} 2^{\gamma_j} L_j(x_{q-\tau}, x_{q-\tau+1}, \dots, x_{q-1}) = \\ &= g_{q,q-\tau} x_{q-\tau} + \dots + g_{q,q-1} x_{q-1} + \dots \\ &\dots + 2^{\gamma_{q+d-1}} g_{q+d-1,q-\tau} x_{q-\tau} + \dots \\ &\dots + 2^{\gamma_{q+d-1}} g_{q+d-1,q-1} x_{q-1} = h_{q-\tau} x_{q-\tau} + \\ &+ h_{q-\tau+1} x_{q-\tau+1} + \dots + h_{q-1} x_{q-1}, \end{aligned}$$

где $\gamma_j = \sum_{\varepsilon=q}^{j-1} (l_\varepsilon + 1)$; $h_i \in Z$;
 $i = q - \tau, q - \tau + 1, \dots, q - 1$.

Запишем ЛЧП следующим образом:

$$\begin{aligned} U &= L(X) = h_0 + \sum_{i=1}^r h_i x_i = \\ &= h_0 + h_1 x_1 + \dots + h_r x_r, \end{aligned} \tag{1}$$

где коэффициенты h_0, h_1, \dots, h_r — целые числа.

Первый способ контроля функционирования генераторов ПСП

В модулярной арифметике (МА) целый неотрицательный коэффициент ЛЧП (1) h_t ($t = 0, 1, \dots, r$) может быть однозначно представлен набором остатков по основаниям МА ($m_1, m_2, \dots, m_n < m_{n+1} < \dots < m_k$ — попарно простые, $M_n = m_1 m_2 \dots m_n > h_t$):

$$h_t = (\alpha_1, \alpha_2, \dots, \alpha_n, \alpha_{n+1}, \dots, \alpha_k)_{MA}, \tag{2}$$

где $\alpha_j = |h_t|_{m_j}$; $j = 1, 2, \dots, n, n + 1, \dots, k$; $| \bullet |_{m_j}$ — наименьший неотрицательный остаток числа \bullet по модулю m .

При этом остатки $\alpha_1, \alpha_2, \dots, \alpha_n$ являются информационными, а $\alpha_{n+1}, \dots, \alpha_k$ — контрольными (избыточными). МА в этом случае называется расширенной, где $M_k = m_1 m_2 \dots m_n m_{n+1} \dots m_k$, и охватывает полное множество состояний, представляемых

всеми k вычетами. Эта область будет являться полным диапазоном МА $[0, M_k)$ и состоять из рабочего диапазона $[0, M_n)$, где $M_n = m_1 m_2 \dots m_n$, и диапазона, определяемого избыточными основаниями $[M_n, M_k)$, представляющего недопустимую область. Это означает, что операции над числами h_t выполняются в диапазоне $[0, M_k)$. Поэтому если результат операции МА выходит за пределы M_n , то делается вывод о возникновении ошибки вычислений.

Для осуществления контроля ошибок при реализации ЛЧП рассмотрим МА, заданную основаниями $m_1, m_2, \dots, m_n, m_{n+1}$. Представим каждый коэффициент ЛЧП h_t в виде (2), получим избыточный код МА, представленный системой ЛЧП:

$$\begin{cases} U^{(1)} = L^{(1)}(X) = \\ = \alpha_0^{(1)} + \alpha_1^{(1)} x_1 + \dots + \alpha_r^{(1)} x_r, \\ U^{(2)} = L^{(2)}(X) = \\ = \alpha_0^{(2)} + \alpha_1^{(2)} x_1 + \dots + \alpha_r^{(2)} x_r, \\ \dots \dots \dots \\ U^{(n)} = L^{(n)}(X) = \\ = \alpha_0^{(n)} + \alpha_1^{(n)} x_1 + \dots + \alpha_r^{(n)} x_r, \\ U^{(n+1)} = L^{(n+1)}(X) = \\ = \alpha_0^{(n+1)} + \alpha_1^{(n+1)} x_1 + \dots + \alpha_r^{(n+1)} x_r, \end{cases} \tag{3}$$

где вектор $X = [x_1 \ x_2 \ \dots \ x_r]$.

Подставив в (3) значения остатков МА по соответствующим основаниям для каждого коэффициента (1), а также значения переменных x_1, \dots, x_r , получим значения ЛЧП системы (3), где $U^{(1)}, U^{(2)}, \dots, U^{(n)}, U^{(n+1)}$ — целые числа. В соответствии с Китайской теоремой об остатках (КТО) решим систему уравнений:

$$\begin{cases} U^* = |U^{(1)}|_{m_1}, \\ U^* = |U^{(2)}|_{m_2}, \\ \dots \dots \dots \\ U^* = |U^{(n)}|_{m_n}, \\ U^* = |U^{(n+1)}|_{m_{n+1}}. \end{cases} \tag{4}$$

Так как основания $m_1, m_2, \dots, m_n, m_{n+1}$ попарно просты, то решением системы (4) является остаток по модулю $M_{n+1} = m_1 m_2 \dots m_{n+1}$:

Из [16] известно, что при отсутствии ошибок вычислений каждое значение b_0, b_1, \dots, b_r , полученное при решении систем сравнений, будет лежать в диапазоне $[0, M^{(z)})$, где $M^{(z)} = m^{(1)}m^{(2)} \dots m^{(z)}$. Например, для b_0 система уравнений имеет вид:

$$\begin{cases} b_0 = |\beta_0^{(1)}|_{m^{(1)}}, \\ b_0 = |\beta_0^{(2)}|_{m^{(2)}}, \\ \dots\dots\dots \\ b_0 = |\beta_0^{(z)}|_{m^{(z)}}. \end{cases}$$

Для корректного применения к системе (7) методов избыточного модулярного кодирования необходимо выполнить масштабирование системы путем введения дополнительного (попарно простого по отношению к другим основаниям) основания $m^{(0)}$ МА, где $m^{(0)} \geq r + 1$ (r – наибольшая длина (количество слагаемых) ЛЧП из системы (7)).

Ввод общего дополнительного основания $m^{(0)}$ в рамках операции масштабирования позволит выполнять операции над числами b_0, b_1, \dots, b_r , лежащими в рабочем диапазоне $[0, M^{(z)})$, в более широком рабочем диапазоне $[0, r(M^{(z)} - 1))$. Поэтому если в результате операции МА полученное число выходит за пределы $r(M^{(z)} - 1)$, то делается вывод об ошибке вычислений.

Система ЛЧП (7) примет вид:

$$\begin{cases} U^{(0)} = L^{(0)}(\mathbf{X}) = h_0^{(0)} + \sum_{i=1}^r h_i^{(0)} x_i, \\ U^{(1)} = L^{(1)}(\mathbf{X}) = h_0^{(1)} + \sum_{i=1}^r h_i^{(1)} x_i, \\ U^{(2)} = L^{(2)}(\mathbf{X}) = h_0^{(2)} + \sum_{i=1}^r h_i^{(2)} x_i, \\ \dots\dots\dots \\ U^{(z)} = L^{(z)}(\mathbf{X}) = h_0^{(z)} + \sum_{i=1}^r h_i^{(z)} x_i. \end{cases}$$

Вычислим значения элементов β_0, β_i ЛЧП $L^{(0)}(\mathbf{X})$, где $\beta_0 = h_0^{(0)}$, $\beta_i = h_i^{(0)} x_i$, решив КТО для отдельных групп остатков по основаниям МА:

$$\begin{cases} b_0 = (h_0^{(1)}, h_0^{(2)}, \dots, h_0^{(z)})_{MA}, \\ b_1 = (h_1^{(1)} x_1, h_1^{(2)} x_1, \dots, h_1^{(z)} x_1)_{MA}, \\ \dots\dots\dots \\ b_r = (h_r^{(1)} x_r, h_r^{(2)} x_r, \dots, h_r^{(z)} x_r)_{MA}. \end{cases}$$

В случае неполного состава элементов ЛЧП (отсутствие некоторых переменных), необходимо выполнить выравнивание имеющихся элементов справа налево, оставшиеся свободные места заполнить нулями.

Полученные значения b_0, b_1, \dots, b_r необходимо взять по введенному модулю $m^{(0)}$, получим:

$$\begin{cases} \beta_0 = |b_0|_{m^{(0)}}, \\ \beta_1 = |b_1|_{m^{(0)}}, \\ \dots\dots\dots \\ \beta_r = |b_r|_{m^{(0)}}. \end{cases}$$

Для осуществления контроля ошибок арифметических вычислений при реализации z -х ЛЧП рассмотрим МА, заданную основаниями $m^{(0)}, m^{(1)}, m^{(2)}, \dots, m^{(z)}, m^{(z+1)}$. Получим избыточный код МА, представленный системой ЛЧП вида:

$$\begin{cases} U^{(0)} = L^{(0)}(\mathbf{X}) = h_0^{(0)} + \sum_{i=1}^r h_i^{(0)} x_i, \\ U^{(1)} = L^{(1)}(\mathbf{X}) = h_0^{(1)} + \sum_{i=1}^r h_i^{(1)} x_i, \\ U^{(2)} = L^{(2)}(\mathbf{X}) = h_0^{(2)} + \sum_{i=1}^r h_i^{(2)} x_i, \\ \dots\dots\dots \\ U^{(z)} = L^{(z)}(\mathbf{X}) = h_0^{(z)} + \sum_{i=1}^r h_i^{(z)} x_i, \\ U^{(z+1)} = L^{(z+1)}(\mathbf{X}) = h_0^{(z+1)} + \sum_{i=1}^r h_i^{(z+1)} x_i. \end{cases} \quad (9)$$

Подставив в (9) известные значения остатков МА по соответствующим основаниям, а также вычислив аналогично с вычислениями для $L^{(0)}(\mathbf{X})$ значения элементов для ЛЧП $L^{(z+1)}(\mathbf{X})$, получим значения ЛЧП системы (9), где $U^{(0)}, U^{(1)}, U^{(2)}, \dots, U^{(z)}, U^{(z+1)}$ – целые числа.

Решим систему:

$$\begin{cases} U^* \equiv |U^{(0)}|_{m^{(0)}}, \\ U^* \equiv |U^{(1)}|_{m^{(1)}}, \\ U^* \equiv |U^{(2)}|_{m^{(2)}}, \\ \dots\dots\dots \\ U^* \equiv |U^{(z)}|_{m^{(z)}}, \\ U^* \equiv |U^{(z+1)}|_{m^{(z+1)}}. \end{cases} \quad (10)$$



Так как основания $m^{(0)}, m^{(1)}, m^{(2)}, \dots, m^{(z)}, m^{(z+1)}$ попарно просты, то решением (10) является наименьший неотрицательный вычет по модулю $M^{(z+1)} = m^{(1)}m^{(2)} \dots m^{(z+1)}$:

$$U^* = \left| \sum_{s=0}^{z+1} M^{(s,z+1)} \mu^{(s,z+1)} U^{(s)} \right|_{M^{(z+1)}}, \quad (11)$$

где $M^{(s,z+1)} = \frac{M^{(z+1)}}{m^{(s)}}$, $\mu^{(s,z+1)} = \left| M^{(s,z+1)-1} \right|_{m^{(s)}}$.

Вхождение результата вычисления (11) в диапазон (контрольное выражение)

$$0 \leq U^* < r(M^{(z)} - 1) \quad (12)$$

означает отсутствие обнаруживаемых ошибок вычислений.

Пример 2. Пусть w -й блок участка двоичной ПСП разбит на v -е подблоки, каждый из которых представлен одним ЛЧП. Система (7) имеет вид:

$$\begin{cases} U^{(1)} = L^{(1)}(\mathbf{X}) = x_1 + 5x_2 + 4x_3, \\ U^{(2)} = L^{(2)}(\mathbf{X}) = x_3 + 5x_4 + 4x_5, \\ U^{(3)} = L^{(3)}(\mathbf{X}) = 5x_1 + x_2 + 4x_3 + x_5. \end{cases}$$

Выберем основания системы: $m^{(1)} = 16$, $m^{(2)} = 17$, $m^{(3)} = 19$.

Вычислим значение рабочего диапазона: $M^{(3)} = m^{(1)}m^{(2)}m^{(3)} = 5168$.

Выполним выравнивание имеющихся элементов ЛЧП справа налево, оставшиеся свободные места заполним нулями. Для наглядности запишем систему ЛЧП следующим образом:

$$\begin{cases} U^{(1)} = L^{(1)}(\mathbf{X}) = 0 + x_1 + 5x_2 + 4x_3, \\ U^{(2)} = L^{(2)}(\mathbf{X}) = 0 + x_3 + 5x_4 + 4x_5, \\ U^{(3)} = L^{(3)}(\mathbf{X}) = 5x_1 + x_2 + 4x_3 + x_5. \end{cases}$$

Выполним операцию масштабирования (введем дополнительное основание $m^{(0)} = 5$) и получим:

$$\begin{cases} U^{(0)} = L^{(0)}(\mathbf{X}) = h_1^{(0)}x_1 + h_2^{(0)}(x_1, x_2, x_3) + \\ + h_3^{(0)}(x_2, x_3, x_4) + h_4^{(0)}(x_3, x_5), \\ U^{(1)} = L^{(1)}(\mathbf{X}) = 0 + x_1 + 5x_2 + 4x_3, \\ U^{(2)} = L^{(2)}(\mathbf{X}) = 0 + x_3 + 5x_4 + 4x_5, \\ U^{(3)} = L^{(3)}(\mathbf{X}) = 5x_1 + x_2 + 4x_3 + x_5. \end{cases}$$

В соответствии с КТО вычислим значения элементов $\beta_1, \beta_2, \beta_3, \beta_4$ ЛЧП $L^{(0)}(\mathbf{X})$ для отдельных групп остатков по основаниям МА:

$$\begin{cases} b_1 = (0, 0, 5x_1)_{MA}, \\ b_2 = (x_1, x_3, x_2)_{MA}, \\ b_3 = (5x_2, 5x_4, 4x_3)_{MA}, \\ b_4 = (4x_3, 4x_5, x_5)_{MA}. \end{cases}$$

Получим:

0	0	$h_1^{(3)}x_1$	β_1
0	0	0	0
0	0	5	3

$h_2^{(1)}x_1$	$h_2^{(2)}x_3$	$h_2^{(3)}x_2$	β_2
0	0	0	0
0	0	1	2
0	1	0	2
0	1	1	1
1	0	0	3
1	0	1	2
1	1	0	2
1	1	1	1

$h_3^{(1)}x_2$	$h_3^{(2)}x_4$	$h_3^{(3)}x_3$	β_3
0	0	0	0
0	0	4	4
0	5	0	4
0	5	4	3
5	0	0	1
5	0	4	0
5	5	0	0
5	5	4	1

$h_4^{(1)}x_3$	$h_4^{(2)}x_5$	$h_4^{(3)}x_5$	β_4
0	0	0	0
0	0	1	2
0	4	0	0
0	4	1	4
4	0	0	1
4	0	1	0
4	4	0	3
4	4	1	2

Пусть $x_1 = x_2 = x_3 = x_4 = x_5 = 1$, тогда система ЛЧП примет вид:

$$\begin{cases} U^{(0)} = L^{(0)}(\mathbf{X}) = 3 + 1 + 1 + 2 = 7, \\ U^{(1)} = L^{(1)}(\mathbf{X}) = 0 + 1 + 5 + 4 = 10, \\ U^{(2)} = L^{(2)}(\mathbf{X}) = 0 + 1 + 5 + 4 = 10, \\ U^{(3)} = L^{(3)}(\mathbf{X}) = 5 + 1 + 4 + 1 = 11. \end{cases}$$

Рабочий диапазон после масштабирования равен $r(M^{(z)} - 1) = 20668$.

Для осуществления контроля ошибок при реализации z -х ЛЧП введем избыточное основание $m^{(4)} = 21$. Получим избыточный код МА, представленный системой ЛЧП вида:

$$\begin{cases} U^{(0)} = L^{(0)}(\mathbf{X}) = 3 + 1 + 1 + 2 = 7, \\ U^{(1)} = L^{(1)}(\mathbf{X}) = 0 + 1 + 5 + 4 = 10, \\ U^{(2)} = L^{(2)}(\mathbf{X}) = 0 + 1 + 5 + 4 = 10, \\ U^{(3)} = L^{(3)}(\mathbf{X}) = 5 + 1 + 4 + 1 = 11, \\ U^{(4)} = L^{(4)}(\mathbf{X}) = 17 + 1 + 2 + 16 = 36. \end{cases}$$

В соответствии с КТО решим систему:

$$\begin{cases} U^* \equiv |7|_5, \\ U^* \equiv |10|_{16}, \\ U^* \equiv |10|_{17}, \\ U^* \equiv |11|_{19}, \\ U^* \equiv |36|_{21}. \end{cases}$$

В соответствии с (11) получим $U^* = 4362$. Так как $0 \leq U^* < 20668$, то согласно (12) делается заключение об отсутствии ошибок.

Таким образом, использование методов МА для реализации логических операций, в частности, при формировании ПСП и ключевых последовательностей, помимо повышения производительности СКЗИ позволяет получить важные преимущества по повышению безопасности их функционирования.

СПИСОК ЛИТЕРАТУРЫ

1. **Yang, B.** Scan Based Side Channel Attack on Data Encryption Standard [Электронный ресурс] / B. Yang, K. Wu, R. Karri // Report. – 2004/083. – Режим доступа <http://eprint.iacr.org> (Дата обращения 2004).
2. **Щербаков, Н.С.** Достоверность работы цифровых устройств [Текст] / Н.С. Щербаков. – М.: Машиностроение, 1989. – 224 с.
3. **Савельев, А.Я.** Прикладная теория цифровых автоматов [Текст] / А.Я. Савельев. – М.: Высш. школа, 1987. – 272 с.
4. **Малюгин, В.Д.** Параллельные логические вычисления посредством арифметических полиномов [Текст] / В.Д. Малюгин. – М.: Физматлит, 1997. – 192 с.
5. **Финько, О.А.** Реализация систем булевых функций большой размерности методами модулярной арифметики [Текст] / О.А. Финько // Автоматика и телемеханика. – 2004. – № 6. – С. 37–60.
6. **Бабаш, А.В.** Криптография [Текст] / А.В. Бабаш, Г.П. Шанкин; под ред. В.П. Шерстюка, Э.А. Применко. – М.: СОЛОН-П Gutmann ПРЕСС, 2007. – 512 с.
7. **Шнайер, Б.** Практическая криптография [Текст] / Б. Шнайер. – М.: ИД «Вильямс», 2005. – 424 с.
8. **Фороузан, Б.А.** Криптография и безопасность сетей: Учеб. пособие [Текст] / Б.А. Фороузан; пер. с англ. под ред. А.Н. Берлина. – М.: Интернет-Университет информационных технологий: БИНОМ. Лаборатория знаний, 2010. – 784 с.
9. **Финько, О.А.** Самопроверяемый специализированный вычислитель систем булевых функций [Текст] / О.А. Финько, С.А. Диченко, А.К. Вишневский // Патент России № 2485575, 20.06.2013.
10. **Финько, О.А.** Арифметический вычислитель систем булевых функций [Текст] / О.А. Финько, А.К. Вишневский, С.А. Диченко, Д.В. Самойленко [и др.] // Патент России № 2461868, 20.09.2012.
11. **Финько, О.А.** Самопроверяемый модулярный вычислитель систем логических функций [Текст] / О.А. Финько, С.М. Сульгин, А.В. Щербаков [и др.] // Патент России № 2417405, 27.04.2011.
12. **Финько, О.А.** Модулярный вычислитель систем логических функций [Текст] / О.А. Финько, А.В. Щербаков // Патент России № 2417303, 16.11.2009.
13. **Диченко, С.А.** Реализация двоичных псевдослучайных последовательностей линейными числовыми полиномами [Текст] / С.А. Диченко, А.К. Вишневский, О.А. Финько // Изв. Южного федерального ун-та. Технические науки. – 2011. – № 12. – С. 130–140.
14. **Диченко, С.А.** Алгоритм генерации блочной ПСП, основанный на применении логико-числовых форм [Текст] / С.А. Диченко, О.А. Финько // Изв. Южного федерального ун-та. Технические науки. – 2012. – № 12. – С. 158–166.
15. **Yanushkevich, L.** Logic design of nano-ICs [Text] / S. Yanushkevich, V. Shmerko, S. Lyshovski. – CRC Press, 2005.
16. **Акушский, И.Я.** Машинная арифметика в остаточных классах [Текст] / И.Я. Акушский, Д.И. Юдицкий. – М.: Сов. радио, 1968. – 440 с.



REFERENCES

1. **Yang B., Wu K., Karri R.** Scan Based Side Channel Attack on Data Encryption Standard / Report. – 2004/083, <http://eprint.iacr.org>, 2004.
2. **Shcherbakov N.S.** Dostovernost' raboty tsifrovyykh ustroystv. – Moscow: Mashinostroenie, 1989. – 224 s. (rus)
3. **Savel'ev A.Ia.** Prikladnaia teoriia tsifrovyykh avtomatov. – Moscow: Vyssh. shkola, 1987. – 272 s. (rus).
4. **Maliugin V.D.** Parallel'nye logicheskie vychisleniia posredstvom arifmeticheskikh polinomov. – Moscow: Fizmatlit, 1997. – 192 s. (rus)
5. **Fin'ko O.A.** Realizatsiia sistem bulevykh funktsii bol'shoi razmernosti metodami moduliarnoi arifmetiki / Avtomatika i telemekhanika. – 2004. – № 6. – S. 37–60. (rus)
6. **Babash A.V., Shankin G.P.** Kriptografiia; pod red. V.P. Sherstiuka, E.A. Primenko. – Moscow: SOLON-P Gutmann PRESS. – 512 s. (rus)
7. **Shnaier B.** Prakticheskaiia kriptografiia. – Moscow: ID «Vil'iams», 2005. – 424 s. (rus)
8. **Forouzan B.A.** Kriptografiia i bezopasnost' setei: Ucheb. posobie; per. s angl.; pod red. A.N. Berlina. – Moscow: Internet-Universitet Informatsonnykh Tekhnologii: BINOM. Laboratoriia znanii, 2010. – 784 s. (rus)
9. **Fin'ko O.A., Dichenko S.A., Vishnevskii A.K.** Samoproveriaemyi spetsializirovannyi vychislitel' sistem bulevykh funktsii / Patent Rossii № 2485575, 20.06.2013. (rus)
10. **Fin'ko O.A., Vishnevskii A.K., Dichenko S.A., Samoilenko D.V. i dr.** Arifmeticheskii vychislitel' sistem bulevykh funktsii / Patent Rossii № 2461868, 20.09.2012. (rus)
11. **Fin'ko O.A., Sul'gin S.M., Shcherbakov A.V. i dr.** Samoproveriaemyi moduliarnyi vychislitel' sistem logicheskikh funktsii / Patent Rossii № 2417405, 27.04.2011. (rus)
12. **Fin'ko O.A., Shcherbakov A.V.** Moduliarnyi vychislitel' sistem logicheskikh funktsii / Patent Rossii № 2417303, 16.11.2009. (rus)
13. **Dichenko S.A., Vishnevskii A.K., Fin'ko O.A.** Realizatsiia dvoichnykh psevdosluhainykh posledovatel'nostei lineinymi chislovymi polinomami / Izv. Iuzhnogo federal'nogo un-ta. Tekhnicheskie nauki. – 2011. – № 12 – S. 130–140. (rus)
14. **Dichenko S.A., Fin'ko O.A.** Algoritm generatsii blochnoi PSP, osnovannyi na primenenii logiko-chislovykh form / Izv. Iuzhnogo federal'nogo un-ta. Tekhnicheskie nauki. – 2012. – № 12. – S. 158–166. (rus)
15. **Yanushkevich L., Shmerko V., Lyshevski S.** Logic design of nanoICs. – CRC Press, 2005.
16. **Akushskiy I.Ya., Yuditskiy D.I.** Mashinnaya arifmetika v ostatochnykh klassakh. – Moscow: Sov. radio, 1968. – 440 s. (rus)

ДИЧЕНКО Сергей Александрович – адъюнкт филиала Военной академии связи (г. Краснодар).
350035, Россия, г. Краснодар, ул. Красина, д. 4.

DICHENKO, Sergey A. *Military Academy of Communications (Krasnodar).*
350035, Krasin Str. 4, Krasnodar, Russia

ЕЛИСЕЕВ Николай Иванович – доцент кафедры специальной техники филиала Военной академии связи (г. Краснодар), кандидат технических наук.
350035, Россия, г. Краснодар, ул. Красина, д. 4.

ELISSEEV, Nikolai I. *Military Academy of Communications (Krasnodar).*
350035, Krasin Str. 4, Krasnodar, Russia

ФИНЬКО Олег Анатольевич – профессор кафедры обеспечения безопасности информации в автоматизированных системах филиала Военной академии связи (г. Краснодар), доктор технических наук, профессор.

350035, Россия, г. Краснодар, ул. Красина, д. 4.
E-mail: ofinko@yandex.ru; URL: <http://финько.рф>

FINKO, Oleg A. *Military Academy of Communications (Krasnodar).*
350035, Krasin Str. 4, Krasnodar, Russia
E-mail: ofinko@yandex.ru; URL: <http://ofinko.ru>

УДК 681.3.07

О.П. Малофей, Ю.О. Щелкунова

ПОДХОД К СОЗДАНИЮ СТРУКТУРНОГО КОДА ДЛЯ ЗАПИСИ И СЧИТЫВАНИЯ ДАННЫХ В ЗАПОМИНАЮЩИХ УСТРОЙСТВАХ ИНФОТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМ

O.P. Malofey, Yu.O. Shchelkunova

APPROACH TO THE CREATION OF A STRUCTURAL CODE TO WRITE AND INCLUDE THE DEPLOYMENT OF DATA IN A STORAGE DEVICE INFOTELECOMMUNICATION SYSTEMS

Предложен подход к созданию структурного кода, на основе которого разработан способ записи и считывания данных в запоминающем устройстве, сокращающий в полтора раза объем задействованной памяти и, следовательно, улучшающий в соответствующее число раз ее надежностные, массо-габаритные и энергоемкостные показатели. Синтезированный код гарантированно исправляет модульные ошибки и применим для контроля работоспособности элементов памяти. Представлены также отказоустойчивые структуры памяти для 8-, 16- и 32-разрядных форматов представления данных.

КОДЫ ИСПРАВЛЕНИЯ ОШИБОК. УСТРОЙСТВО ВВОДА/ВЫВОДА. КОНЕЧНЫЕ ПОЛЯ.

The paper presents way for generating of a structural code, the basis for treating of the method for recording and reading of data in the storage, which can decrease the size of memory in one and a half time and, consequently, increase its indexes of reliability, dimension and capacity of the energy in few times. The synthesis code securely corrects modular errors and can be applied for the control of workability of the units of memory. Also we can see here fault-tolerant structures of the memory for 8-, 16- and 32-positional formats of the data notation.

ERROR CORRECTION CODES. DEVICE INPUT / OUTPUT. FINITE FIELDS.

Рост объема массивов данных и интеграции элементной базы с одной стороны и уменьшение времени обращения к запоминающему устройству (ЗУ) с другой – накладывают жесткие рамки на вероятностные характеристики, тракта «записи – хранение – воспроизведение» информации в автоматизированной системе управления (АСУ). При этом очевидно, что для ЗУ имеют место как случайные неповторяющиеся сбои (ошибки), так и систематические ошибки различной кратности. Причем интенсивность сбоев интегральных схем (ИС) ЗУ на один–два порядка выше чем интенсивность отказов [1]. Борьба с данными ошибками наиболее сложна и разделяется на методы аппаратного и программного

контроля. Из всего многообразия методов борьбы с ошибками в ЗУ для АСУ, не допускающей факта наличия катастрофической ошибки, широкое применение нашел метод мажоритарной обработки хранимой информации.

Очевидно, что при известной избыточности данного метода он полностью оправдывает себя с точки зрения исправления ошибок [2], однако его применение в ЗУ АСУ связано со следующими трудностями:

во-первых, при $N = 2m - 1$ -кратном повторении сообщения и $m > 2$ сложность второй решающей схемы, реализующей мажоритарный метод повышения достоверности (МПД), возрастает [3];

во-вторых, мажоритарный метод аппа-

ратурного резервирования абсолютно неприемлем для контроля технического состояния ЗУ, т. к. при исправлении ошибки невозможно определить место ее нахождения.

Решение данной проблемы возможно, если использовать подход создания структурного кода, удовлетворяющего следующим условиям:

код должен надежно защищать каждый байт информации (машинного слова) от сбоев и групповых ошибок длины до $j = 8$;

должен иметь избыточность, существенно меньшую, чем троирование. Процедуры кодирования и декодирования не должны требовать дополнительного тактирования и вносить задержку во временную диаграмму «запись – считывание»;

код должен быть приемлем для контроля технического состояния ЗУ.

Алгоритм кодирования

Вход: машинное слово, состоящее из двух байтов $W = A|B$, где $A = (a_1, a_2, a_3, \dots, a_8)$, $B = (b_1, b_2, b_3, \dots, b_8)$.

Выход: слово памяти из четырех байтов, являющееся кодовым вектором кода V :

$$V = A|B|C|D.$$

Метод: для байтов входного машинного слова AIB и матрицы P размера 8×8 вида

$$P = \begin{bmatrix} \vdots & I \\ Y & \dots\dots\dots \\ \vdots & X \end{bmatrix},$$

где I – единичная матрица размера 7×7 ; Y – вектор-столбец из единиц; X – вектор-строка из нулей. Определим проверочные байты C и D как

$$C = A \oplus B, D = A \oplus BP, \quad (1)$$

где $BP = (b_0, b_1, \dots, b_7)$, $b_0 = \sum_{i=1}^8 b_i \bmod 2$.

Алгоритм декодирования

Вход: слово памяти из четырех байтов, возможно, содержащее ошибки:

$$V^* = A^*|B^*|C^*|D^*.$$

Выход: машинное слово из двух байтов с исправленными ошибками $W = A|B$ или

сигнал отказа от декодирования $q = 1$.

Метод: полагая $q = 0$ (т. е. процедура декодирования выполняется), вычислим синдром ошибки $S = S_1|S_2$, где

$$S_1 = A^* \oplus B^* \oplus C^*, S_2 = A^* \oplus B^*P \oplus D^*. \quad (2)$$

Рассмотрим возможные значения синдромов при возникновении ошибок в различных байтах, а также докажем, что предложенный алгоритм удовлетворяет исходному условию 2, т. е. код исправляет любой одиночный модуль ошибки длиной $j = 8$ в слове памяти.

Доказательство: положим

$$A^* = A \oplus e_a, B^* = B \oplus e_b, C^* = C \oplus e_c, \\ D^* = D \oplus e_d,$$

где e_x – модуль ошибки в байте X , возможно, имеющей место.

Тогда из (1) и (2) следует:

$$S_1 = e_a \oplus e_b \oplus e_c, S_2 = e_a \oplus e_bP \oplus e_d. \quad (3)$$

Анализ значения синдромов S_1 и S_2 дает следующие ситуации:

1. Если $e_a = e_b = e_c = e_d = 0$, то $S_1 = S_2 = 0$, что соответствует отсутствию ошибок.

2. Если $e_a = e_b = e_c = 0$, а $e_d \neq 0$, то $S_1 = 0$, $S_2 \neq 0$, что соответствует ошибке в байте D . Исправление имеющейся ошибки e_d имеет вид:

$$D = D^* \oplus S_2. \quad (4)$$

3. Если $e_a = e_b = e_d = 0$, а $e_c \neq 0$, то $S_1 \neq 0$, $S_2 = 0$, что соответствует ошибке в байте C . Исправление имеющейся ошибки e_c имеет вид:

$$C = C^* \oplus S_1. \quad (5)$$

4. Если $e_b = e_c = e_d = 0$, а $e_a \neq 0$, то $S_1 = S_2 \neq 0$, что соответствует ошибке в байте A . Исправление имеющейся ошибки e_a имеет вид:

$$A = A^* \oplus S_1. \quad (6)$$

5. Если $e_a = e_c = e_d = 0$, а $e_b \neq 0$, то $S_1 = e_b$, $S_2 = e_bP$, то есть $S_1P = S_2 \neq 0$, что соответствует ошибке в байте B . Исправление имеющейся ошибки e_b имеет вид:

$$B = B^* \oplus S_1. \quad (7)$$

6. Все остальные случаи ведут к отказу

от декодирования ($q \geq 1$), т. е. вызваны отказом более одного корпуса микросхем памяти ЗУ, при этом групповыми ошибками поражены два и более байта слова памяти.

Таким образом, случаи 1...5 охватывают все возможные виды ошибок в одном модуле слова памяти. Для завершения доказательства осталось лишь показать, что при одиночном модуле ошибок ситуации из пунктов 4 и 5 не могут иметь место одновременно, т. е. нужно показать, что $X = XP$ тогда и только тогда, когда $X = 0$.

Пусть $X = (x_1, x_2, x_3, \dots, x_8)$, тогда

$XP = (x_0, x_1, x_2, \dots, x_7)$, где $x_0 = \sum_{i=1}^8 x_i \bmod 2$.

Если $(x_1, x_2, x_3, \dots, x_8) = (x_0, x_1, x_2, \dots, x_7)$, то $x_0 = x_1 = x_2 = \dots = x_7 = x_8$.

Это возможно лишь при $x_i = 0$, так как при $x_i = 1$ сумма $x_0 = 0$, где $i = \overline{1, 8}$.

Определим избыточность полученного кода как отношение информационных символов к сумме информационных и проверочных:

$$\xi = \frac{|A| + |B|}{|A| + |B| + |C| + |D|} = \frac{16}{32} = 0,5,$$

где $|A| = |B| = |C| = |D|$ – длина информационного байта.

Рассмотрим теперь предложенный код и процедуру декодирования с точки зрения дополнительных условий, заданных вначале.

1. Код защиты двухбайтового слова имеет избыточность 16 бит (0,5) и исправляет модуль ошибки длины 8, согласно [3] такой код удовлетворяет условию минимальной избыточности. Проведем сравнение известного и предложенного алгоритмов записи по величине объема задействованной памяти. Для выполнения трижды дублированной записи информационного слова размера n двоичных символов известный способ требовал наличия памяти объемом $\Pi_1 = 3n$. Предложенный способ имеет избыточность 0,5, т. е. для записи информационного слова той же длины потребуется объем памяти $\Pi_2 = 2n$, при этом сохранится возможность мажоритарной обработки информации. Сокращение ресурса задействованной памяти составляет

$$\xi = \frac{\Pi_1}{\Pi_2} = \frac{3n}{2n} = 1,5.$$

2. Как было доказано выше, декодер исправляет сбои и ошибки в каждом машинном слове, независимо от их характера и причин возникновения [4].

3. Схема кодера и декодера строится из стандартных логических элементов без памяти и поэтому не нуждается в тактировании, а вносимая ими задержка не превышает время переключения данных элементов.

4. Оценка частоты возникновения ненулевых синдромов S_1 и S_2 позволяет сделать вывод о случайности или систематичности возникновения ошибок, что применимо для контроля технического состояния элементов (модулей) ЗУ.

Следовательно, синтезированный код удовлетворяет заданным нами условиям и применим для исправления случайных и систематических групповых ошибок, возникающих в ЗУ модульного типа, а также для технического контроля элементов данных ЗУ в резервированных системах памяти [5], что позволяет достичь сокращения задействованного объема ЗУ в 1,5 раза и во столько же раз улучшить его надежность и массогабаритные показатели.

Структурная схема реализации предложенного алгоритма приведена на рисунке.

Рассмотренная отказоустойчивая структура для 8-разрядных ЭВМ строится из модулей памяти вида $I \times NK$. Однако постоянный рост скорости обработки массивов информации требует ее дальнейшего распараллеливания, т. е. применения 16- и 32-разрядных шин и соответствующего формата представления данных. Выполним математическое описание предложенного способа и построим отказоустойчивую структуру памяти для 16- и 32-разрядных ЭВМ из модулей вида $4 \times NK$ [6].

Отказоустойчивую структуру будем представлять четверкой неотрицательных чисел (v, N, m, d) , где v – разрядность модуля памяти; N – число модулей памяти, покрывающее слово избыточной структуры; m – число модулей, покрывающее машинное слово; d – допустимое число отказов модулей. В этих обозначениях (1, 16, 16, 0)

и (1, 32, 16, 1) представляют безубыточную структуру и структуру с кодом Хэмминга из модулей вида $I \times NK$, а четверка $(v, 3m, m, 1)$ – троированную структуру из модулей вида $v \times NK$. Модульный код для отказоустойчивой структуры памяти должен также удовлетворять четырем условиям, указанным ранее.

Введем обозначения: a – машинное слово, вектор над GF [3]; $|a|$ – длина вектора a ; $wt(a)$ – вес Хэмминга вектора a ; $a = |a_0|a_1|...|a_{m-1}|$ – представление вектора a в виде последовательности подблоков длины v ; $(a)(i)$ – циклический сдвиг вектора a на i разрядов вправо; $a^>$ – вектор длины $|a|-1$, полученный из вектора a удалением крайней правой координаты.

Определим код V как множество векторов вида:

$$V = |a_0|a_1|...|a_{m-1}|c_0|c_1|, \quad (8)$$

где a_i – подблоки машинного слова; c_i – проверочные подблоки

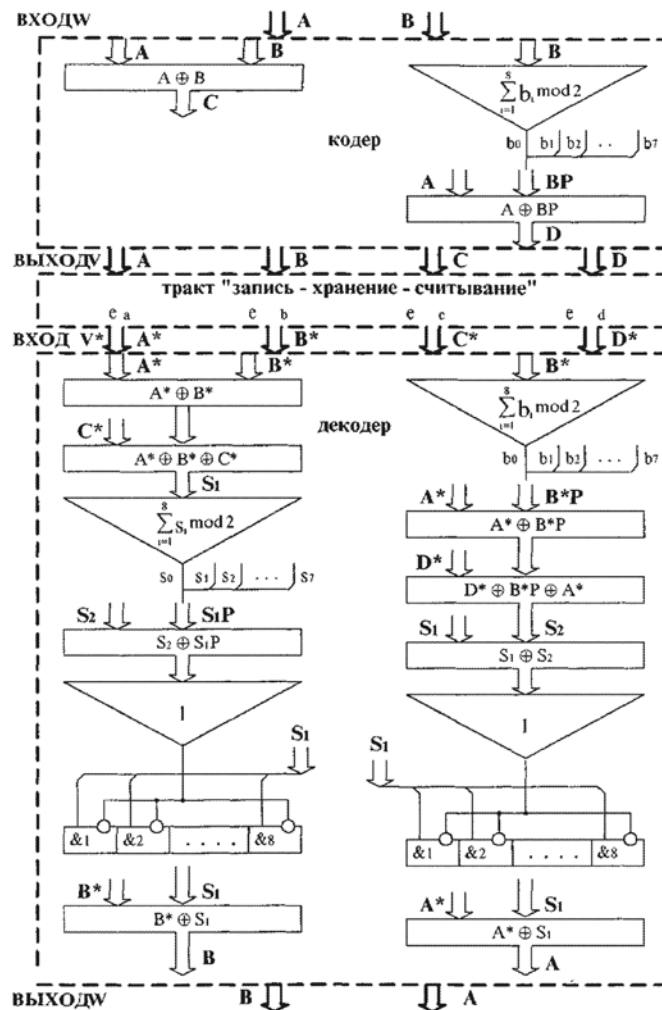
$$c_0 = \sum_{i=0}^{m-1} a_i, \quad c_1 = \left(\sum_{i=0}^{m-1} (|a_i|0|(i)) \right)^>. \quad (9)$$

Соотношения (8) и (9) задают процедуру кодирования в систематической форме для кода V с параметрами $(v(m+2), vm)$. При считывании информации из памяти на вход декодера поступает слово

$$V \oplus e = |a_0 \oplus e_0|a_1 \oplus e_1|... \quad (10)$$

$$... |a_{m-1} \oplus e_{m-1}|c_0 \oplus e_m|c_1 \oplus e_{m+1}|,$$

где e – вектор ошибки, который в наших



Способ записи и считывания информации с исправлением модульных ошибок

предположениях содержит ровно один ненулевой подблок e_i . По (10) вычисляется вектор синдрома ошибки:

$$\left\{ \begin{aligned} S &= |S_0 |S_1| \\ S_0 &= c_0 \oplus e_m \oplus \sum_{i=0}^{m-1} (a_i \oplus e_i) = \\ &= e_m \oplus \sum_{i=0}^{m-1} e_i \\ S_1 &= c_1 \oplus e_{m+1} \oplus \left(\sum_{i=0}^{m-1} (|a_i \oplus e_i |0\rangle)(i) \right)^{\triangleright} = \\ &= e_{m+1} \oplus \left(\sum_{i=0}^{m-1} (|e_i |0\rangle)(i) \right)^{\triangleright} \end{aligned} \right. \quad (11)$$

Из (11) следует, что

$$S_0 = \begin{cases} e_m, & \text{если } e_m \neq 0 \\ 0, & \text{если } e_{m+1} \neq 0 \\ e_i, & \text{если } e_i \neq 0, 0 \leq i \leq m \end{cases} \quad (12)$$

$$S_1 = \begin{cases} 0, & \text{если } e_m \neq 0 \\ e_{m+1}, & \text{если } e_{m+1} \neq 0 \\ ((|e_i |0\rangle)(i))^{\triangleright}, & \text{если } e_i \neq 0, 0 \leq i \leq m \end{cases} \quad (13)$$

Таким образом, если S_0 или S_1 равны нулю, то информационные подблоки слова V не содержат ошибок.

При $S_0, S_1 \neq 0$, S_0 дает значение модуля ошибки, а S_1 позволяет найти номер i подблока, в котором эта ошибка произошла. Число i может быть определено перебором, как решение уравнения:

$$S_1 = ((|S_0 |0\rangle)(i))^{\triangleright}, i = \overline{0, m-1}. \quad (14)$$

Рассмотрим условия, при которых (14) имеет единственное решение.

Теорема. Если $v + 1$ простое число и $m \leq v + 1$, то уравнение (14) имеет единственное решение, отличное от нулевого.

Доказательство.

1. Для того чтобы уравнение $x = (x)(i)$ имело решение, необходимо, чтобы x был периодической последовательностью с длиной периода, равной общему делителю чисел $v + 1$ и i , но при простом $v + 1$ имеется только тривиальный делитель 1. Имеются две последовательности, состоящие из последовательностей периода 1:

$x = (0 0 \dots 0)$ – соответствует отсутствию ошибок;

$x = (1 1 \dots 1)$ – такой вид вектора не допускается по построению (см. (14)).

2. $w(x - (x)(i)) \geq 2$, так как веса x и $(x)(i)$ одинаковые.

3. $w(x^* - ((x)(i))^*) \geq 1$. Соотношения 1, 2 и 3 выполняются при $0 < i < v$, что и требовалось доказать.

Для 16-разрядной ЭВМ и модулей памяти вида $4 \times \text{НК}$ код V непосредственно приводит к структуре (4, 6, 4, 1) [6].

Для 32-разрядной ЭВМ непосредственное использование кода V невозможно, т. к. $m = 8$, что больше $v + 1$.

Определим код W как множество, состоящее из векторов

$$W = |a_0 |a_1 | \dots |a_7 |c_0 |c_1 |c_2|,$$

где

$$c_0 = \sum_{i=0}^7 a_i; \quad c_1 = \left(\sum_{i=0}^3 (|a_i |0\rangle)(i) \right)^{\triangleright}; \quad (15)$$

$$c_2 = \left(\sum_{i=0}^3 (|a_{4+i} |0\rangle)(i) \right)^{\triangleright}.$$

Вектор синдрома ошибки $S = |S_0 |S_1 |S_2|$, при этом

$$\left\{ \begin{aligned} S &= |e_8 |0 |0\rangle, & \text{если } e_8 \neq 0; \\ S &= |0 |e_9 |0\rangle, & \text{если } e_9 \neq 0; \\ S &= |0 |0 |e_{10}\rangle, & \text{если } e_{10} \neq 0; \\ S &= |e_i |((|e_i |0\rangle)(i))^{\triangleright} |0\rangle, & \text{если } e_i \neq 0; 0 \leq i \leq 3; \\ S &= |e_i |0 |((|e_i |0\rangle)(i))^{\triangleright}, & \text{если } e_i \neq 0; 4 \leq i \leq 7. \end{aligned} \right. \quad (16)$$

Таким образом, для 32-разрядных ЭВМ и модулей памяти вида $4 \times \text{НК}$ код W приводит к структуре (4, 11, 8, 1) [6].

Осталось лишь показать, что предложенный способ допускает реализацию мажоритарной обработки информации при декодировании, для чего выполним с учетом (1) следующие действия:

$$\left. \begin{aligned} A &= A_1; \\ C \oplus B &= A \oplus B \oplus B = A_2; \\ D \oplus BP &= A \oplus BP \oplus BP = A_3 \end{aligned} \right\} \quad (17)$$

$$\left. \begin{aligned} B &= B_1; \\ C \oplus A &= A \oplus B \oplus A = B_2; \\ D \oplus A &= A \oplus BP \oplus A = B_3P. \end{aligned} \right\} (18)$$

То есть по значениям декодируемых слов кодового вектора $W = |A|B|C|D$ каждое из исходных слов A и B может быть трижды восстановлено независимыми путями, после чего возможна их мажоритарная обработка по методу «два из трех».

В статье предложен подход к созданию структурного кода, который гарантированно исправляет модульные ошибки и может применяться для контроля работоспособности памяти.

Разработан способ записи и считывания данных в запоминающем устройстве.

Представлены отказоустойчивые структуры памяти для 8-, 16- и 32-разрядных форматов представления данных.

СПИСОК ЛИТЕРАТУРЫ

1. **Бородин, Г.А.** Методы расчета надежности запоминающих устройств [Текст] / Г.А. Бородин, В.А. Иванов // Зарубежная радиоэлектроника. –1989. –№ 2. –С. 92–111 с.
2. **Саганович, Ю.Л.** Кодовая защита оперативной памяти ЭВМ от ошибок [Текст] / Ю.Л. Саганович // Автоматика и телемеханика. –1991. –№ 5. –С. 3–45.
3. **Бояринов, Н.М.** Помехоустойчивое кодирование числовой информации [Текст] / Н.М. Бояринов. –М.: Наука, 1983.
4. **Малофей, О.П.** Алгоритм записи и считывания

информации в мажоритарно резервированной памяти [Текст] / О.П. Малофей // Системы обработки информации. – Харьков. –2006. –Вип. 8 (57). – С. 3–6.

5. **Баринов, В.В.** Сверхбольшие интегральные микросхемы оперативных запоминающих устройств [Текст] / В.В. Баринов, А.С. Березин, В.Д. Вернер [и др.]. –М.: Радио и связь, 1991.

6. **Камыш, А.В.** Отказоустойчивая структура памяти для 8-, 16- и 32-разрядных ЭВМ [Текст] / А.В. Камыш, А.В. Ткаченко // Сб. Тезисов III НТК КВВКИУ РВ 19-20.09.1995. –Краснодар, 1996.

REFERENCES

1. **Borodin G.A., Ivanov V.A.** Metody rascheta nadezhnosti zapominaiushchikh ustroystv / Zarubezhnaia radioelektronika. – № 2. – 1989. – 92–111 s. (rus)
2. **Saganovich Iu.L.** Kodovaiia zashchita operativnoi pamiati EVM ot oshibok / Avtomatika i telemekhanika. – № 5. – 1991. –S. 3–45. (rus)
3. **Boiarinov N.M.** Pomekhoustoichivoe kodirovanie chislovoi informatsii. – Moscow: Nauka, 1983. (rus)
4. **Malofei O.P.** Algoritm zapisi i schityvaniia

informatsii v mazhoritarno rezervirovannoi pamiati / Sistemi obrobki informatsii. – Khar'kov, 2006. – Vipusk 8(57). – S. 3–6.

5. **Barinov V.V., Berezin A.S., Verner V.D. i dr.** Sverkhbol'shie integral'nye mikroskhemy operativnykh zapominaiushchikh ustroystv. – Moscow: Radio i sviaz', 1991. (rus)

6. **Kamysh A.V., Tkachenko A.V.** Otkazo-ustoichivaia struktura pamiati dlia 8-, 16- i 32-razriadnykh EVM / Sb. Tezisov III NTK KVVKIU RV 19-20.09.1995. – Krasnodar, 1996. (rus)

МАЛОФЕЙ Олег Павлович – профессор кафедры высшей алгебры и геометрии Института математики и естественных наук Северо-Кавказского федерального университета, кандидат технических наук. 355000, Россия, г. Ставрополь, пр. Кулакова, д. 2.

MALOFEY, Oleg P. North-Caucasian Federal University. 355000, prosp. Kulakova 2, Stavropol, Russia.

ЩЕЛКУНОВА Юлия Олеговна – доцент кафедры прикладной математики и информатики Ставропольского государственного аграрного университета, кандидат технических наук. 355000, Россия, г. Ставрополь, ул. Шеболдаева, д. 11а.

SHCHELKUNOVA, Yulia O. Stavropol State Agrarian University. 355000, Sheboldaeva Str. 11a, Stavropol, Russia.

УДК 517.518.82

П.К. Корнеев, И.А. Журавлёва

ПОСТРОЕНИЕ НАИЛУЧШИХ СРЕДНЕКВАДРАТИЧЕСКИХ ПОЛИНОМОВ, ПРИБЛИЖАЮЩИХ ФУНКЦИЮ И ЕЕ ПРОИЗВОДНЫЕ

P.K. Korneyev, I.A. Zhuravleva

CONSTRUCTING BEST MEAN-SQUARE POLYNOMIALS APPROXIMATING A FUNCTION AND ITS DERIVATIVES

Предложен прямой метод построения полинома m -й степени наилучшего среднеквадратического приближения, аппроксимирующий одновременно и функцию, и ее производные до m -го порядка. Эти полиномы близки к полиномам наилучших равномерных приближений.

АППРОКСИМАЦИЯ. ПОЛИНОМЫ НАИЛУЧШИХ РАВНОМЕРНЫХ ПРИБЛИЖЕНИЙ. ПОЛИНОМЫ СРЕДНЕКВАДРАТИЧЕСКИХ ПРИБЛИЖЕНИЙ. ЧЕБЫШЁВСКАЯ НОРМА ФУНКЦИЙ.

The paper proposes a direct method for constructing a polynomial of degree m -best mean approximation that approximates both the function and its derivatives up to the m -th order. These polynomials are close to polynomials of best uniform approximations.

APPROXIMATION. BEST UNIFORM POLYNOMIAL APPROXIMATION. POLYNOMIAL APPROXIMATIONS OF THE RMS. CHEBYSHEV NORM FUNCTIONS.

Вопросу приближения функций полиномами наилучшего приближения посвящена журнальная [1], монографическая [2–4], справочная [5–6] и учебная [7] литература.

Алгоритмы построения полиномов наилучшего приближения носят итерационный характер [1, 2, 7]. Но как оказалось, полиномы наилучшего равномерного приближения плохо приближают производные функций [6], а полином наилучшего среднеквадратического приближения, построенный предлагаемым методом, не только хорошо приближает данную функцию на отрезке, но и ее производные.

1. Пусть аналитически заданную функцию $y = f(x)$, имеющую на отрезке $[a, b]$ m производных $f'(x)$, $f''(x)$, ..., $f^{(m)}(x)$ ($f^{(m)}(x)$ может иметь разрыв первого рода), необходимо приблизить вместе с этими производными при помощи многочлена степени m :

$$P_m(x) = \sum_{j=0}^m a_j x^j. \quad (1)$$

Это значит, что функция $f(x)$ приближается многочленом (1), производная $f'(x)$ приближается производной от многочлена (1), производная $f''(x)$ приближается второй производной от многочлена (1) и т. д. Многочлен, приближающий i -ю производную, будет иметь вид:

$$P_m^{(i)}(x) = \sum_{j=0}^m \frac{j!}{(j+1-i)!} a_j x^{j-1}. \quad (2)$$

В частности, производная $f^{(m)}(x)$ будет приближаться постоянным числом $P_m^{(m)}(x) = m! a_m$.

2. Введем обозначения:

$$S_j = \int_a^b (f^{(j)}(x) - P_m^{(j)}(x))^2 dx, \quad j = 1, 2, \dots, m, \quad (3)$$

$$f^{(0)}(x) = f(x), \quad P_m^{(0)}(x) = P_m(x); \quad (4)$$

$$\Phi_j = \Phi_j(a_0, a_1, \dots, a_m) = S_j, \quad j = 1, 2, \dots, m.$$

Поставим задачу: среди многочленов m -й степени (1) найти такой, который реализует минимум каждого из выражений (4) одновременно.

Указанный минимум существует, т. к. выражения Φ_j как функции от a_0, a_1, \dots, a_m

представляют собой многочлены второй степени, кроме того, $\Phi_j \geq 0$.

Для определения тех значений a_j , при которых Φ_j обращается в минимум, составим следующую систему линейных уравнений:

$$\frac{\partial \Phi_j}{\partial a_j} = 0, \quad j = 0, 1, \dots, m. \quad (5)$$

В развернутом виде система (5) имеет вид:

$$\frac{\partial \Phi_j}{\partial a_j} = -2 \int_a^b (f^{(j)}(x) - P_m^{(j)}(x)) j! dx = 0$$

$$A = \begin{bmatrix} \int_a^b dx & \int_a^b x dx & \int_a^b x^2 dx & \int_a^b x^3 dx & \int_a^b x^4 dx & \dots & \int_a^b x^m dx \\ 0 & 1! \int_a^b dx & 2! \int_a^b x dx & 3! \int_a^b x^2 dx & 4! \int_a^b x^3 dx & \dots & m! \int_a^b x^{m-1} dx \\ 0 & 0 & 2! \int_a^b dx & \frac{3!}{1!} \int_a^b x dx & \frac{4!}{2!} \int_a^b x^2 dx & \dots & \frac{m!}{(m-2)!} \int_a^b x^{m-2} dx \\ 0 & 0 & 0 & 3! \int_a^b dx & \frac{4!}{1!} \int_a^b x dx & \dots & \frac{m!}{(m-3)!} \int_a^b x^{m-3} dx \\ 0 & 0 & 0 & 0 & 4! \int_a^b dx & \dots & \frac{m!}{(m-4)!} \int_a^b x^{m-4} dx \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & m! \int_a^b dx \end{bmatrix}$$

является верхней треугольной матрицей и $A_{jj} \neq 0$ ($j = 0, 1, \dots, m$);

$$a = [a_0, a_1, \dots, a_m]^T, \\ f = \left[\int_a^b f(x) dx, \int_a^b f'(x) dx, \dots, \int_a^b f^{(m)}(x) dx \right]^T.$$

Таким образом, система (7) имеет единственное решение a_0, a_1, \dots, a_m , которое будет давать наименьшие значения интегралам S_j ($j = 0, 1, \dots, m$). Подставляя найденное решение в формулу (1), получим искомый полином, решающий поставленную выше задачу.

3. Приведем расчетные формулы определения коэффициентов a_0, a_1, \dots, a_m для некоторых частных случаев.

3.1. В случае аппроксимации функции $f(x)$ многочленом первой степени $P_1(x) = a_0 + a_1x$ система (7) будет иметь вид

или

$$\int_a^b (f^{(j)}(x) - P_m^{(j)}(x)) dx = 0, \quad (6)$$

то есть

$$\int_a^b f^{(j)}(x) dx = \int_a^b P_m^{(j)}(x) dx, \quad j = 0, 1, \dots, m.$$

Последняя система представляет систему $m+1$ линейных уравнений относительно параметров a_0, a_1, \dots, a_m :

$$Aa = f, \quad (7)$$

где матрица

$$(b-a)a_0 + \frac{b^2 - a^2}{2} \cdot a_1 = \int_a^b f(x) dx,$$

$$(b-a)a_1 = \int_a^b f'(x) dx.$$

Решая ее, мы получаем искомые коэффициенты a_1, a_0 :

$$a_1 = \frac{1}{b-a} \int_a^b f'(x) dx,$$

$$a_0 = \frac{1}{b-a} \int_a^b f(x) dx - \frac{b+a}{2} \cdot a_1.$$

3.2. При аппроксимации функции $f(x)$ многочленом второй степени $P_2(x) = a_0 + a_1x + a_2x^2$ система (7) будет иметь вид

$$(b-a)a_0 + \frac{b^2 - a^2}{2} \cdot a_1 + \frac{b^3 - a^3}{3} \cdot a_2 = \int_a^b f(x) dx,$$

$$(b - a)a_1 + (b^2 - a^2)a_2 = \int_a^b f'(x)dx,$$

$$2(b - a)a_2 = \int_a^b f''(x)dx.$$

Решая ее, мы получим искомые коэффициенты a_2, a_1, a_0 :

$$a_2 = \frac{1}{2(b - a)} \int_a^b f''(x)dx,$$

$$a_1 = \frac{1}{b - a} \int_a^b f'(x)dx - (b + a) \cdot a_2,$$

$$a_0 = \frac{1}{b - a} \int_a^b f(x)dx - \frac{b + a}{2} \cdot a_1 - \frac{b^2 + ba + a^2}{3} \cdot a_2.$$

3.3. При аппроксимации $f(x)$ многочленом третьей степени $P_2(x) = a_0 + a_1x + a_2x^2 + a_3x^3$ получим следующие значения коэффициентов a_3, a_2, a_1, a_0 :

$$a_3 = \frac{1}{6(b - a)} \int_a^b f'''(x)dx,$$

$$a_2 = \frac{1}{2(b - a)} \int_a^b f''(x)dx - \frac{3}{2} \cdot (b + a) \cdot a_3,$$

$$a_1 = \frac{1}{b - a} \int_a^b f'(x)dx - (b + a) \times a_2 - (b^2 + ba + a^2) \cdot a_3,$$

$$a_0 = \frac{1}{b - a} \int_a^b f(x)dx - \frac{b + a}{2} \cdot a_1 - \frac{b^2 + ba + a^2}{3} \cdot a_2 - \frac{(b + a)(b^2 + a^2)}{4} \cdot a_3.$$

4. Приведем значения коэффициентов приближающих полиномов для функции $y = \sin(x)$ на отрезке $\left[0; \frac{\pi}{4}\right]$.

$$a) \sin(x) \approx \sum_{k=0}^3 a_k x^k, \quad \|r\|_{\left[0; \frac{\pi}{4}\right]} < 2,1 \cdot 10^{-4}$$

a_0	$2,0002 \cdot 10^{-4}$	a_2	$-9,6849 \cdot 10^{-3}$
a_1	1,0005	a_3	-0,1501

$$б) \sin(x) \approx \sum_{k=0}^4 a_k x^k, \quad \|r\|_{\left[0; \frac{\pi}{4}\right]} < 5,5 \cdot 10^{-5}$$

a_0	$2,93988 \cdot 10^{-6}$	a_3	-0,17446
-------	-------------------------	-------	----------

a_1	1,00048	a_4	0,01554
a_2	$-1,00011 \cdot 10^{-4}$		

$$в) \sin(x) \approx \sum_{k=0}^5 a_k x^k, \quad \|r\|_{\left[0; \frac{\pi}{4}\right]} < 3 \cdot 10^{-6}$$

a_0	$2,939875 \cdot 10^{-6}$	a_3	-0,166747
a_1	1,000007	a_4	$8,071007 \cdot 10^{-4}$
a_2	$-1,000112 \cdot 10^{-4}$	a_5	$7,502623 \cdot 10^{-3}$

$$г) \sin(x) \approx \sum_{k=0}^6 a_k x^k, \quad \|r\|_{\left[0; \frac{\pi}{4}\right]} < 5,6 \cdot 10^{-7}$$

a_0	$4,5352811 \cdot 10^{-8}$	a_4	$8,358343 \cdot 10^{-6}$
a_1	1,0000071	a_5	$8,7230117 \cdot 10^{-3}$
a_2	$-1,4703024 \cdot 10^{-6}$	a_6	$-5,1794893 \cdot 10^{-4}$
a_3	-0,1667472		

$$д) \sin(x) \approx \sum_{k=0}^7 a_k x^k, \quad \|r\|_{\left[0; \frac{\pi}{4}\right]} < 4,6 \cdot 10^{-8}$$

a_0	$4,5352811 \cdot 10^{-8}$	a_4	$8,3335025 \cdot 10^{-6}$
a_1	1,0000001	a_5	$8,3373575 \cdot 10^{-3}$
a_2	$-1,4703024 \cdot 10^{-6}$	a_6	$-2,6902553 \cdot 10^{-5}$
a_3	-0,1666678	a_7	$-1,7863419 \cdot 10^{-4}$

Оценим в этом случае качество приближений производных данной функции построенным полиномом седьмой степени.

Пусть $f(x)$ – данная функция; $P(x)$ – приближающий полином седьмой степени; $f'(x), f''(x), f'''(x), f^{(4)}(x)$ – производные данной функции; $P'(x), P''(x), P'''(x), P^{(4)}(x)$ – производные приближающего полинома $P(x)$; $r_i = f^{(i)}(x) - P^{(i)}(x)$ ($i = 0, 1, 2, 3, 4$) – погрешности. Тогда

$$\|f'(x) - P'(x)\|_{\left[0; \frac{\pi}{4}\right]} < 3,8 \cdot 10^{-7},$$

$$\|f''(x) - P''(x)\|_{\left[0; \frac{\pi}{4}\right]} < 3 \cdot 10^{-6},$$

$$\|f'''(x) - P'''(x)\|_{\left[0; \frac{\pi}{4}\right]} < 2,5 \cdot 10^{-5},$$

$$\|f^{(4)}(x) - P^{(4)}(x)\|_{\left[0; \frac{\pi}{4}\right]} < 2,1 \cdot 10^{-4}.$$

5. Сравним качество аппроксимаций полиномом седьмой степени наилучшего приближения и полиномом седьмой степени, построенным описанным выше методом.

Имеем

$$\sin\left(\frac{\pi}{4} \cdot x\right) \approx p_7(x) = \sum_{k=0}^3 b_{2k+1} x^{2k+1},$$



$$\left\| \sin\left(\frac{\pi}{4} \cdot x\right) - p_7(x) \right\|_{[-1;1]} = 1,2 \cdot 10^{-9},$$

где $p_7(x)$ – полином наилучшего приближения [1].

В нашем случае

$$\left\| \sin(x) - P_7(x) \right\|_{\left[0; \frac{\pi}{4}\right]} < 4,6 \cdot 10^{-8}.$$

Наилучшие среднеквадратические полиномы строятся просто, однако качество приближения в этом случае хуже приближения полиномами наилучшего равномерного приближения примерно на порядок.

СПИСОК ЛИТЕРАТУРЫ

1. Дзядык, В.К. Об эффективном построении многочленов, которые осуществляют близкое к наилучшему приближение функции, e^x , $\sin x$ и др. [Текст] / В.К. Дзядык. – Укр. мат. журн. – 1973. – Вып. 25. – № 4. – С. 435–453.
2. Ремез, Е.Я. Основы численных методов чебышёвского приближения [Текст] / Е.Я. Ремез. – Киев: Наук. думка, 1969. – 623 с.
3. Дзядык, В.К. Введение в теорию приближения функций полиномами [Текст] / В.К. Дзядык. – М.: Наука, 1977. – 512 с.
4. Натансон, И.П. Конструктивная теория функций [Текст] / И.П. Натансон. – М. – Л.: Гос. изд-во технико-теоретич. лит-ры, 1949. – 688 с.
5. Справочная математическая библиотека. Математический анализ. Вычисление элементарных функций [Текст] / Под общ. ред. Л.А. Люстерника, А.Р. Янпольского. – М.: Физматгиз, 1963. – 248 с.
6. Справочная математическая библиотека. Элементы теории функций. Функции действительного переменного. Приближение функций. Почти периодические функции [Текст] / Под общ. ред. П.Л. Ульянова, А.Р. Янпольского. – М.: Физматгиз, 1963. – 244 с.
7. Волков, Е.А. Численные методы [Текст] / Е.А. Волков. – 2-е изд. испр. – М.: Наука. Гл. ред. Физматлит, 1987. – 248 с.

REFERENCES

1. Dziadyk V.K. Ob effektivnom postroenii mnogochlenov, kotorye osushchestvliaiut blizkoe k nailuchshemu priblizhenie funktsii e^x , $\sin x$ i dr. / Ukr. mat. zhurn. – 1973. – Vip. 25. – № 4. – S. 435–453.
2. Remez E.Ia. Osnovy chislennykh metodov chebyshevskogo priblizheniia. – Kiev: Nauk. dumka, 1969. – 623 s.
3. Dziadyk V.K. Vvedenie v teoriu priblizheniia funktsii polinomami. – Moscow: Nauka, 1977. – 512 s. (rus)
4. Natanson I.P. Konstruktivnaia teoriia funktsii. Moscow-Leningrad: Gos. Izd-vo tekhniko-teoreticheskoi literatury, 1949. – 688 s. (rus)
5. Spravochnaia matematicheskaia biblioteka. Matematicheskii analiz. Vychislenie elementarnykh funktsii; Pod obshch. red. L.A. Liusternika, A.R. Ianpol'skogo. – Moscow: Fizmatgiz, 1963. – 248 s. (rus)
6. Spravochnaia matematicheskaia biblioteka. Elementy teorii funktsii. Funktsii deistvitel'nogo peremennogo. Priblizhenie funktsii. Pochti periodicheskie funktsii; Pod obshch. red. P.L. Ul'ianova, A.R. Ianpol'skogo. – Moscow: Fizmatgiz, 1963. – 244 s. (rus)
7. Volkov E.A. Chislennye metody. – 2-e izd. ispr. – Moscow: Nauka. Gl. red. Fizmatlit, 1987. – 248 s. (rus)

КОРНЕЕВ Петр Кириллович – доцент кафедры прикладной математики и математического моделирования Института математики и естественных наук Северо-Кавказского федерального университета, кандидат физико-математических наук.

355000, Россия, г. Ставрополь, пр. Кулакова, д. 2.

KORNEYEV, Petr K. North-Caucasian Federal University.

355000, prosp. Kulakova 2, Stavropol, Russia.

ЖУРАВЛЁВА Ирина Александровна – доцент кафедры прикладной математики и математического моделирования Института математики и естественных наук Северо-Кавказского федерального университета, кандидат педагогических наук.

355000, Россия, г. Ставрополь, пр. Кулакова, д. 2.

ZHURAVLEVA, Irina A. North-Caucasian Federal University.

355000, prosp. Kulakova 2, Stavropol, Russia.

НАУЧНОЕ ИЗДАНИЕ
«НАУЧНО-ТЕХНИЧЕСКИЕ ВЕДОМОСТИ
САНКТ-ПЕТЕРБУРГСКОГО
ГОСУДАРСТВЕННОГО ПОЛИТЕХНИЧЕСКОГО УНИВЕРСИТЕТА.
ИНФОРМАТИКА. ТЕЛЕКОММУНИКАЦИИ. УПРАВЛЕНИЕ»
«ST. PETERSBURG STATE POLYTECHNICAL UNIVERSITY JOURNAL.
COMPUTER SCIENCE. TELECOMMUNICATIONS AND CONTROL SYSTEMS»

№ 4 (176) 2013

Учредитель – Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Санкт-Петербургский государственный политехнический университет»

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор).
Свидетельство о регистрации ПИ № ФС77-51457 от 19.10.2012 г.

Редакция журнала

канд. техн. наук, д-р экон. наук, профессор *А.В. Бабкин* – научный редактор
Е.А. Калинина – литературный редактор, корректор
Г.А. Пышкина – ответственный секретарь, выпускающий редактор

Телефон редакции (812)552-62-16, 297-18-21

E-mail: infocom@spbstu.ru

Компьютерная верстка *А.Н. Смирнов*

Директор Издательства Политехнического университета *А.В. Иванов*

Лицензия ЛР № 020593 от 07.08.97

Подписано в печать 29.08.2013. Формат 60×84 1/8. Бум. тип. № 1.
Печать офсетная. Усл. печ. л. 18,83. Уч.-изд. л. 18,83. Тираж 1000. Заказ

Санкт-Петербургский государственный политехнический университет
Издательство Политехнического университета
член Издательско-полиграфической ассоциации университетов России
Адрес университета и издательства: 195251, Санкт-Петербург, ул. Политехническая, д. 29.

УСЛОВИЯ ПУБЛИКАЦИИ СТАТЕЙ

в журнале «Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление»

1. ОБЩИЕ ПОЛОЖЕНИЯ

Журнал «Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление» является периодическим печатным научным рецензируемым изданием. Зарегистрировано Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ПИ № ФС77-51457 от 19 октября 2012 г. С 2008 года выпускался в составе сериального периодического издания «Научно-технические ведомости СПбГПУ» (ISSN 1994-2354).

Издание с 2002 года входит в Перечень ведущих научных рецензируемых журналов и изданий (перечень ВАК) и принимает для печати материалы научных исследований, а также статьи для опубликования основных результатов диссертаций на соискание ученой степени доктора наук и кандидата наук по следующим основным научным направлениям: **ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА, РАДИОТЕХНИКА И СВЯЗЬ, ЭЛЕКТРОНИКА, ИЗМЕРИТЕЛЬНАЯ ТЕХНИКА, УПРАВЛЕНИЕ В СОЦИАЛЬНЫХ И ЭКОНОМИЧЕСКИХ СИСТЕМАХ**. Научные направления журнала учитываются ВАК Минобрнауки РФ при защите докторских и кандидатских диссертаций в соответствии с Номенклатурой специальностей научных работников.

Сведения о публикации представлены в Реферативном журнале ВИНТИ РАН, в международной справочной системе «Ulrich's Periodical Directory».

Периодичность выхода журнала – 6 номеров в год.

2. ТРЕБОВАНИЯ К ПРЕДОСТАВЛЯЕМЫМ МАТЕРИАЛАМ

2.1. Оформление материалов

1. Рекомендуемый объем статей для авторов с ученой степенью доктора наук, званием профессора, соискателей ученой степени доктора наук (докторантов) 12–20 страниц формата А-4 с учетом графических вложений. Количество графических вложений (диаграмм, графиков, рисунков, таблиц, фотографий и т. п.) не должно превышать 4.

2. Рекомендуемый объем статей для преподавателей, авторов без ученой степени, соискателей ученой степени кандидата наук – 8–15 страниц формата А-4; аспирантов – 8 страниц формата А-4 с учетом графических вложений. Количество графических вложений (диаграмм, графиков, рисунков, таблиц, фотографий и т. п.) не должно превышать 3.

3. Авторы должны придерживаться следующей обобщенной структуры статьи: вводная часть (0,5–1 стр., актуальность, существующие проблемы); основная часть (постановка и описание задачи, изложение и суть основных результатов); заключительная часть (0,5–1 стр., предложения, выводы), список литературы (оформление по ГОСТ 7.1-2003).

4. Число авторов статьи не должно превышать трех человек.

5. Набор текста осуществляется в редакторе **MS Word**, формул – в редакторе **MythType**. Таблицы набираются в том же формате, что и основной текст.

6. Шрифт – **TNR**, размер шрифта основного текста – 14, интервал – 1,5; таблицы большого размера могут быть набраны 12 кеглем. Параметры страницы: поля слева – 3 см, сверху, снизу – 2,5 см, справа – 2 см, текст размещается без переносов. Абзацный отступ – 1 см.

2.2. Предоставление материалов

Вместе с материалами статьи должны быть обязательно предоставлены:

- номер УДК в соответствии с классификатором (в заголовке статьи);
- аннотация на русском и английском языках;
- ключевые слова (5–7) на русском и английском языках;
- сведения об авторах на русском и английском языках: ФИО, место работы, должность, ученое звание, ученая степень, контактные телефоны, e-mail;
- аспиранты представляют документ отдела аспирантуры, заверенный печатью;
- рецензия на имя зам. главного редактора, подписанная специалистом, имеющим ученую степень доктора наук и/или ученое звание профессора. Рецензия должна быть **ОБЯЗАТЕЛЬНО** заверена в отделе кадров. Рецензент несет ответственность за содержание статьи, достоверность представленных материалов;
- акт экспертизы о возможности опубликования материалов в открытой печати.

С авторами статей заключается издательский лицензионный договор.

Предоставление всех материалов осуществляется в электронном виде через личный кабинет **ЭЛЕКТРОННОЙ РЕДАКЦИИ** по адресу <http://journals.spbstu.ru>

2.3. Рассмотрение материалов

Предоставленные материалы (п. 2.2) первоначально рассматриваются редакционной коллегией и передаются для рецензирования. После одобрения материалов, согласования различных вопросов с автором (при необходимости) редакционная коллегия сообщает автору решение об опубликовании статьи. В случае отказа в публикации статьи редакция направляет автору мотивированный отказ.

При отклонении материалов из-за нарушения сроков подачи, требований по оформлению или как не отвечающих тематике журнала материалы не публикуются и не возвращаются.

Редакционная коллегия не вступает в дискуссию с авторами отклоненных материалов.

Публикация материалов аспирантов очной бюджетной формы обучения осуществляется бесплатно в соответствии с очередностью.

При поступлении в редакцию значительного количества статей их прием в очередной номер может закончиться **ДОСРОЧНО**.

Более подробную информацию можно получить:

на сайте журнала <http://ntv.spbstu.ru>

по телефону редакции +7(812) 552-62-16 с 10⁰⁰ до 18⁰⁰ Галина Александровна

или по e-mail: infocom@spbstu.ru