

ISSN 2078-9181

DOI 10.15622/sp.50

РОССИЙСКАЯ АКАДЕМИЯ НАУК
Отделение нанотехнологий и информационных технологий

САНКТ-ПЕТЕРБУРГСКИЙ
ИНСТИТУТ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ РАН

ТРУДЫ СПИИРАН

proceedings.spiras.nw.ru



ВЫПУСК 1(50)



Санкт-Петербург
2017

18+

SPIIRAS PROCEEDINGS

Issue № 1(50), 2017

Scientific, educational, and interdisciplinary journal primarily specialized
in computer science, automation, and applied mathematics

Trudy SPIIRAN ♦ Founded in 2002 ♦ Труды СПИИРАН

Founder and Publisher

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences

Editor-in-Chief

R. M. Yusupov, Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia

Editorial Board Members

A. A. Ashimov ,	Prof., Dr. Sci., Academician of the National Academy of Sciences of the Republic of Kazakhstan, Almaty, Kazakhstan
S. N. Baranov ,	Prof., Dr. Sci., St. Petersburg, Russia
N. P. Veselkin ,	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
V. I. Gorodetski ,	Prof., Dr. Sci., St. Petersburg, Russia
O. Yu. Gusikhin ,	Ph. D., Dearborn, USA
V. Delic ,	Prof., Dr. Sci., Novi Sad, Serbia
A. Dolgui ,	Prof., Dr. Habil., St. Etienne, France
M. Zelezny ,	Assoc. Prof., Ph.D., Plzen, Czech Republic
I. A. Kalyaev ,	Prof., Dr. Sci., Academician of RAS, Taganrog, Russia
D. A. Ivanov ,	Prof., Dr. Habil., Berlin, Germany
G. A. Leonov ,	Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia
K. P. Markov ,	Assoc. Prof., Ph.D., Aizu, Japan
Yu. A. Merkurjev ,	Prof., Dr. Habil., Academician of the Latvian Academy of Sciences, Riga, Latvia
R. V. Meshcheryakov ,	Prof., Dr. Sci., Tomsk, Russia
N. A. Moldovian ,	Prof., Dr. Sci., St. Petersburg, Russia
V. E. Pavlovskiy ,	Prof., Dr. Sci., Moscow, Russia
A. A. Petrovsky ,	Prof., Dr. Sci., Minsk, Belarus
V. A. Putilov ,	Prof., Dr. Sci., Apatity, Russia
V. K. Pshikhopov ,	Prof., Dr. Sci., Taganrog, Russia
A. L. Ronzhin	(Deputy Editor-in-Chief), Prof., Dr. Sci., St. Petersburg, Russia
A. I. Rudskoi ,	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
V. Sgurev ,	Prof., Dr. Sci., Academician of the Bulgarian academy of sciences, Sofia, Bulgaria
V. Skormin ,	Prof., Ph.D., Binghamton, USA
A. V. Smirnov ,	Prof., Dr. Sci., St. Petersburg, Russia
B. Ya. Sovetov ,	Prof., Dr. Sci., Academician of RAE, St. Petersburg, Russia
V. A. Soyfer ,	Prof., Dr. Sci., Academician of RAS, Samara, Russia
B. V. Sokolov ,	Prof., Dr. Sci., St. Petersburg, Russia
L. V. Utkin ,	Prof., Dr. Sci., St. Petersburg, Russia
A. L. Fradkov ,	Prof., Dr. Sci., St. Petersburg, Russia
L. B. Sheremetov ,	Assoc. Prof., Dr. Sci., Mexico, Mexico

Editor: E. P. Miroshnikova

Technical editor: A. I. Motienko

Translator: P. N. Emeleva

Editorial Board's address

14-th line VO, 39, SPIIRAS, St. Petersburg, 199178, Russia,
e-mail: publ@iias.spb.su, web: <http://www.proceedings.spiiras.nw.ru/>

The journal is indexed in Scopus

© St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences, 2017

ТРУДЫ СПИИРАН

Выпуск № 1(50), 2017

Научный, научно-образовательный, междисциплинарный журнал с базовой специализацией в области информатики, автоматизации и прикладной математики
Журнал основан в 2002 году

Учредитель и издатель

Федеральное государственное бюджетное учреждение науки
Санкт-Петербургский институт информатики и автоматизации Российской академии наук
(СПИИРАН)

Главный редактор

Р. М. Юсупов, чл.-корр. РАН, д-р техн. наук, проф., С-Петербург, РФ

Редакционная коллегия

- А. А. Ашимов**, академик национальной академии наук Республики Казахстан д-р техн. наук, проф., Алматы, Казахстан
С. Н. Баранов, д-р физ.-мат. наук, проф., С.-Петербург, РФ
Н. П. Веселкин, академик РАН, д-р мед. наук, проф., С.-Петербург, РФ
В. И. Городецкий, д-р техн. наук, проф., С.-Петербург, РФ
О. Ю. Гусихин, Ph.D., Диаборн, США
В. Делич, д-р техн. наук, проф., Нови-Сад, Сербия
А. Б. Долгий, Dr. Habil., проф., Сент-Этьен, Франция
М. Железны, Ph.D., доцент, Пльзень, Чешская республика
Д. А. Иванов, д-р экон. наук, проф., Берлин, Германия
И. А. Каляев, академик РАН, д-р техн. наук, профессор, Таганрог, РФ
Г. А. Леонов, член-корр. РАН, д-р физ.-мат. наук, проф., С.-Петербург, РФ
К. П. Марков, Ph.D., доцент, Аизу, Япония
Ю. А. Меркурьев, академик Латвийской академии наук, Dr. Habil., проф., Рига, Латвия
Р. В. Мещеряков, д-р техн. наук, профессор, Томск, РФ
Н. А. Молдовян, д-р техн. наук, проф., С.-Петербург, РФ
В. Е. Павловский, д-р физ.-мат. наук, профессор, Москва, РФ
А. А. Петровский, д-р техн. наук, проф., Минск, Беларусь
В. А. Путилов, д-р техн. наук, проф., Апатиты, РФ
В. Х. Шихопов, д-р техн. наук, профессор, Таганрог, РФ
А. Л. Ронжин (зам. главного редактора), д-р техн. наук, проф., С.-Петербург, РФ
А. И. Рудской, академик РАН, д-р техн. наук, проф., С.-Петербург, РФ
В. С. Струев, академик Болгарской академии наук, д-р техн. наук, проф., София, Болгария
В. А. Скормин, Ph.D., проф., Бингемптон, США
А. В. Смирнов, д-р техн. наук, проф., С.-Петербург, РФ
Б. Я. Советов, академик РАН, д-р техн. наук, проф., С.-Петербург, РФ
В. А. Соيفер, академик РАН, д-р техн. наук, проф., Самара, РФ
Б. В. Соколов, д-р техн. наук, проф., С.-Петербург, РФ
Л. В. Уткин, д-р техн. наук, проф., С.-Петербург, РФ
А. Л. Фрадков, д-р техн. наук, проф., С.-Петербург, РФ
Л. Б. Шереметов, д-р техн. наук, Мехико, Мексика

Редактор: Е. П. Мирошникова

Технический редактор: А. И. Мотиенко

Переводчик: П. Н. Емелева

Адрес редакции

199178, Санкт-Петербург, 14-я линия, д. 39,
e-mail: publ@iias.spb.su, сайт: <http://www.proceedings.spiras.nw.ru/>

Журнал индексируется в международной базе данных Scopus

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук»

© Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук, 2017
Разрешается воспроизведение в прессе, а также сообщение в эфир или по кабелю опубликованных в составе печатного периодического издания—журнала «Труды СПИИРАН» статей по текущим экономическим, политическим, социальным и религиозным вопросам с обязательным указанием имени автора статьи и печатного периодического издания—журнала «Труды СПИИРАН»

CONTENTS

Means of Computer Facilities and Control Systems

Kalyaev I.A., Dordopulo A.I., Levin I.I., Fedorov A.M. EVOLUTION DOMESTIC OF MULTICHIP RECONFIGURABLE COMPUTER SYSTEMS: FROM AIR TO LIQUID COOLING: FROM AIR TO LIQUID COOLING	5
Zheleznov D.V., Tarasov E.M., Isaycheva A.G, Mikheyeva T.I. DEVELOPMENT OF THE LEARNING CLASSIFIER OF RAIL LINES STATES WITH MULTIVARIATE INFORMATIVE FEATURES	32

Methods of Information Processing and Management

A.A. Petrovsky, A.A. Petrovsky SCALABLE SPEECH AND AUDIO CODERS BASED ON ADAPTIVE TIME-FREQUENCY SIGNAL ANALYSIS	55
Popov A.I., Svinyin S.F. APPLICATION OF FINITE BASIS SPLINES IN RECONSTRUCTING ELECTROGASTROENTEROGRAPHY SIGNALS	93
Zuenko A.A., Lomov P.A., Oleynik A.G. APPLICATION OF CONSTRAINT PROPAGATION TECHNIQUES TO SPEED UP PROCESSING OF QUERIES TO ONTOLOGIES	112
Sapozhnikov V.V., Sapozhnikov V.I., Efanov D.V., Kotenko A.G. MODULO CODES WITH SUMMATION OF WEIGHTED TRANSITIONS WITH NATURAL NUMBER SEQUENCE OF WEIGHTS	137

Algorithms and Software

Marchenkov S.A., Vdovenko A.S., Korzun D.G. ENHANCING THE OPPORTUNITIES OF COLLABORATIVE WORK IN AN INTELLIGENT ROOM USING E-TOURISM SERVICES	165
---	-----

Theoretical and Applied Mathematics

Kotov Yu.A. APPROXIMATION OF DISTRIBUTIONS OF TEXT CHARACTERS BIGRAMS FREQUENCIES FOR ALPHABETIC CHARACTERS IDENTIFICATION	190
Balonin Yu.N., Vostrikov A.A., Sergeev A.M., Egorova I.S. ON RELATIONSHIPS AMONG QUASI-ORTHOGONAL MATRICES CONSTRUCTED ON THE KNOWN SEQUENCES OF PRIME NUMBERS	209

СОДЕРЖАНИЕ

Средства вычислительной техники и систем управления

- Каляев И.А., Дордопуло А.И., Левин И.И., Федоров А.М.
РАЗВИТИЕ ОТЕЧЕСТВЕННЫХ МНОГОКРИСТАЛЬНЫХ РЕКОНФИГУРИРУЕМЫХ
ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ: ОТ ВОЗДУШНОГО К ЖИДКОСТНОМУ ОХЛАЖДЕНИЮ 5
- Железнов Д.В., Тарасов Е.М., Исайчева А.Г., Михеева Т.И.
РАЗРАБОТКА ОБУЧАЕМОГО КЛАССИФИКАТОРА СОСТОЯНИЙ РЕЛЬСОВЫХ ЛИНИЙ
С МНОГОМЕРНЫМИ ИНФОРМАТИВНЫМИ ПРИЗНАКАМИ 32

Методы управления и обработки информации

- Петровский А.А., Петровский А.А.
МАСШТАБИРУЕМЫЕ АУДИОРЕЧЕВЫЕ КОДЕРЫ НА ОСНОВЕ АДАПТИВНОГО
ЧАСТОТНО-ВРЕМЕННОГО АНАЛИЗА ЗВУКОВЫХ СИГНАЛОВ 55
- Попов А.И., Свинын С.Ф.
ПРИМЕНЕНИЕ ФИНИТНЫХ БАЗИСНЫХ СПЛАЙНОВ ПРИ ВОССТАНОВЛЕНИИ
СИГНАЛОВ ЭЛЕКТРОГАСТРОЭНТЕРОГРАФИИ 93
- Зуенко А.А., Ломов П.А., Олейник А.Г.
ПРИМЕНЕНИЕ МЕТОДОВ РАСПРОСТРАНЕНИЯ ОГРАНИЧЕНИЙ ДЛЯ УСКОРЕНИЯ
ОБРАБОТКИ ЗАПРОСОВ К ОНТОЛОГИЯМ 112
- Сапожников В.В., Сапожников Вл.В., Ефанов Д.В., Котенко А.Г.
МОДУЛЬНЫЕ КОДЫ С СУММИРОВАНИЕМ ВЗВЕШЕННЫХ ПЕРЕХОДОВ С
ПОСЛЕДОВАТЕЛЬНОСТЬЮ ВЕСОВЫХ КОЭФФИЦИЕНТОВ, ОБРАЗУЮЩЕЙ
НАТУРАЛЬНЫЙ РЯД ЧИСЕЛ 137

Алгоритмы и программные средства

- Марченков С.А., Вдовенко А.С., Корзун Д.Ж.
РАСШИРЕНИЕ ВОЗМОЖНОСТЕЙ СОВМЕСТНОЙ ДЕЯТЕЛЬНОСТИ В
ИНТЕЛЛЕКТУАЛЬНОМ ЗАЛЕ НА ОСНОВЕ СЕРВИСОВ ЭЛЕКТРОННОГО ТУРИЗМА 165

Теоретическая и прикладная математика

- Котов Ю.А.
АППРОКСИМАЦИЯ РАСПРЕДЕЛЕНИЙ ЧАСТОТ БУКВЕННЫХ БИГРАММ ТЕКСТА ДЛЯ
ИДЕНТИФИКАЦИИ БУКВ 190
- Балонин Ю.Н., Востриков А.А., Сергеев А.М., Егорова И.С.
О ВЗАИМОСВЯЗЯХ КВАЗИОРТОГОНАЛЬНЫХ МАТРИЦ, ПОСТРОЕННЫХ НА
ИЗВЕСТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЯХ ЧИСЕЛ 209

И.А. КАЛЯЕВ, А.И. ДОРДОПУЛО, И.И. ЛЕВИН, А.М. ФЕДОРОВ
**РАЗВИТИЕ ОТЕЧЕСТВЕННЫХ МНОГОКРИСТАЛЬНЫХ
РЕКОНФИГУРИРУЕМЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ: ОТ
ВОЗДУШНОГО К ЖИДКОСТНОМУ ОХЛАЖДЕНИЮ**

Каляев И.А., Дордопуло А.И., Левин И.И., Федоров А.М. Развитие отечественных многокристалльных реконфигурируемых вычислительных систем: от воздушного к жидкостному охлаждению.

Аннотация. В статье рассматривается история развития технологий построения реконфигурируемых вычислительных систем на основе кристаллов ПЛИС различных семейств. Описано пять поколений реконфигурируемых вычислительных систем с высокой плотностью компоновки, созданных на основе ПЛИС, начиная с Xilinx Virtex-E и заканчивая современными ПЛИС Virtex UltraScale. Представлены последние достижения в области проектирования энергоэффективных реконфигурируемых вычислительных систем с высокой реальной производительностью — разработанная система жидкостного охлаждения для кристаллов Virtex UltraScale, которая обеспечивает автономную циркуляцию охлаждающей жидкости в вычислительном модуле с размерами 3U на 19" для охлаждения 96-128 кристаллов ПЛИС с выделяемой суммарной тепловой мощностью 9,6-12,8 кВт. Отличительными характеристиками разработанной погружной системы жидкостного охлаждения являются высокая эффективность охлаждения с резервом мощности для проектируемых перспективных семейств ПЛИС, нечувствительностью к протечкам и их последствиям и совместимость с традиционными системами водяного охлаждения на базе промышленных чиллеров.

Ключевые слова: реконфигурируемые вычислительные системы, ПЛИС, энергоэффективность, жидкостное охлаждение, промышленные чиллеры

1. Введение. Одним из перспективных способов достижения высокой реальной производительности вычислительной системы является адаптация ее архитектуры под структуру решаемой задачи, а также создание специализированного вычислительного устройства, аппаратно реализующего все вычислительные операции информационного графа задачи с наименьшими задержками. Естественным требованием к современной вычислительной системе является аппаратная поддержка возможности изменения как алгоритма решаемой задачи, так и самой задачи, что обуславливает применение программируемых логических интегральных схем (ПЛИС) в качестве основного вычислительного ресурса реконфигурируемых вычислительных систем [1].

Основными преимуществами ПЛИС являются: возможность реализации сложных параллельных алгоритмов, возможность адаптации конфигурации вычислительной системы под структуру решаемой задачи, высокая производительность получаемого решения, совместимость различных проектов на языках описания аппаратуры — VHDL, AHDL, Verilog и др. Следует учитывать, что вычислительные системы на основе ПЛИС не лишены и недостатков: сложность проектирования, изго-

товления и программирования, необходимость эффективного охлаждения, приобретающая особое значение для многокристалльных вычислительных систем, поскольку рабочие температуры ПЛИС, как правило, на 30-40% ниже, чем у микропроцессоров. Поэтому в настоящее время разработчикам реконфигурируемых вычислительных систем необходимо обеспечить эффективное охлаждение вычислительных модулей, содержащих не менее восьми кристаллов ПЛИС с высокой плотностью компоновки, поскольку воздушные системы охлаждения, успешно использовавшиеся для ПЛИС с момента их появления, практически достигли своего предела для охлаждения суперкомпьютеров.

2. История развития ПЛИС как элементной базы для РВС. История развития архитектур ПЛИС берет свое начало с конца 1970-х годов, когда на рынке появились ПЛИС с программируемыми матрицами «И» и «ИЛИ» — архитектуры FPLA (Field Programmable Logic Array) и FPLS (Field Programmable Logic Sequencers) [2]. Их недостатком было слабое использование ресурсов программируемой матрицы «ИЛИ».

Идя по пути совершенствования архитектуры, в конце 1980-х годов разработчики ПЛИС предложили более простую и изящную архитектуру программируемой матричной логики PAL (Programmable Array Logic) и GAL (Gate Array Logic) — ПЛИС фирм Intel, Altera, AMD, Lattice и других с небольшой степенью интеграции, имеющую программируемую матрицу «И» и фиксированную матрицу «ИЛИ» [3]. Другим подходом к уменьшению избыточности программируемой матрицы «ИЛИ» являлась программируемая макрологика. Построенные на основе данной архитектуры микросхемы содержали единственную программируемую матрицу «И-НЕ» или «ИЛИ-НЕ», но за счет многочисленных инверсных обратных связей были способны формировать сложные логические функции.

В начале 1980-х годов на мировой рынок микроэлектронных изделий выходят три ведущие фирмы-производители ПЛИС. В июне 1983 года основана фирма Altera Corporation [4], в феврале 1984 — компания Xilinx Inc. [5], в 1985 году — Actel Corporation [6]. Эти три компании в настоящее время занимают до 80% всего рынка ПЛИС и являются основными идеологами их применения. Если ранее ПЛИС являлись одним из множества продуктов, выпускаемых такими гигантами как Intel, AMD и другие, то начиная с середины 1980-х годов на рынке ПЛИС происходит специализация, и законодателями мод становятся фирмы, специализирующиеся только на разработке и производстве ПЛИС.

С появлением новых производителей возникли и новые архитектуры. ПЛИС CPLD (Complex Programmable Logic Devices) [7] имеют довольно высокую степень интеграции и содержат несколько логиче-

ских блоков, объединенных коммутационной матрицей. Каждый логический блок представляет собой программируемую матрицу «И» и фиксированную матрицу «ИЛИ». К этому классу относятся ПЛИС семейств MAX7000 фирмы Altera, XC9500 фирмы Xilinx, а также большое число микросхем других производителей (Atmel, Vantis, Lucent и др.).

Настоящая революция в устройствах программируемой логики произошла с появлением ПЛИС, имеющих архитектуру Field Programmable Gate Array (FPGA) [8]. К FPGA относятся ПЛИС XC2000, XC3000, XC4000, Spartan фирмы Xilinx, АСТ1, АСТ2 фирмы Actel, а также семейства FLEX8000 фирмы Altera, некоторые ПЛИС Atmel и Vantis.

В ПЛИС FPGA множество конфигурируемых логических блоков (Configurable Logic Blocks (CLBs)) объединяется с помощью матрицы соединений. Логические блоки таких ПЛИС состоят из одного или нескольких относительно простых логических элементов, в основе которых лежат: четырехходовая таблица преобразования (Look-up table, LUT), программируемый мультиплексор, D-триггер, а также цепи управления. Характерными для FPGA-архитектур являются элементы ввода-вывода, позволяющие реализовать двунаправленный ввод/вывод, третье состояние и т.п. Особенностью современных ПЛИС FPGA является возможность тестирования узлов с помощью порта JTAG (B-scan), а также наличие внутреннего генератора (Osc) и схем управления последовательной конфигурацией.

Дальнейшее развитие архитектуры FPGA привело к появлению ПЛИС, имеющих встроенные модули памяти, позволяющие использовать ПЛИС без применения внешних запоминающих устройств. Благодаря большой логической емкости, удобной архитектуре, достаточно высокой надежности и удачному соотношению цена/логическая емкость данные ПЛИС удовлетворяют разнообразным требованиям, возникающим у разработчика.

В последние годы, наряду с заказными SoC, ПЛИС также становятся основой для «систем-на-кристалле». IP-ядра этих систем разрабатываются отдельно и могут многократно использоваться в различных проектах. Окончательная структура SoC-микросхемы на базе ПЛИС выполняется на основе IP-ядер с помощью программ автоматизации проектирования электронных устройств.

Идеология построения SoC подстегнула ведущих производителей ПЛИС к выпуску в конце 1998 – начале 1999 года изделий с эквивалентной емкостью миллион эквивалентных вентилях и более. Примером новых семейств ПЛИС, пригодных для реализации SoC, является семейство ArEX20K фирмы Altera.

Сходной архитектурой обладает и семейство Virtex фирмы Xilinx [9], которое характеризуется широким разнообразием высокоскоростных трассировочных ресурсов, наличием выделенного блочно-го ОЗУ, развитой логикой ускоренного переноса. ПЛИС данной серии обеспечивают высокие скорости межкристального обмена — до 200 МГц (стандарт HSTL IV). Кристаллы серии Virtex за счет развитой технологии производства и усовершенствованного процесса верификации имеют достаточно низкую стоимость (до 40 % от эквивалентной стоимости серии XC4000XL).

Увеличение эквивалентной логической емкости ПЛИС привело к тому, что в 1998–1999-х годах началось изменение отношения к программному обеспечению САПР ПЛИС со стороны как разработчиков программного обеспечения, так и пользователей. Если до конца 1990-х годов основным средством описания проекта являлся ввод схемы с помощью графических редакторов с использованием библиотек стандартных логических примитивов — логических элементов, простейших комбинационных и последовательностных функциональных узлов, аналогов стандартных интегральных схем малой и средней степени интеграции, то в настоящее время разработчиками широко используются языки описания аппаратуры (Hardware Description Languages) для реализации алгоритмов на ПЛИС. Причем в современных САПР поддерживаются как стандартизированные языки описания аппаратуры типа VHDL, Verilog, HDL, так и языки описания аппаратуры, разработанные компаниями-производителями ПЛИС специально для использования только в своих САПР и учитывающие архитектурные особенности конкретных семейств ПЛИС. Примером может служить AHDL (Altera Hardware Description Languages), поддерживаемый САПР MAX PLUS II и Quartus компании Altera. HDL-языки являются очень удобным средством описания различного рода интерфейсов, однако при реализации сложных вычислений разработчик не имеет возможности влиять на процедуру отображения написанного на HDL кода на физические ресурсы кристалла, что отрицательно сказывается на эффективности реализации. Кроме того, очевидна наглядность схемотехнического описания, поэтому при реализации сложного вычислительного алгоритма представляется предпочтительным использование графического редактора.

Фирма Xilinx предлагает IP-ядра для часто используемых элементов, включая блоки цифровой обработки сигналов, шинные интерфейсы, процессоры и периферийные устройства. Использование пакета LogiCORE™ фирмы Xilinx и аналогичных программ сторонних производителей для создания IP-ядер сокращает время разработки

проектов, уменьшает риски и позволяет получать наивысшую производительность. В дополнение к этому система генерации IP-ядер CORE Generator™ позволяет разработчикам реализовать собственные IP-ядра с предсказуемыми и воспроизводимыми временными характеристиками. Данный продукт имеет простой пользовательский интерфейс для генерации параметризованных IP-ядер, оптимизированных для FPGA Xilinx.

3. Построение вычислительных систем на основе ПЛИС. В то же время появление на рынке высокопроизводительных кристаллов ПЛИС с FPGA-архитектурой открывает широкие перспективы по их использованию в качестве элементной базы для построения высокопроизводительных вычислительных систем (суперкомпьютеров).

В настоящее время высокопроизводительные вычислительные системы, использующие ПЛИС в качестве элементной базы, принято разбивать на два типа. К первому типу относятся так называемые гибридные вычислительные системы, представляющие собой классические кластерные вычислители, в микропроцессорных узлах которых ПЛИС используются в качестве ускорителей вычислений. В качестве примера гибридных суперкомпьютеров можно привести XT4 компании Cray и RASC компании SiliconGraphics. В этих системах блоки программируемых сопроцессоров реализованы на базе ПЛИС, связанных высокоскоростными магистралями с основными процессорами и между собой.

Второй тип вычислительных систем, использующих ПЛИС в качестве элементной базы, представляют реконфигурируемые вычислительные системы (PBC). В PBC ПЛИС используются в качестве основного вычислительного элемента, в то время как процессорам общего назначения отводится второстепенная функция сопровождения и контроля над работой реконфигурируемой части системы.

Признанным лидером в области создания высокопроизводительных реконфигурируемых систем на основе вычислительных полей ПЛИС является таганрогская школа, основанная академиком А.В. Каляевым и представленная в настоящее время разработками НИИ многопроцессорных вычислительных систем Таганрогского радиотехнического университета (ныне Южного федерального университета) и Научно-исследовательского центра супер-ЭВМ и нейрокомпьютеров. Научным коллективом разрабатываются и производятся PBC класса супер-ЭВМ, в которых основным вычислительным ресурсом являются не микропроцессоры, а множество кристаллов ПЛИС, объединенных в вычислительные поля высокоскоростными каналами передачи данных. Спектр выпускаемых и проектируемых изделий до-

статочно широк: от полностью автономных малогабаритных реконфигурируемых ускорителей (вычислительных блоков), вычислительных модулей в настольном или стоечном конструктивном исполнении (на основе Xilinx Virtex-6, Virtex-7 и Virtex UltraScale) до вычислительных систем, состоящих из нескольких вычислительных шкафов, размещаемых в специально оборудованном машинном зале[10].

Начиная с 2001 года произошла смена четырех поколений реконфигурируемых вычислительных систем, построенных на базе ПЛИС, которая обусловлена выходом новых семейств ПЛИС, повышением вычислительной сложности задач, требующих постоянного роста производительности РВС. Первым поколением РВС были РВС с макропроцессорной архитектурой (РВС МПА), которая включала в свой состав группу базовых модулей, реализованных с помощью ПЛИС, и персональную ЭВМ (ПЭВМ). Каждый базовый модуль (БМ) представлял собой реконфигурируемое вычислительное устройство, построенное по тем же архитектурным принципам, что и система в целом. Подобный подход обеспечивал естественную реализацию структурно-процедурных параллельных программ для различных степеней параллелизма и конвейеризации вычислений.

В 2001 году был разработан и создан одноплатный базовый модуль РВС МПА на ПЛИС Xilinx Virtex E. Плата была изготовлена по 12-слойной технологии с двухсторонним монтажом элементов, при этом шесть слоев — сигнальных и шесть слоев — потенциальных: два слоя земли и по два слоя каждого напряжения питания 1.8 В и 3.3 В. На рисунке 1 показаны лицевая и тыльная стороны платы базового модуля РВС МПА.

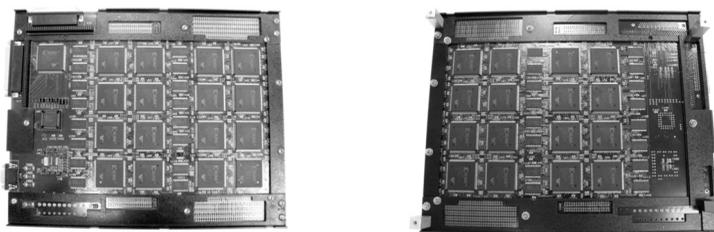


Рис. 1. Лицевая и тыльная стороны платы базового модуля РВС МПА

Плата БМ содержала 32 ПЛИС и 32 микросхемы оперативной памяти, конструктивно расположенные на обеих ее сторонах. Элементы на плате были размещены для обеспечения минимальной длины связей между элементами. Производительность базового модуля РВС МПА составляла $2,5 \cdot 10^{10}$ оп/сек, количество элементарных процессоров обработки данных — 64, а потребляемая мощность 30 Вт. На ос-

нове платы базового модуля РВС МПА был разработан и создан ряд модульно-наращиваемых многопроцессорных систем [1]. На смену РВС МПА пришли реконфигурируемые вычислительные системы второго поколения, к которым относят РВС с макрообъектной архитектурой. Программирование РВС с макрообъектной архитектурой обеспечивает пользователю возможность двухуровневого программирования архитектуры РВС. Такая организация программирования архитектуры РВС позволяет на порядок повысить эффективность вычислительного процесса по сравнению с РВС МПА. Первым представителем семейства РВС второго поколения с макрообъектной архитектурой была модульно-наращиваемая РВС «Медведь», базовый модуль которой представлен на рисунке 2.

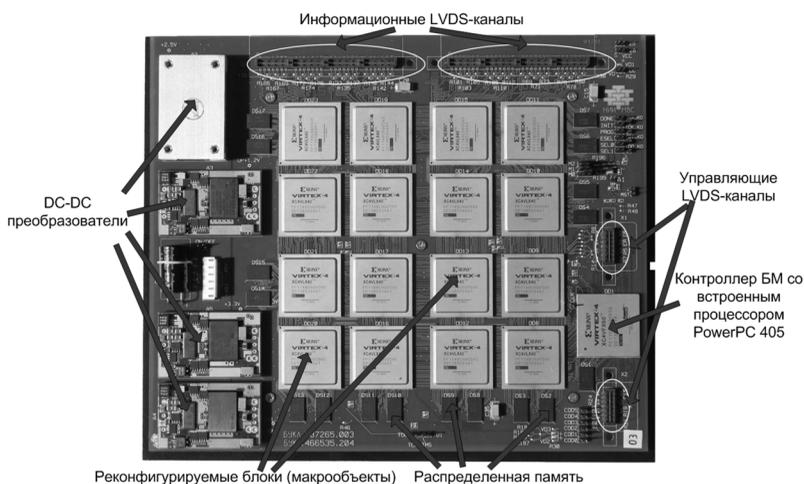


Рис. 2. Общий вид базового модуля РВС «Медведь»

Основу базового модуля составляют: 20-слойная печатная плата с двухсторонним монтажом элементов, вычислительное поле из 16 ПЛИС Xilinx XC4VLX40-10FF1148, контроллер обращений, 17 микросхем динамической памяти SDRAM, программируемый тактовый генератор, четыре малогабаритных DC-DC преобразователя напряжения, LVDS-разъемы для соединения базовых модулей по информационным каналам и другие элементы поверхностного монтажа. Производительность базового модуля РВС «Медведь» составляла 50 Гфлопс, частота базового модуля — 160 МГц, а потребляемая мощность не превышала 150 Вт.

Принципы макрообъектной архитектуры также были использованы при создании малогабаритного реконфигурируемого ускорителя персонального компьютера, предназначенного для решения вычисли-

тельно трудоемких фрагментов задач различных предметных областей. На рисунке 3 представлен внешний вид малогабаритного реконфигурируемого ускорителя персонального компьютера.

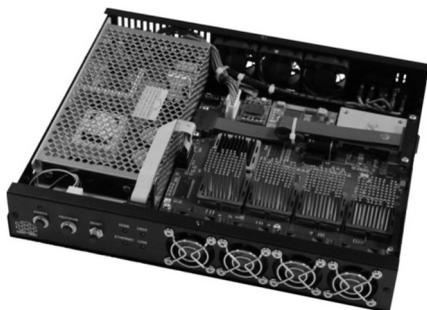


Рис. 3. Малогабаритный реконфигурируемый ускоритель персонального компьютера

В основе ускорителя лежал базовый модуль 4V4-25, выполненный на 18-слойной печатной плате размером 150×190 мм с двухсторонним расположением элементов. Производительность базового модуля 4V4-25 составляла 25 Гфлопс, частота базового модуля — 160 МГц, а потребляемая мощность не превышала 145 Вт.

Накопленный при создании РВС МПА и РВС «Медведь» опыт позволил перейти к реализации РВС третьего поколения, первым представителем которого является семейство РВС «Большая Медведица», которое строилось на основе трех типов базовых модулей: 16V5-75, 16V5-50 и 16S3-25. Базовый модуль 16V5-75 (как самый производительный) использовался в старших моделях семейства: системах РВС-5, РВС-1Р и РВС-0.2-РС. Базовые модули 16V5-50 и 16S3-25 входили в состав ускорителей персональных компьютеров РУПК-50 и РУПК-25.

Решающее поле базового модуля 16V5-75 выполнено на 16-ти ПЛИС Virtex-5 XC5VLX110-2FF1153 фирмы Xilinx, содержащих 11 миллионов эквивалентных вентилей. ПЛИС расположены в узлах двумерной решетки 4×4 и соединены между собой ортогональной системой связей по близкодействию. Такая система связей позволяет существенно упростить печатную плату и улучшить ее частотные характеристики, поскольку связи между соседними микросхемами не превышают четыре сантиметра. Данные между удаленными ПЛИС передаются по транзитным каналам через промежуточные микросхемы, используя систему ортогональных связей. К периферийным ПЛИС решающего поля, расположенным на краях двумерной решетки 4×4, подключено 20 микросхем динамических SDRAM DDR2,

которые образуют распределенную память общим объемом 1,25 Гбайт. На рисунке 4 показан внешний вид платы базового модуля 16V5-75 с установленными электронными элементами и в сборе с системой охлаждения.

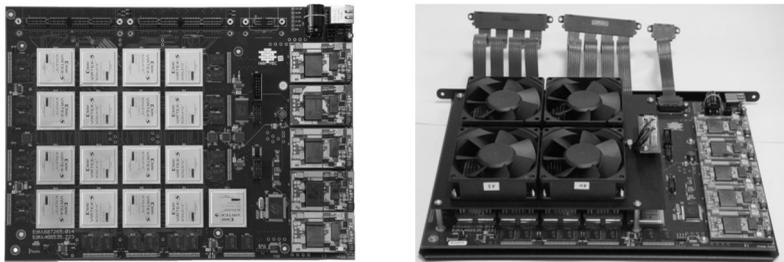


Рис. 4. Плата базового модуля 16V5-75

Для отвода тепла и поддержания необходимых температурных режимов в микросхемах базового модуля предусмотрена комбинированная система охлаждения, которая включает радиаторы на ПЛИС решающего поля и вентиляторы для их обдува. В целом базовый модуль 16V5-75 представляет собой мощный вычислительный узел производительностью свыше 75 (140) Гфлопс. На основе базового модуля 16V5-75 были разработаны рабочая станция PBC-0.2-PC и вычислительный блок PBC-0.2-ВБ производительностью 300 Гфлопс. На рисунке 5 показан общий вид рабочей станции PBC-0.2-ВБ со снятой верхней крышкой и вычислительный блок PBC-0.2-В.



Рис. 5. Рабочая станция PBC-0.2-PC и вычислительный блок PBC-0.2-ВБ

На основе вычислительных блоков PBC-0.2-ВБ была создана реконфигурируемая вычислительная система PBC-1P с пиковой производительностью 1200 Гфлопс, предназначенная для оснащения научных центров, проводящих исследования в области физики, химии, биологии,

космоса, построения информационно-управляющих систем для управления потенциально опасными производствами, решения задач аэрокосмической и автомобильной промышленности, энергетики и т.п.

Старшим представителем РВС третьего поколения являлась система РВС-5 (рисунок 6) с пиковой производительностью 6000 Гфлопс, содержащая пять стоек РВС-1Р, соединенных между собой коммутаторами Ethernet с общим управлением.

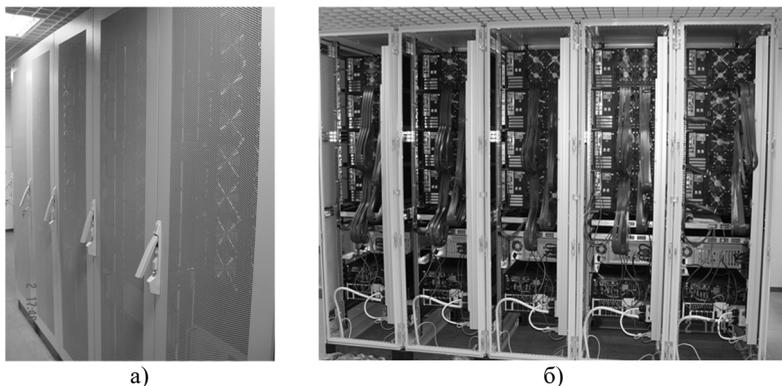


Рис. 6. РВС-5: а) вид спереди; б) вид сзади

РВС-5, установленная в НИВЦ МГУ им. М.В. Ломоносова, содержит 20 вычислительных блоков РВС-0.2-ВБ, 80 базовых модулей 16V5-75, 16384 элементарных процессора (IEEE-754), осуществляющих обработку 64-разрядных данных по стандарту IEEE-754 на тактовой частоте 330 МГц с производительностью более 6000 Гфлопс. Обмен между блоками осуществляется по внешним интерфейсам LVDS, Gigabit Ethernet на частоте 640 МГц.

Растущие требования к производительности РВС обусловили четырехкратное повышение плотности компоновки вычислительных систем и создание в 2012-2014 годах на основе ПЛИС Xilinx Virtex-7 РВС четвертого поколения на базе вычислительных модулей (ВМ) 24V7-750 («Плеяда») и «Тайгета» [11].

Вычислительный модуль 24V7-750, представленный на рисунке 7, содержит четыре платы 6V7-180; управляющий модуль УМ-7; подсистему питания; подсистему охлаждения и другие подсистемы. Производительность ВМ 24V7-750 составляет 2,58 Тфлопс при обработке 32-разрядных данных с плавающей запятой.



Рис. 7. Вычислительный модуль 24V7-750: а) со снятой верхней крышкой; б) с установленной верхней крышкой)

ВМ 24V7-750 использовался для создания по государственному контракту №14.527.12.0004 от 03.10.2011 реконфигурируемой вычислительной системы PBC-7, которая содержит 24 ВМ 24V7-750 с вычислительным полем из 576 микросхем ПЛИС Virtex-7 XC7V585T-FFG1761 объемом 58 миллионов эквивалентных вентилях каждая, конструктивно объединенных в один вычислительный шкаф высотой 47U с пиковой производительностью 10^{15} операций с фиксированной запятой в секунду. Производительность PBC-7 при комплектации от 24 до 36 ВМ 24V7-750 составит от 62 до 93 Тфлопс при обработке 32-разрядных данных с плавающей запятой и 19,4–29,4 Тфлопс при обработке 64-разрядных данных с плавающей запятой. Области применения PBC-7 и вычислительных комплексов на ее основе являются решение задач цифровой обработки сигналов и многоканальная цифровая фильтрация.

На основе ПЛИС Virtex-7 также был разработан ВМ «Тайгета» в конструктивном исполнении высотой 2U, предназначенный для установки в стандартную 19" вычислительную стойку. ВМ «Тайгета», представленный на рисунке 8а, содержит четыре ПВМ 8V7-200, представленных на рисунке 8б, соединенных быстрыми LVDS-каналами; встроенную управляющую ЭВМ; систему питания; систему управления; систему охлаждения и другие подсистемы.

ПВМ 8V7-200, лежащая в основе ВМ «Тайгета», представляет собой 20-слойную печатную плату с двухсторонним монтажом элементов, на которой располагаются 8 ПЛИС типа XC7VX485T-1FFG1761, содержащих 48,5 миллионов эквивалентных вентилях, 16 микросхем распределенной памяти SDRAM типа DDR2 общим объемом 2 Гбайт, интерфейсы LVDS и Ethernet и другие компоненты. Производительность ВМ «Тайгета» составляет 2,66 Тфлопс при обработке 32-разрядных данных с плавающей запятой. Производительность

ность PBC на основе ВМ «Тайгета» при комплектации 18 ВМ «Тайгета» составляет 48 Тфлопс при обработке 32-разрядных данных с плавающей запятой и 23 Тфлопс при обработке 64-разрядных данных с плавающей запятой.

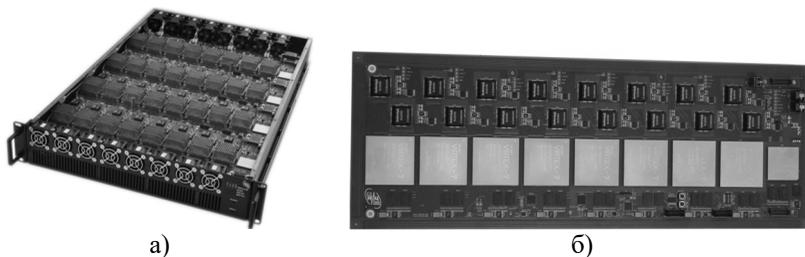


Рис. 8. ВМ «Тайгета»: а) со снятой верхней крышкой, б) плата 8V7-200

Переход к четвертому поколению PBC был обусловлен проблемами системы охлаждения. Согласно полученным экспериментальным данным переход от семейства ПЛИС Virtex-6 к следующему семейству ПЛИС Virtex-7 ведет к росту максимальной температуры ПЛИС на 11...15°C. Поэтому дальнейшее развитие технологий изготовления ПЛИС и переход на следующее семейство ПЛИС Virtex Ultra Scale приведет к росту величины перегрева ПЛИС на дополнительные 10...15°C, что сместит зону их рабочей температуры до 80...85°C с выходом за допустимые пределы рабочей температуры ПЛИС (65...70°C), что негативно скажется на их надежности.

Практический опыт эксплуатации больших вычислительных комплексов на основе ВМ PBC показывает, что воздушные системы охлаждения подошли к своему тепловому пределу. Поэтому использование воздушного охлаждения при переходе к следующему поколению ПЛИС Virtex UltraScale, содержащих порядка 100 млн. эквивалентных вентиляей и обладающих энергопотреблением не менее 100 Вт на кристалл ПЛИС, не позволяет обеспечить стабильную и надежную работу PBC при заполнении кристаллов на 85-95% от доступного аппаратного ресурса.

Это обстоятельство потребовало перехода к принципиально иному способу охлаждения, позволившему сохранить темпы роста производительности PBC для перспективных проектируемых семейств ПЛИС фирмы Xilinx: Virtex UltraScale, Virtex UltraScale+, Virtex UltraScale2 и др.

Большинство разработчиков вычислительной техники видит выход в применении систем жидкостного охлаждения, которые на сегодняшний день представляются наиболее перспективной областью

разработки для охлаждения современных высоконагруженных электронных компонентов вычислительных систем.

Существенными преимуществами всех жидкостных схем охлаждения являются лучшая теплоемкость жидкостей по сравнению с воздухом (от 1500 до 4000 раз) и больший коэффициент теплоотдачи (увеличение до 100 раз). Для охлаждения одного современного кристалла ПЛИС необходимо наличие 1 м^3 воздуха или $0,00025 \text{ м}^3$ (250 мл) воды в минуту. На перекачивание 250 мл воды тратится намного меньше электроэнергии, чем для перекачивания 1 м^3 воздуха.

4. Жидкостное охлаждение для реконфигурируемых вычислительных систем. Развитие вычислительных технологий приводит к разработке все более производительной в плане вычислительной мощности, а значит, и более тепловыделяющей вычислительной техники. Утилизация выделяемого тепла осуществляется системой охлаждения электронных компонентов, которая реализует перенос тепла от более горячего тела (охлаждаемый объект) к менее горячему (система охлаждения). При постоянном нагреве охлаждаемого объекта температура системы охлаждения повышается и со временем сравняется с температурой охлаждаемого объекта, передача тепла прекратится, что вызовет его перегрев. Система охлаждения защищается от перегрева с помощью холодного вещества — хладагента (теплоносителя), характеристиками эффективности охлаждения которого являются теплоемкость и теплоотдача. Перенос тепла, как правило, осуществляется либо с помощью механизма теплопроводности, требующего физического контакта теплоносителя с охлаждаемым объектом, либо с помощью механизма конвективного теплообмена с хладагентом, который связан с физическим переносом свободно циркулирующего охлаждающего вещества.

Для организации переноса тепла к хладагенту необходимо организовать тепловой контакт системы охлаждения с хладагентом. Для этого применяют различные *радиаторы* — устройства для рассеивания тепла в хладагенте, которые устанавливают на наиболее нагреваемые компоненты вычислительных систем. Для повышения эффективности переноса тепла от электронного компонента к радиатору между ними устанавливается *термоинтерфейс* — слой теплопроводящего состава (обычно многокомпонентного) между охлаждаемой поверхностью и отводящим тепло устройством, применяемый для уменьшения термического сопротивления между двумя соприкасающимися поверхностями. Современные процессоры и ПЛИС нуждаются в охлаждающих устройствах с как можно более низким термическим сопротивлением, так как в настоящий момент даже самые продвинутые ра-

диаторы и термоинтерфейсы не справляются с этой задачей при воздушном охлаждении.

Системы воздушного охлаждения достаточно успешно использовались для охлаждения суперкомпьютеров до 2012 года. С ростом мощности и степени интеграции применяемых кристаллов микропроцессоров и ПЛИС использование воздушных систем охлаждения для разрабатываемых перспективных суперкомпьютеров, в том числе вычислительных систем гибридного типа, практически достигло своего предела. Поэтому большинство разработчиков вычислительной техники видит выход в применении систем жидкостного охлаждения, которые на сегодняшний день представляются наиболее перспективной областью разработки для охлаждения современных высоконагруженных электронных компонентов вычислительных систем.

Существенными преимуществами всех жидкостных схем охлаждения являются лучшая теплоемкость жидкостей по сравнению с воздухом (от 1500 до 4000 раз) и больший коэффициент теплоотдачи (увеличение до 100 раз). Для охлаждения одного современного кристалла ПЛИС необходимо наличие 1 м^3 воздуха или $0,00025 \text{ м}^3$ (250 мл) воды в минуту. На перекачивание 250 мл воды тратится намного меньше электроэнергии, чем для перекачивания 1 м^3 воздуха. Тепловой поток, передаваемый одинаковыми поверхностями при традиционных скоростях теплоносителя, становится при жидкостном охлаждении в 70 раз большим, чем при воздушном охлаждении. Дополнительным преимуществом является использование традиционных, достаточно надежных и недорогих рабочих элементов (насосы, теплообменники, клапаны, механизмы управления и т.д.). Для организаций, которые работают с высокой плотностью оборудования при высоких тепловых режимах, жидкостное охлаждение, по существу, оказывается единственным решением проблемы организации охлаждения современных вычислительных систем. Дополнительным резервом повышения эффективности жидкостного охлаждения является улучшение исходных параметров теплоносителя: увеличение скорости, снижение температуры, создание турбулентного режима течения, повышение теплоемкости, снижение вязкости.

В жидкостных системах охлаждения вычислительной техники роль хладагента выполняет жидкость (вода либо диэлектрические жидкости). Нагретые электронные компоненты отдают тепло постоянно циркулирующему хладагенту — жидкости, которая после ее охлаждения во внешнем теплообменнике вновь поступает для охлаждения нагретых электронных компонентов. Жидкостные системы охлаждения можно разделить на системы закрытого типа, где нет прямого контакта

между жидкостью и электронными компонентами на печатных платах [12, 13], и системы открытого типа (погружные), в которых жидкость непосредственно омывает электронные компоненты. Каждый тип систем жидкостного охлаждения имеет свои достоинства и недостатки.

В системах закрытого типа все тепловыделяющие элементы печатной платы накрываются одним или несколькими плоскими пластинами с каналом для прокачки жидкости [14, 15, 16]. Так, например, в суперкомпьютере «СКИФ-Аврора» использовался принцип «одна охлаждающая пластина на одну печатную плату». Соответственно пластина имела сложный рельеф, чтобы иметь плотный тепловой контакт с каждой микросхемой. В суперкомпьютере IBM Aquasar используется принцип «одна охлаждающая пластина на одну (горячую) микросхему». В любом случае каналы пластин объединяются коллекторами в единый контур, соединенный с общим радиатором (или другим теплообменником), обычно расположенным за пределами корпуса и/или шкафа или даже машинного зала. С помощью насоса через пластины прокачивается теплоноситель, который отводит тепло от элементов вычислителя и рассеивает его через теплообменник. В такой системе требуется обеспечить доступ теплоносителя к каждому тепловыделяющему элементу вычислителя, что означает довольно сложную «трубопроводную систему», большое количество герметичных соединений. При этом требования к возможности обслуживания печатных плат без серьезного демонтажа в системе охлаждения неизбежно означают применение специальных жидкостных разъемов, которые обеспечивают герметичность соединений с одновременной легкостью операций по сборке/разборке системы.

В системах жидкостного охлаждения закрытого типа можно в качестве хладагента использовать обычную воду или растворы гликолей. Однако утечка теплоносителя может привести к возможному попаданию электропроводной жидкости на незащищенные контакты рабочих плат охлаждаемого компьютера, что, в свою очередь, может быть фатальным как для отдельных электронных компонентов, так и для вычислительной системы в целом. Для устранения неисправности требуются остановка всего комплекса, проверка системы электропитания и ее просушка. В системах управления и мониторинга таких компьютеров всегда предусматривают многочисленные внутренние датчики влажности и протечек. Для решения проблемы протечек часто используется метод, основанный на использовании в системе охлаждения отрицательного давления воды, когда вода не накачивается под давлением, а высасывается, что практически исключает утечку влаги. При нарушении герметичности в систему охлаждения попадет

воздух, а вода оттуда не вытечет. Для детектирования утечек используются специальные датчики, а модульная конструкция позволяет осуществлять ремонт без прерывания работы системы в целом. Однако все эти мероприятия значительно усложняют конструкцию гидравлической системы.

Еще одна проблема систем жидкостного охлаждения закрытого типа — проблема «точки росы». Воздух центра обработки данных (ЦОД) входит в контакт с пластинами охлаждения. Это значит, что если какие-то участки этих пластин слишком холодные, а воздух в ЦОД более теплый и не слишком сухой, то на пластинах может сконденсироваться влага. Последствия этого процесса аналогичны последствиям протечек. Проблема решается либо переходом на охлаждение горячей водой, что неэффективно, либо контролем и поддержанием параметров температуры и влажности воздуха в ЦОД, что сложно и дорого.

Конструкция еще более усложняется, когда необходимо охладить несколько компонентов потоком воды, пропорциональным их тепловыделению. Не считая разветвленных трубок, приходится применять сложные регулирующие приборы (простыми тройниками и крестовинами не обойдешься). Альтернативный вариант — использовать конструкцию с заводскими регулировками потоков; но в этом случае пользователь лишен возможности существенно изменить конфигурацию охлаждаемых вычислительных модулей.

К достоинствам систем жидкостного охлаждения закрытого типа относятся:

- возможность использования в качестве хладагента воды или водных растворов, преимуществами которых являются доступность, прекрасные теплотехнические свойства (теплопроводность, теплоемкость, вязкость), простота и сравнительная безопасность эксплуатации;
- наличие большого числа унифицированных механизмов, узлов и деталей для систем водоснабжения, которые можно использовать;
- большой опыт эксплуатации систем с водяным охлаждением в промышленности.

При этом системы жидкостного охлаждения закрытого типа обладают целым рядом существенных недостатков, сдерживающих их массовое применение:

- сложностью определения места утечки воды;
- зачастую катастрофическими последствиями не обнаруженных своевременно протечек;
- технологическими сложностями ликвидации протечек (необходимость отключения всего вычислительного шкафа, что не всегда возможно и удобно);

- необходимостью поддержания микроклимата помещения вычислительного зала (проблема «точки росы»);
- проблемой охлаждения остальных элементов на плате вычислительного модуля PBC. Даже при небольшом изменении конфигурации PBC необходимо изготавливать новый теплообменник;
- проблемой электрохимической коррозии при использовании алюминиевых теплообменников либо массогабаритными ограничениями при использовании более устойчивых медных теплообменников (алюминий в три раза легче меди);
- необходимостью удаления воздуха из системы охлаждения как при пуско-наладке, так и в процессе эксплуатации;
- сложностью компоновки вычислительных модулей в шкафу при большом количестве фитингов, необходимых для подключения каждого вычислительного модуля;
- необходимостью использования специализированного вычислительного шкафа со значительными массогабаритными характеристиками.

В системах жидкостного охлаждения открытого типа основным компонентом является хладагент, представляющий собой диэлектрическую жидкость, как правило, на основе белого минерального масла, благодаря которому теплоаккумулирующая способность хладагента гораздо выше, чем у воздуха при том же объеме. Конструктивно такие системы представляют собой заполненную хладагентом ванну (в том числе устанавливаемую в вычислительном шкафу), в которой размещаются печатные платы и серверы вычислительного оборудования. Выделяемое электронными компонентами тепло рассеивается циркулирующим в пределах всего объема ванны хладагентом. Достоинствами погружной системы охлаждения являются простота конструкции и адаптация под меняющуюся геометрию печатных плат, отсутствие сложной системы коллекторов и жидкостных разъемов, отсутствие проблем управления потоками жидкости и проблемы «точки росы», повышение надежности и снижение стоимости изделия.

Основная проблема систем жидкостного охлаждения открытого типа заключается в химическом составе используемого хладагента, который должен удовлетворять жестким требованиям по теплопроводности, электропроводности, вязкости, токсичности, пожаробезопасности, стабильности основных параметров, и при этом стоимость жидкости должна быть разумной.

Системы жидкостного охлаждения открытого типа обладают следующими преимуществами:

- нечувствительностью к протечкам и их последствиям, возможностью эксплуатации системы даже при наличии местных протечек хладагента;
- нечувствительностью к климатическим характеристикам помещения вычислительного зала;
- решением проблемы охлаждения остальных элементов РВС, так как плата вычислительного модуля погружена в хладагент;
- возможностью изменения конфигурации платы вычислительного модуля без изменения системы охлаждения;
- простотой гидравлической балансировки системы благодаря отсутствию сложной системы коллекторов;
- возможностью использования как унифицированных механизмов, узлов и деталей для гидравлических систем общего машиностроения, так и накопленного опыта эксплуатации электрооборудования с использованием трансформаторных масел;
- повышением общей надежности системы жидкостного охлаждения.

Недостатками систем жидкостного охлаждения открытого типа являются:

- необходимость дополнительного насосного и теплообменного оборудования для улучшения теплотехнических свойств (теплопроводность, теплоемкость, вязкость) хладагента, в роли которого используются специальные диэлектрические органические жидкости;
- необходимость обучения обслуживающего персонала и соблюдение повышенных мер безопасности при работе с хладагентом;
- необходимость более частой очистки помещения машинного зала из-за высокой проникающей способности хладагента, особенно в случае протечки;
- необходимость специальной оснастки для регламентных и аварийных эксплуатационных операций (монтаж/демонтаж вычислительного модуля, залив/слив хладагента и т.д.);
- повышение стоимости эксплуатации из-за необходимости регулярной замены хладагента по истечении срока службы и необходимость организации учета (транспортирование, прием, учет, хранение, выдача, утилизация хладагента и т.д.) хладагента в организации.

Оценивая приведенные достоинства и недостатки двух систем жидкостного охлаждения, можно с достаточной степенью уверенности отметить более весомые преимущества систем жидкостного охлаждения электронных компонентов вычислительных систем открытого типа. Поэтому для вычислительных модулей РВС, проектируемых на основе перспективных семейств ПЛИС, целесообразно использовать жидкостное

охлаждение, в частности непосредственное погружение плат вычислительных модулей в жидкостный хладагент на основе минерального масла.

Особенностями выпускаемых в Научно-исследовательском центре супер-ЭВМ и нейрокомпьютеров РВС являются наличие не менее 6-8 кристаллов ПЛИС на одной печатной плате и высокая плотность компоновки платы, что существенно увеличивает число тепловыделяющих элементов по сравнению с микропроцессорными модулями. Это затрудняет применение существующих готовых решений погружных систем [17-19] и требует дополнительных технических и конструктивных решений для эффективного непосредственного жидкостного охлаждения вычислительных модулей РВС.

5. Реконфигурируемая вычислительная система на основе ПЛИС Xilinx UltraScale. Поэтому с 2013 года в НИЦ СЭ и НК активно развивается направление по созданию РВС нового поколения на основе собственной системы жидкостного охлаждения для печатных плат с высокой степенью компоновки и большим количеством тепловыделяющих электронных компонентов. В основу критериев проектирования вычислительного модуля (ВМ) нового поколения для РВС с жидкостным охлаждением открытого типа были положены следующие принципы:

- основой конфигурации вычислительного шкафа является вычислительный модуль с размерами 3U на 19" и автономной циркуляцией охлаждающей жидкости;

- в одном вычислительном шкафу стандартной высоты 47U должно размещаться не менее 12 вычислительных модулей с жидкостным охлаждением;

- в одном вычислительном модуле должно размещаться 12-16 плат с установленными на них кристаллами ПЛИС;

- на каждой печатной плате должно быть размещено до восьми ПЛИС с выделяемым тепловым потоком порядка 100 Вт от каждой ПЛИС;

- для охлаждения жидкости должна использоваться традиционная система водяного охлаждения на базе промышленных чиллеров.

Основным элементом модульной реализации погружной системы жидкостного охлаждения электронных компонентов вычислительных систем открытого типа является реконфигурируемый вычислительный модуль нового поколения, технологический образец которого представлен на рисунке 9. ВМ нового поколения состоит из вычислительной секции, теплообменной секции, корпуса, насоса, теплообменника и штуцера. В корпусе, образующем основу вычислительной секции, размещены герметичный контейнер с диэлектрической охлаждающей жидкостью, содержащий электронные устройства с компонен-

тами, которые выделяют тепло в процессе работы. В качестве электронных устройств могут быть вычислительные модули (не менее 12-16 шт.), платы управления, ОЗУ, блоки питания, накопители, дочерние платы и так далее. Сверху вычислительная секция закрыта крышкой.

К вычислительной секции примыкает теплообменная секция, содержащая насос и теплообменник. Насос обеспечивает циркуляцию хладагента в ВМ по следующему замкнутому контуру: из вычислительного модуля нагретый хладагент поступает в теплообменник, охлаждается там, из теплообменника охлажденный хладагент вновь поступает в вычислительный модуль, охлаждает там нагретые электронные компоненты, там же нагревается и затем вновь поступает в теплообменник и так далее. Теплообменник подключен к внешнему теплообменному контуру через штуцеры и предназначен для охлаждения хладагента с помощью вторичной охлаждающей жидкости. В качестве теплообменника можно использовать пластинчатый теплообменник, в котором первичный и вторичный контуры разделены между собой. Тогда в качестве вторичной охлаждающей жидкости можно использовать обыкновенную воду, охлаждаемую промышленным чиллером. При этом чиллер может располагаться вне здания серверной и соединяться с реконфигурируемыми вычислительными модулями через стационарную систему инженерных коммуникаций.

Вычислительная и теплообменная секции механически соединены между собой, образуя единый реконфигурируемый вычислительный модуль. Для эксплуатации реконфигурируемого вычислительного модуля требуется подключение к источнику вторичной охлаждающей жидкости (через краны), к источнику энергоснабжения и к сетевому концентратору (через электрические разъемы).

В корпусе вычислительного шкафа ВМ размещены друг над другом. Их количество ограничивается линейными размерами шкафа, техническими возможностями помещения вычислительного зала и подводимых инженерных коммуникаций. Каждый ВМ вычислительного шкафа подключается к источнику вторичной охлаждающей жидкости с помощью подающих обратных коллекторов через штуцеры (или балансировочные клапаны) и гибкие трубопроводы; подключение к источнику энергоснабжения и сетевому концентратору осуществляется через электрические разъемы. Через штуцеры (или балансировочные клапаны) подают холодную и отводят нагретую вторичную охлаждающую жидкость в стационарную систему инженерных коммуникаций, подключенную к шкафу. Множество вычислительных шкафов, размещенных в одном или нескольких помещениях, образует вычислительный комплекс. Для эксплуатации вычислительный ком-

плекс также подключается к источнику вторичной охлаждающей жидкости, к источнику энергоснабжения, модулю управления вычислительным комплексом.

Рассмотренная модульная реализация системы жидкостного охлаждения электронных компонентов вычислительных систем открытого типа, помимо свойственных системам открытого типа достоинств, дополнительно обладает рядом преимуществ:

- платы вычислительных модулей и реконфигурируемые вычислительные модули идентичны, относительно автономны и взаимозаменяемы. При выходе из строя одного ВМ или при необходимости проведения профилактических работ нет необходимости полностью отключать вычислительный шкаф и останавливать рабочую задачу;
- достигается повышенная плотность компоновки ПЛИС в ВМ;
- предложенное техническое решение позволяет при необходимости наращивать производительность реконфигурируемых вычислительных модулей без кардинального увеличения габаритных размеров (более производительные насос и теплообменник позволяют вписаться в выбранные габариты). Увеличение количества плат вычислительных модулей незначительно увеличит размер реконфигурируемого вычислительного модуля по глубине при неизменной плотности компоновки.

В силу простоты конструкции теплообменной секции реконфигурируемого вычислительного модуля значительно повышается его надежность.

Вычислительный 19” шкаф суперкомпьютера имеет следующие технические характеристики:

- стандартный вычислительный шкаф высотой 47U;
- 12 вычислительных модулей высотой 3U с жидкостным охлаждением;
- каждый вычислительный модуль содержит 12 плат мощностью 800 Вт каждая;
- каждая плата содержит восемь кристаллов ПЛИС Kinetix UltraScale - XCKU095-1FFVB2104C по 95 млн. эквивалентных вентилей (134 400 логических блоков) каждый;
- производительность ВМ нового поколения — 105 Тфлопс;
- производительность вычислительного шкафа с 12 ВМ — 1 Пфлопс;
- потребляемая мощность вычислительного шкафа с 12 ВМ — 124 кВт.

Производительность одного вычислительного шкафа с системой жидкостного охлаждения, содержащего 12 ВМ с 12 платами каждый, в 6,55 раз превышает производительность аналогичного шкафа с ВМ

«Тайгета». При этом производительность одного ВМ нового поколения по сравнению с ВМ «Тайгета» повышена в 8,74 раза. Такое качественное повышение удельной производительности системы обеспечено более чем трехкратным повышением плотности компоновки системы за счет оригинальных конструктивных решений, а также повышением тактовой частоты и увеличением числа вентиляторов в кристалле.

Для проверки технических, технологических решений и определения ожидаемых технико-экономических и эксплуатационных характеристик разрабатываемой высокопроизводительной реконфигурируемой вычислительной системы с жидкостным охлаждением был создан ряд макетов, опытных и технологических образцов. На рисунке 9 представлен технологический образец ВМ нового поколения, для которого были разработаны новые конструкции печатных плат и вычислительных модулей, характеризующиеся высокой плотностью компоновки.



Рис. 9. Технологический образец ВМ нового поколения

Для создания эффективной погружной системы охлаждения был разработан диэлектрический хладагент, обладающий наилучшей электрической прочностью, высокой теплопроводностью, максимально возможной теплоемкостью при низкой вязкости. На основе трансформаторного масла ГК методом вакуумной перегонки было получено масло с пониженной вязкостью «Масло маловязкое диэлектрик МД-4,5 для охлаждения электронных компонентов ЭВМ», на него разработаны технические условия ТУ 38.401-58-421-2015 и получены рекомендации для его применения. Масло МД-4,5 прошло всестороннее исследование в теплотехнической лаборатории НИЦ СЭ и НК на технологическом образце вычислительного модуля с погружной системой жидкостного охлаждения открытого типа. Цикл проведенных лабораторных и эксплуатационных испытаний подтвердил целесообразность использования масла МД-4,5 для охлаждения электронных компонен-

тов ЭВМ и применение маломощных насосов для его циркуляции (вследствие пониженной вязкости).

При разработке ВМ нового поколения получен ряд прорывных технических решений: разработаны погружной блок питания для напряжения 380В и преобразователь DC/DC 380/12 В, обеспечена минимальная высота платы ВМ 100 мм, разработана погружная плата управления собственного производства. Для подсистемы охлаждения ВМ нового поколения определены необходимые элементы системы охлаждения: выбран оригинальный термоинтерфейс, разработана собственная конструкция радиатора ПЛИС малой высоты для конвективного теплообмена, выбраны оптимальный для используемого хладагента насос и теплообменник. Также определены конструкция компенсатора объемного расширения хладагента и элементы контроля подсистемы охлаждения: оптические датчики уровня и датчик протока. Разработанные решения по конструкции системы охлаждения и циркуляции хладагента позволяют эффективно решать проблему теплоотвода с наиболее нагретых компонентов ВМ.

Комплекс разработанных решений погружной системы жидкостного охлаждения позволил обеспечить в рабочем режиме ВМ температуру хладагента не более 33 °С при потребляемой мощности 91 Вт для каждой ПЛИС (8736 Вт для ВМ). При этом максимальная температура ПЛИС не превышает 55°С, что свидетельствует о резерве мощности разработанной погружной системы жидкостного охлаждения, позволяющей обеспечить эффективное охлаждение для проектируемых перспективных семейств ПЛИС фирмы Xilinx (UltraScale+, UltraScale 2 и т.д.).

6. Заключение. Использование воздушных систем охлаждения для разрабатываемых суперкомпьютеров практически достигло своего предела, что обусловлено снижением эффективности охлаждения с ростом потребляемой (и отдаваемой) мощности, вызванным увеличением степени интеграции кристаллов микропроцессоров и других микросхем. Поэтому использование жидкостного охлаждения для организации работы современных вычислительных систем является приоритетным направлением совершенствования систем охлаждения с широкими перспективами дальнейшего развития. Жидкостное охлаждение вычислительных модулей PBC, содержащих не менее восьми кристаллов ПЛИС с высокой плотностью компоновки, обладает спецификой по сравнению с охлаждением кристаллов микропроцессоров, что требует разработки специализированной погружной системы охлаждения. Разработанная оригинальная жидкостная система охлаждения для вычислительного модуля PBC нового поколения позволяет обеспечить высокие эксплуатационные показатели системы — макси-

мальную температуру ПЛИС, не превышающую 55 °С при температуре хладагента не более 33 °С в рабочем режиме. Полученные прорывные решения погружной системы жидкостного охлаждения позволяют в пределах одного вычислительного шкафа высотой 47U разместить не менее 12 ВМ нового поколения с суммарной производительностью свыше 1 Пфлопс. Резерв мощности системы жидкостного охлаждения ВМ нового поколения позволяет обеспечить эффективное охлаждение не только для существующих, но и для проектируемых перспективных семейств ПЛИС фирмы Xilinx UltraScale+ и UltraScale 2.

Литература

1. *Kalyaev I.A., Levin I.I., Semernikov E.A., Shmoilov V.I.* Reconfigurable multipipeline computing structures // New York: Nova Science Publishers. USA. 2012. 330 p.
2. *Carter W.S.* The evolution of programmable logic // VLSI Circuits. Digest of Technical Papers. 1991 Symposium on IEEE. 1991. pp. 43–46.
3. *Wilson R.* In the Beginning. URL: https://www.altera.com/solutions/technology/system-design/articles/_2013/in-the-beginning.html (дата обращения: 04.04.2016).
4. *Chartrand L.* Advanced Digital Systems: Experiments and Concepts with CPLDs // Cengage Learning. 2005. 268 p.
5. *Power Methodology Guide.* UG786 (v14.5). 2013. 54 p. URL: http://www.xilinx.com/support/documentation/sw_manuals/xilinx13_1/ug786_PowerMethodology.pdf (дата обращения 04.04.2016).
6. *User guide Microsemi.* 158 p. URL: http://www.actel.ru/files/uploads/Novosti/Download/Libero_UG_rus.pdf (дата обращения 04.04.2016).
7. *Barkalov A., Titarenko L.* Evolution of programmable logic // Logic Synthesis for FSM-Based Control Units. Springer Berlin Heidelberg. 2009. LNEE 53. pp. 53–75.
8. *Hutton M.* Architecture and CAD for FPGAs // Proceedings of the 17th Symposium on Integrated Cuits and Systems Design (SBCCI2004). 2004. pp. 3.
9. *Kuon I., Tessier R., Rose J.* FPGA Architecture: Survey and Challenges // Foundations and Trends in Electronic Design Automation. 2008. vol. 2. no. 2. pp. 135–253.
10. *Levin I.I., Dordopulo A.I., Doronchenko Y.I., Raskladkin M.K.* Reconfigurable computer system on the base of Virtex UltraScale FPGAs with liquid cooling // Proceedings of international scientific conference “Parallel computer technologies” (PaCT’2016). 2016. pp. 221–230.
11. *Kalyaev I.A., Levin I.I., Dordopulo A.I., Slasten L.M.* Reconfigurable Computer Systems Based on Virtex-6 and Virtex-7 FPGAs // IFAC Proceedings Volumes. 2013. vol. 12. no. 1. pp. 210–214.
12. *Li L., Zheng W., Wang X., Wang X.* Data center power minimization with placement optimization of liquid-cooled servers and free air cooling // Sustainable Computing: Informatics and Systems. 2016. vol. 11. no. 1. pp. 3–15.
13. *Sun L. et al.* Corrosion investigation of the inlet section of REAC pipes in the refinery // Engineering Failure Analysis. 2016. vol. 66. no. 1. pp. 468–478.
14. *Абрамов С.М. и др.* Суперкомпьютеры ряда 4 семейства “СКИФ”: Штурм вершины суперкомпьютерных технологий // Вестник Нижегородского университета им. Н.Н. Лобачевского. 2009. № 5. С. 200–210.
15. *Левшин И., Орлов Д., Суомин А., Челищев А.* Не революция, но переворот // М.: Суперкомпьютеры. 2012. № 2(10). С. 6–21. URL: <http://admin.clus-ter.sfu>

kras.ru/data/uploads/2015/02/12/ Supercomputers_10-2012.pdf. (дата обращения: 04.04.2016).

16. Новое суперкомпьютерное решение «Т-Платформы». А-Класс. Пресс-релиз от 25.06.2014. URL: https://www.pcweek.ru/infrastructure/news-company/detail_print.php?ID=164557&print=Y. (дата обращения 04.04.2016).
17. *Shah J.M., Eiland R., Siddarth, A. Agonafer D.* Effects of mineral oil immersion cooling on IT equipment reliability and reliability enhancements to data center operations // Proceedings of the 15th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm 2016). 2016. pp. 316–325.
18. *Gess J.L., Bhavnani S.H., Johnson R.W.* Experimental Investigation of a Direct Liquid Immersion Cooled Prototype for High Performance Electronic Systems // IEEE Transactions on Components, Packaging and Manufacturing Technology. 2015. vol. 5. no. 10. pp. 1451–1464.
19. *Gess J., Dreher T., Bhavnani S., Johnson W.* Effect of flow guide integration on the thermal performance of high performance liquid cooled immersion server modules // ASME 2015 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems collocated with the ASME 2015 13th International Conference on Nanochannels, Microchannels, and Minichannels. American Society of Mechanical Engineers. 2015. pp. V001T09A047–V001T09A047.

Каляев Игорь Анатольевич — д-р техн. наук, профессор, академик РАН, главный научный сотрудник научно-исследовательского института многопроцессорных вычислительных систем имени академика А.В. Каляева, Южный федеральный университет (НИИ МВС ЮФУ). Область научных интересов: многопроцессорные вычислительные и информационно-управляющие системы. Число научных публикаций — 330. kaliaev@mvs.sfedu.ru; ул. Чехова, 2, ГСП-28, Таганрог, Ростовская область, 347928; р.т.: +7(8634)360-757, Факс: +7(8634)360-376.

Дордопуло Алексей Игоревич — к-т техн. наук, начальник отдела математического и алгоритмического обеспечения, ООО «Научно-исследовательский центр супер-ЭВМ и нейрокомпьютеров» (ООО «НИЦ СЭ и НК»). Область научных интересов: языки программирования высокого уровня, программное обеспечение многопроцессорных вычислительных систем. Число научных публикаций — 153. scorpio@mvs.tsure.ru, <http://www.superevm.ru>; пер. Итальянский, 106, Таганрог, 347900; р.т.: +7(8634)612-111, Факс: +7(8634)477-415.

Левин Илья Израилевич — д-р техн. наук, профессор, директор, ООО «Научно-исследовательский центр супер-ЭВМ и нейрокомпьютеров» (ООО «НИЦ СЭ и НК»). Область научных интересов: высокопроизводительные многопроцессорные вычислительные системы, языки программирования высокого уровня, программное обеспечение. Число научных публикаций — 313. levin@superevm.ru, <http://www.superevm.ru>; пер. Итальянский, 106, Таганрог, 347900; р.т.: +7(8634)612-111, Факс: +7(8634)477-415.

Федоров Александр Михайлович — главный теплотехник, ООО «Научно-исследовательский центр супер-ЭВМ и нейрокомпьютеров» (ООО «НИЦ СЭ и НК»). Область научных интересов: системы охлаждения высокопроизводительных многопроцессорных вычислительных систем. Число научных публикаций — 3. ss24@mail.ru; пер. Итальянский, 106, Таганрог, 347900; р.т.: +7(8634)612-111, Факс: +7(8634)477-415.

Поддержка исследований. Работа выполнена при частичной финансовой поддержке Министерства образования и науки РФ по Соглашению о предоставлении субсидии № 14.5178.21.0006 от 05.06.2014, уникальный идентификатор RFMEFI57814X0006.

I.A. KALYAEV, A.I. DORDOPULO, I.I. LEVIN, A.M. FEDOROV
**EVOLUTION OF DOMESTIC MULTICHIP RECONFIGURABLE
 COMPUTER SYSTEMS: FROM AIR TO LIQUID COOLING**

Kalyaev I.A., Dordopulo A.I., Levin I.I., Fedorov A.M. Evolution Domestic of Multichip Reconfigurable Computer Systems: from Air to Liquid Cooling.

Abstract. In the paper we consider the evolution of design technologies of reconfigurable computer systems based on FPGAs of various families. Five FPGA-based generations of reconfigurable computer systems with high placement density, from Xilinx Virtex-E to modern Virtex UltraScale, are described. We show results of design of high real performance energy-efficient reconfigurable computer systems. The main contribution is a liquid cooling system designed for Virtex UltraScale FPGAs. It provides independent circulation of the cooling liquid in the 19" 3U computational module for cooling of 96-128 FPGA chips that generate 9.6-12.8 kWatt of heat in total. The distinctive features of the designed immersion liquid cooling system are high cooling efficiency with power reserve for the designed perspective FPGA families, resistance to leaks and their consequences, and compatibility with traditional water cooling systems based on industrial chillers.

Keywords: reconfigurable computer systems, FPGA, immersion liquid cooling system, industrial chillers.

Kalyaev Igor Anatolievich — Ph.D., Dr. Sci., professor, academician of RAS, chief researcher of Scientific Research Institute of Multiprocessor Computer Systems, Southern Federal University (SFedU). Research interests: multiprocessor computer and data control systems. The number of publications — 330. kaliaev@mvs.sfedu.ru; 2, Chehov st., GSP-284, Taganrog, 347928, Russia; office phone: +7(8634)360-757, Fax: +7(8634)360-376.

Dordopulo Alexey Igorevich — Ph.D., head of the mathematical and algorithmic support department, "Scientific Research Centre of Supercomputers and Neurocomputers" Co Ltd ("SRC SC & NC" Co Ltd). Research interests: high level programming languages, high-performance multiprocessor computer systems software. The number of publications — 153. scorpio@mvs.tsure.ru, <http://www.superevm.ru>; 106, Italyansky lane, Taganrog, 347900, Russia; office phone: +7(8634)612-111, Fax: +7(8634)477-415.

Levin Ilya Izrailevich — Ph.D., Dr. Sci., professor, director, "Scientific Research Centre of Supercomputers and Neurocomputers" Co Ltd ("SRC SC & NC" Co Ltd). Research interests: high-performance multiprocessor computer systems, high level programming languages, software. The number of publications — 313. levin@superevm.ru, <http://www.superevm.ru>; 106, Italyansky lane, Taganrog, 347900, Russia; office phone: +7(8634)612-111, Fax: +7(8634)477-415.

Fedorov Alexander Michailovich — chief heat engineer, "Scientific Research Centre of Supercomputers and Neurocomputers" Co Ltd ("SRC SC & NC" Co Ltd). Research interests: cooling systems for multiprocessor computer systems. The number of publications — 3. ss24@mail.ru; 106, Italyansky lane, Taganrog, 347900, Russia; office phone: +7(8634)612-111, Fax: +7(8634)477-415.

Acknowledgements. This paper was financially supported in part by the Ministry of Education and Science of the Russian Federation under Grant № 14.578.21.0006 from 05.06.2014, ID RFMEFI57814X0006.

References

1. Kalyaev I.A., Levin I.I., Semernikov E.A., Shmoilov V.I. Reconfigurable multipipeline computing structures. New York: Nova Science Publishers. USA. 2012. 330 p.

2. Carter W.S. The evolution of programmable logic. VLSI Circuits. Digest of Technical Papers. 1991 Symposium on IEEE. 1991. pp. 43–46.
3. Wilson R. In the Beginning. Available at: https://www.altera.com/solutions/technology/system-design/articles/_2013/in-the-beginning.html. Available at: https://www.altera.com/solutions/technology/system-design/articles/_2013/in-the-beginning.html (accessed 04.04.2016).
4. Chartrand L. Advanced Digital Systems: Experiments and Concepts with CPLDs. Cengage Learning. 2005. 268 p.
5. Power Methodology Guide. UG786 (v14.5). 2013. 54 p. Available at: http://www.xilinx.com/support/documentation/sw_manuals/xilinx13_1/ug786_PowerMethodology.pdf (accessed 04.04.2016).
6. User guide Microsemi. 158 p. Available at: http://www.actel.ru/files/uploads/Novosti/Download/Libero_UG_rus.pdf (accessed 04.04.2016).
7. Barkalov A., Titarenko L. Evolution of programmable logic. Logic Synthesis for FSM-Based Control Units. Springer Berlin Heidelberg. 2009. LNEE 53. pp. 53–75.
8. Hutton M. Architecture and CAD for FPGAs. Proceedings of the 17th Symposium on Integrated Circuits and Systems Design (SBCCI2004). 2004. pp. 3.
9. Kuon I., Tessier R., Rose J. FPGA Architecture: Survey and Challenges. Foundations and Trends in Electronic Design Automation. 2008. vol. 2. no. 2. pp. 135–253.
10. Levin I.I., Dordopulo A.I., Doronchenko Y.I., Raskladkin M.K. Reconfigurable computer system on the base of Virtex UltraScale FPGAs with liquid cooling. Proceedings of international scientific conference “Parallel computer technologies” (PaCT’2016). 2016. pp. 221–230.
11. Kalyaev I.A., Levin I.I., Dordopulo A.I., Slasten L.M. Reconfigurable Computer Systems Based on Virtex-6 and Virtex-7 FPGAs. IFAC Proceedings Volumes. Programmable Devices and Embedded Systems. 2013. vol. 12. no. 1. pp. 210–214.
12. Li L., Zheng W., Wang X., Wang X. Data center power minimization with placement optimization of liquid-cooled servers and free air cooling. *Sustainable Computing: Informatics and Systems*. 2016. vol. 11. no. 1. pp. 3–15.
13. Sun L. et al. Corrosion investigation of the inlet section of REAC pipes in the refinery. *Engineering Failure Analysis*. 2016. vol. 66. no. 1. pp. 468–478.
14. Abramov S.M. et al. [Supercomputers series 4 family SKIF: summit supercomputer technologies]. *Vestnik Nizhegorodskogo universiteta im. N. N. Lobachevsky – Vestnik of Lobachevsky University of Nizhni Novgorod*. 2009. vol. 5. pp. 200–210. (In Russ.).
15. Lyovshin I., Orlov D., Syomin A., Chelyshev A.J. [Not a revolution, but a coup]. *Superkomp'yutery – Supercomputers*. 2012. vol. 2(10). pp. 6–21. Available at: http://admin.clus-ter.sfu-kras.ru/data/uploads/2015/02/12/Supercom-puters_10-2012.pdf. (accessed 04.04.2016). (In Russ.).
16. Novoe superkomp'yuternoe reshenie «T-Platformy» A-Class [A new supercomputing solution "T-Platforms" A-Class]. Press-reliz ot 25.06.2014. https://www.pcweek.ru/infrastructure/news-company/detail_print.php?ID=164557&print=Y. (accessed 04.04.2016). (In Russ.).
17. Shah J.M., Eiland R., Siddarth, A., Agonafer D. Effects of mineral oil immersion cooling on IT equipment reliability and reliability enhancements to data center operations. Proceedings of the 15th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm 2016). 2016. pp. 316–325.
18. Gess J.L., Bhavnani S.H., Johnson R.W. Experimental Investigation of a Direct Liquid Immersion Cooled Prototype for High Performance Electronic Systems IEEE Transactions on Components, Packaging and Manufacturing Technology. 2015. vol. 5. no. 10. pp. 1451–1464.
19. Gess J., Dreher T., Bhavnani S., Johnson W. Effect of flow guide integration on the thermal performance of high performance liquid cooled immersion server modules. ASME 2015 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems collocated with the ASME 2015 13th International Conference on Nanochannels, Microchannels, and Minichannels. American Society of Mechanical Engineers. 2015. pp. V001T09A047–V001T09A047.

Д.В. ЖЕЛЕЗНОВ, Е.М. ТАРАСОВ, А.Г. ИСАЙЧЕВА, Т.И. МИХЕЕВА
**РАЗРАБОТКА ОБУЧАЕМОГО КЛАССИФИКАТОРА
СОСТОЯНИЙ РЕЛЬСОВЫХ ЛИНИЙ С МНОГОМЕРНЫМИ
ИНФОРМАТИВНЫМИ ПРИЗНАКАМИ**

Железнов Д.В., Тарасов Е.М., Исайчева А.Г., Михеева Т.И. Разработка обучаемого классификатора состояний рельсовых линий с многомерными информативными признаками.

Аннотация. Значительное количество отказов в системах интервального управления движением поездов связано с воздействием возмущающих факторов в широком диапазоне изменения на единственный информационный признак, характеризующий состояние рельсовых линий. В работе предложено определять состояния объекта контроля принципами распознавания образов с многомерными информативными признаками. В качестве признаков предложено использовать амплитуды напряжений и токов на входе и выходе рельсового четырехполосника. В качестве полинома решающей функции – ортогональный многочлен Эрмита, позволяющий путем усложнения порядка и размерности увеличить глубину распознавания и обеспечить относительную инвариантность к возмущающим воздействиям. При решении задачи распознавания состояний рельсовых линий в качестве критерия качества использована относительная погрешность вычисления границ классов решающими функциями.

Работоспособность предложенной методики демонстрируется результатами исследования распознавания состояний рельсовых линий с «обученной» решающей функцией.

Ключевые слова: классификатор состояний, решающая функция, множество признаков.

1. Введение. В алгоритме функционирования технологического процесса на транспорте используется информация первичных датчиков — рельсовых цепей, в которых чувствительным элементом датчика служат рельсовые линии (РЛ). Поэтому надежность работы первичных датчиков определяет надежность функционирования всей системы автоматики и телемеханики на железнодорожном транспорте [1].

В настоящее время устройства контроля состояний рельсовых линий проектируются по одноканальной схеме с использованием при классификации одного признака — амплитуды напряжения на выходе рельсовой линии. Значительное увеличение интенсивности движения поездов в последние годы привели к существенному повышению возмущающих воздействий на рельсовые линии, и оценка их состояний по одному признаку приводит к фиксации состояния «ложной занятости» при фактически свободном или «ложной свободности» при фактически занятом поездом участке контроля [2]. Наиболее эффективным методом решения проблемы является способ оценки состояния рельсовых линий, реализованный принципами

распознавания образов с множеством информативных признаков и решающих функций.

Распознавание устанавливает зависимость соответствия i -го образа, задаваемого в виде совокупности признаков к конкретному классу из множества взаимоисключающих классов. Это означает, что существует однозначное отображение множества признаков, являющихся числовым множеством $\{X\}$ на множестве классов $\{S\}=\{s_1, s_2, \dots, s_k\}$, размерность которого известна заранее. В случае распознавания состояний рельсовых линий классы можно заменить номерами $1, 2, \dots, k$, и представлять последние в виде натуральных чисел, что описывает распознавание как отображение признакового пространства на конечное множество натуральных чисел [3, 4].

В распознавании состояний рельсовых линий множество признаков может быть представлено в виде измеренных p -значений характеристик электрических параметров рельсовых линий x_1, x_2, \dots, x_p . При этом размерность признаков равна определенному числу n . Тогда распознавание состояний РЛ сводится к pn -аргументной функции вида $k=d(x_{11}, x_{12}, \dots, x_{ij}, x_{pn})$, где x_{ij} — j -ое измеренное значение i -го признака, $i=1, 2, \dots, p$, $j=1, 2, \dots, n$, причем данная функция, как правило, называется решающей функцией и однозначна при конкретных p и n .

Важнейшей особенностью распознавания состояний РЛ является то, что признаки $\{x_{ij}\}_{p \times n}$ подвержены возмущениям, вероятностный характер которых отражается на всех этапах: от процесса получения самих измерений и заканчивая вычислением значений решающей функции $d(x_{11}, x_{12}, \dots, x_{ij}, x_{pn})$ [2]. Дестабилизирующие факторы представляются в распознавании состояний РЛ как погрешности аналогово-цифровых преобразователей, первичных датчиков съема информации; как ошибки округления при вычислениях, что является следствием малой размерности разрядной сетки микропроцессоров, установленных у объектов измерений. Взаимодействие возмущений между собой приводит к тому, что признаки x_{ij} также имеют погрешности с различными знаками, и в результате обучения решающей функции разработчикам, как правило, приходится учитывать границы изменений этих возмущений и при синтезе решающих функций использовать признаки с различными знаками возмущений, что помогает добиться относительной нечувствительности на начальном признаковом уровне обработки информации [5].

Использование методов распознавания образов перед другими методами, используемыми в настоящее время для определения

состояний рельсовых линий, позволяет расширить эксплуатационные возможности классификатора состояний рельсовых линий за счет специфичной обработки измеренной информации и априорного формирования кластеров состояний.

2. Основные задачи, возникающие при создании обучаемого классификатора состояний рельсовых линий. В общем случае состояния рельсовых линий могут быть кластеризованы на классы: свободного и исправного состояний рельсовой линии; занятого состояния; неисправного (излома рельсовой линии) состояний, а также диагностический класс, который, в свою очередь, может подразделяться на ряд подклассов. Возможности подхода с обучаемыми классификаторами состояний позволяют организовать внутреннюю логику развития распознающей системы с иерархической структурой, которая путем формирования в процессе обучения решающих функций оптимальной сложности и разумной селекцией признаков повышает точность распознавания; а в целом обеспечивает более высокую надежность и безопасность перевозочного процесса, что особенно важно для систем автоматизации технологических процессов на железнодорожном транспорте.

Основными задачами, требующими решения при создании обучаемого классификатора состояний объекта с многомерными информативными признаками, функционирующего в условиях воздействия значительных дестабилизирующих факторов, являются:

- определение характеристик электрических параметров, описывающих состояния рельсовых линий и служащих первичными информативными признаками;
- формирование априорного словаря признаков;
- выбор вида решающей функции;
- определение сложности решающей функции;
- разработка алгоритма «обучения» решающей функции;
- исследование инвариантных свойств решающей функции обученного классификатора.

3. Процедура выбора оптимального набора информативных признаков. Процедура формирования адекватного множества признаков, учитывающая трудности, которые связаны с реализацией процессов выделения и селекции признаков из априорного словаря, и обеспечивающая в то же время необходимое качество распознавания, представляет собой одну из наиболее трудных задач при построении классификаторов состояний рельсовых линий.

В априорный словарь признаков обычно включаются все признаки, характеризующие состояние объекта распознавания,

которые могут быть представлены в виде набора данных, полученных путем измерения и формирования n признаков, принадлежащих образам различных классов состояний [5-7]. При этом возникает задача селекции признаков и исключения из априорного словаря малоинформативных признаков. Задача выбора $n_1 < n$ признаков эквивалентна задаче выбора минимального числа признаков из множества m , обеспечивающих заданное качество распознавания.

Решение задачи попутно должно обеспечивать селекцию признаков по информативности. В случае независимых признаков, вычисляя качество распознавания для совокупности из множества $\{m\}$ признаков, включающего исследуемый признак, а затем для совокупности, не содержащего данный признак, можно определить, сравнивая качество определения информативности (или распознавания) каждого признака, следует выбрать те $n_1 < n$ признаков, информативность которых больше [8]. Эта процедура решения поставленной задачи обеспечивает оптимальное решение только на уровне гипотезы о независимости признаков.

В случае зависимых признаков подобная процедура состоит из следующих этапов:

- ряд признаков исключается поочередно, как в описанной выше процедуре;

- в пространстве оставшихся $(n-1)$ признаков определяется качество распознавания;

- исключению подлежит тот признак, отсутствие которого наименее сильно изменило качество распознавания;

- поочередно исключаются остальные $(n-1)$ признаков, а из группы в $(n-1)$ исключается второй признак;

- процедура повторяется $(n-n_1)$ раз.

Подобная процедура не позволяет оценивать качество распознавания при одновременном исключении нескольких признаков. Некоторые авторы считают оптимальной процедуру полного перебора по той причине, что малоинформативные сами по себе, но сильно коррелированные признаки могут составлять малоинформативную систему [8]. Априорные матрицы признаков, образов и классов имеют вид:

$$M_N = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_i^T \\ \vdots \\ X_m^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}, \quad (1)$$

$$M_S = \begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_i^T \\ \vdots \\ X_m^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}, \quad (2)$$

$$M_K = \underbrace{\begin{bmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_i^T \\ \vdots \\ X_m^T \end{bmatrix}}_{\text{классы (n) образы (m)}} = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mj} & \cdots & x_{mn} \end{bmatrix}}_{\text{матрицы признаков}}, \quad (3)$$

где X_j^T — транспонированный вектор столбец X_i ; x_{mn} — значения информативных признаков; M_i — классы состояний, $i \in N, S, K$.

Очевидно, совокупность признаков должна в наибольшей степени характеризовать те свойства объекта распознавания (РЛ), которые принципиальны для их распознавания. При этом от размерности p -признакового пространства в значительной степени зависит процедура «обучения», классификации, достоверность распознавания, затраты на измерения характеристик объекта распознавания.

Уменьшение количества информативных признаков снижает затраты на измерение и вычисление, но также может снизить

точность распознавания. Но если время на обучение или корректировку коэффициентов решающей функции в реальном времени жестко ограничено (особенно в динамике движения поездов), то повышение размерности пространства признаков оказывается единственным средством увеличения точности и достоверности до необходимого уровня.

Рельсовая линия, как многополюсник с распределенными параметрами, характеризуется входными и выходными электрическими параметрами [9], в качестве которых возможно применить [2]:

– комплексные амплитуды напряжения на входе рельсового четырехполюсника;

– комплексные амплитуды тока на входе рельсового четырехполюсника;

– комплексные амплитуды напряжения на выходе рельсового четырехполюсника.

Множества образов, сформированных с использованием указанных шести признаков, имеют вид:

$$m_{iN} = \{U_{1N}, \varphi_{1N}, I_{1N}, \psi_{1N}, U_{2N}, \varphi_{2N}\}, i = 1, 2, \dots, n, \quad (4)$$

$$m_{jS} = \{U_{1S}, \varphi_{1S}, I_{1S}, \psi_{1S}, U_{2S}, \varphi_{2S}\}, j = 1, 2, \dots, l, \quad (5)$$

$$m_{pK} = \{U_{1K}, \varphi_{1K}, I_{1K}, \psi_{1K}, U_{2K}, \varphi_{2K}\}, p = 1, 2, \dots, q, \quad (6)$$

где $n \neq l \neq q$, m_N, m_S, m_K – соответственно, образы NSK классов.

4. Определение вида и порядка полинома решающей функции классификатора. Следующая задача, связанная с разработкой обучаемого классификатора состояний рельсовых линий, состоит в отыскании оптимальных решающих процедур, необходимых при классификации. Решающие функции (РФ) можно получить рядом способов. В тех случаях, когда о распознаваемых образах имеются полные априорные сведения, РФ могут быть точно определены на основе этой информации, но если об образах имеются немногочисленные априорные сведения, то при построении распознающей системы используются обучающие процедуры. На первом этапе выбираются решающие функции полиномами минимального порядка, а затем, в процессе выполнения итеративных шагов, РФ доводятся до оптимального либо приемлемого вида.

Наиболее удобными РФ, имеющими наименьшую чувствительность, являются линейные РФ, которые можно обобщить и

на случай нелинейных границ классов. Обобщить линейные РФ удобнее введением РФ, вида [10]:

$$d_i(X) = C_1 f_1(X) + C_2 f_2(X) + \dots + C_k f_k(X) + C_{k+1} = \sum_{i=1}^{k+1} C_i f_i(X), \quad (7)$$

где $\{f_i(X)\}, i=1,2,\dots,k$ — действительные однозначные функционалы образа (X) , $f_{k+1}(X)=1$, а $(k+1)$ — число членов разложения. Соотношение (7) представляет бесконечное множество решающих функций, и применение несложного преобразования позволяет работать с ними как с линейными. Если представить (7) в виде $d(X)=C^T X^*$, где

$$C=(C_1, C_2, \dots, C_k, C_{k+1})^T, \mathbf{a},$$

$$X^* = \left\{ \begin{array}{c} f_1(X) \\ f_2(X) \\ \cdot \\ \cdot \\ f_k(X) \\ 1 \end{array} \right\}, \quad (8)$$

функция $\{f_i(X)\}$, после того как их значения вычислены, будет представлять собой просто набор чисел, а вектор X^* — обычный k -мерный вектор, дополненный единицей.

Таким образом, хотя в K -мерном пространстве РФ можно считать линейными, в n -мерном пространстве исходных образов они полностью сохраняют нелинейный характер.

Количество членов в полиноме РФ (порядок полинома) определяется по критерию качества распознавания, то есть сравнением общего разброса вычисленных значений РФ с показателем классов. Если он больше вычисленного ранее, то ранее принятый порядок полинома РФ достаточен, если же нет, то требуется выбрать другой вид РФ.

Привлекателен подход с использованием в качестве решающих функций $d_i(X)$, где i -индекс класса, полиномов Эрмита [11], обеспечивающих требуемое качество классификации состояний рельсовой линии при минимальной сложности решающих функций.

Для получения многочленов Эрмита используется рекуррентное соотношение:

$$H_{k+1}(x) - 2xH_k(x) + 2kH_{k-1}(x) = 0, \quad k \geq 1. \quad (9)$$

Эти функции ортогональны относительно весовой функции $u(x) = \exp(-x^2)$, причем интервал ортогональности составляет $-\infty < x < \infty$; это обстоятельство делает использование таких функций чрезвычайно удобным, поскольку освобождает от забот относительно диапазона изменения переменных, который у принятых выше информативных признаков рельсовых линий значительный.

Определив коэффициент $A_k = \int_a^b u(x) \phi_k^{*2}(x) dx$ и подставив его в

соотношение $\phi_k(x) = \sqrt{\frac{u(x)}{A_k}} \phi_k^*(x)$, можно показать, что

ортонормированные многочлены Эрмита определяются следующим соотношением:

$$\phi_k(x) = \frac{\exp(-x^2/2) H_k(x)}{\sqrt{2^k k! \sqrt{\pi}}}, \quad k = 0, 1, 2, \dots \quad (10)$$

Формула (9) представлена для одной переменной, а решающая функция многих переменных классификатора состояний рельсовых линий с использованием многочленов Эрмита с учетом (10) имеет вид:

$$d_j(X) = \sum_{i=1}^n C_{ij} \phi_j(X), \quad (11)$$

где $\phi_j(X) = \prod \phi_k(x_i)$ — ортонормированные функции Эрмита.

Для определения состояний рельсовых линий удобно воспользоваться последовательным алгоритмом классификации [12] основанным на информации множества решающих функций, когда каждому классу соответствует своя решающая функция и показатель класса:

$$d_j(X) \equiv \left\{ \begin{array}{l} d_N(X) \\ d_S(X) \\ d_K(X) \end{array} \right\},$$

где $d_N(X)$ — решающая функция класса свободного и исправного состояний рельсовых линий, $d_S(X)$ — решающая функция класса занятого и исправного состояний рельсовых линий, $d_K(X)$ — решающая функция класса неисправного состояния рельсовых линий (излом рельсовой линии).

Последовательный алгоритм определения состояний рельсовых линий с множеством решающих функций приведен на рисунке 1.

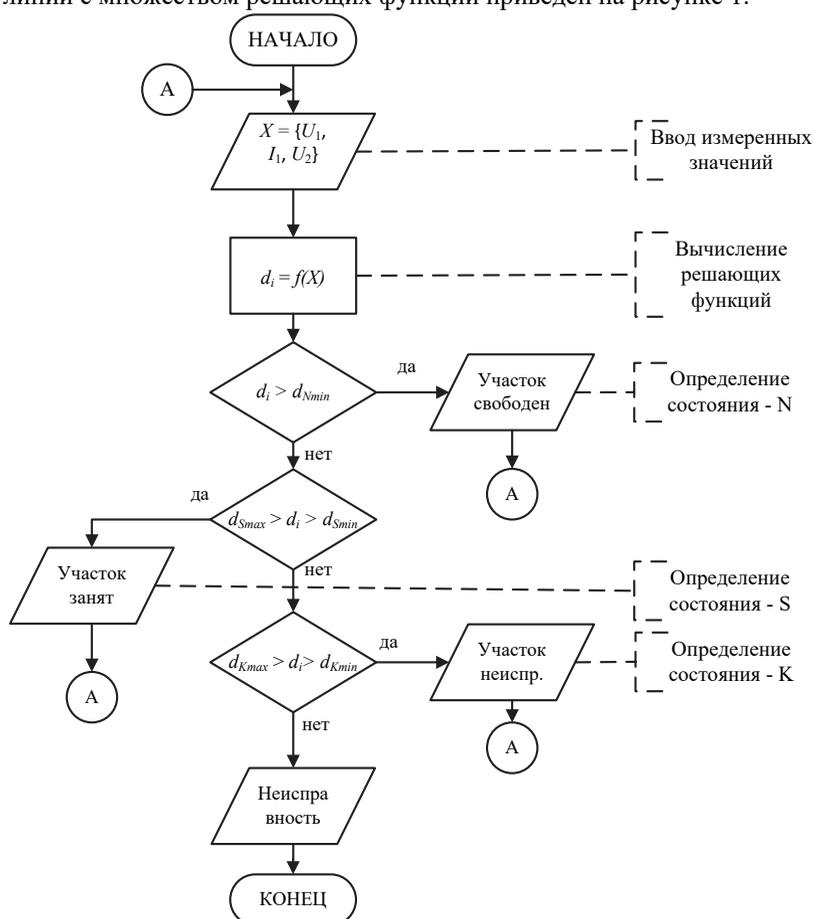


Рис. 1. Блок-схема алгоритма определения состояний рельсовых линий с множеством решающих функций: d_i — текущая решающая функция, d_N , d_S , d_K — решающие функции классов, соответствующих состояниям рельсовых линий

Выбрав вид решающих функций с использованием многочленов Эрмита, необходимо провести процедуру их «обучения», в процессе которой определяются коэффициенты и сложность (порядок полинома) решающих функций $d_N(x_{N_i})$, $d_S(x_{S_i})$, $d_K(x_{K_i})$. Для определения коэффициентов РФ в работе используется репрезентативная «обучающая» выборка образов, и этот этап является этапом «обучения» решающей функции классификатора [13]. В этом процессе используется «обучающая» выборка образов $L = \{X_1^j, X_2^j, \dots, X_n^j, Y_j, j = 1, 2, \dots, l\}$, в которую входят как образы, принадлежащие к одному классу X_n^j , так и образы, принадлежащие другим классам Y_j . Известно, что при полном отсутствии информации вероятностного характера общее количество образов, отбираемых для разбиения на n классов, удовлетворительных к обобщению, должно быть, по крайней мере, равно утроенной размерности векторов образов, то есть $V_k = 3(n+1)$ [14]. Если $M > V_k$, где M — число образов обучающей выборки, и обучающее множество распознается правильно, то полученное решение классификатора и, соответственно, его способность к обобщению можно считать удовлетворительной. Эмпирическим правилом служит выбор числа образов $M \approx 10V_k$ [15].

Способность классификатора к обобщению проверяется при предъявлении ему данных, не вошедших в обучающую выборку образов. Следует отметить, что признаки, вошедшие в обучающую выборку, должны обладать репрезентативностью по отношению к «полевым» данным [3, 16].

На точность классификации состояния участка контроля существенное влияние оказывает порядок полинома решающих функций $d_N(x_{N_i})$, $d_S(x_{S_i})$, $d_K(x_{K_i})$ — чем он выше, тем выше становится и распознающая возможность решающих функций. При этом на начальном этапе порядок полинома решающих функций $d_N(x_{N_i})$, $d_S(x_{S_i})$, $d_K(x_{K_i})$ выбирается минимальной, а затем, в процессе итерации, доводится до оптимального либо приемлемого вида. Выбранный порядок полинома решающих функций $d_N(x_{N_i})$, $d_S(x_{S_i})$, $d_K(x_{K_i})$ на промежуточном этапе проверяется критерием Фишера, то есть производится сравнение общего разброса вычисленных значений решающих функций $d_N(x_{N_i})$, $d_S(x_{S_i})$, $d_K(x_{K_i})$ с соответствующими показателями классов. Если он больше

вычисленного ранее, то полученный порядок полинома решающих функций $d_N(x_{N_i})$, $d_S(x_{S_i})$, $d_K(x_{K_i})$ достаточен, если же нет, то требуется выбрать другой, более высокий порядок полинома решающих функций. Следующий порядок полинома соответствует функциям второго порядка, затем третьего и т.д. В результате выбора порядка полинома решающих функций получаем множество, состоящее из $(y+1)$ ортонормированных функций $\varphi_0(x), \varphi_1(x), \dots, \varphi_y(x)$, где y — порядок полинома решающей функции. Исследования, проведенные авторами, показали, что для поставленной задачи, достаточным является первая степень полинома решающих функций.

Ввиду того, что в каждое из трех выбранных состояний рельсовых линий входит подмножество состояний, зависящих от изменения первичных параметров рельсовой линии от изменения возмущающих воздействий, то получаем множество $d_i(x_i)$ ($i = N, S, K$) в виде линейной комбинации функции $\varphi(x)$ с неизвестными коэффициентами C , то есть:

$$d_j(X) = \sum_{i=1}^n C_{ij} \phi_j(X),$$

$$d_{N_i}(x_{N_i}) = \sum_{j=1}^{(r+1)^p} C_{N_j} \cdot \varphi_{N_j}(x_{N_i}),$$

$$i=1..nI,$$

$$d_{S_i}(x_{S_i}) = \sum_{j=1}^{(r+1)^p} C_{S_j} \cdot \varphi_{S_j}(x_{S_i}),$$

$$i=1..sI,$$

$$d_{K_i}(x_{K_i}) = \sum_{j=1}^{(r+1)^p} C_{K_j} \cdot \varphi_{K_j}(x_{K_i}),$$

$$i=1..kI,$$

где nI , sI , kI — диапазон подсостояний (образов) в классах состояний, соответственно, p — количество информативных признаков рабочего

алфавита (равняется трем: амплитуды входных и выходных параметров рельсового четырехполосника — U_1, I_1 и U_2), $r=1$ порядок полинома решающей функции, $C_{N_j}, C_{S_j}, C_{K_j}$ — коэффициенты, $\varphi_{N_j}(x_{N_i}), \varphi_{S_j}(x_{S_i}), \varphi_{K_j}(x_{K_i})$ — ортогональные функции.

Для представления полиномиальной решающей функции число необходимых членов зависит от порядка полинома y и размерности (числа признаков-аргументов) h :

$$Q_y^h = \frac{(h+y)!}{y!h!}, \quad (12)$$

где h — размерность полинома. Так как рабочий алфавит априорно состоит из трех признаков (U_1, I_1, U_2), то количество сочетаний $Q_y^h = 6$.

Поэтому ортогональные функции $\varphi_{N_j}(x_{N_i}), \varphi_{S_j}(x_{S_i}), \varphi_{K_j}(x_{K_i})$ определяются как:

класс N:

$$\begin{aligned} \varphi_{N_1}(x_{N_i}) &= \phi_1(U_{1N_i}) \cdot \phi_1(I_{1N_i}) \cdot \phi_1(U_{2N_i}) \\ \varphi_{N_2}(x_{N_i}) &= \phi_1(U_{1N_i}) \cdot \phi_2(I_{1N_i}) \cdot \phi_1(U_{2N_i}) \\ \varphi_{N_3}(x_{N_i}) &= \phi_1(U_{1N_i}) \cdot \phi_1(I_{1N_i}) \cdot \phi_2(U_{2N_i}) \\ \varphi_{N_4}(x_{N_i}) &= \phi_2(U_{1N_i}) \cdot \phi_2(I_{1N_i}) \cdot \phi_1(U_{2N_i}) \\ \varphi_{N_5}(x_{N_i}) &= \phi_2(U_{1N_i}) \cdot \phi_1(I_{1N_i}) \cdot \phi_1(U_{2N_i}) \\ \varphi_{N_6}(x_{N_i}) &= \phi_2(U_{1N_i}) \cdot \phi_1(I_{1N_i}) \cdot \phi_2(U_{2N_i}) \end{aligned}, \quad (13)$$

класс S:

$$\begin{aligned} \varphi_{S_1}(x_{S_i}) &= \phi_1(U_{1S_i}) \cdot \phi_1(I_{1S_i}) \cdot \phi_1(U_{2S_i}) \\ \varphi_{S_2}(x_{S_i}) &= \phi_1(U_{1S_i}) \cdot \phi_2(I_{1S_i}) \cdot \phi_1(U_{2S_i}) \\ \varphi_{S_3}(x_{S_i}) &= \phi_1(U_{1S_i}) \cdot \phi_1(I_{1S_i}) \cdot \phi_2(U_{2S_i}) \\ \varphi_{S_4}(x_{S_i}) &= \phi_2(U_{1S_i}) \cdot \phi_2(I_{1S_i}) \cdot \phi_1(U_{2S_i}) \\ \varphi_{S_5}(x_{S_i}) &= \phi_2(U_{1S_i}) \cdot \phi_1(I_{1S_i}) \cdot \phi_1(U_{2S_i}) \\ \varphi_{S_6}(x_{S_i}) &= \phi_2(U_{1S_i}) \cdot \phi_1(I_{1S_i}) \cdot \phi_2(U_{2S_i}) \end{aligned}, \quad (14)$$

класс K:

$$\begin{aligned}
\varphi_{K_1}(x_{K_i}) &= \phi_1(U_{1K_i}) \cdot \phi_1(I_{1K_i}) \cdot \phi_1(U_{2K_i}) \\
\varphi_{K_2}(x_{K_i}) &= \phi_1(U_{1K_i}) \cdot \phi_2(I_{1K_i}) \cdot \phi_1(U_{2K_i}) \\
\varphi_{K_3}(x_{K_i}) &= \phi_1(U_{1K_i}) \cdot \phi_1(I_{1K_i}) \cdot \phi_2(U_{2K_i}) \\
\varphi_{K_4}(x_{K_i}) &= \phi_2(U_{1K_i}) \cdot \phi_2(I_{1K_i}) \cdot \phi_1(U_{2K_i}) \\
\varphi_{K_5}(x_{K_i}) &= \phi_2(U_{1K_i}) \cdot \phi_1(I_{1K_i}) \cdot \phi_1(U_{2K_i}) \\
\varphi_{K_6}(x_{K_i}) &= \phi_2(U_{1K_i}) \cdot \phi_1(I_{1K_i}) \cdot \phi_2(U_{2K_i})
\end{aligned} \tag{15}$$

Ортонормированные многочлены $\phi(U_1)$, $\phi(U_2)$ и $\phi(I_1)$ полиномов решающих функций для признаков U_1 , U_2 и I_1 определяются по (10).

Последним этапом формирования решающих функций состояний рельсовых линий является проверка их репрезентативности на образах, не вошедших в обучающие выборки. Если в результате формирования получим решающую функцию с большой погрешностью классификации, то необходимо повысить порядок полинома решающей функции либо выбрать другой вид решающей функции [17, 18].

В качестве номера классов P_j в процессе обучения принимаются числовые значения с различными знаками: для класса N — $P_N = +1$; для класса S — $P_S = -0,5$; для класса K — $P_K = -1$. Выбранный принцип назначения номера класса $d_N(x_{N_i}) > 0$, $d_S(x_{S_i})$ и $d_K(x_{K_i}) \leq 0$ позволяет обеспечить безопасность движения поездов, и появляется возможность двукратной проверки состояния объекта контроля [19, 20] по знаку, и значению решающей функции:

- если $d_N(x_i) > 0$, движение разрешено;
- если $d_S(x_i)$ и $d_K(x_i) \leq 0$, движение запрещено.

Решающее правило при этом имеет вид:

$$\left. \begin{aligned}
&\text{если } P_N > d_N(x_i) \geq P_N + \Delta P \text{ и } d_N(x_i) > 0, \text{ то } x_i \in N; \\
&\text{если } P_S > d_S(x_i) \geq P_S + \Delta P \text{ и } d_S(x_i) \leq 0, \text{ то } x_i \in S \\
&\text{если } P_K > d_K(x_i) \geq P_K + \Delta P \text{ и } d_K(x_i) \leq 0, \text{ то } x_i \in K
\end{aligned} \right\}, \tag{16}$$

где ΔP — допустимый предел отклонения величины $d_j(x_i)$ от показателя класса P_j .

5. Исследование обучаемого классификатора с многомерными информативными признаками. Исходными данными при формировании решающих функций являются:

- а) длина контролируемой рельсовой линии;
- б) частота тока опроса рельсовых линий;
- в) величины сопротивлений в начале и в конце рельсовой линии;
- г) диапазон варьирования проводимости изоляции $g_{из}$;
- д) априорный порядок решающей функции.

Информативные признаки и их математическое описание формируются математическими моделями электрических входных и выходных параметров рельсового четырехполюсника, представленного в форме матрицы «А» параметров и уравнений состояний рельсовой линии:

а) для класса N

$$\begin{cases} U_{1N} = \frac{E(A_N Zn + B_N)}{(C_N Zn + D_N)Zo + A_N Zn + B_N} \\ U_{2N} = \frac{EZn}{(C_N Zn + D_N)Zo + A_N Zn + B_N}, \\ I_{1N} = \frac{E(C_N Zn + D_N)}{(C_N Zn + D_N)Zo + A_N Zn + B_N} \end{cases}, \quad (17)$$

б) для класса S

$$\begin{cases} U_{1S} = \frac{E(A_S Zn + B_S)}{(C_S Zn + D_S)Zo + A_S Zn + B_S} \\ U_{2S} = \frac{EZn}{(C_S Zn + D_S)Zo + A_S Zn + B_S}, \\ I_{1S} = \frac{E(C_S Zn + D_S)}{(C_S Zn + D_S)Zo + A_S Zn + B_S} \end{cases}, \quad (18)$$

в) для класса K

$$\begin{cases} U_{1K} = \frac{E(A_K Zn + B_K)}{(C_K Zn + D_K)Zo + A_K Zn + B_K} \\ U_{2K} = \frac{EZn}{(C_K Zn + D_K)Zo + A_K Zn + B_K}, \\ I_{1K} = \frac{E(C_K Zn + D_K)}{(C_K Zn + D_K)Zo + A_K Zn + B_K} \end{cases}, \quad (19)$$

где U_{1i} , U_{2i} , I_{1i} — амплитуды напряжений на входе и выходе и тока входе четырехполюсника рельсовой линии, A_i , B_i , C_i , D_i , $i \in N, S, K$ — коэффициенты четырехполюсника рельсовой линии в трех классах

состояний; Z_0 , Z_n — двухполюсники на входе и выходе рельсового четырехполосника; E — напряжение источника питания.

Погрешность классификации состояний участка контроля решающими функциями $d_j(x_{j_i})$ для комбинации первичных информативных признаков оценивается формулой относительной погрешности:

$$\delta G_j = \left| \max \left[\left(d_j(x_{j_i}) - P_j \right) / P_j \right], i=1..n, \right. \quad (20)$$

где P_j — истинный показатель j -го класса, $d_j(x_{j_i})$ — вычисленное значение класса, n — размерность множества образов.

Оптимальной считается та решающая функция $d_N(x_{N_i})$, $d_S(x_{S_i})$, $d_K(x_{K_i})$, которая дает наименьшую относительную погрешность распознавания состояний рельсовой линии.

В работе сформированы решающие функции и исследовано качество распознавания состояний рельсовой линии при трех признаках U_1 , I_1 , U_2 для участка контроля рельсовой линии длиной 2.5 км, при частоте тока опроса рельсовых линий $f_e=50$ Гц и $Z_n = 0.2 \cdot e^{j50^\circ}$ Ом, $Z_o = 0.1 \cdot e^{j0^\circ}$ Ом, при $0.02 < g_{из} < 1$ См/км.

а) *формирование РФ класса N*. Решающая функция $d_N(X)$ имеет вид:

$$d_N = 2.6791 \cdot \varphi_1 \cdot (U_{1N_i}) \cdot \varphi_1(I_{1N_i}) - 2.2196 \cdot \varphi_1 \cdot (U_{1N_i}) \cdot \varphi_2(I_{1N_i}) - 1.6249 \cdot \varphi_2 \cdot (U_{1N_i}) \cdot \varphi_1(I_{1N_i}) - 7.0483 \cdot \varphi_2 \cdot (U_{1N_i}) \cdot \varphi_2(I_{1N_i}) \quad (21)$$

На рисунке 2 приведены результаты исследования в среде Mathcad решающей функции (21) в виде области существования решающей функции в классе свободного и исправного состояний рельсовых линий при изменении проводимости изоляции на участке контроля от 1 до 0.02 См/км.

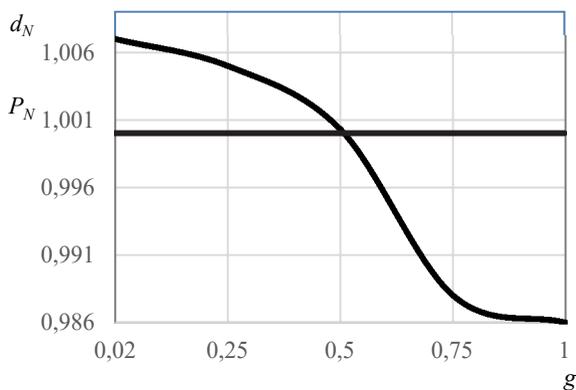


Рис. 2. Область существования решающей функции d_N : показатель класса $P_N=1$

Из графиков следует, что максимальное отклонение решающей функции, вычисленное по формуле (14), от показателя класса P_N составляет 1.473%, что указывает на высокую инвариантность решающей функции к изменению проводимости изоляции на участке контроля. Минимальное возможное значение решающей функции d_N составляет 0.986, а максимальное — $d_N = 1.007$. Эти величины определяют минимальные и максимальные границы решающей функции в классе N , и используются в микропроцессорном классификаторе как пороговые значения.

б) *формирование РФ класса S*. Решающая функция $d_S(X)$ имеет вид:

$$d_S = -6.1226 \cdot \varphi_1 \cdot (U_{1S_i}) \cdot \varphi_1(I_{1S_i}) - 0.8259 \cdot \varphi_1 \cdot (U_{1S_i}) \cdot \varphi_2(I_{1S_i}) + 41.6314 \cdot \varphi_2 \cdot (U_{1S_i}) \cdot \varphi_1(I_{1S_i}) + 28.2436 \cdot \varphi_2 \cdot (U_{1S_i}) \cdot \varphi_2(I_{1S_i}). \quad (22)$$

На рисунке 3 приведены результаты исследования в среде Mathcad решающей функции (22) в виде области существования решающей функции в классе занятого поездом и исправного состояний рельсовых линий при изменении проводимости изоляции на участке контроля от 1 до 0.02 См/км.

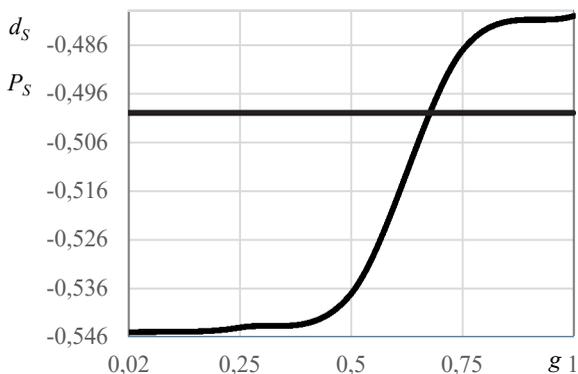


Рис. 3. Область существования решающей функции d_S : показатель класса $P_S = -0.5$

Из графиков следует, что максимальное отклонение решающей функции, вычисленное по формуле (14), от показателя класса P_S составляет 1.371%, что показывает на высокую инвариантность к изменению проводимости изоляции на участке контроля. Минимально возможное значение решающей функции d_S составляет -0.545 , а максимальное возможное значение $d_S = -0.48$. Эти величины определяют минимальные и максимальные границы решающей функции в классе S , и используются в микропроцессорном классификаторе, как пороговые значения границы области существования решающей функции d_S .

в) *формирование РФ класса K*. Решающая функция $d_K(X)$ имеет вид:

$$d_K = -7.4613 \cdot \varphi_1 \cdot (U_{1K_i}) \cdot \varphi_1(I_{1K_i}) + 2.1029 \cdot \varphi_1 \cdot (U_{1K_i}) \cdot \varphi_2(I_{1K_i}) + 56.8385 \cdot \varphi_2 \cdot (U_{1K_i}) \cdot \varphi_1(I_{1K_i}) + 35.7538 \cdot \varphi_2 \cdot (U_{1K_i}) \cdot \varphi_2(I_{1K_i}) \quad (23)$$

На рисунке 4 приведены результаты исследования в среде Mathcad решающей функции (23) в виде области существования решающей функции в классе неисправного состояний рельсовых линий (излом рельсов) при изменении проводимости изоляции на участке контроля от 1 до 0.02 См/км.

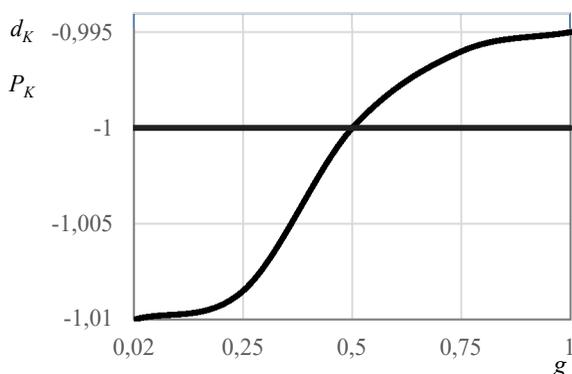


Рис. 4. Область существования решающей функции d_K : показатель класса $P_K = -1$

Из графиков следует, что максимальное отклонение решающей функции, вычисленное по формуле (14), от показателя класса P_K составляет 1.1%, что показывает на высокую инвариантность к изменению проводимости изоляции на участке контроля. Минимальное возможное значение решающей функции d_K в контрольном режиме составляет -1.01 , а максимальное возможное значение $d_K = -0.995$. Минимально возможное значение РФ d_S составляет -0.545 , а максимальное возможное значение $d_S = -0.48$. Эти величины определяют минимальные и максимальные границы решающей функции в классе K , и используются в микропроцессорном классификаторе как пороговые значения границы области существования решающей функции d_K .

6. Заключение. Одним из эффективных путей повышения устойчивости информационных систем управления технологическими процессами на транспорте является применение основополагающих принципов распознавания образов при создании классификаторов состояний рельсовых линий, функционирующих в условиях воздействия значительных возмущений. Использование многочленов Эрмита в качестве полинома решающей функции позволяет существенно расширить диапазон правильного распознавания состояний рельсовых линий, синтезировать полиномы любого порядка и размерности. Использование такой функции объясняется тем условием, что ее легко воспроизвести, и она удовлетворяет условиям теоремы Вейерштрасса об аппроксимирующих свойствах полиномиальных ортогональных функций. Предложенный в работе алгоритм формирования

распознающего классификатора реализован в виде программы в среде Mathcad, удовлетворительные результаты, полученные в ходе исследований, подтверждают функциональные возможности подхода по разработке обучаемого классификатора состояний рельсовых линий с многомерными информативными признаками.

Литература

1. *Лябах Н.Н., Умрихин Н.Г.* Автоматизация процесса классификации интеллектуальных транспортных комплексов на основе экспертных систем // Вестник Ростовского государственного университета путей сообщения. 2013. № 2(50). С. 98–103.
2. *Тарасов Е.М.* Принципы распознавания в классификаторах состояний рельсовых линий // М.: Маршрут. 2003. 156 с.
3. *Ивахненко А.Г.* Самообучающиеся системы // Киев: АН УССР. 1963. 328 с.
4. *Ту Дж., Гонсалес Р.* Принципы распознавания образов / Пер. с англ. И.Б. Гуревича / под ред. Ю.И. Журавлева // М.: Мир. 1978. 412 с.
5. *Михеева Т.И., Сапрыкин О.Н.* Нейросетевое управление пространственно-координированными объектами транспортной инфраструктуры // Самара: D.S. Style. 2011. 217 с.
6. *Фомин Я.А., Тарловский Г.Р.* Статистическая теория распознавания образов // М.: Радио и связь. 1986. 286 с.
7. *Баранов Р.П., Белоконь А.В., Фаворская М.Н.* Определение и приоритизация признаков объектов на изображении в системах распознавания // Актуальные проблемы авиации и космонавтики. 2012. Т. 1. № 8. С. 328–329.
8. *Борисова И.А., Загоруйко Н.Г., Кутненко О.А.* Критерии информативности и пригодности подмножества признаков, основанные на функции сходства // Заводская лаборатория. Диагностика материалов. 2008. Т.74. №1. С. 68–71.
9. *Галушкин А.И.* Синтез многослойных систем распознавания образов // М.: Энергия. 1974. 367 с.
10. *Tarasov E.M., Zhelezov D.V.* Development of an Intelligent System of Determinating the Coordinates and the Speed of the Train // Transport and Telecommunication Journal. 2015. vol. 17. no. 2. pp. 138–143.
11. *Потапов А.С.* Распознавание образов и машинное восприятие: Общий подход на основе минимальной длины описания // СПб.: Политехника. 2007. 548 с.
12. *Суетин П. К.* Классические ортогональные многочлены: 3-е изд., перераб. и доп. // М.: ФИЗМАТЛИТ. 2005. 480 с.
13. *Орлов А.И.* Математические методы теории классификации // Научный журнал КубГАУ. 2014. №95(01). С. 1–37.
14. *Tarasov E.M., Isaicheva A.G.* Technique of measurement of ultralow resistance of current conductive junction of rail lines as the problem of states object identification // Proceedings of Information Technology and Nanotechnology (ITNT - 2015). 2015. vol. 1490. pp. 397–401.
15. *Журавлёв Ю.И., Рязанов В.В., Сенько О.В.* Распознавание. Математические методы. Программная система. Практические применения // М.: ФАЗИС. 2006. 176 с.
16. *Белоногов А.С., Куров М.Б.* Выбор решающей функции в задаче оценки электрических параметров рельсовых цепей // Наука и образование транспорту. Самара: СамГУПС. 2011. № 1. С. 69–71.

17. *Загоруйко Н. Г., Кутненко О. А.* Количественная мера компактности образов и метод ее повышения // Интеллектуализация обработки информации: Сб. докл. Девятой Международной конф. М.: Торус Пресс. 2012. С. 29–32.
18. *Неделько В.М.* Некоторые вопросы оценивания качества методов построения решающих функций // Вестник Томского государственного университета. 2013. №3(24). С. 123–132.
19. *Куров М.Б., Белоногов А.С., Смирнова Л.Б.* Метод удаленного мониторинга проводимости изоляции рельсовой линии // Наука и образование транспорту. Самара: СамГУПС. 2011. № 1. С. 100–101.
20. *Волик В.Г., Гуменников В.Б., Шорохов Н.С.* Измерение распределенных параметров рельсовой линии // Перспективные информационные технологии (ПИТ 2015): Сб. науч. тр. международной научно-технической конференции. Самара: СамНЦ РАН. 2015. С. 48–51.

Железнов Дмитрий Валерианович — д-р техн. наук, доцент, ректор, Федеральное государственное бюджетное образовательное учреждение высшего образования "Самарский государственный университет путей сообщения" (СамГУПС). Область научных интересов: разработка и создание методов и принципов инвариантности и распознавания состояний систем интервального управления движением поездов в условиях значительных возмущающих воздействий. Число научных публикаций — 85. zheleznov_dim@mail.ru, <https://www.samgups.ru/>; ул. Свободы, 2В, Самара, 443066; р.т.: +7(846)2624112, Факс: +7(846)2623076.

Тарасов Евгений Михайлович — д-р техн. наук, профессор, проректор, Федеральное государственное бюджетное образовательное учреждение высшего образования "Самарский государственный университет путей сообщения" (СамГУПС). Область научных интересов: разработка принципов инвариантности и распознавания состояний систем интервального управления движением поездов в условиях значительных возмущающих воздействий; разработка методов удаленного контроля состояния рельсовых линий и технической диагностики параметров рельсовой сети на магистральных железных дорогах. Число научных публикаций — 230. tarasov53em@yandex.ru, <https://www.samgups.ru/>; ул. Свободы, 2В, Самара, 443066; р.т.: +7(846)255-6735, Факс: +7(846)262-3076.

Исайчева Алевтина Геннадьевна — к-т техн. наук, доцент, доцент кафедры автоматики, телемеханики и связи на железнодорожном транспорте, Федеральное государственное бюджетное образовательное учреждение высшего образования "Самарский государственный университет путей сообщения" (СамГУПС). Область научных интересов: разработка методов удаленного контроля состояния рельсовых линий и технической диагностики параметров рельсовой сети на магистральных железных дорогах. Число научных публикаций — 43. aspirantsamit@rambler.ru; ул. Свободы, 2В, Самара, 443066; р.т.: +7(846)2556706, Факс: +7(846)2623076.

Михеева Татьяна Ивановна — д-р техн. наук, доцент, профессор кафедры организации и управления перевозками на транспорте, Федеральное государственное автономное образовательное учреждение высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» (Самарский университет). Область научных интересов: структурно-параметрический анализ интеллектуальных транспортных систем, разработка принципов нейросетевого управления пространственно-координированными объектами транспортной инфраструктуры. Число научных публикаций — 125. mikheevati@mail.ru, <https://www.ssau.ru/>; Московское шоссе, 34, Самара, 443086; р.т.: +7(846)2674674, Факс: +7(846)2674674.

D.V. ZHELEZNOV, E.M. TARASOV, A.G. ISAYCHEVA, T.I. MIKHEEVA
DEVELOPMENT OF THE LEARNING CLASSIFIER OF RAIL
LINES STATES WITH MULTIVARIATE INFORMATIVE
FEATURES

Zheleznov D.V., Tarasov E.M., Isaycheva A.G., Mikheeva T.I. **Development of the Learning Classifier of Rail Lines States with Multivariate Informative Features.**

Abstract. A considerable number of failures in the system of interval control of train traffic are associated with the impact of disturbing actions in a wide range of change in the only informative feature characterizing the rail lines states. In the paper, it is proposed to determine the state of the control object by the principles of the pattern recognition with multivariate informative features. It is suggested to use as the features the voltages and currents at the input and output of a quadripole, and as a polynomial of the decisive function — the Hermit's orthogonal polynomial, which allows increasing the depth of recognition and ensuring the relative invariance of disturbing actions by amplifying the order and dimension. In recognition of the rail lines states the relative error of calculating class boundaries by decisive functions is used as a quality criterion.

Serviceability of the proposed method is demonstrated by the results of rail lines states recognition with a "trained" decisive function.

Keywords: classifier of states, decisive function, multivariate features.

Zheleznov Dmitry Valerianovich — Ph.D., Dr. Sci., associate professor, rector, Samara State Transport University. Research interests: design and development of methods and principles of invariance and recognition of system states of interval train control in major disturbing actions. The number of publications — 85. zheleznov_dim@mail.ru, <https://www.samgups.ru/>; 2B, Svoboda str., Samara, 443066, Russia; office phone: +7(846)2624112, Fax: +7(846)2623076.

Tarasov Evgeniy Mikhailovich — Ph.D., Dr. Sci., professor, vice-rector, Samara State Transport University. Research interests: development of the principles of invariance and recognition of systems states of interval train control in major disturbing actions, development of remote control of rail lines and technical diagnostics parameters of the rail network on the railways. The number of publications — 230. tarasov53em@yandex.ru, <https://www.samgups.ru/>; 2B, Svoboda str., Samara, 443066, Russia; office phone: +7(846)255-6735, Fax: +7(846)262-3076.

Isaycheva Alevtina Genadievna — Ph.D., associate professor, associate professor of automation, telemechanics and communication on railway transport department, Samara State Transport University. Research interests: development of remote control of rail lines and methods of technical diagnostics parameters of the rail network of trunk railways. The number of publications — 43. aspirantsamiit@rambler.ru; 2B, Svoboda str., Samara, 443066, Russia; office phone: +7(846)2556706, Fax: +7(846)2623076.

Mikheyeva Tatyana Ivanovna — Ph.D., Dr. Sci., associate professor, professor of organization of transportation and management on transport department, Samara University. Research interests: structural-parametric analysis of intelligent transport systems, the development of the principles of neural network control of space-coordinated transport infrastructure objects. The number of publications — 125. mikheevati@mail.ru,

References

1. Lyabakh N.N., Umrihin N.G. [Automation of the process of classification of intelligent transport complexes on basis of expert systems]. *Vestnik Rostovskogo gosudarstvennogo universiteta putej soobshhenija – Herald of the Rostov State Transport University*. 2013. vol. 2(50). pp. 98–103. (In Russ.).
2. Tarasov E.M. *Principy raspoznavanija v klassifikatorah sostojanij rel'sovyh linij* [Principles of recognition of rail lines states classifiers]. Moscow: Marshrut. 2003. 156 p. (In Russ.).
3. Ivakhnenko A.G. *Samoobuchajushhiesja sistemy* [Self-learning system]. Kiev: Publishing House of the USSR Academy of Sciences. 1963. 328 p. (In Russ.).
4. Tou J. T., Gonzalez R. C. *Pattern recognition principles*. Addison-Wesley Pub. Co. 1974. 395 p. (Russ. ed.: Tu Dzh., Gonsales R. *Principy raspoznavanija obrazov*. M.: Mir. 1978. 412 p.).
5. Mikheeva T.I., Saprykin O.N. *Nejrosetevoe upravlenie prostranstvenno koordinirovannymi ob'ektami transportnoj infrastruktury*. [Neural network management space-coordinated transport infrastructure objects]. Samara: D.S. Style. 2011. 217 p. (In Russ.).
6. Fomin J.A., Tarlovskii G.R. *Statisticheskaja teorija raspoznavanija obrazov* [Statistical theory of pattern recognition]. Moscow: Radio and Communications. 1986. 286 p. (In Russ.).
7. Baranov R.P., Belokon A.V., Favorskaja M.N. [Definition and prioritization features of objects in the image of recognition systems]. *Aktual'nye problemy aviacii i kosmonavтики – Actual problems of aviation and astronautics*. 2012. Issue 1. vol. 8. pp. 328–329. (In Russ.).
8. Borisova I.A., Zagoruiko N.G., Kutnenko O.A. [Criteria informative and suitability of subsets based on similarity of function]. *Zavodskaja laboratorija. Diagnostika materialov – Factory Laboratory. Diagnosis materials*. 2008. Issue 74. vol. 1. pp. 68–71. (In Russ.).
9. Galushkin A.I. *Sintez mnogoslojnyh sistem raspoznavanija obrazov* [Synthesis of multilayer pattern recognition systems]. Moscow: Energy. 1974. 367 p. (In Russ.).
10. Tarasov E.M., Zheleznov D.V. Development of an Intelligent System of Determinating the Coordinates and the Speed of the Train. *Transport and Telecommunication Journal*. 2016. vol. 17. no. 2. pp. 138–143.
11. Potapov A.S. *Raspoznavanie obrazov i mashinnoe vosprijatie: Obshhij podhod na osnove minimal'noj dliny opisaniya* [Pattern recognition and machine perception: The general approach on the basis of a minimum description length]. Sankt-Peterburg: University of Technology. 2007. 548 p. (In Russ.).
12. Suetin P.K. *Klassicheskie ortogonal'nye mnogochleny* [Classical orthogonal polynomials]. Moscow: FIZMATLIT. 2005. 480 p. (In Russ.).
13. Orlov A.I. [Mathematical methods of the theory of classification]. *Nauchnyj zhurnal KubGAU – Scientific Journal KubGAU*. 2014. vol. 95(01). pp. 1–37. (In Russ.).
14. Tarasov E.M., Isaicheva A.G. [Technique of measurement of ultralow resistance of current conductive junction of rail lines as the problem of states object identification]. *Proceedings of Information Technology and Nanotechnology (ITNT – 2015)*. 2015. vol. 1490. pp. 397–401.
15. Zhuravlev J.I., Ryazanov V.V., Senko O.V. *Raspoznavanie. Matematicheskie metody. Programmnaja sistema. Prakticheskie primenenija* [Recognition. Mathematical methods. Software system. Practical applications]. Moscow: Fazis. 2006. 176 p. (In Russ.).

16. Belonogov A.S., Kurov M.B. [Selecting a decisive function in the problem of estimation of the electrical parameters of track circuits]. *Nauka i obrazovanie transportu – Science and education for Transport*. 2011. vol. 1. pp. 69–71. (In Russ.).
17. Zagoruiko N.G., Kutnenko O.A. [Quantitative measure of a compact manner and method of its increase] *Intellektualizacija obrabotki informacii: Sb. dokl. Devjatoj Mezhdunarodnoj konf.* [Intellectualization information processing. Coll. papers 9 Int. Conf]. Moscow: Torus Press. 2012. pp 29–32. (In Russ.).
18. Nedelko V.M. [Some questions of estimation of quality of construction methods of decision functions]. *Vestnik Tomskogo gosudarstvennogo universiteta – Bulletin of the Tomsk State University*. 2013. vol. 3(24). pp 123–132. (In Russ.).
19. Kurov M.B., Belonogov A.S., Smirnova L.B. [The remote monitoring method isolation conduction of rail line]. *Nauka i obrazovanie transportu – Science and Education for Transport*. 2011. vol. 1. pp. 100–101. (In Russ.).
20. Volik V.G., Gumennikov V.B., Shorohov N.S. [The measurement of distributed parameters of the rail line] *Perspektivnye informacionnye tehnologii (PIT 2015): Sb. nauch. tr. mezhdunarodnoj nauchno-tehnicheskoy konferencii* [Prospective information technology (PIT 2015) studies of the International scientific and technical conference]. Samara: SamNC RAN. 2015. pp. 48–51. (In Russ.).

А.А. ПЕТРОВСКИЙ, А.А. ПЕТРОВСКИЙ
**МАСШТАБИРУЕМЫЕ АУДИОРЕЧЕВЫЕ КОДЕРЫ НА
ОСНОВЕ АДАПТИВНОГО ЧАСТОТНО-ВРЕМЕННОГО
АНАЛИЗА ЗВУКОВЫХ СИГНАЛОВ**

Петровский А.А., Петровский А.А. Масштабируемые аудиоречевые кодеры на основе адаптивного частотно-временного анализа звуковых сигналов.

Аннотация. Рассматриваются методы перцептуальной субполосной обработки звуковых сигналов с динамической трансформацией частотно-временного плана на основе пакетного дискретного вейвлет-преобразования (ПДВП), достоинством которых является то, что рост дерева осуществляется сверху вниз, без возвратов на меньшие масштабные уровни преобразования и необходимости построения полного дерева ПДВП, что соответствует концепции реализации масштабируемых аудиоречевых кодеров в реальном масштабе времени. Приводятся объективные оценки качества предлагаемых кодеров на основе методики *PEMO-Q* и сравнения с широко распространенными кодерами *Opus* и *Vorbis*, которые показывают, что реконструированный сигнал соответствует требованиям стандарта ITU-R PEAQ при высокой степени компрессии в 18 и более раз, не содержит артефактов: отношение мощности шума к порогу маскирования NMR_{total} меньше ≈ -9 дБ.

Ключевые слова: масштабируемый аудиоречевой кодер, пакет дискретного вейвлет-преобразования, согласованная подгонка.

1. Введение. Термин «мультимедиа» в настоящее время приобрел необычайную популярность. Это объясняется тем, что появилась возможность связывать в единое целое визуальную, акустическую и текстовую информацию, то есть сделан большой шаг от простых массовых продуктов, например, таких как сотовый телефон и MP3-проигрыватель, к амбициозным проектам — человеко-машинным интерфейсам. Стремительное внедрение цифровых технологий в такие приложения, как аудиокниги, системы радиовещания и телевидения, непрерывное увеличение передач через Интернет, информационное наполнение которых может быть различно и не ограничено только речью или музыкой, определило необходимость в объединенном аудиоречевом кодере с низкой скоростью цифрового потока, который одинаково хорошо работает со всеми типами звукового информационного наполнения [1-3].

В большинстве своем техника обработки аудиосигналов связана с тремя областями: компрессией, классификацией и защитой информации. Компрессия речевых сигналов традиционно базируется на определенных моделях речеобразования, в то время как в методах высококачественного кодирования аудиосигналов используется свойство шумового маскирования человеческого слуха [4-6]. Общая философия перцептуального кодера взаимосвязана с выбором метода частотно-временного анализа — банка цифровых фильтров [7]. Развитие этих работ в настоящее время

идет по пути построения перцептуальных субполосных аудиокодеров с постоянным частотно-временным планом, что приводит к высокой скорости цифрового потока для достижения высокого качества реконструированного сигнала [8]. Увеличение степени компрессии обуславливает применение техник частотно-временного анализа с динамически изменяемым частотно-временным планом для соответствующего фрейма обработки аудиосигнала, например, наиболее гибкого в смысле управления частотно-временным разрешением, пакета дискретного вейвлет-преобразования (ПДВП) [9]. Адаптивная частотно-временная аппроксимация аудиосигнала на основе ПДВП с перцептуально-оптимизированным частотно-временным планом позволяет исключить избыточность в сигнале, что обусловлено невосприимчивостью определенных частотных компонент человеком, вследствие маскирующего эффекта [10, 11].

В настоящее время для решения задачи компрессии звуковых сигналов часто применяются алгоритмы разреженной аппроксимации на основе согласованной подгонки (СП) [12, 13], заключающейся в поиске наилучшей проекции входного сигнала на избыточный словарь базисных функций — атомов [14]. В работе [15] представлен метод разреженной аппроксимации на основе СП со словарем, состоящим из функций Габора. Данный алгоритм позволяет достичь почти перфективной реконструкции входного сигнала, но при большом числе итераций. Кроме того, словарь атомов имеет фиксированную структуру, что влечет за собой увеличение его размера и вычислительной сложности алгоритма. В подходе [16] используется два словаря, переключение между которыми зависит от энергии остаточного сигнала. Это дает большую гибкость алгоритму и уменьшает его вычислительную сложность, однако параметр переключения между словарями фиксированный, что не всегда адекватно обрабатываемому фрейму входного сигнала. Гибридная модель представления аудиосигнала на основе трех частей: гармонической, шумовой и переходной применяется в масштабируемых аудиокодерах [12, 17]. Переходная составляющая параметризуется с помощью СП. Данный вариант кодирования требует трех различных алгоритмов для параметризации сигнала, что обуславливает большую вычислительную сложность метода. В [13, 18, 19] решается задача параметрического анализа аудиосигналов на основе разреженной аппроксимации с перцептуально-оптимизированным ПДВП-словарем вейвлет-коэффициентов, применение которого, как показано в [19], дает возможность построения универсального аудиоречевого кодера.

Цель данной работы — показать методы перцептуальной субполосной обработки звуковых сигналов на основе ПДВП с динамической трансформацией частотно-временного плана анализа для построения масштабируемых аудиоречевых кодеров.

2. Частотно-временной анализ с использованием пакетного дискретного вейвлет-преобразования.

2.1. Пакетное дискретное вейвлет-преобразование. Пусть $\{\psi_n(t) : n \in Z\}$ определяет множество структур деревьев ПДВП и пусть $E \in \{(l, n) : 0 \leq l \leq L\}$ — масштабный уровень, $0 \leq n \leq 2^l$ — номер узла на масштабном уровне представляет собой узлы дерева ПДВП, тогда отрезок $[0, 1)$ разделяется на диадические интервалы $I_{l,n} = [n2^{-l}, (n+1)2^{-l}]$, которые соответствуют специфическому множеству узлов E . В частности, $\{\psi_{l,n,k}(t) : (l, n) \in E, k \in Z\}$, где $\psi_{l,n,k}(t) \triangleq 2^{-l/2} \psi_n(2^{-l}t - k)$ является базовой формой в пространстве сигнала $\overline{\text{span}\{\psi_0(t - k) : k \in Z\}}$. Узел $(l, n) \in E$ дерева ПДВП ассоциируется с частотной полосой, у которой центральная частота и полоса пропускания приблизительно задаются следующими соотношениями: $f_{l,n} = 2^{-l}(GC^{-1}(n) + 0.5) \cdot f_s / 2$, $\Delta f_{l,n} = 2^{-l} \cdot f_s / 2$, где GC^{-1} — обратный код перестановок Грея, f_s — частота дискретизации сигнала. В отличие от дискретного вейвлет-преобразования, представляющего собой дерево декомпозиции области нижних частот, ПДВП это декомпозиция обеих частей дерева преобразования: области нижних и верхних частот [20].

На рисунке 1 (а, в, д, ж) показаны возможные варианты построения второй ступени ПДВП. Как видно из рисунка 1, структура дерева преобразования позволяет варьировать разрешающую способность в частотной и временной областях. Выполнение более детальной декомпозиции, например, в области нижних частот, приводит к увеличению разрешающей способности в области частот и ее уменьшению во временной области, в то время как менее детальная декомпозиция в области верхних частот, обеспечивает высокое разрешение во временной области и уменьшает частотное разрешение. Соответствующие частотно-временные планы показаны на рисунке 1 (б, г, е, з), где кружками отмечены коэффициенты преобразования $X_{l,n,k}(t)$. Здесь индексы имеют следующее назначение: l — уровень преобразования, n — номер узла на уровне l ; k — это временная позиция в блоке обработки. Анализ входного сигнала $x(t)$ на базе ПДВП с переменным частотно-временным разрешением (многомасштабный анализ [20]) выполняется на основе рекурсивных масштабной и вейвлет-функций. При цифровой реализации это коэффициенты цифровых фильтров с конечной импульсной характеристикой.

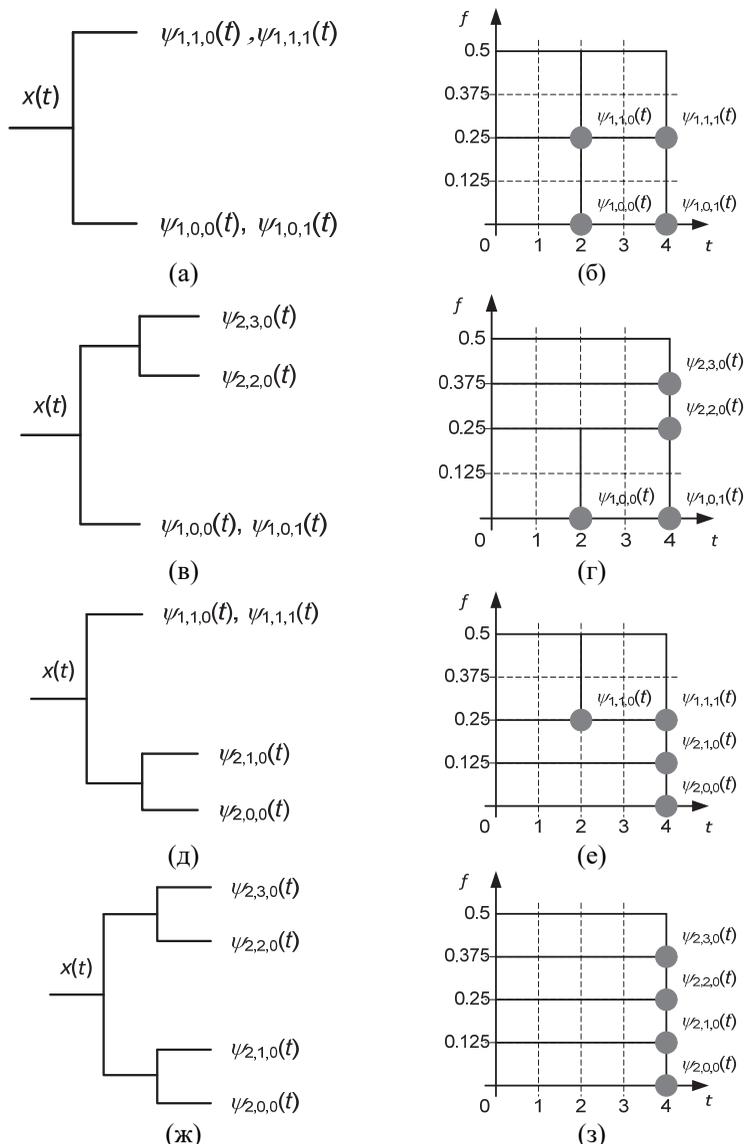


Рис. 1. Структуры деревьев ПДВП: (а), (в), (д), (ж) – возможные варианты построения второй ступени ПДВП; (б), (г), (е), (з) – частотно-временные планы соответствующие вариантам ПДВП

Таким образом, в отличие от дискретного преобразования Фурье (ДФФ) и банков ДПФ или косинусно модулированных фильтров [7] ПДВП обеспечивает двумерную развертку кодируемого звукового сигнала, при этом частота и время рассматриваются как независимые переменные. В результате появляется возможность анализировать свойства сигнала одновременно в частотном и временном пространствах.

2.2. ПДВП с управляемым временным сдвигом. Декомпозиция ПДВП может быть расширена в корень из двух раз путем варьирования временным сдвигом в каждом из узлов n структуры дерева ПДВП $l, n \in E$. Увеличение числа возможных комбинаций не изменяет древовидную структуру ПДВП. Временной сдвиг модифицирует ортогональные проекции ортонормальных масштабных функций на подпространстве $V_{l,n}$ [7, 21]. Данное изменение описывает как уже существующие, так и вновь образованные подмножества:

$$V_{l,n,m} \equiv \left\{ \psi_{l,2n+1,m}(t) = 2^{l/2} \sum_{k \in Z} \psi_{l,n,m}(2(t-m_c) - k) \right\}, \quad (1)$$

где $m_c = \{m, m + 2^{-l}\}$ — временной сдвиг, $l, n \in E$. Тройные индексы каждого узла (l, n, m) структуры дерева E ПДВП определяют номер узла n , уровень l , временной сдвиг m . Декомпозиция сигнала на основе ПДВП без временного сдвига, например, (рисунок 1в и 1г), а соответствующая ей структура и частотно-временной план ПДВП для выполнения декомпозиции с временным сдвигом показаны на рисунке 2, где структура ПДВП получается путем каскадного соединения блоков декомпозиций удовлетворяющих записи в выражении (1).

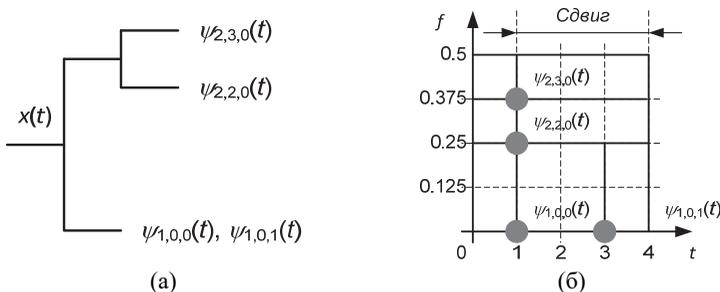


Рис. 2. Декомпозиция ПДВП с временным сдвигом: а) вариант построения второй ступени ПДВП; б) частотно-временной план соответствующий варианту ПДВП

Как видно из рисунка 2, структура дерева ПДВП позволяет не только варьировать разрешающую способность в частотной и временной областях, но и выполнять сдвиг по времени для более детальной декомпозиции входного сигнала [21, 22]. На рисунках 3 и 4 показаны структуры деревьев ПДВП и соответствующие частотно-временные планы для тестового сигнала, содержащего короткие импульсы тона с нелинейной частотной модуляцией для адаптивного ПДВП без временного сдвига и с временным сдвигом. Как видно из данного примера, адаптивный ПДВП с временным сдвигом лучше локализовал неопределенность (информативность) сигнала, чем без сдвига.

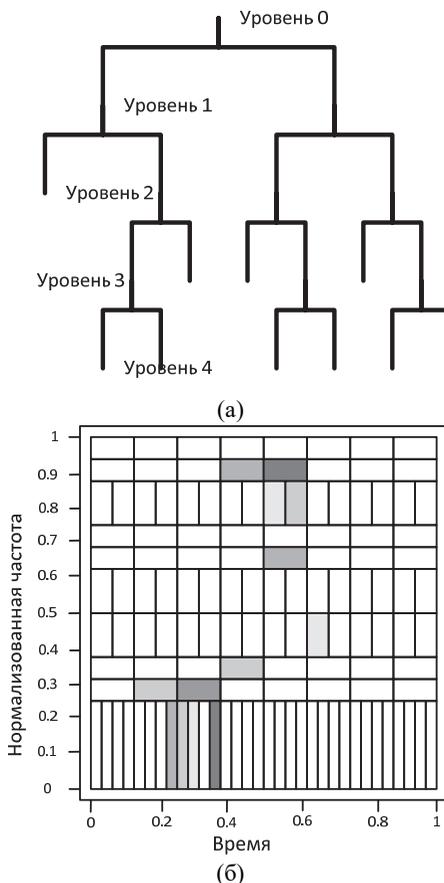
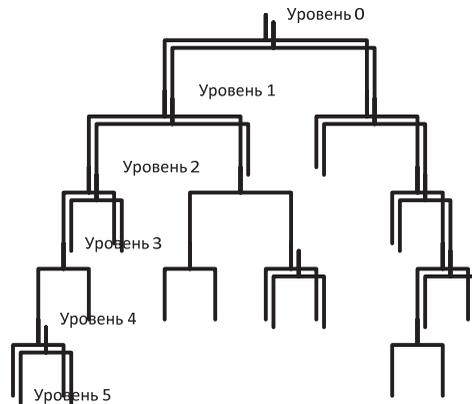
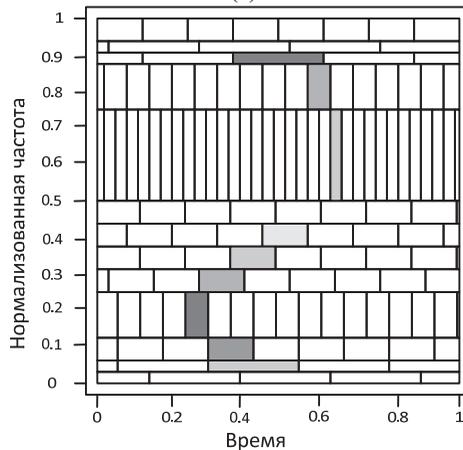


Рис. 3. Декомпозиция сигнала на основе ПДВП без временного сдвига:
(а) структура ПДВП; (б) частотно-временной план ПДВП



(а)



(б)

Рис. 4. Декомпозиция сигнала на основе ПДВП с временным сдвигом: (а) структура ПДВП; (б) частотно-временной план ПДВП

2.3. ПДВП согласованный с критической частотной шкалой.

Построение частотно-временного плана для перцептуальной обработки сигналов речи и звука, согласованного с критической частотной шкалой, при минимальном множестве узлов ПДВП, определенной глубине декомпозиции структуры дерева, заданной частоте дискретизации сигнала, минимальной ошибке аппроксимации критических частотных полос в области барков [11] осуществляется в соответствии с психоакустической моделью Zwicker [23], где:

— расстояние между центральными частотами соседних критических частотных полос в барках:

$$Z = F(f) = 13 \cdot \arctan(0.00076 \cdot f) + 3.5 \cdot \arctan\left(\left(\frac{f}{7500}\right)^2\right), \quad (2)$$

где f — частота в герцах, единица измерения в данном масштабе — 1 барк;

— ширина критических частотных полос в герцах.

$$CBW(f) = 25 + 75 \cdot (1 + 1.4 \cdot (f / 1000)^2)^{0.69}. \quad (3)$$

При этом, также вычисляются характеристики психоакустической модели восприятия человеком акустической информации, такие как пороги маскирования:

— абсолютный порог слышимости АТН (absolute threshold of hearing), частотная зависимость которого аппроксимируется выражением:

$$ATH_{SPL}(f) = 3.64 \cdot (10^{-3} \cdot f)^{-0.8} - 6.5 \cdot \exp(-0.6 \cdot (10^{-3} \cdot f - 33)^2) + 10^{-3} \cdot (10^{-3} \cdot f)^4, [\text{дБ}] \quad (4)$$

где f — частота в герцах;

— частотное маскирование (simultaneous masking), проявляющееся при воздействии маскера в течение некоторого времени одновременно на разных частотах сигнала;

— маскирование во временной области (temporal masking): если громкий звук маскирует следующий за ним слабый звук, то такое явление называется маскированием вперед (post-masking), которое может продолжаться от 5 до 300 мс. в зависимости от силы и длительности звука; маскирование назад (pre-masking), когда громкий звук маскирует звук, воспроизводимый до него, длительность которого составляет примерно до 20 мс.

Для того чтобы получить аппроксимацию шкалы критических частотных полос с помощью ПДВП, необходимо осуществить декомпозицию дерева ПДВП таким образом, чтобы расстояние между центрами одной субполосы и другой субполосы составляло 1 барк. Следует отметить, что ширина критических частотных полос $CBW(f)$ монотонно увеличивающаяся функция частоты (3). Для формирования низкоча-

стотных полос требуется интенсивная декомпозиция ПДВП в сравнении с характером изменения дерева ПДВП для аппроксимации высокочастотных полос.

Интегральная перцептуально взвешенная ошибка аппроксимации шкалы критических частотных полос деревом $(l, n) \in E_m$ ПДВП в области Барков может быть определена следующим образом:

$$Q_E = \frac{1}{L} \sum_{\substack{\text{для} \\ \forall (l, n) \in E_m}} \left[\widehat{CBW}_{Z_w}(Z) - \widehat{CBW}_{E_m}(Z_{(l, n)}) \right]^2 \cdot \widehat{W}(Z). \quad (5)$$

Здесь ширина критических частотных полос $\widehat{CBW}(Z)$ в Гц как функция центральных частот соседних критических частотных полос заданных в барках, то есть:

$$\widehat{CBW}(Z) = CBW(F^{-1}(Z)), [\Gamma u]. \quad (6)$$

$(\widehat{CBW}_{Z_w} Z)$ определяет шкалу критических частотных полос в модели Zwicker [23], $\widehat{CBW}_{E_m}(Z_{(l, n)})$ — аппроксимация критических частотных полос деревом ПДВП $(l, n) \in E_m$, центр $Z_{(l, n)}$ в барках полосы (l, n) дерева ПДВП E_m вычисляется для центральной частоты $f_{(l, n)}$, заданной в Гц, как $Z_{(l, n) \in E_m} = F(f_{(l, n)})$, где F — преобразование (2). Перцептуальная взвешивающая функция $\widehat{W}(Z)$, учитывающая определенные частотные свойства наружного и среднего уха, задает распределение ошибки аппроксимации шкалы критических частотных полос меньше в области средних частот по сравнению с низкочастотным и высокочастотным диапазонами, и определяется в шкале дБ, как функция частоты [24]:

$$W_{\text{дБ}}(f) = -0.6 \cdot 3.64 \cdot (10^{-3} \cdot f)^{-0.8} + 6.5 \cdot \exp(-0.6 \cdot (10^{-3} \cdot f - 3.3)^2) - 10^{-3} (10^{-3} \cdot f)^4, \quad (7)$$

а также $\widehat{W}(Z)$ может быть переопределена для шкалы барков, как:

$$\widehat{W}(Z) = \widehat{W}(F(f)) = W(F^{-1}(Z)) = W(f), \quad (8)$$

где $W(f) = 10 \left(\frac{W_{\text{дБ}}(f)}{20} \right)$. Минимизация ошибки Q_E может позволить автоматизировать процесс построения оптимального дерева

ПДВП $(l, n) \in E_{CB}$ для шкалы критических частотных полос. На рисунке 5 показано дерево ПДВП (Critical Band Wavelet Packet Decomposition (CB-WPD)), полученное эмпирически [22, 25, 26].

Дерево CB-WPD делит частотный диапазон $[0 - 22,05 \text{ кГц}]$ на 25 неравномерных полос $CBW(f)$ [26], то есть на 25 барков. Корневой узел $(l, n) = (0, 0)$ данного дерева соответствует всему частотному диапазону сигнала. Каждый внутренний узел дерева $(l, n) \in E$, названный узлом предка, делится на два потомка: 1-й потомок и 2-й потомок, ассоциируемые с высокочастотной и низкочастотной фильтрацией, выходные сигналы (вейвлет-коэффициенты) которых децимируются в соотношении 2:1:

$$X_{l, n, k}(t) = \langle x(t), \varphi_{l, n, k}(t) \rangle, (l, n) \in E_{CB}, k \in \mathbb{Z}, \quad (9)$$

где l — номер масштабного уровня преобразования ($0 \leq l \leq 8$), n — номер узла масштабного уровня преобразования, k — вейвлет-коэффициентов в полосе (узле (l, n) дерева E). Банк вейвлет-фильтров (CB-WPD: $(l, n) \in E_{CB}$), согласованный с критической шкалой частот восприятия акустической информации человеком, является предельной структурой для метода перцептуального кодирования сигнала звука [26-28].

На рисунках 6 и 7 показаны аппроксимации центральной частоты и ширины каждой частотной полосы критической шкалы частот деревом ПДВП, структура которого приведена на рисунке 5.

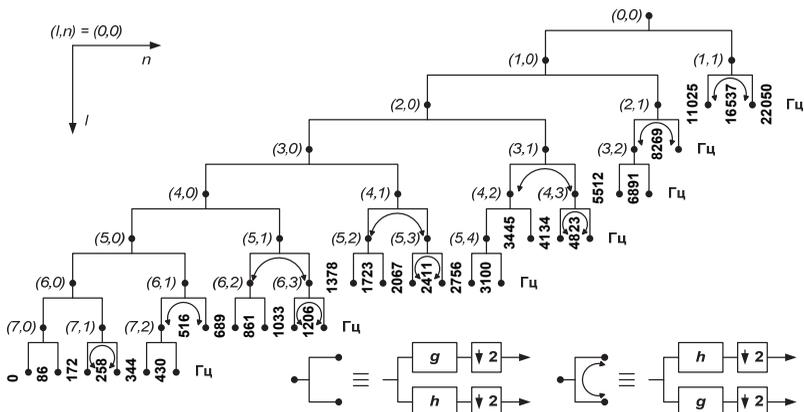


Рис. 5. Структура критического дерева ПДВП $CB - WPD : l = \overline{0, 8}$

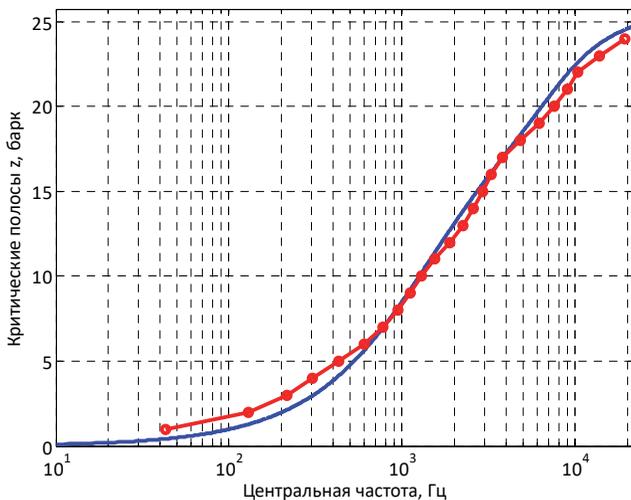


Рис. 6. Аппроксимация центральных частот $CB-WPD:l=\overline{0,8}$

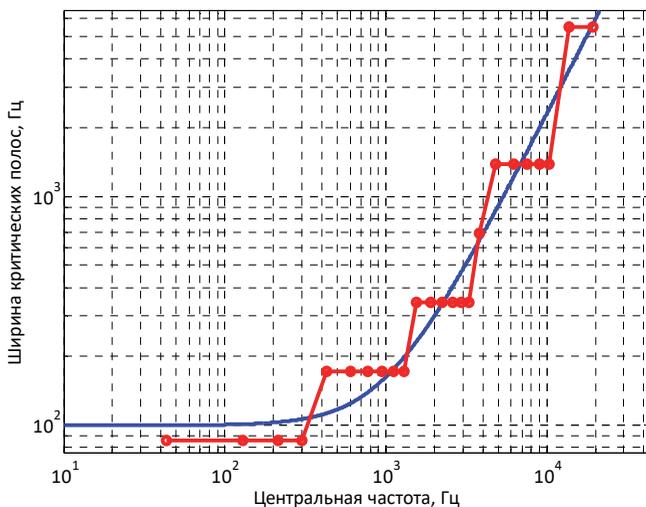


Рис. 7. Аппроксимация ширины критических частотных полос $CB-WPD:l=\overline{0,8}$

Частотно-временной план для структуры дерева ПДВП $CB-WPD:(l, n) \in E_{CB}, l = \overline{0,8}, f_s = 44.1 \text{ кГц}$ (рисунок 5) [26, 29] представлен на рисунке 8.

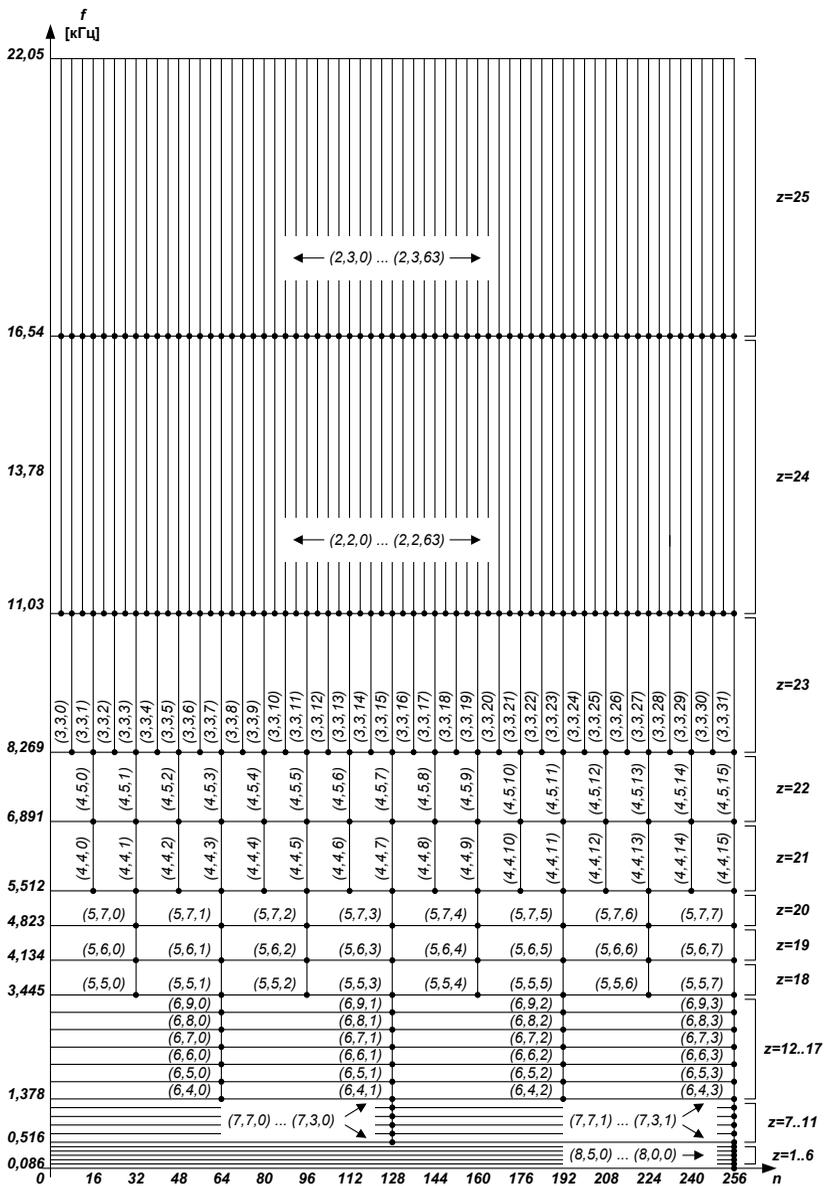


Рис.8. Частотно-временной план структуры дерева ПДВП

$$CB - WPD : l = 0.8, f_s = 44.1 \text{ кГц}$$

Ширина каждой клеточки есть длина фрейма, и определяется как $F_l = 2^l \cdot (F_{min} = 2 \text{ отсчетов и } F_{max} = 256 \text{ отсчетов})$. Следовательно, длина анализируемого окна равна $W = (P-1)(F_{l-1})+1$ отсчетов. Для первого уровня $l = 1$ преобразования определяющей является область верхних частот и длина окна $W = 40$ отсчетов при длине фильтра прототипа $P = 40$. Для уровня $l = 8$ преобразования наибольшая частотная разрешающая способность в области нижних частот, а окно $W = 9946$ отсчетов.

3. Динамическая трансформация частотно-временного плана. В перцептуальном кодировании аудиосигналов выбирается такая декомпозиция ПДВП, при которой минимизируется скорость передачи с сохранением высокого качества восприятия человеком декодированного сигнала. Декомпозиция «лучшего» дерева преобразования выбирается как можно ближе к шкале барков, то есть к шкале критических частотных полос, а банк вейвлет-фильтров, согласованный с критической шкалой частот восприятия акустической информации человеком, является предельной структурой для метода перцептуального кодирования аудио сигнала.

На основе введенной функции стоимости можно определить наилучший путь по этому дереву [30]. Если исходный блок вейвлет-фильтров был ортогональным, то и схема, соответствующая любой конфигурации, будет ортогональной, так как она есть не что иное, как каскадное соединение ортогональных блоков, то есть получается базис, адаптированный к сигналу. В работе [26] предложен метод роста дерева ПДВП или, другими словами, динамическая трансформация частотно-временного плана, позволяющий определить субоптимальную структуру декомпозиции ПДВП. Динамическая трансформация алгоритма ПДВП осуществляется в процессе работы ПДВП-кодера, то есть «на ходу».

Стоимостная функция $PE_{l,n}$ декомпозиции узлов $(l, n) \in E_j$ дерева ПДВП (роста структуры ПДВП) определяется как перцептуальная энтропия узла $(l, n) \in E_j$ и показывает требуемое число двоичных разрядов для кодирования звукового сигнала в частотной полосе, определяемой узлом (l, n) . Функция $PE_{l,n}(l, n) \in E_j$ представляет собой функ-

цию перцептуальной энтропии Джонсона [10], однако, в отличие от известных решений, вычисляемую для действительных коэффициентов и в вейвлет-области для текущего дерева E_j ПДВП:

$$PE_{l,n} = \sum_{k=0}^{K_{l,n}-1} \log_2(2[\text{int}(SMR_{l,n,k})] + 1), \left[\frac{\text{бит}}{(l,n)} \right],$$

где $(l,n) \in E_j$, $k \in \mathbb{Z}$, SMR — отношение среднеквадратического значения вейвлет-коэффициентов $X_{l,n,k}$ в полосе узла (l,n) дерева E_j к соответствующему маскирующему порогу $T_{l,n}$, равномерно распределенному между $K_{l,n}$ коэффициентами $X_{l,n,k}$, $k = \overline{1, K_{l,n}}$ узла (l,n) , определяется следующим образом:

$$SMR_{l,n,k} = |X_{l,n,k}| / \sqrt{12 \cdot T_{l,n} / K_{l,n}}.$$

где знаменатель $\sqrt{12 \cdot T_{l,n} / K_{l,n}}$ представляет собой максимальный шаг квантователя $\Delta_{l,n}$ вейвлет-коэффициентов в узле $(l,n) \in E_j$, а величина $SMR_{l,n,k}$ задает минимальное количество уровней квантования.

Меру информативности структуры дерева ПДВП предлагается конструировать следующим образом:

$$H_{E_i} = - \sum_{\forall (l,n) \in E_i} \sum_k \frac{|X_{l,n,k}|}{\sum_{\forall (l,n) \in E_i} |X_{l,n,k}|} \ln \left(\frac{|X_{l,n,k}|}{\sum_{\forall (l,n) \in E_i} |X_{l,n,k}|} \right),$$

где $X_{l,n,k} \in (l,n)$ — коэффициенты узла (l,n) дерева E_i , $(l,n) \in E_j$, $k \in \mathbb{Z}, i = 1 \dots 8$. Данная стоимостная функция характеризует изменение во времени информативности дерева E_i ПДВП, отсюда и название временная энтропия вейвлет-коэффициентов H_E (wavelet time entropy), и представляет собой сумму энтропий $H_{l,n}$ вейвлет-коэффициентов $X_{l,n,k}$ в узлах (l,n) дерева E_i :

$$H_{l,n} = - \sum_k \frac{|X_{l,n,k}|}{\sum_{(l,n) \in E_i} |X_{l,n,k}|} \ln \left(\frac{|X_{l,n,k}|}{\sum_{(l,n) \in E_i} |X_{l,n,k}|} \right), \quad (10)$$

где $X_{l,n,k} \in (l,n)$ — коэффициенты узла (l,n) дерева E_i , $(l,n) \in E_j$, $k \in \mathbb{Z}, i=1...8$. Правило отбора наилучшей декомпозиции для узла базируется на выборе той декомпозиции значение энтропии вейвлет-коэффициентов, которой будет меньше. Минимизация стоимостной функции H_E ведет к сокращению неопределенности и соответственно к увеличению информативности дерева ПДВП, описывающего входной сигнал. Декомпозиция ПДВП, то есть «рост» дерева преобразования или, другими словами, динамическая трансформация частотно-временного плана, может осуществляться на основании следующего алгоритма [22, 26].

Алгоритм 1. Алгоритм динамической трансформации частотно-временного плана на основе перцептуального критерия и оценки временной энтропии вейвлет-коэффициентов.

Пусть решение о декомпозиции узла (l,n) дерева E_i ПДВП будет обозначаться как $split(l,n)$, где l — уровень декомпозиции, то есть масштабный уровень преобразования, а n есть n -й узел на уровне l . Пусть текущий узел (предок) будет (l,n) , а его потомки определяются как $(l+1, 2n)$ и $(l+1, 2n+1)$, где $l=0,1,2,3..., n=0,1,2,3...$

Шаг 1. Пусть $l=0$, $split(l,n) = YES$, то есть задан корневой узел $(0,0)$ дерева преобразования $E_0 = E_j$, где $j=0$ — входной фрейм звукового сигнала, перцептуальная энтропия которого равна $PE_{0,0}$ и информативность дерева равна $H_{E_0} \cdot PE_{1,2,n}$

Шаг 2. Осуществляется декомпозиция узлов предков (l,n) входного сигнала на основе банка из двух ортонормальных вейвлет-фильтров с временным сдвигом 2^0 и без него 2^1 .

Шаг 3. Вычисление информативности $H_{l,n}$ в каждом узле $(l,n,2^0)$ и $(l,n,2^1)$. Выбор узла удовлетворяющего требованию:

ЕСЛИ $H_{l,n,2^0} > H_{l,n,2^1}$,

ТОГДА $(l,n) = (l,n,2^1)$

ИНАЧЕ $(l,n) = (l,n,2^0)$

Шаг 4. $l=l+1, j=j+1$.

ЕСЛИ $l-1 >$ максимального масштабного уровня предельного дерева $CB-WPD$,

ТОГДА STOP — конец роста дерева ПДВП.

Шаг 5. Вычисляется перцептуальная энтропия $PE_{l,2n}$ и $PE_{l,2n+1}$ в узлах декомпозиции $(l, 2n)$ и $(l, 2n + 1)$ соответственно для всех n .

Шаг 6. Для каждого узла n уровня l рост дерева E_i ПДВП осуществляется следующим образом:

ЕСЛИ $PE_{l,n} \geq PE_{l+1,2n} + PE_{l+1,2n+1}$,

ТОГДА $split(l, n) = YES$,

ИНАЧЕ $split(l, n) = NO$,

Шаг 7. Оценивается информативность H_{E_j} дерева E_i :

ЕСЛИ $H_{E_j} > H_{E_{j-1}}$,

ТОГДА STOP — конец роста дерева, результирующее дерево ПДВП E_{j-1} .

Шаг 8. Переход к шагу 2.

Достоинством алгоритма 1 является то, что рост дерева осуществляется сверху вниз, без возвратов на меньшие масштабные уровни преобразования и необходимости построения полного дерева ПДВП [26], что соответствует концепции обработки сигналов в реальном времени [27]. Данный алгоритм роста дерева ПДВП позволяет определить субоптимальную структуру декомпозиции ПДВП при минимальном числе бит на отсчет звукового сигнала без воспринимаемых на слух искажений, вносимых в процессе кодирования входного сигнала. На рисунке 9 выполнена наглядная иллюстрация конструирования дерева ПДВП на уровне $l = 2$.

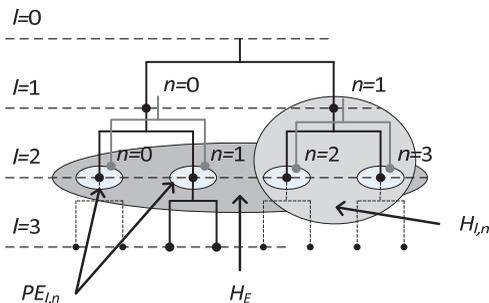


Рис. 9. Иллюстрация конструирования эффективного дерева ПДВП

Процесс роста дерева ПДВП E , согласно алгоритму динамической трансформации частотно-временного плана, и формируемый им частотно-временной план в соответствии со структурой дерева ПДВ E для некоторого фрейма входного сигнала показан на рисунке 10.

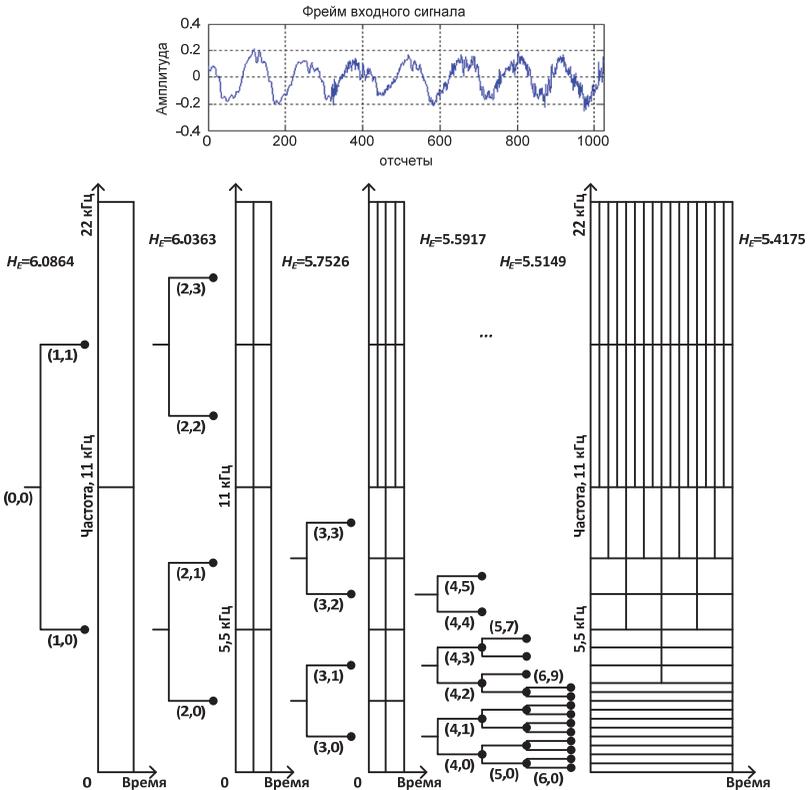


Рис. 10. Формирование дерева ПДВП и частотно-временного плана

Здесь на каждом масштабном уровне l дерева ПДВП E определяется его информативность (частотно-временного разрешения) H_E , и оценивается перцептуальная энтропия $PE_{l,n}$ для каждого узла n уровня l дерева ПДВП E . В соответствии с алгоритмом принимается решение о дальнейшем росте дерева ПДВП или о нахождении субоптимального частотно-временного разрешения для входного фрейма аудиосигнала в соответствии со структурой дерева ПДВП. Дерево $E_{1,6}$ наиболее полно

удовлетворяет требованиям обработки, то есть значение меры информативности H_{E_j} минимально. По мере увеличения декомпозиции дерева ПДВП порог маскирования — порог едва различимых искажений приближается к такому же порогу, который был построен для предельного дерева преобразования $CB - WPD$ (рисунок 11).

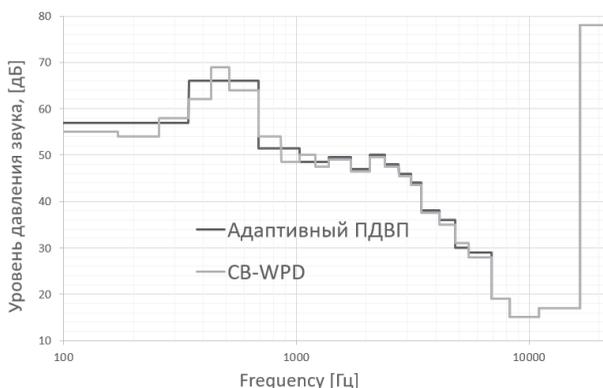


Рис. 11. Порог маскирования аудиосигнала

4. Параметрическое описание звуковых сигналов на основе согласованной подгонки.

4.1. Разреженная аппроксимация сигнала с полным ПДВП-словарем вейвлет-коэффициентов. Использование частотно-временного преобразования на основе алгоритма согласованной подгонки со словарем частотно-временных функций позволяет получить разреженную частотно-временную аппроксимацию аудиосигналов, а следовательно, лучшую, чем на основе вейвлет-преобразования и ПДВП локализацию нестационарности в сигнале [14, 31]. Это в конечном итоге позволит уменьшить цифровой поток в схеме компрессии [12, 13, 19].

Положим, что $D = (g_\gamma)_{\gamma \in \Gamma}$ — есть полный словарь частотно-временных функций, в дальнейшем называемых атомами, относительно которых параметризуется частотно-временные функции, например, масштабный фактор, частота, время индексируются $\gamma \in \Gamma$, где Γ — случайное множество индексов. Пусть дана некоторая частотно-временная функция g_γ , параметризованная относительно индекса γ , то

наилучшая возможная аппроксимация $x(n)$ получается как ортогональная проекция аудио сигнала $x(n)$ на стянутое подпространство частотно-временной функции g_γ . Декомпозиция аудиосигнала может быть представлена как $x(n) = \langle x(n), g_\gamma \rangle \cdot g_\gamma + r(n)$, где $r(n)$ — сигнал-остаток после вычитания проекции $\langle x(n), g_\gamma \rangle \cdot g_\gamma$. На основании ортогональности $r(n)$ и g_γ , следует, что $\|x(n)\|^2 = \left| \langle x(n), g_\gamma \rangle \right|^2 + \|r(n)\|^2$. Данная декомпозиция делается для любого и каждого элемента словаря и лучшее согласование будет найдено путем выбора элемента g_γ , для которого $\|r(n)\|$ минимальна, или, что эквивалентно, для которого $\left| \langle x(n), g_\gamma \rangle \right|$ — максимальный. Математически это можно сформулировать следующим образом $\gamma' = \arg \sup_{\gamma \in \Gamma} \left| \langle x(n), g_\gamma \rangle \right|$. Таким образом, сигнал $x(n)$ проецируется на избыточный словарь частотно-временных функций со всеми возможными комбинациями масштабов, переходных и модуляционных параметров. Когда $x(n)$ реальный и дискретный, как аудиосигнал, тогда используется словарь реальных и дискретных функций. Благодаря избыточности словаря это дает возможность максимально гибко выбрать лучшую частотно-временную функцию для локальной структуры сигнала (локальная оптимизация). Максимальная гибкость модели сигнала позволяет осуществлять компактную аппроксимацию сигнала с максимально возможной точностью минимально возможным числом частотно-временных функций.

Структура алгоритма разреженной аппроксимации, используя алгоритм согласованной подгонки со словарем на основе полного дерева ПДВП, показана на рисунке 12 [31]. Для N отсчетов входного сигнала словарь D содержит N векторов $g_\gamma \in D$, где $\gamma = (l, n, k)$ индекс вектора для каждого уровня $0 < l, \log_2(N)$ узла $0 < n < 2^l$ и коэффициента $0 < k < 2^l N$ декомпозиции. Каждый вектор имеет частотно-временную локализацию схожую с дискретной функцией окна, которая увеличивается на 2^l и имеет центр в $2^l(k + 1/2)$.

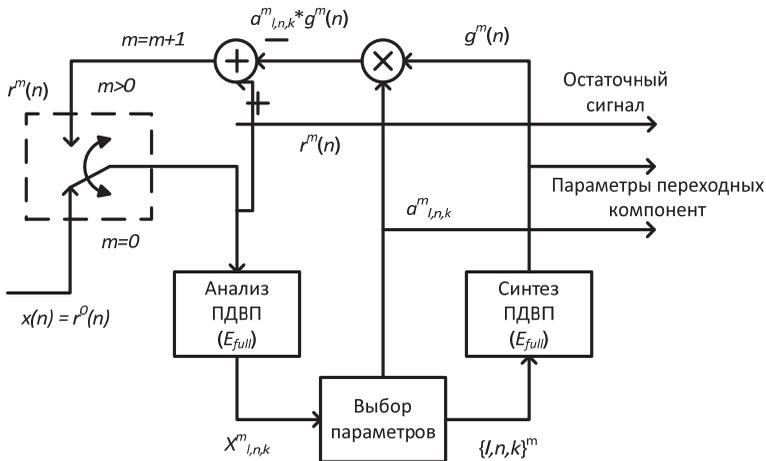


Рис. 12. Структура алгоритма согласованной подгонки со словарем, построенным на основе полного дерева ПДВП

Алгоритм согласованной подгонки реализуется итерационным повторением следующих шагов:

Шаг 1. Декомпозиция сигнала-остатка $r^m(n)$ банком фильтров на основе полного дерева ПДВП E_{full} , для $m = 1$ итерации $r^m(n) = x(n)$;

Шаг 2. Выбор наиболее значимого вейвлет-коэффициента X_γ , то есть коэффициента с абсолютным максимальным значением весового коэффициента a_γ ;

Шаг 3. Наиболее значимому вейвлет-коэффициенту X_γ ставиться в соответствие атом g_γ из словаря D ;

Шаг 4. Формирование результирующего вектора $g_\gamma^m(n)$ декомпозиции выполняется при помощи обратного ПДВП преобразования на основании структуры дерева ПДВП E_{full} ;

Шаг 5. Получение результирующего сигнала, как умножение результирующего вектора $g_\gamma^m(n)$ на весовой коэффициент a_γ^m ;

Шаг 6. Получение сигнала-остатка $r^{m+1}(n)$ путем вычитания результирующего сигнала $a_\gamma^m \cdot g_\gamma^m(n)$ из остаточного сигнала $r^m(n)$.

Процесс повторяется и на следующей итерации алгоритма входным сигналом будет сигнал-остаток с предыдущей итерации. На m -й итерации сигнал-остаток $r^m(n)$ вычисляется как:

$$r^m(n) = \begin{cases} x(n) & m = 0 \\ r^{m+1}(n) + a_\gamma^m \cdot g_\gamma^m(n) & m \neq 0 \end{cases} \quad (1)$$

где a_γ^m весовой коэффициент оптимального вектора $g_\gamma^m(n)$ на m -й итерации, а γ индекс словаря D на m -й итерации. Оптимальным вектором считается тот вектор, у которого получается наибольшее значение произведения с сигналом-остатком $\langle x(n), g_\gamma \rangle$.

Минимизация вычислительной сложности метода напрямую связана также со структурой и размером словаря. Количество атомов словаря D находится в прямой зависимости от числа уровней дерева ПДВП и длины анализируемого фрейма входного сигнала. С точки зрения, например, перцептуального восприятия акустической информации человеком, наиболее критичные компоненты сигнала расположены в области нижних, нежели в области верхних частот, что сказывается на качестве выбора перцептуально значимых полос для анализа.

4.2. Согласованная подгонка с перцептуально оптимизированным словарем частотно-временных функций. В [32] предлагается задачу параметрического анализа аудиосигналов решать на основе разреженной аппроксимации с перцептуально оптимизированным ПДВП-словарем вейвлет-коэффициентов. Алгоритм согласованной подгонки требует значительных вычислительных затрат от итерации к итерации при формировании субоптимального решения, которое в конце концов может быть не применимо в некоторых приложениях обработки звука и речи. Задача поиска оптимального словаря D , построенного на основе ПДВП, сводится к поиску лучшей структуры дерева декомпозиции ПДВП [30]. Формирование оптимального набора частотно-временных функций в словаре выполняется на основе применения перцептуально адаптированного к текущему фрейму входного сигнала ПДВП, что позволяет уменьшить размер формируемого словаря и сделать его динамическим, то есть зависимым от структуры аудиосигнала [25, 32].

Структурная схема параметрического анализа аудиосигналов на основе алгоритма согласованной подгонки с перцептуально адаптированным к фрейму входного сигнала ПДВП словарем состоит из двух частей:

— первая часть — схема оптимизации структуры дерева ПДВП под фрейм входного сигнала согласно алгоритму динамической трансформации частотно-временного плана входного сигнала на основе перцептуальных критериев [26];

— вторая часть представляет собой схему модифицированного алгоритма согласованной подгонки с перцептуально-оптимизированным ПДВП словарем.

Структура модифицированного алгоритма согласованной подгонки с динамически оптимизированным словарем показана на рисунке 13.

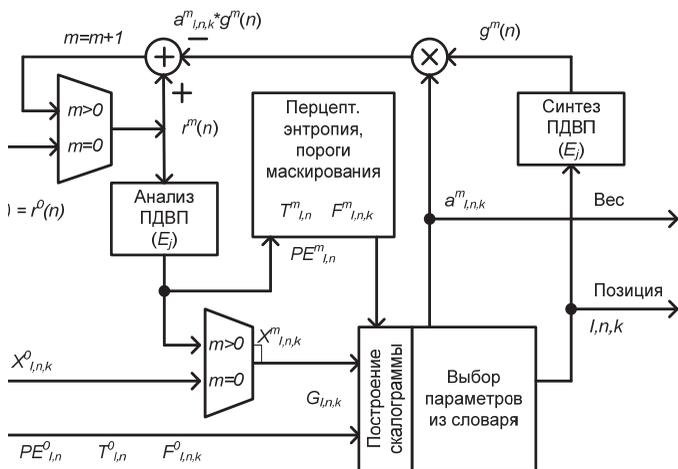


Рис. 13. Схема параметрического анализа аудиосигналов на основе разреженной аппроксимации с перцептуально оптимизируемым ПДВП-словарем

На итерации $m = 0$ коммутация выполнена так, чтобы остаточный сигнал $r^m(n)$ стал равным фрейму входного сигнала $x(n)$, а вейвлет-коэффициенты $X^0_{l,n,k}$ попали напрямую в блок скалярного произведения и скалограммы возбуждения соответственно. Первоначальное значение перцептуальной энтропии $PE^0_{l,n}$, порогов маскирования $T^0_{l,n}$ и временных маскеров $F^0_{l,n,k}$ берется из схемы декомпозиции входного сигнала адаптивным ПДВП. В блоке скалярного произведения и скалограммы возбуждения выполняется формирование скалограммы возбуждения для текущего входного фрейма сигнала $G^0_{l,n,k}$, отбор по ней перцептуально значимого вейвлет-коэффициента, формирование весового параметра атома

$\alpha_{l,n,k}^0$ и позиции вейвлет-коэффициента. В блоке обратного ПДВП выполняется синтез единичной атомарной функции $g^0(n)$ на основании знания позиции вейвлет-коэффициента в структуре дерева E_j ПДВП, затем умножение ее на весовой коэффициент $\alpha_{l,n,k}^0$. Остаточный сигнала, он же входной сигнал для следующей итерации, формируется вычитанием результирующего сигнала $\langle \alpha_{l,n,k}^0 \cdot g^0(n) \rangle$ из сигнала $r^0(n)$. На каждой последующей итерации $m > 0$ пороги маскирования $T_{l,n}^m$ и временные маскеры $T_{l,n,k}^m$ вычисляются для формирования скалограммы возбуждения $G_{l,n,k}^m$, на основании которой ведется отбор перцептуально-значимых компонент сигнала [17]. Работа схемы параметрического анализа аудиосигналов на основе разреженной аппроксимации с перцептуально оптимизируемым ПДВП-словарем выполняется согласно следующему алгоритму [32].

Алгоритм 2. Алгоритм согласованной подгонки с динамически оптимизируемым ПДВП словарем на основе психоакустического критерия.

Исходные данные алгоритма согласованной подгонки: E_j — оптимизированная структура ПДВП для фрейма входного сигнала $x(n)$; $T_{l,n}$ — порог маскирования для каждого оконечного узла $(l,n) \in E_j$ ПДВП; $F_{l,n}$ — временный маскер для каждого оконечного узла $(l,n) \in E_j$ ПДВП; $G_{l,n,k}$ — скалограмма возбуждения соответствующая фрейму входного сигнала $x(n)$.

Шаг 1. Установить номер итерации $m = 0$.

Шаг 2. Разместить $G_{l,n,k}$ и установить $G_{l,n,k} = 0$ для всех l,n,k в соответствии со структурой дерева ПДВП E_j .

Шаг 3. Вычислить $PE_{l,n}^m$ для всех узлов (l,n) , используя $T_{l,n}$;

ЕСЛИ $PE_{l,n}^m = 0 \forall (l,n,k) \in E_j$

ТОГДА СТОП;

ЕСЛИ $PE_{l,n}^m \neq 0$,

ТОГДА $X_{l,n}^m = 0$ для $k = \{0, K_{l,n} - 1\}$ узла (l, n) .

Шаг 4. Выбрать из $X_{l,n,k}^m$ значимые коэффициенты $X_{l,n,k}^{*m}$, имеющие наибольший вес возбуждения.

Шаг 5. Создать скалограмму возбуждения, соответствующую моделируемому сигналу, используя $T_{l,n}$ и $F_{l,n}^{m-1}$ для выполненной итерации и каждого нового значимого коэффициента $X_{l,n,k}^{*m}$.

Шаг 6. Выбрать вес $\alpha_{l,n,k}^m = X_{l,n,k}^{*m}$, который улучшает соответствие между скалограммами (исходной и моделируемой).

Шаг 7. Получить позицию коэффициента в структуре дерева ПДВП: $l^* = l, n^* = n, k^* = k$.

Шаг 8. Установить 1 в позиции $(l^*, n^*, k^*) : G_{l^*, n^*, k^*} = 1$.

Шаг 9. Синтезировать атом $g^m(n)$ из G_{l^*, n^*, k^*} , используя обратный ПДВП со структурой дерева E_j , соответствующей ПДВП-словарю.

Шаг 10. Вычислить сигнал-остаток $r^m(n)$ из $g^m(n)$ и $\alpha_{l,n,k}^m$ в соответствии с (11).

Шаг 11. Применить оптимизированный на основе текущего фрейма ПДВП со структурой дерева E_j к сигналу-остатку $r^m(n)$.

Шаг 12. Увеличить номер итерации $m = m + 1$.

Шаг 13. Перейти к шагу 2.

Число итераций алгоритма 2 определяется числом перцептуально значимых коэффициентов ПДВП в сигнале остатке. В алгоритм введена процедура расчета перцептуальной энтропии $PE_{l,n}$ для каждой полус (l, n), на основании которой ведется селекция перцептуально-значимых частотных полос. Знание не только частотной составляющей сигнала, то есть порогов маскирования $T_{l,n}$, но и временного маскера $F_{l,n,k}$ позволяет обеспечить оптимальный выбор значимых компонент сигнала. Данный подход позволит уменьшить число итераций алгоритма согласованной подгонки и производить выбор только значимых компонент сигнала, что будет соответствовать оптимальным векторам

декомпозиции. Однако поиск наилучшей аппроксимации анализируемого входного сигнала $x(n)$ векторами $g_{l,n,k}$ из словаря D , построенного на основе перцептуально-адаптированного ПДВП, сложная в вычислительном плане задача.

На рисунке 14 показан пример аппроксимации аудиосигнала на основе предложенного метода. Функция Db20 из семейства Добеши с 40 коэффициентами использовалась для анализа входного сигнала в схеме перцептуально адаптированного ПДВП. Скалограммы возбуждения для входного сигнала (рисунок 14а) и синтезированного сигнала на основе 5 атомов (рисунок 14б) приведены соответственно на рисунке 15а и рисунке 15б, где перцептуально значимые компоненты показаны как положительные элементы.

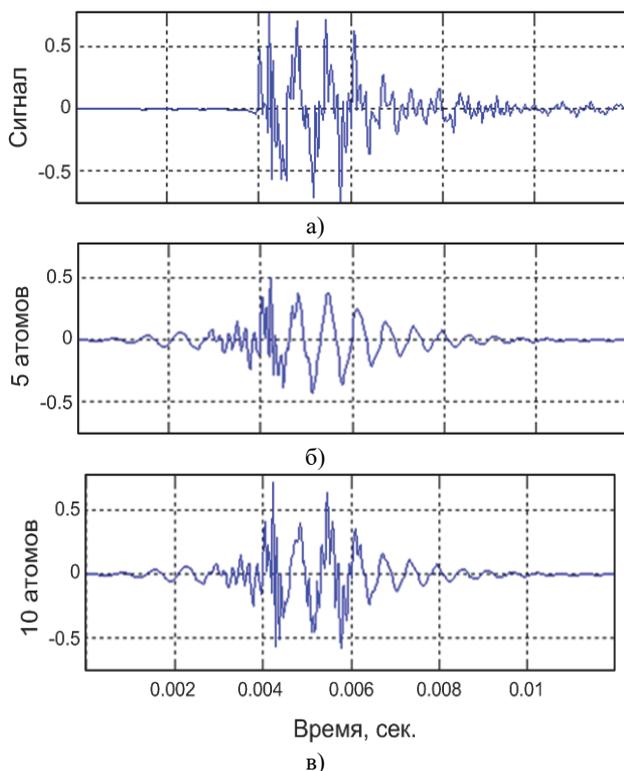
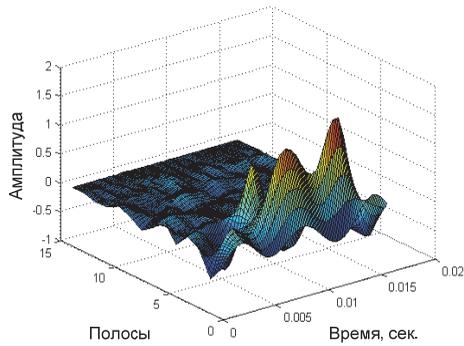
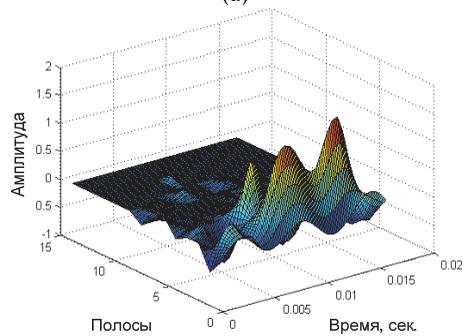


Рис. 14. Пример разреженной аппроксимации с перцептуально оптимизированным ПДВП словарем вейвлет-функций: (а) исходный сигнал; (б) 5 атомов; (в) 10 атомов



(а)



(б)

Рис. 15. Скалограммы возбуждения: (а) входного сигнала; (б) синтезированного сигнала, на основе 5 атомов

Сравнение сходимости алгоритма согласованной подгонки на основе трех различных словарей аппроксимирующих функций показано на рисунке 16.

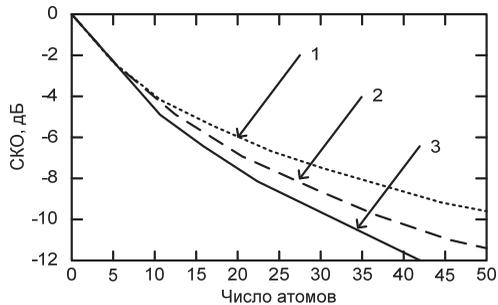


Рис. 16. Среднеквадратичное отклонение ошибки аппроксимации:

1 – затухающие синусоиды; 2 – вейвлет-пакет с избыточным словарем;

Алгоритм 2 согласованной подгонки на базе перцептуально оптимизированного ПДВП-словаря вейвлет-функций [32] имеет меньшее среднеквадратичное отклонение ошибки аппроксимации, по сравнению с алгоритмами на основе ПДВП-словаря с полным набором вейвлет-функций [31] и с использованием затухающих синусоид [33].

Таким образом, в алгоритме 2 формирование оптимального набора частотно-временных функций в словаре выполняется на основе применения перцептуально адаптированного к текущему фрейму входного сигнала ПДВП, что позволяет уменьшить размер формируемого словаря и сделать его динамическим, то есть зависимым от структуры фрейма аудиосигнала. Преимуществом алгоритма 2 разреженной аппроксимации с перцептуально оптимизированным ПДВП-словарем вейвлет-коэффициентов является высокая скорость сходимости и минимизация перцептуальных искажений, обусловленная построением перцептуально оптимизированного частотно-временного плана соответствующей декомпозиции вейвлет-пакета текущего сигнала-остатка для выбора оптимальных частотно-временных функции для каждой итерации подгонки. Более того, существует точный психоакустический критерий остановки описанной процедуры.

5. Параметрический аудиоречевой кодер с разреженной аппроксимацией сигнала и перцептуально-оптимизированным словарем частотно-временных функций.

5.1. Структура параметрического аудиоречевого кодера.

Структура параметрического перцептуального аудиоречевого кодера показана на рисунке 17, который состоит из блока динамической трансформации частотно-временного плана на базе алгоритма 1 (отмечен пунктирной линией) и блока разреженной аппроксимации сигнала на основе алгоритма 2, согласованной подгонки с динамически оптимизированным ПДВП словарем частотно-временных функций (серая область, выделенная пунктирной линией). Результатом адаптивного анализа являются вейвлет-коэффициенты $X_{l,n,k}$, трехмерная скалограмма возбуждения обрабатываемого фрейма сигнала, построенная на их основе, и структура дерева ПДВП E_j . Перцептуально значимые вейвлет-коэффициенты $X_{l,n,k}$, определенные в алгоритме согласованной подгонки и их позиции в дереве ПДВП из блока разреженной аппроксимации сигнала поступают в блок квантования и кодирования. В [22] разработано правило оптимального распределения битов при квантовании

вейвлет-коэффициентов в кодере с учетом реконструкции сигнала в декодере при переменном коэффициенте децимации в каналах банков фильтров кодера и декодера. Так, «эффект просачивания» энергии шума квантования в смежные полосы банка фильтров декодера не может пренебрегаться без внесения ущерба в качество восстановленного аудиосигнала. Показано в [22], что выбирая достаточно большой порядок (20-й и более) вейвлет-функции (высокую частотную избирательность канальных фильтров), можно ограничиться одинаковыми фильтрами как в анализирующем (кодере), так и синтезирующем (декодере) банках фильтров.

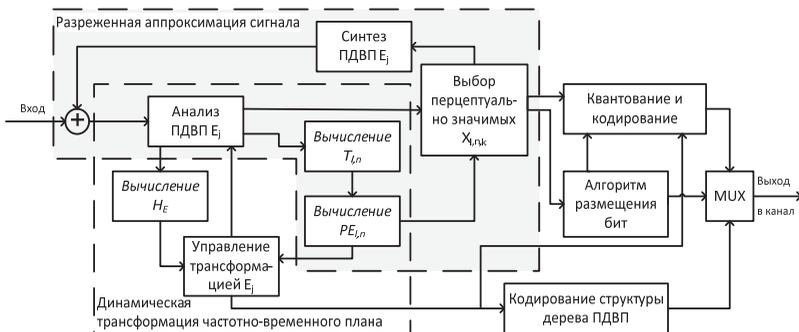


Рис. 17. Структура параметрического аудиоречевого кодера

Структура закодированных перцептуально значимых компонент — вейвлет-коэффициентов $X_{l,n,k}$ такова, что на стороне декодера после деквантования, параметры записываются в уже подготовленную структуру критического дерева ПДВП – *CB WPD* и выполняется синтез сигнала на основе обратного ПДВП – *CB WPD*. Структура параметрического аудиоречевого декодера показана на рисунке 18. Работа декодера выполняется в следующем порядке: входная информация декодируется и восстанавливается в блоке декодирования и восстановления. Полученные параметры содержат значения перцептуально значимых компонент — вейвлет-коэффициентов $X_{l,n,k}$ и информацию о их местоположении в структуре критического дерева ПДВП – *CB WPD*. Синтезированные фреймы сигнала умножаются на треугольное окно и складываются для формирования реконструированного сигнала $\hat{x}(n)$.

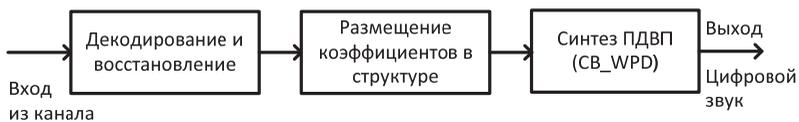


Рис. 18. Структура параметрического аудиоречевого декодера

Масштабирование цифрового потока осуществляется выбором условия остановки разреженной аппроксимации сигнала на основе алгоритма 2 согласованной подгонки с динамически оптимизированным ПДВП словарем частотно-временных функций. В данном кодере используются два варианта: фиксированное число атомов, либо определенный порог отобранной алгоритмом согласованной подгонки энергии или по уровню энергии остаточного сигнала $r^m(n)$. Для условия остановки алгоритма в виде фиксированного числа атомов задается определенное число итераций m , по достижению которого прекращается работа алгоритма кодирования, и реконструкция сигнала осуществляется на основе отобранных атомов. Это дает дополнительные возможности масштабирования цифрового потока кодера, так как количество атомов для каждого фрейма может быть различным, следовательно, есть возможность динамического добавления дополнительных пакетов атомов от фрейма к фрейму при наличии возможности канала передачи данных либо по запросу потребителя. Таким образом, этот факт дает возможность выстроить многоуровневую структуру аудиокодера относительно качества реконструированного сигнала. В проводимых экспериментах, за базовый уровень, при котором искажения восстановленного сигнала не являются раздражающими (на основе объективной оценки качества), бралось значение в 200 атомов. Добавляя к базовому количеству атомов дополнительные пакеты (в тестах — 50 атомов) достигается увеличение качества при росте скорости цифрового потока.

В силу того, что структура кодера (рисунок 17) работает по принципу анализа через синтез, и на каждой итерации производятся вычисления сигнала остатка $r^m(n)$ текущего фрейма, то без дополнительных изменений схемы кодирования можно отслеживать затухание энергии остаточного сигнала. При достижении определенного уровня процедуру согласованной подгонки можно остановить. Это дает возможность динамически изменять скорость цифрового потока в зависимости от структуры входных данных и отобранной для передачи информации.

5.2. Экспериментальные результаты. Для объективной оценки качества восстановленного аудиоречевого сигнала была использована модель $PEMO-Q$ [34]. Данная методика имеет низкую степень специализации относительно оцениваемого сигнала, что хорошо подходит для оценки качества различных по своей природе аудио образцов. Шкала оценки (*Objective Difference Grade* — ODG) в зависимости от степени искажения выходного сигнала формируется следующим образом: не воспринимаемое искажение («*imperceptible*») = 0; воспринимаемое, но не раздражающее («*perceptible but not annoying*») = -1.0; не-

много раздражающее («*slightly annoying*») = -2.0 ; раздражающее («*annoying*») = -3.0 ; очень раздражающее («*very annoying*») = -1.0 . Входной тестовой последовательностью служили одноканальные образцы звуковых сигналов (таблица 1). Частота дискретизации 44.1 кГц, разрядность представления отсчетов сигнала – 16 бит.

Таблица 1. Тестовые образцы

Образец	Описание	Образец	Описание
<i>es01</i>	Вокал (Suzan Vega)	<i>si01</i>	Клавеcин
<i>es02</i>	Речь на немецком языке	<i>si02</i>	Каcтаньеты
<i>es03</i>	Речь на английском языке	<i>si03</i>	Pitch pipe
<i>sc01</i>	Соло на трубе и оркестр	<i>sm01</i>	Волынка
<i>sc02</i>	Оркестровая композиция	<i>sm02</i>	Металлофон
<i>sc03</i>	Современная поп-музыка	<i>sm03</i>	Plucked strings

В ходе экспериментов было рассчитано, что бюджет бит для варианта с 200 атомами и учетом затрат на кодирование местоположения атомов в структуре критического дерева ПДВП — $CB - WPD$ составляет 37 кбит/с, а каждые дополнительные 50 атомов увеличивают битрейт на 8,6 кбит/с. Объективные оценки качества ODG каждого испытуемого образца показаны на рисунке 19 [18].

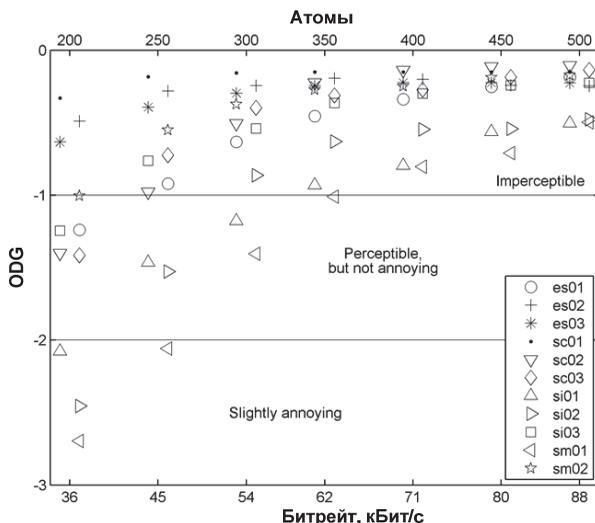


Рис. 19. Распределение объективных оценок качества ODG

Как видно из рисунка 19, ни для одного из исследуемых образцов объективная оценка ODG не находится ниже «*slightly annoying*». На самом низком битрейте (200 атомов) только три образца в зоне «*slightly annoying*»: речевые образцы (*es02*, *es03*). Речевой образец *es01* (вокал,

Suzan Vega) имеет отметку в диапазоне «*perceptible, but not annoying*». С ростом скорости передачи распределение объективных оценок ODG сдвигается в область «*imperceptible*». Для 350 атомов (или 62 кбит) все образцы имеют значения ODG от -1 до 0 (только *sm01* находится на границе -1). Применение схемы масштабируемости позволяет подтянуть оценки до «*imperceptible*» для всех тестовых образцов, изменяя число атомов на ходу для каждого кадра.

Анализ оригинального сигнала состоящего из смеси тестовых образцов *es03*, *sc01*, *si02*, *sm01* и восстановленных сигналов для разного числа атомов разреженной аппроксимации в кодере, а также их спектрограмм (рисунок 20) показывает коррелированность результатов с объективными оценками ODG (рисунок 19). Таким образом, анализ экспериментальных результатов показывает, что предложенный универсальный аудиоречевой кодек обеспечивает хорошее качество восстановленного сигнала и эффективно работает с различными типами входного звукового содержимого.

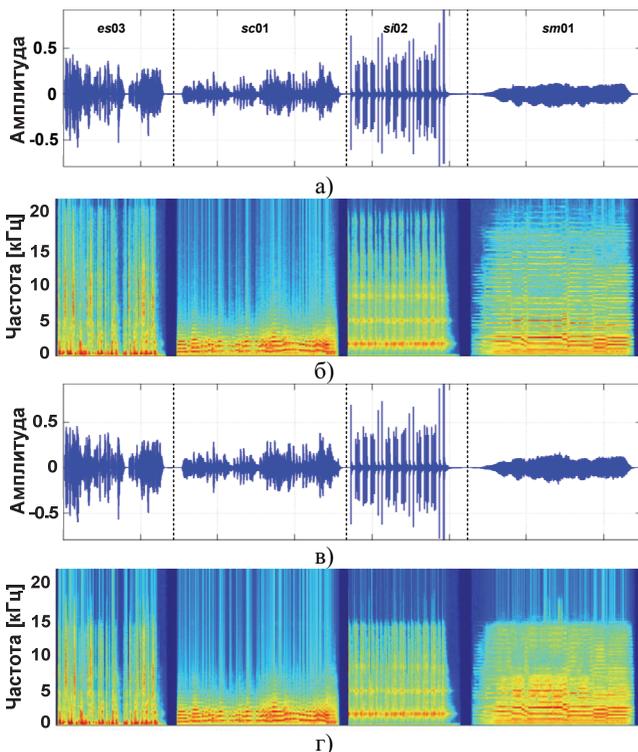


Рис. 20. Оригинальный и восстановленные сигналы для разного числа атомов разреженной аппроксимации: а, б) – оригинальный сигнал; в, г) восстановленный сигнал, число атомов 200

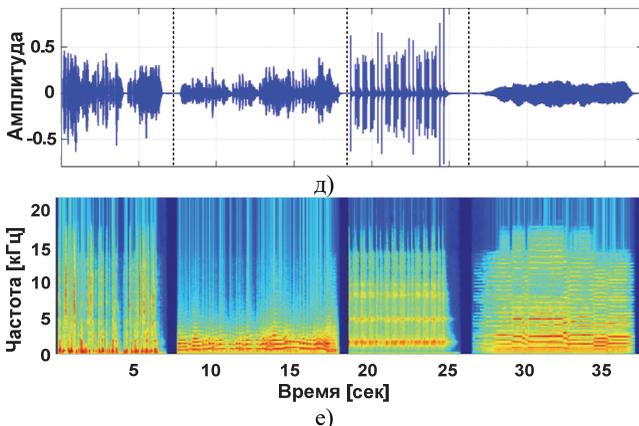


Рис. 20. Оригинальный и восстановленные сигналы для разного числа атомов разреженной аппроксимации: д, е) восстановленный сигнал, число атомов 500

Тестовые образцы (таблица 1) также были сжаты алгоритмами *Vorbis* и *Opus*[35, 36]. Результаты сравнения оценок объективного качества ODG данного масштабируемого аудиоречевого кодера (MP coding scheme) с современными кодерами (*Opus* и *Vorbis*) показано на рисунке 21 [18, 19].

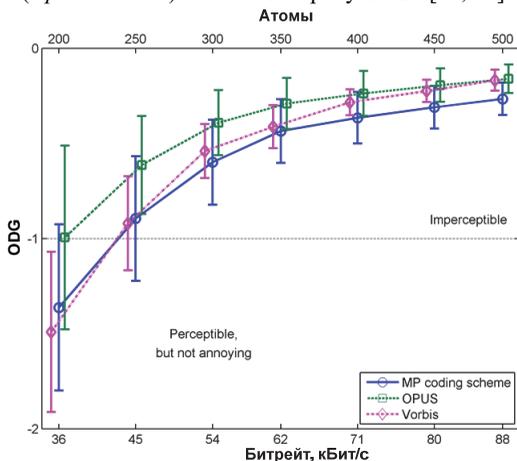


Рис. 21. Объективные оценки качества ODG кодеров MP coding scheme, *Opus* и *Vorbis*

Анализ результатов эксперимента показывает, что от 45 кбит и выше все три кодера находятся в «*imperceptible*» диапазоне. Оценка качества ODG для *Opus* (36 кбит) находится на границе «*imperceptible*» области. Оценка качества ODG для алгоритма *Vorbis* немного меньше и

попадает в область «*perceptible, but not annoying*». Стоит отметить, что *Opus* является составным кодером, который состоит из двух различных алгоритмов для кодирования речи и музыки, в то время как разрабатываемый аудиокодер использует одну модель для работы со всеми типами входных сигналов.

6. Заключение. Два алгоритма адаптивного частотно-временного анализа звуковых сигналов предложены в работе: алгоритм динамической трансформации частотно-временного плана, позволяющий определить субоптимальную структуру декомпозиции ПДВП, достоинством которого является то, что рост дерева осуществляется сверху вниз, без возвратов на меньшие масштабные уровни преобразования и необходимости построения полного дерева ПДВП, что соответствует концепции обработки в реальном масштабе времени; алгоритм параметрического анализа сигналов на основе применения разреженной аппроксимации. Формирование оптимального набора частотно-временных функций в словаре выполняется на основе применения перцептуально адаптированного к текущему фрейму входного сигнала ПДВП, что позволяет уменьшить размер формируемого словаря и сделать его динамическим, то есть зависимым от структуры аудиосигнала. Предложено структурное решение масштабируемых перцептуальных аудиоречевых кодеров на базе ПДВП с динамической реконфигурацией частотно-временного плана: параметрический аудиоречевой кодер на основе разреженной аппроксимации с перцептуально оптимизированным словарем частотно-временных функций. При сравнении предложенного масштабируемого кодера с известными универсальными кодерами *Vorbis* и *Opus* можно сделать вывод, что часть тестовых звуковых образцов эквивалентна по качеству при равных скоростях цифрового потока, масштабирование которого, в разработанном кодере, позволяет при небольшом увеличении скорости цифрового потока добиться эквивалентного качества и для остальных образцов.

Литература

1. *Kahrs M., Brandenburg K.* Application of digital signal processing to audio and acoustics // USA Boston: Kluwer Academic Publishers. 1998. 545 p.
2. *Valin J.-M., Terriberry T. B., Montgomery C., Maxwell G.*, A high-quality speech and audio codec with less than 10-ms delay, // IEEE Transaction on audio, speech, and language processing. 2010. vol. 18. pp. 58–67.
3. *Umaphathy K., Ghoraani B., Krishnan S.*, Audio signal processing using time-frequency approaches: coding, classification, fingerprinting, and watermarking // EURASIP Journal on Advances in Signal Processing. 2010. vol. 2010. no. 1. pp. 451695.
4. *Painter T., Spanias A.* Perceptual Coding of Digital Audio // Proceedings of IEEE. 2000. vol. 88. no. 4. pp. 451–513.
5. *Brandenburg K.* Introduction to perceptual coding // Collected Papers on Digital Audio Bit-Rate Reduction. Eds. 1996. pp. 23–30.

6. *Spanias A., Painter T., Atti V.* Audio signal processing and coding // John Wiley & Sons, Inc. New Jersey. USA. 2007. 464 p.
7. *Вашкевич М.И., Азаров П.С., Петровский А.А.* Косинусно-модулированные банки фильтров с фазовым преобразованием: реализация и применение в слуховых аппаратах // Горячая линия-Телеком. Москва. 2014. 210 с.
8. *Bosi M., Goldberg R.E.* Introduction to digital audio coding and standards // Springer Science+Business Media. USA. 2003. 434 p.
9. *Wickerhauser M.V.* Adaptive Wavelet Analysis from Theory to Software // Massachusetts: A.K. Peters Ltd. 1994. 486 p.
10. *Johnston J.D.* Transform coding of audio signals using perceptual noise criteria // IEEE Transaction on Selected Areas of Communication. 1988. vol. 6. pp. 314–323.
11. *Ковалгин Ю.А., Вологдин Э.И.* Аудиотехника // Горячая линия-Телеком. 2013. 742 с.
12. *Reyes N.R., Candeas P.V.* Adaptive signal modeling based on sparse approximations for scalable parametric audio coding // IEEE Transactions on audio, speech, and language processing. 2010. vol. 18. pp. 447–460.
13. *Petrovsky Al., Herasimovich V., Petrovsky A.* Scalable parametric audio coder using sparse approximation with frame-to-frame perceptually optimized wavelet packet based dictionary // 138th AES Convention. 2015. paper 9264. 10 p.
14. *Mallat S., Zhang Z.* Matching pursuits with time-frequency dictionaries // IEEE Transaction on Signal Processing. 1993 vol. 41. no. 12. pp. 3397–3415.
15. *Chardon G., Neccari T., Balazs P.* Perceptual matching pursuit with Gabor dictionaries and time-frequency masking // Proceedings of IEEE ICASSP'2014. 2014. pp. 3126–3130.
16. *Ravelli E., Gaeul R., Daudet L.*, Matching pursuit in adaptive dictionaries for scalable audio coding // Proceedings of EUSIPCO'2008. 2008. pp. 1–5.
17. *Petrovsky Al., Azarov E., Petrovsky A.* Hybrid signal decomposition based on instantaneous harmonic parameters and perceptually motivated wavelet packets for scalable audio coding // Signal Processing. Special issue “Fourier Related Transforms for Non-Stationary Signals”. 2011. vol. 91. Issue 6. pp. 1489–1504.
18. *Petrovsky Al., Herasimovich M., Petrovsky A.* Bio-inspired sparse representation of speech and audio using psychoacoustic adaptive matching pursuit // Proceedings of 18th International Conference of SPECOM 2016. 2016. pp. 156–164.
19. *Petrovsky Al., Herasimovich V., Petrovsky A.* Audio/speech coding using frame-based psychoacoustic optimized time-frequency dictionaries and its performance evaluation // IEEE conference proceedings “Signal processing: algorithms, architectures, arrangements, and applications” (SPA-2016). 2016. pp. 225–229.
20. *Burrus C.S., Gopinath R.A., Guo H.* Introduction to wavelets and wavelet transforms // N.J.: Prentice Hall. 1998. 298 p.
21. *Cohen I., Raz S., Malah D.* Orthonormal shift-invariant adaptive packet decomposition and representation // Signal Processing. 1997. vol. 57. Issue 3. pp. 251–270.
22. Анализаторы речевых и звуковых сигналов: методы, алгоритмы и практика / под ред. проф. А.А. Петровского // Минск: Бестпринт. 2009. 455 с.
23. *Zwicker E., Fastl H.* Psychoacoustics: Facts and Models // Berlin, Germany: Springer-Verlag. 1990. 380 p.
24. ITU-R Recommendation BS.1387, Method for Objective Measurements of Perceived Audio Quality. 1998.
25. *Petrovsky Al.* A multiresolution auditory model using adaptive WP excitation scalograms // Polska akademia nauk “Elektronika”. 2008. vol. 49. no 4. pp. 65–70.
26. *Petrovsky Al., Krahe D., Petrovsky A.* Real-time wavelet packet-based low bit rate audio coding on a dynamic reconfigurable system // Proc. 138th Convention. 2003. paper 5778. 22p.

27. *Petrovsky Al., Rodionov M., Petrovsky A.* Dynamic reconfigurable on the lifting steps wavelet packet processor with frame-based psychoacoustic optimized time-frequency tiling for real-time audio applications // Design and architectures for digital signal processing. InTech. 2013. pp. 3–30.
28. *Karmakar A., Kumar A., Patney R.K.* Synthesis of an optimal wavelet based on auditory perception criterion // EURASIP Journal on Advance in Signal Processing. 2011. vol. 2011. no. 1. pp. 170927.
29. *Петровский Ал.А.* Построение психоакустической модели в области вейвлет коэффициентов для перцептуальной обработки звуковых и речевых сигналов // Научно-практический журнал «Речевые технологии». Москва. 2008. № 4. С. 61–71.
30. *Cotfman R., Wickerhauser M.V.* Entropy-Based Algorithms for Best Basis Selection // IEEE Transaction on Information Theory. 1992. vol. 38. no. 2. pp. 713–718.
31. *Vera-Candeas P., Ruiz-Reyes N., Roza-Zurera M.* Transient modelling by Matching-Pursuits with a wavelet dictionary for parametric audio coding // IEEE Signal Processing Letters. 2004. vol. 11. no. 3. pp. 349–352.
32. *Petrovsky Al., Petrovsky A.* Matching pursuit algorithm with frame-based auditory optimized WP-dictionary for audio transient modeling // Polska academia nauk “Elektronika”. 2008. vol. 49. no. 4. pp. 74–79.
33. *Heusdens R., Vafin R., Kleijn W.B.* Sinusoidal modeling using psychoacoustic-adaptive matching pursuits // IEEE Signal Processing Letters. 2002. vol. 9. no. 8. pp. 262–265.
34. *Huber R., Kollmeier BPEMO-Q – A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception // IEEE Transactions on audio, speech, and language processing.* 2006. vol. 14. pp. 1902–1911.
35. *Vos K., Sørensen K. V., Jensen S. S., Valin J.-M.* Voice coding with Opus // Proc. AES 135th Convention. 2013. paper 8941.10 p.
36. *Valin J.-M., Maxwell G., Terriberry T.B., Vos K*High-quality, low-delay music coding in the Opus codec // Proc. AES 135th Convention. 2013. paper 8942. 10 p.

Петровский Алексей Александрович — д-р техн. наук, доцент, руководитель группы по обработке аудио сигналов и речи в отделе мультимедийных приложений департамента исследований и разработок, ООО «Техкомпания Хуawei». Область научных интересов: цифровая обработка сигналов, системы кодирования аудиоданных, редактирования шума и устранения акустического эха, а также обработка сигналов в реальном времени для мультимедиа приложений. Число научных публикаций — 80. alexey@petrovsky.eu; Алтуфьевское шоссе 1/7, Москва, 127106, Россия; р.т.: +7(495)660-44-59.

Петровский Александр Александрович — д-р техн. наук, профессор, заведующий кафедрой электронных вычислительных средств, Белорусский государственный университет информатики и радиоэлектроники (БГУИР). Область научных интересов: цифровая обработка сигналов, системы кодирования аудио и видео данных, мультипроцессорные системы реального времени для мультимедиа приложений. Число научных публикаций — 662. palex@bsuir.by; ул. П. Бровки, 6, Минск, 220013, Беларусь; р.т.: +3751729323-40.

AL.A. PETROVSKY, A.A. PETROVSKY
**SCALABLE SPEECH AND AUDIO CODERS BASED ON
 ADAPTIVE TIME-FREQUENCY SIGNAL ANALYSIS**

Petrovsky, Al.A, Petrovsky A.A. Scalable Speech and Audio Coders based on Adaptive Time-Frequency Signal Analysis.

Abstract. The paper discusses the methods of perceptual sub-band audio signal processing with the dynamic time-frequency map transformation based on the discrete wavelet packet (DWPT) transform. The advantage of these methods is a top-down construction of the DWPT tree without returning to smaller scale levels of decomposition and needing to build a complete DWPT tree. This corresponds to the concept of scalable audio/speech coders implementation in real time. The objective quality assessment of the proposed coders, based on techniques PEMO-Q, and comparisons with the widespread encoders *Opus* and *Vorbis* are given. It shows that the reconstructed signal complies with ITU-R PEAQ at a high compression ratio of up to 18 times or more, does not contain artifacts: noise-to-mask ratio NMR_{total} is less ≈ -9 dB.

Keywords: scalable audio/speech coder, wavelet packet, matching pursuit

Petrovsky Alexey Aleksandrovich — Ph.D., Dr. Sci., associate professor, head of the speech and audio processing group of the media engineering department, Russian Research Center of Huawei Technologies. Research interests: acoustic signal processing, such as wideband speech and audio processing, perceptual coding, psychoacoustics, noise reduction and echo cancellation, and real-time signal processing. The number of publications — 80. alexey@petrovsky.eu; 1/7, Altufevskoe shosse, Moscow, 127106, Russia; office phone: +7(495)660-44-59.

Petrovsky Alexander Alexandrovich — Ph.D., Dr. Sci., professor, head of the Computer Engineering Department, Belarusian State University of Informatics and Radioelectronics (BSUIR). Research interests: digital signal processing, speech and audio coding, multiprocessor real-time systems for multimedia applications. The number of publications — 662. palex@bsuir.by; 6, P.Brovky str. Minsk, 220013, Republic of Belarus; office phone: +37517293-23-40.

References

1. Kahrs M., Brandenburg K. Application of digital signal processing to audio and acoustics. USA Boston: Kluwer Academic Publishers. 1998. 545 p.
2. Valin J.-M., Terriberry T. B., Montgomery C., Maxwell G., A high-quality speech and audio codec with less than 10-ms delay. *IEEE Transaction on audio, speech, and language processing*. 2010. vol. 18. pp. 58–67.
3. Umaphathy K., Ghoraani B., Krishnan S., Audio signal processing using time-frequency approaches: coding, classification, fingerprinting, and watermarking. *EURASIP Journal on Advances in Signal Processing*. 2010. vol. 2010. no. 1. pp. 451695.
4. Painter T., Spanias A. Perceptual Coding of Digital Audio. Proceedings of IEEE. 2000. vol. 88. no. 4. pp. 451–513.
5. Brandenburg K. Introduction to perceptual coding. Collected Papers on Digital Audio Bit-Rate Reduction. Eds. 1996. pp. 23–30.
6. Spanias A., Painter T., Atti V. Audio signal processing and coding. John Wiley & Sons, Inc. New Jersey, USA. 2007. 464 p.
7. Vashkevich M.I., Azarov I.S., Petrovsky A.A. *Kosinusno-modulirovanie banki filtrov s phasovim preobrasovaniem: realizatsia i primenenie v sluchovykh apparatach* [Cosine

- modulated filter banks with phase transform: implementation and application to the hearing aids]. Moscow. Hotline-Telecom. 2014. 210 p. (In Russ.).
8. Bosi M., Goldberg R.E. Introduction to digital audio coding and standards. Springer Science+Business Media. USA. 2003. 434 p.
 9. Wickerhauser M.V., Adaptive Wavelet Analysis from Theory to Software. Massachusetts: A.K. Peters Ltd. 1994. 486 p.
 10. Johnston J.D. Transform coding of audio signals using perceptual noise criteria. *IEEE Transaction on Selected Areas of Communication*. 1988. vol. 6. pp. 314–323.
 11. Kovalgin J.A., Vologdin E.I. *Audiotechnika* [Audio techniques]. Moscow. Hotline-Telecom. 2013. 742 p. (In Russ.).
 12. Reyes N.R., Candeas P.V. Adaptive signal modeling based on sparse approximations for scalable parametric audio coding. *IEEE Transactions on audio, speech, and language processing*. 2010. vol. 18. pp. 447–460.
 13. Petrovsky Al., Herasimovich V., Petrovsky A. Scalable parametric audio coder using sparse approximation with frame-to-frame perceptually optimized wavelet packet based dictionary. 138th AES Convention. 2015. paper 9264. 10 p.
 14. Mallat S., Zhang Z. Matching pursuits with time-frequency dictionaries. *IEEE Transaction on Signal Processing*. 1993. vol. 41. no. 12. pp. 3397–3415.
 15. Chardon G., Necciarri T., Balazs P. Perceptual matching pursuit with Gabor dictionaries and time-frequency masking. Proceedings of IEEE ICASSP'2014. 2014. pp. 3126–3130.
 16. Ravelli E., Gaeul R., Daudet L. Matching pursuit in adaptive dictionaries for scalable audio coding. Proceedings of EUSIPCO'2008. 2008. pp. 1–5.
 17. Petrovsky Al., Azarov E., Petrovsky A. Hybrid signal decomposition based on instantaneous harmonic parameters and perceptually motivated wavelet packets for scalable audio coding. Signal Processing, Special issue “Fourier Related Transforms for Non-Stationary Signals”. 2011. vol. 91. Issue 6. pp. 1489–1504.
 18. Petrovsky Al., Herasimovich M., Petrovsky A. Bio-inspired sparse representation of speech and audio using psychoacoustic adaptive matching pursuit. Proceedings of 18th International Conference of SPECOM 2016. 2016. pp. 156–164.
 19. Petrovsky Al., Herasimovich V., Petrovsky A. Audio/speech coding using frame-based psychoacoustic optimized time-frequency dictionaries and its performance evaluation. IEEE conference proceedings “Signal processing: algorithms, architectures, arrangements, and applications” (SPA-2016). 2016. pp. 225–229.
 20. Burrus C.S., Gopinath R.A., Guo H. Introduction to wavelets and wavelet transforms. N.J.: Prentice Hall. 1998. 298 p.
 21. Cohen I., Raz S., Malah D. Orthonormal shift-invariant adaptive packet decomposition and representation. Signal Processing. 1997. vol. 57. Issue 3. pp. 251–270.
 22. *Analizatory rechevyh i zvukovyh signalov: metody, algoritmy i praktika / pod red. prof. A.A. Petrovskogo* [Analyzers of speech and audio signals: methods, algorithms and practices. Edited by A.A. Petrovsky]. Minsk: Bestprint. 2009. 455 p.
 23. Zwicker E., Fastl H. Psychoacoustics: Facts and Models. Berlin, Germany: Springer-Verlag. 1990. 380 p.
 24. ITU-R Recommendation BS.1387, Method for Objective Measurements of Perceived Audio Quality. 1998.
 25. Petrovsky Al. A multiresolution auditory model using adaptive WP excitation scalograms. *Polska akademia nauk “Elektronika”*. 2008. vol. 49. no 4. pp. 65–70.
 26. Petrovsky Al., Krahe D., Petrovsky A. Real-time wavelet packet-based low bit rate audio coding on a dynamic reconfigurable system. Proc. 138th Convention. 2003. paper 5778. 22p.
 27. Petrovsky Al., Rodionov M., Petrovsky A. Dynamic reconfigurable on the lifting steps wavelet packet processor with frame-based psychoacoustic optimized time-frequency

- tiling for real-time audio applications. *Design and architectures for digital signal processing*. InTech. 2013. pp. 3–30.
28. Karmakar A., Kumar A., Patney R.K. Synthesis of an optimal wavelet based on auditory perception criterion. *EURASIP Journal on Advance in Signal Processing*. 2011. vol. 2011. no. 1. pp. 170927.
 29. Petrovsky A.I. [A building the psychoacoustic model in the wavelet domain for the perceptual processing of speech and sound signals]. *Nauchno-prakticheskij zhurnal «Reshevye tehnologii» – Scientific journal "Speech Technologies"*. 2008. vol. 4. pp. 61–71. (In Russ.).
 30. Coifman R., Wickerhauser M.V. Entropy-Based Algorithms for Best Basis Selection. *IEEE Transaction on Information Theory*. 1992. vol. 38. no. 2. pp. 713–718.
 31. Vera-Candeas P., Ruiz-Reyes N., Roza-Zurera M. Transient modelling by Matching-Pursuits with a wavelet dictionary for parametric audio coding. *IEEE Signal Processing Letters*. 2004. vol. 11. no. 3. pp. 349–352.
 32. Petrovsky A.I., Petrovsky A. Matching pursuit algorithm with frame-based auditory optimized WP-dictionary for audio transient modeling. *Polska academia nauk "Elektronika"*. 2008. vol. 49. no. 4. pp. 74–79.
 33. Heusdens R., Vafin R., Kleijn W.B. Sinusoidal modeling using psychoacoustic-adaptive matching pursuits. *IEEE Signal Processing Letters*. 2002. vol. 9. no. 8. pp. 262–265.
 34. Huber R., Kollmeier B. PEMO-Q – A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception. *IEEE Transactions on audio, speech, and language processing*. 2006. vol. 14. pp. 1902–1911.
 35. Vos K., Sørensen K. V., Jensen S. S., Valin J.-M., Voice coding with Opus. Proc. AES 135th Convention. 2013. paper 8941.10 p.
 36. Valin J.-M., Maxwell G., Terriberry T.B., Vos K. High-quality, low-delay music coding in the Opus codec. Proc. AES 135th Convention. 2013. paper 8942. 10 p.

А.И. Попов, С.Ф. Свинын
**ПРИМЕНЕНИЕ ФИНИТНЫХ БАЗИСНЫХ СПЛАЙНОВ ПРИ
ВОССТАНОВЛЕНИИ СИГНАЛОВ
ЭЛЕКТРОГАСТРОЭНТЕРОГРАФИИ**

Попов А.И., Свинын С.Ф. **Применение финитных базисных сплайнов при восстановлении сигналов электрогастроэнтерографии.**

Аннотация. Электрогастроэнтерография — перспективный метод обследования сократительной активности органов пищеварительной системы, основанный на съеме и последующей обработке биоэлектрических сигналов. Прогресс в развитии электрофизиологических методов диагностики в последние годы во многом обеспечен возможностью компьютерной обработки измерительных сигналов. Данная статья посвящена вопросам организации измерений в электрогастроэнтерографии. Дано введение в предметную область, проанализирован информационный состав измерительных сигналов электрогастроэнтерографии, приведены диагностические показатели, получаемые в результате спектрального анализа сигналов, рассмотрены задачи автоматизации диагностики.

Предложен новый метод выбора частоты отсчетов гастроэнтерограмм, учитывающий фактор конечной длительности измерительных сеансов и спектральных свойств сигнала. В предлагаемом методе применяется разложение сигнала в виде конечной суммы финитных кардинальных В-сплайнов целых степеней. Проведен вычислительный эксперимент для определения точности восстановления сигнала при используемых в электрогастроэнтерографии параметрах измерительного сеанса.

Ключевые слова: базисный сплайн, конечная энергия, компактный носитель, биосигнал, автоматизированная обработка биомедицинских сигналов, электрофизиология, электрогастроэнтерография.

1. Введение. Электрофизиологические обследования [1] представляют собой важный компонент современной медицинской диагностики. При таких обследованиях количественные параметры функционирования различных органов получают в результате математической обработки электрических сигналов.

Для исследования сократительной активности органов желудочно-кишечного тракта (ЖКТ) в настоящее время применяются такие электрофизиологические методы, как электрогастрография (ЭГГ) [2, 3] и электрогастроэнтерография (ЭГЭГ) [4, 5]. В современном состоянии данные методы основаны на неинвазивном (т.е. без вмешательства в организм человека) сборе электрических сигналов, содержащих информацию о моторике ЖКТ, и их последующем анализе. В ЭГГ исследуется только желудок, в ЭГЭГ — желудок и различные отделы кишечника.

Для широкого внедрения в медицинскую практику ЭГЭГ еще нуждается в развитии. Качественное развитие ЭГЭГ обеспечивается целенаправленным изменением технологий сбора и обработки данных.

Можно выделить следующие уровни работы с электрофизиологическими данными: съем сигналов, их цифровая обработка (фильтрация, вычисление значений диагностических показателей), диагностика на основе наборов значений показателей. Тенденция современной электрофизиологии — автоматизация диагностики [1] — непосредственно относится к последнему, верхнему уровню обработки данных. Однако вопросы ввода сигналов и их цифровой обработки носят фундаментальный характер, и исследования на этих уровнях не менее важны.

Обрабатываемые в ЭГЭГ сигналы отличаются наиболее низкочастотными спектрами (0,01–0,25 Гц) среди всех электрофизиологических сигналов и, соответственно, большей длительностью, обычно порядка 40 минут. Применительно к ЭГЭГ в большинстве случаев можно говорить об отсутствии разрывов в уровнях энергии внутри интервалов записей и, имея в виду их конечность во времени, рассматривать записи ЭГЭГ как сигналы, заданные на компактных носителях. Функциями с компактным носителем являются все непрерывные сигналы, обнуляющиеся при $|t| > C$, где t — вещественная прямая. Создается возможность обратиться к разложениям сигналов в ряды по финитным ортогональным базисам для повышения точности цифровых моделей [6, 7].

Роль обобщенных спектральных методов анализа ЭГЭГ-сигналов, опирающихся на теорию энергетических пространств сигналов, неуклонно возрастает. Цели данного исследования ставятся как повышение значимости диагностических показателей, полученных в результате автоматизации обработки электрофизиологических сигналов, и как обоснование достаточности длин цифровых выборок сигналов с учетом их спектрального состава.

2. Современное состояние электрофизиологических методов обследования ЖКТ. Зарубежные исследования сосредоточены на ЭГГ, то есть на изучении моторики желудка. Достаточно полное представление о современном состоянии ЭГГ дают обзорные работы [3] и [8]. К настоящему времени разработаны различные способы размещения датчиков на теле пациента — схемы отведений. Применяются как двухполюсные, так и однополюсные отведения. В качестве датчиков используются кардиографические электроды. В ходе обследования обычно последовательно проводятся два измерительных сеанса: один — натощак (30 минут), второй — после еды (60 минут). Практикуется также длительная ЭГГ (до 24 часов) с применением портативных измерительных модулей, однако в этом случае особенно остро проявляется проблема артефактов движения. Ввиду

низкочастотной природы исследуемого сигнала используются низкие частоты дискретизации. Например, в [8] описана амбулаторная измерительная система, в которой входной сигнал оцифровывается с частотой 1 Гц. Указывается на различие измерительных средств, ориентированных на медицинскую практику и на исследования. В исследовательских целях применяются многоканальные измерительные системы (до 8 каналов). Они обеспечивают возможность анализа распространения волн вдоль желудка. Непосредственная интерпретация ЭГГ-сигналов во временной области затруднена, диагностика базируется на вычислении набора параметров сигналов. Прежде всего, выявляются такие нарушения ритма, как тахикастрия (повышенная частота сокращений), брадикастрия (пониженная частота сокращений) и аритмия. Установлен ряд синдромов, с которыми связаны эти нарушения. Вопрос о том, являются нарушения ритма причиной синдромов или они представляют собой сопутствующий эффект, остается открытым. С точки зрения клинического значения ЭГГ сейчас рассматривается не как самостоятельный метод диагностики конкретных заболеваний, а как полезное звено в совокупности методов.

Указывается, что одной из насущных проблем ЭГГ является ее нестандартизированность. До сих пор не выработаны стандарты на схемы расположения датчиков, параметры измерительного сеанса, вид воздействия между измерениями, функциональные требования на программное обеспечение для анализа сигналов. Кроме того, определения диагностических показателей допускают различные способы их вычисления с различными итоговыми результатами.

Уникальность отечественных разработок состоит в том, что они позволяют исследовать функционирование различных отделов ЖКТ, а не только желудка. Российской промышленностью выпускается аппаратно-программный комплекс «Гастроскан-ГЭМ» (НПП «Исток-система», г. Фрязино) [9]. Портативный измерительный блок «Гастроскана-ГЭМ» позволяет проводить съем ЭГГ и ЭГЭГ-сигналов с частотой дискретизации 2 Гц. В качестве датчиков при различных схемах отведений могут использоваться как корпусные, так и наконечностные электроды. Основной протокол обследования включает два измерительных сеанса продолжительностью по 40 минут [8].

3. Информационный состав ЭГЭГ-сигнала. В таблице 1 приведены полосы частот сокращений, характерные для различных отделов ЖКТ [9]. В компонентах измерительного сигнала, имеющих частоты в указанных полосах, сосредоточена информация о

сократительной активности соответствующих органов. Эти компоненты будем называть гастроэнтерограммами.

Таблица 1. Полосы частот сокращений отделов ЖКТ

Название отдела	Номер отдела в ЖКТ	Полоса частот (Гц)
Толстая кишка	5	0,01 – 0,03
Желудок	1	0,03 – 0,07
Подвздошная кишка	4	0,08 – 0,12
Тощая кишка	3	0,13 – 0,17
Двенадцатиперстная кишка	2	0,18 – 0,22

На рисунке 1а представлен 10-минутный фрагмент входного ЭГЭГ-сигнала здорового добровольца. Запись выполнена при помощи измерительного блока комплекса «Гастроскан-ГЭМ» с частотой дискретизации 2 Гц. Путем полосовой фильтрации с применением цифрового фильтра Баттерворта получены гастроэнтерограммы. Они показаны на рисунке 1б – 1е. Границы полос пропускания фильтров соответствуют частотам, приведенным в таблице 1. Фильтр Баттерворта характеризуется гладкостью АЧХ в переходной зоне. При использовании других фильтров графики гастроэнтерограмм будут несколько отличаться в силу свойств этих фильтров.

Порядок фильтра для каждого отдела ЖКТ (таблица 2) выбирался таким образом, чтобы при требуемой полосе пропускания и имеющейся частоте дискретизации амплитудно-частотная характеристика (АЧХ) фильтра была наиболее близка к идеальной АЧХ. Например, для полосы пропускания 0,03 Гц – 0,07 Гц (желудок) при частоте дискретизации 2 Гц данное условие достигается, когда порядок фильтра равен 5. Это проиллюстрировано на рисунке 2. Сплошной линией показан график АЧХ фильтра Баттерворта порядков 4, 5 и 6, пунктирной линией — график идеальной АЧХ. Числом n обозначен порядок фильтра.

Таблица 2. Порядок фильтра Баттерворта для отделов ЖКТ

Отдел ЖКТ	Порядок фильтра
Толстая кишка	4
Желудок	5
Подвздошная кишка	6
Тощая кишка	6
Двенадцатиперстная кишка	7

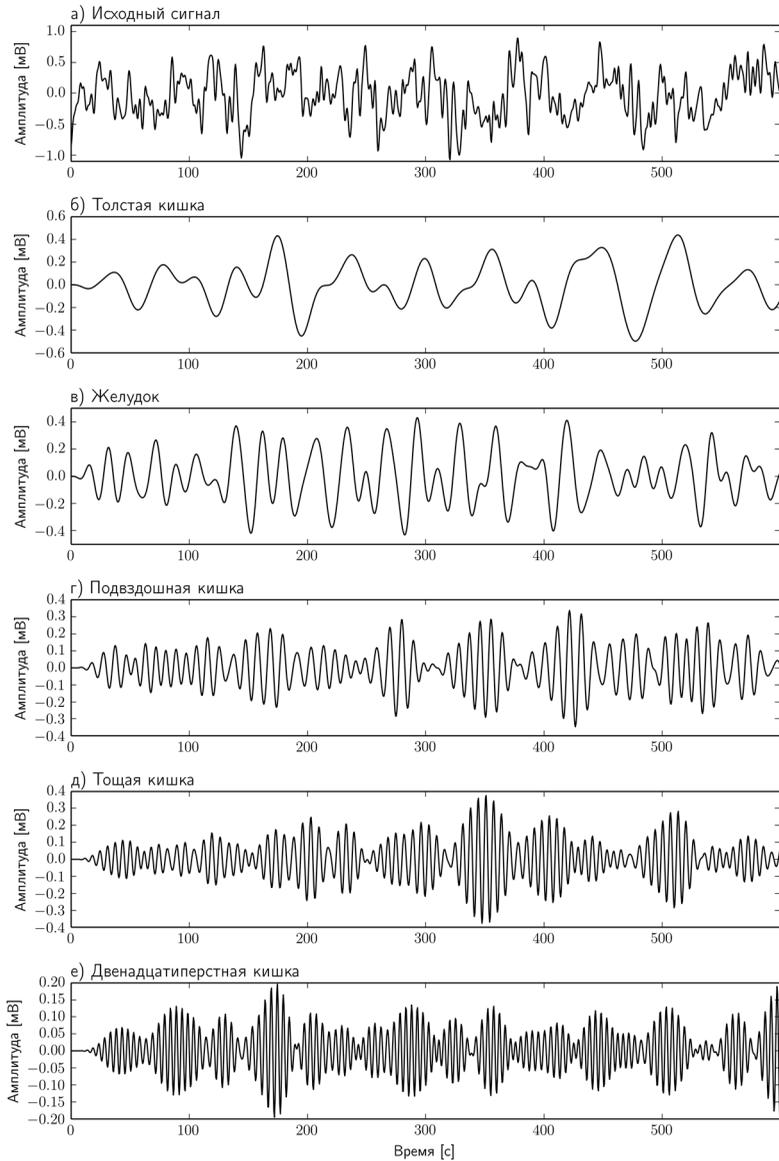


Рис. 1. Цифровая фильтрация ЭГЭГ-сигнала: а) исходный сигнал; б, в, г, д, е) гастроэнтерограммы

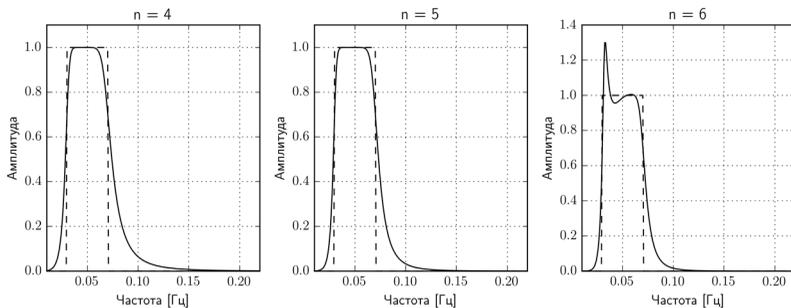


Рис. 2. Выбор порядка фильтра Баттерворта для получения гастрোগраммы из суммарного ЭГЭГ-сигнала

4. Диагностические показатели в ЭГЭГ. Существующее аппаратное обеспечение ЭГЭГ позволяет проводить анализ измерительных сигналов только в отложенном режиме.

Обработка сигналов ЭГЭГ в основном базируется на спектральном анализе. На рисунке 3 приведен график амплитудного спектра гастрোগраммы здорового добровольца.

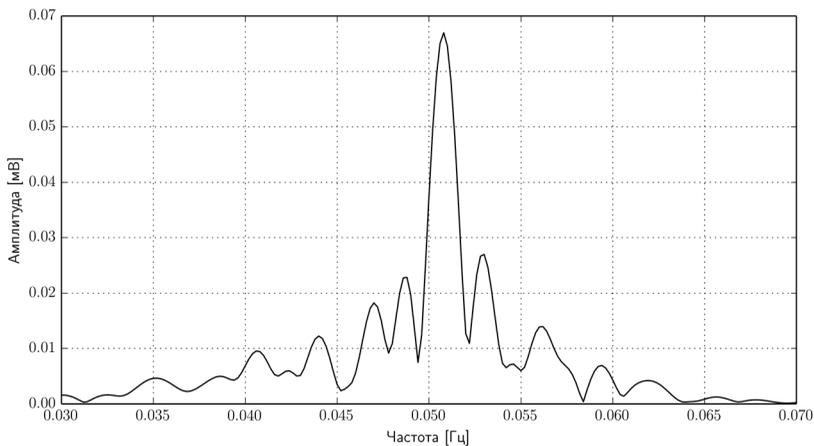


Рис. 3. Амплитудный спектр гастрোগраммы

Пусть N — количество отсчетов в ЭГЭГ-сигнале, длина сигнала; X — массив значений его амплитудного спектра, модуль результата дискретного преобразования Фурье сигнала; a и b — индексы в массиве X , соответствующие границам диапазона частот сокращений исследуемого отдела ЖКТ, $a < b$; f_d — частота дискретизации в Гц.

Приведем формулы для вычисления некоторых параметров сигналов, отражающих силу, частоту и ритмичность сокращений [9, 10].

Средняя мощность:

$$P = \frac{1}{N} \sum_{k=a}^b X[k]^2. \quad (1)$$

Доминантная частота:

$$DF = \frac{f_d}{N} \arg \max_k X[k], k \in [a, b]. \quad (2)$$

Другими словами, DF — это частота, которой соответствует максимальное значение амплитуды в анализируемой части спектра.

Коэффициент ритмичности:

$$K = \frac{\sum_{k=a}^{b-1} \left| |X[k+1]| - |X[k]| \right|}{\max(|X[a:b]|)(b-a-1)}. \quad (3)$$

Здесь $X[a:b]$ обозначает фрагмент массива X , соответствующий значениям индексов от a до b включительно.

Для анализа динамики показателей в рамках одного измерительного сеанса используется оконное преобразование Фурье. Например, вычислив ряд значений доминантной частоты DF на разных участках сигнала, можно получить значение так называемого коэффициента нестабильности доминантной частоты [8]:

$$DFIC = \frac{\sigma_{DF}}{DF}, \quad (4)$$

где σ_{DF} — среднеквадратическое отклонение доминантной частоты, а DF — средняя доминантная частота.

На рисунке 4 приведен пример ЭГЭГ-сигнала и его спектрограммы, полученной в результате оконного преобразования Фурье исходного сигнала с шириной окна, равной 10 мин. Более высоким значениям амплитудного спектра на спектрограмме соответствует более темный цвет. При визуальном анализе спектрограммы прослеживается стабильная доминантная частота со средним значением, несколько превышающим 0,05 Гц.

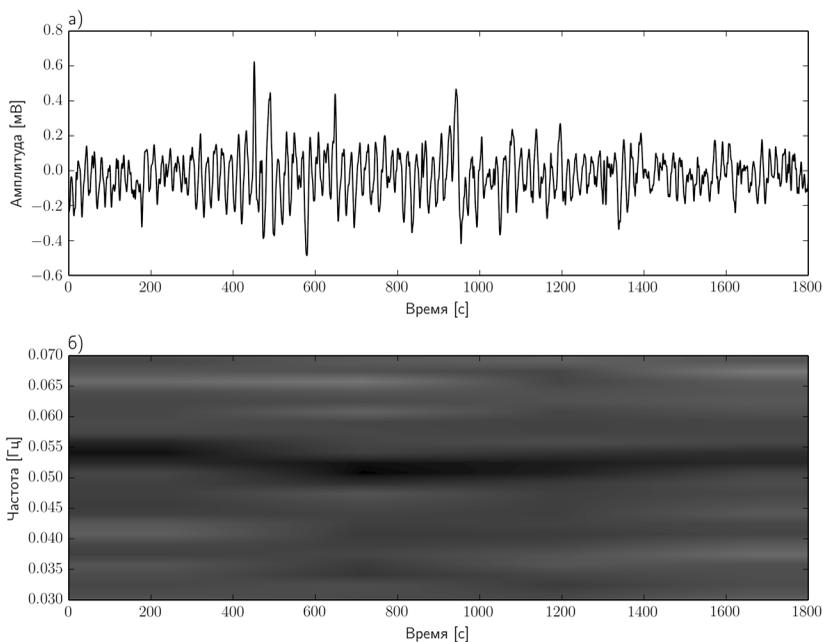


Рис. 4. Оконное преобразование Фурье ЭГЭГ-сигнала: а) исходный сигнал; б) спектрограмма

Также перспективно применение математического аппарата вейвлет-анализа для учета нестационарности ЭГЭГ-сигналов. Например, в [11] предложены варианты использования вейвлета Морле при расчете параметров (местоположение, продолжительность, частота) всплесков активности на гастроэнтерограммах для получения новых динамических характеристик исследуемых сигналов.

Для изучения реакции ЖКТ на воздействие (например, принятие пищи) в рамках одного обследования последовательно проводят два измерительных сеанса: до и после воздействия. Затем вычисляются отношения значений диагностических показателей, например, P_2/P_1 , где P_1 и P_2 — средние мощности сигналов, полученных до и после воздействия.

При автоматизации диагностики параметры сигналов и обследований используются как признаки в методах классификации. Наиболее часто решается задача классификации с целью установления наличия нарушений в функционировании ЖКТ либо отнесения обследуемого к группе «норма». Примерами таких исследований являются [12] (искусственные нейронные сети, точность 85%) и [13] (метод опорных векторов с различными ядрами,

точность 87%). К другому типу относятся задачи уточнения диагноза при наличии заболевания, например, диагностика нарушений ритмичности сокращений желудка у больных сахарным диабетом [14] (нейронные сети, точность 97%), уточнение степени компенсации язвенного пилородуоденального стеноза [15, 16] (дискриминантный анализ, точность 97%).

5. Оптимизация длин выборки отсчетов сигналов. Для гастроэнтерограмм характерны большие длительности во времени (десятки минут), значительно превосходящие величины периодов их частотных составляющих. Но они являются финитными сигналами и, более того, если исключить из рассмотрения участки, соответствующие низкой сократительной активности органов, их можно назвать сигналами с компактными носителями. Напомним, что носителем функции $x(t)$ называется замыкание множества точек, где $x(t) \neq 0$ [17]. Так как диапазон таких сигналов ограничен, то они принадлежат классу непрерывных функций с конечной энергией, то есть выполняется условие:

$$E = \int_{-\infty}^{+\infty} |x(t)|^2 dt < \infty. \quad (5)$$

Малыми буквами $x(t)$ обозначены значения непрерывного сигнала во времени. Большой буквой E обозначена энергия сигнала на всей оси.

В широко известной статье [18] предлагался подход к решению общей проблемы финитности или инфинитности функций как непосредственно непрерывного сигнала, так и ширины полосы его частот. В ней введены понятия, облегчающие выработку оценки точности восстановленного сигнала. Первое из этих понятий определено как «сигнал, ограниченный во времени на уровне ε », а второе — как «полоса частот спектра $(-\omega_\varepsilon, \omega_\varepsilon)$, ограниченная на уровне ε ». Дополним эту пару определений понятием *энергии спектра сигнала E_ε , ограниченной на уровне ε* . Такой уровень можно определить как значение интеграла спектральной энергии для функций, интегрируемых с квадратом, отличающееся на величину ε от полной энергии E .

Среди базисных функций, заданных на компактных носителях, особую роль играют полиномиальные базисные сплайны (В-сплайны). Доказано [19], что с их помощью можно добиться высокой степени аналитической точности и эффективной сходимости при решении задач численного анализа. С позиций использования в теории выборки

они интересны тем, что результаты их преобразований Фурье отвечают аналитическим выражениям следующего вида [20]:

$$X_B(\omega) = B(0) \left(\frac{\sin\left(\frac{\omega h}{2}\right)}{\frac{\omega h}{2}} \right)^{m+1}, \quad (6)$$

где $B(0)$ — значение В-сплайна при $t = 0$, h — расстояние между узлами сплайна (если сетка равномерная), m — степень сплайна (целое число), ω — круговая частота. Эти выражения имеют много общего с общим членом кардинального ряда Уиттекера-Котельникова-Шеннона. Отличает их в первую очередь то, что независимой переменной является частота, а не время, и кроме того, степень m может быть больше 1.

График модуля спектральной плотности $X_{B3}(\omega)$ кубического В-сплайна приведен на рисунке 5.

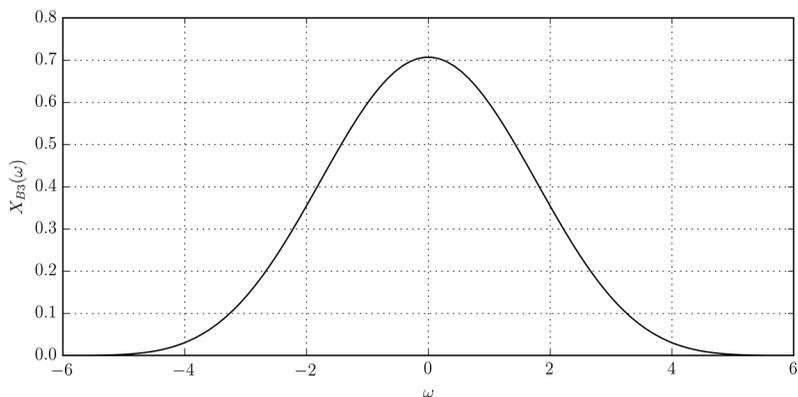


Рис. 5. Спектр одномерного кубического В-сплайна

Рассмотрим конкретную запись электрогастрограммы длительностью около 20 минут и с ограничением по уровню $|x(t)| < 0,51$ мВ. Ее график приведен на рисунке 6.

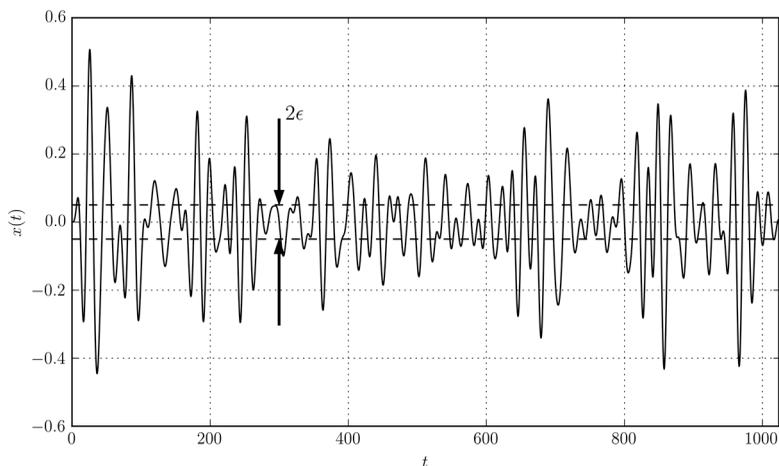


Рис. 6. Пример электрогастрограммы

Функция достаточной степени гладкости (например, подобная изображенной) может быть приближенно представлена в виде суммы «взвешенных» В-сплайнов степени m дефекта 1:

$$x(t) \cong \sum_{i=-m}^{n+m} b_i B_i(t), \quad (7)$$

где n — полное число отсчетов сигнала, b_i — значения коэффициентов ($i = 0, 1, \dots, n$), $B_i(t)$ — значение В-сплайна на текущем интервале длины h . Графики последовательностей В-сплайнов 1-й и 3-й степени показаны на рисунке 7 при шаге между узлами $h=1$.

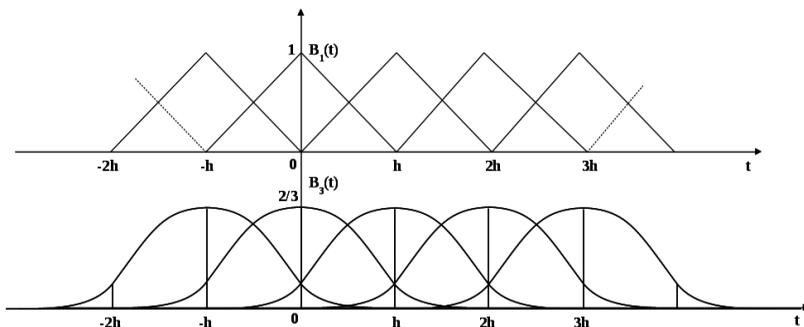


Рис. 7. Последовательности В-сплайнов 1-й и 3-й степени

Сошлемся на аналитическое выражение для спектральной плотности $X_{as}(\omega)$ аппроксимирующей последовательности В-сплайнов, приведенное в монографии [20]. Оно получено в виде:

$$X_{as}(\omega) = \left| X_{B0}(\omega) \right| \left| \sum_{i=-m}^{n+m} b_i \exp(-ji\omega h) \right|, \quad (8)$$

Опираясь на него, можем перейти к интегральному соотношению вида:

$$\int_0^{\infty} |X(\omega)|^2 d\omega \equiv \int_0^{\infty} |X_{as}(\omega)|^2 d\omega. \quad (9)$$

Здесь $X(\omega)$ — функция спектральной плотности сигнала.

Спектры $X(\omega)$ и $X_{as}(\omega)$ инфинитны, а энергия последовательности, заданной на конечном интервале $[a, b]$, конечна. Спектральную энергию как интеграл от квадрата модуля $X_{as}(\omega)$ можно разбить на две части — низкочастотную (НЧ) и высокочастотную (ВЧ):

$$\left(\frac{1}{\pi} \right) \int_0^{\infty} |X_{as}(\omega)|^2 d\omega = \left(\frac{1}{\pi} \right) \int_0^{\omega_c} |X_{as}(\omega)|^2 d\omega + \left(\frac{1}{\pi} \right) \int_{\omega_c}^{\infty} |X_{as}(\omega)|^2 d\omega \quad (10)$$

Частоту ω_c назовем граничной частотой эффективной ширины полосы НЧ-спектра последовательности. Эта полоса может быть рассчитана по энергии НЧ-части, совпадающей «с точностью до ε » со значением полной энергии.

На основании известной из математического анализа теоремы об интегральных неравенствах [17] определим уровень энергии ВЧ-части последовательности:

$$\varepsilon = \frac{1}{\pi} \int_{\omega_c}^{\infty} (X_{as}(\omega))^2 d\omega < C_1 h^2 \int_{\omega_c}^{\infty} \left(\frac{\sin(\omega h / 2)}{\omega h / 2} \right)^{2m+2} d\omega < \\ C_1 h^2 \int_{\omega_c}^{\infty} \left(\frac{2}{\omega h} \right) d\omega = \frac{2^{m+2} C_1}{(2m+1)\pi^{2m+1}} h, \quad (11)$$

где C_1 — коэффициент, зависящий от количества узлов сплайна на $[a, b]$, то есть шагов выборки.

Из выражения (11) следует, что энергия высокочастотных составляющих последовательности В-сплайнов, аппроксимирующей

непрерывный сигнал $x(t)$, пропорциональна значению шага выборки h с коэффициентом, зависящим от степени сплайна m .

Для вычисления коэффициентов интерполяционных сплайнов требуется решение систем линейных алгебраических уравнений (СЛАУ). Часто они имеют большую размерность. Для вычисления коэффициентов в узлах пограничных областей может быть применен метод коллокации, для которого достаточно применение СЛАУ малой размерности [21].

Вместе с тем в теории сплайнов разработаны так называемые «локальные» (сглаживающие) методы вычисления b -коэффициентов, не требующие решения СЛАУ, что ведет к значительному сокращению вычислительных затрат [22]. Эти методы основаны на преобразованиях конечных разностей функций. Формулы сглаживающих приближений аналогичны формулам дискретной фильтрации, используемым в частотной области («окна» Хэннинга, Кайзера, Баттерворта и др.). Они сохраняют свойства гладкости аппроксимирующих сплайнов и тот же порядок оценок точности восстановления сигналов, что и интерполяционные приближения.

В монографии [20] приведены примеры локальных сглаживающих формул вычисления b -коэффициентов кубических сплайнов для внутренних узлов с конкретными значениями весовых множителей:

— усреднение по трем текущим отсчетам функции $x(t_i) = x_i$:

$$b_i = \frac{1}{6}(-x_{i-1} + 8x_i - x_{i+1});$$

— усреднение по пяти текущим отсчетам:

$$b_i = \frac{1}{36}(x_{i-2} - 10x_{i-1} + 54x_i - 10x_{i+1} + x_{i+2});$$

— усреднение по семи текущим отсчетам:

$$b_i = \frac{1}{216}(-x_{i-3} + 12x_{i-2} - 75x_{i-1} + 344x_i - 75x_{i+1} + 12x_{i+2}).$$

Оценку частоты отсчетов конкретной дискретной выборки можно выполнить в одном из двух вариантов: либо при заданной частоте рассчитать полученную ошибку, либо осуществить выборку с такой необходимой частотой, чтобы ошибка восстановления сигнала не превысила заданную величину ε .

Рассмотрим случай, когда измерения произведены по всем пяти отделам ЖКТ. Шаги отсчетов задаются как целые значения, b -коэффициенты вычисляются по формуле трех отсчетов и восстановление сигнала производится по энергетическому критерию.

Интегралы энергии во времени вычисляются по формуле трапеции:

$$E_t = h \left(\frac{x_0^2 + x_n^2}{2} + \sum_{i=1}^{n-1} x_i^2 \right), \quad (12)$$

интегралы спектральной энергии — по формуле:

$$E_{sp} = \left(\frac{1}{\pi} \right) \int_0^{\omega_s} |X_{as}(\omega)|^2 d\omega, \quad (13)$$

а модуль ошибки квантования во времени — по формуле:

$$E_\varepsilon = |E_t - E_{sp}|. \quad (14)$$

Результаты расчетов в процессе уменьшения величин шагов квантования по двоичному закону ($h = 4, 2, 1$) сведены в таблицу 3. Отделы в таблице расположены в порядке возрастания частоты сокращений.

Таблица 3. Зависимость энергии отсчетов ЭГЭГ-сигналов от величин шагов выборки

Отдел ЖКТ	E_{sp}	E_t	E_ε	Относительное отклонение, %
$h = 4$				
Толстая	29,05033	29,06758	0,0172	0,06
Желудок	20,71735	21,14948	0,4317	2,04
Подвздошная	5,91264	6,68488	0,7723	11,5
Тошяя	1,45773	2,40063	0,9429	–
12-перстная	0,60609	2,09175	–	–
$h = 2$				
Толстая	23,28067	23,35799	0,0773	0,33
Желудок	16,20993	16,33852	0,1285	0,79
Подвздошная	4,8168	4,88408	0,0673	1,38
Тошяя	1,54968	1,62103	0,0713	4,40
12-перстная	0,99455	1,16445	0,1699	14,6
$h = 1$				
Толстая	29,0755	29,0843	0,008	0,03
Желудок	20,98009	21,0437	0,0636	0,30
Подвздошная	6,66409	6,67302	0,009	0,13
Тошяя	2,38623	2,3967	0,0105	0,43
12-перстная	2,03972	2,06419	0,0245	1,19

По таблице можно проследить, как влияет изменение частоты дискретизации на показатели точности восстановления гастроэнтерограмм кубическими сплайнами. Значения спектральных энергий при шаге $h = 4$ (это соответствует одному отсчету за 2 секунды) для двух наиболее высокочастотных энтерограмм (тощая кишка и 12-перстная кишка) свидетельствуют о том, что данный шаг вообще для них не годится. Это можно объяснить рассогласованием частоты дискретизации с собственными частотами сигналов, вследствие чего происходят потери информации и значительное уменьшение вычисленной спектральной мощности. Положение улучшается при $h = 2$. Относительные отклонения для данных энтерограмм получаются соответственно порядка 4% и 14%. Приемлемые результаты получаются при $h = 1$, что соответствует частоте дискретизации 2 отсчета в секунду.

6. Заключение. В работе предлагается новый метод выбора частоты отсчетов сигналов ЭГЭГ, учитывающий фактор конечной длительности сеансов измерений и спектральных свойств сигнала. Применяется разложение сигнала не в классический кардинальный ряд бесконечной длительности, а в виде конечной суммы финитных кардинальных В-сплайнов целых степеней. Известно, что оптимальной степенью в смысле наилучшей гладкости является третья [23]. Приведенные в статье результаты вычислительного эксперимента по нескольким каналам с применением аппроксимации кубическими сплайнами показали, что используемое в системе «Гастроскан-ГЭМ» отношение длительности сеанса и периода выборки сигналов позволяет снизить значения ошибок квантования во времени в некоторых каналах до уровня 0,03%.

Роль изложенного в данной статье принципа конечности энергии для решения проблемы оценки выбора шагов дискретизации непрерывных сигналов будет возрастать при анализе процессов с явными признаками нестационарности. При современных методах функциональной диагностики ЖКТ, позволяющих анализировать визуально и численно картины процессов в трехмерных пространствах «время — частота — уровень сигнала» повышается объективность оценок. Особую важность приобретают характеристики, отражающие точность оценок в области частотного аргумента. Этому способствуют математические методы оконного преобразования Фурье, а также вейвлет-преобразования.

Литература

1. *Зайченко К.В., Жаринов О.О., Кулин А.Н.* Съём и обработка биоэлектрических сигналов // СПб: РИО ГУАП. 2001. 140 с.
2. *Alvarez W.C.* The electrogastragram and what it shows // JAMA. 1922. vol. 78. pp. 1116–1119.

3. *Yin J., Chen J. D. Z.* Electrogastrography: Methodology, Validation and Applications // Journal of Neurogastroenterology and Motility. 2013. vol. 19. no. 1. pp. 5–17.
4. *Ребров В.Г.* Возможности электрогастроинтестинографии при ряде заболеваний желудка и кишечника // Терапевтический архив. 1981. № 10. С. 66–70.
5. *Kosenko P.M., Vavrinchuk S.A.* Electrogastroenterography in patients with complicated peptic ulcer // Science Book Publishing House. Yelm. USA. 2013. 164 p.
6. *Khan S., Ahmad M.* A study on B-spline wavelets and wavelet packets // Applied Mathematics. 2014. vol. 5. pp. 3001–3010.
7. *Свиных С.Ф., Попов А.И.* Финитные базисные функции в задачах формирования выборок сигналов конечной протяженности // Труды СПИИРАН. 2015. № 6. С. 50–67.
8. *Riezzo G., Russo F., Indrio F.* Electrogastrography in adults and children: the strength, pitfalls, and clinical significance of the cutaneous recording of the gastric electrical activity // BioMed research international. 2013. pp. 1–14.
9. Электрогастроэнтерография: исследование электрической активности желудка и кишечника. URL: <http://www.gastroscan.ru/physician/egg/> (дата обращения: 20.08.2016).
10. *Попов А.И., Тюльпин А.А., Рудалёв А.В.* Программная библиотека для цифровой обработки сигналов электрогастроэнтерографии // Информационно-измерительные и управляющие системы. 2014. №6. С. 40–45.
11. *Свиных С.Ф., Попов А.И., Рудалёв А.В.* Вейвлет-анализ и информационные технологии в задачах обработки электрогастроэнтерограмм // Труды СПИИРАН. 2013. № 27. С.129–143.
12. *Chen J., Lin Z., McCallum R.W.* Noninvasive feature-based detection of delayed gastric emptying in humans using neural networks // Biomedical Engineering, IEEE Transactions. 2000. vol. 47. no. 3. pp. 409–412.
13. *Liang H.* Application of support vector machine to the detection of delayed gastric emptying from electrogastrograms // Support Vector Machines: Theory and Applications. 2005. pp. 399–412.
14. *Kara S., Dirgenali F., Okkesim Ş.* Detection of gastric dysrhythmia using WT and ANN in diabetic gastroparesis patients // Computers in biology and medicine. 2006. vol. 36. no. 3. pp. 276–290.
15. *Косенко П.М. и др.* Математическое моделирование моторно-эвакуаторных нарушений желудочно-кишечного тракта у пациентов с язвенным пилородуоденальным стенозом // Новости хирургии. 2014. Т. 22. № 2. С. 224–229.
16. *Попов А.И., Косенко П.М.* Программа для диагностики степени компенсации стеноза по данным электрогастроэнтерографии. Свидетельство об официальной регистрации программы для ЭВМ № 2016614008. 2016.
17. *Никольский С.М.* Курс математического анализа // М.: Наука. 1973. Т. 2 392 с.
18. *Слепян А.Д.* О ширине полосы // ТИИЭР. 1976. Т. 64. № 3. С. 4–14.
19. *Yang Z., Chen X., He Z.* Wave propagation modeling in one-dimension structures be the B-spline wavelet on interval finite elements // Applied Mechanics and Materials. 2012. vol. 105. pp. 3–8.
20. *Свиных С.Ф.* Теория и методы формирования выборок сигналов с инфинитными спектрами // СПб: Наука. 2016. 72 с.
21. *Li X.* Numerical solution of fractional partial diffeferential equations using cubic B-spline wavelets collocation method // Australian Communications and Media Authority. 2012. vol. 1. no. 3. pp. 159–164.
22. *Гребенников А.И.* Метод сплайнов и решение некорректных задач теории приближений // М.: Изд-во Московского ун-та. 1983. 208 с.
23. *Rakowski W.* Prefiltering in Wavelet Analysis Applying Cubic B-Splines // Intern. journal of electronics and telecommunications. 2014. vol. 60. no. 4. pp. 331–340.

Попов Александр Игоревич — к-т техн. наук, доцент кафедры прикладной информатики, Северный (Арктический) федеральный университет имени М.В. Ломоносова (САФУ). Область научных интересов: цифровая обработка сигналов в электрофизиологии, информационные системы, автоматизация научных исследований. Число научных публикаций — 43. aleneus@gmail.com, <http://dsplab.narfu.ru>; Набережная Северной Двины, 17, Архангельск, 163000; р.т.: 8(8182)21-61-00, Факс: 8(8182)28-76-14.

Свиньин Сергей Федорович — д-р техн. наук, профессор, ведущий научный сотрудник лаборатории автоматизации научных исследований, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: цифровая обработка сигналов. Число научных публикаций — 160. svinyins@mail.ru; 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7(812)323-5139, Факс: +7(812)328-4450.

A.I. POPOV, S.F. SVINYIN
**APPLICATION OF FINITE BASIS SPLINES IN
 RECONSTRUCTING ELECTROGASTROENTEROGRAPHY**

Popov A.I., Svinyin S. Application of Finite basis Splines in Reconstructing Electrogastroenterography Signals.

Abstract. Electrogastroenterography is the promising method of examination of the motion activity of the digestive system. It is based on the measurement and further processing of bioelectric signals. During last years the progress in the development of electrophysiological methods of diagnostics is due to the computer processing of measuring signals. This paper is devoted to the aspects of organization of measurements in electrogastroenterography. In the paper, we present an introduction to the problem domain; analyze the information structure of a measuring signal; review the diagnostics parameters obtained as a result of spectral analysis of electrogastroenterography signals; discuss the tasks of automation of diagnostics.

We propose the new method of sampling of gastroenterograms. It considers the factor of the finite length of measurement sessions and spectral properties of signals. Representation of a signal as the finite sum of finite cardinal B-splines with integer degrees is used in the method. The computer experiment for testing the accuracy of signal reconstruction with parameters of measuring session used in electrogastroenterography was conducted.

Keywords: basis spline, finite energy, compact support, biosignal, automated processing of biomedical signals, electrophysiology, electrogastroenterography.

Popov Aleksandr Igorevich — Ph.D., associate professor, associate professor of applied informatics department, Northern (Arctic) Federal University. Research interests: digital signal processing in electrophysiology, information systems, automation of resresearches. The number of publications — 43. aleneus@gmail.com, <http://dsplab.narfu.ru>; 17, Northern Dvina Embankment, Arkhangelsk, 163000, Russia; office phone: 8(8182)21-61-00, Fax: 8(8182)287614.

Svinyin Sergey Fedorovich — Ph.D., Dr. Sci., associate professor, leading researcher of laboratory for research automation, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: digital processing of biomedical signals. The number of publications — 160. svinyins@mail.ru; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)323-5139, Fax: +7(812)328-4450.

References

1. Zajchenko K.V., Zharinov O.O., Kulin A.N. *Shem i obrabotka biojelektricheskikh signalov* [Bioelectric signal acquisition and processing]. SPb.: RIO GUAP. 2001. 140 p. (In Russ.).
2. Alvarez W.C. The electrogastrogram and what it shows. *JAMA*. 1922. vol. 78. pp. 1116–1119.
3. Yin J., Chen J. D. Z. *Electrogastrography: Methodology, Validation and Applications. Journal of Neurogastroenterology and Motility*. 2013. vol. 19. no. 1. pp. 5–17.
4. Rebrov V.G. [Abilities of electrogastrointestinography with some diseases of stomach and intestinal tract]. *Terapevticheskij arhiv – Therapeutic archives*. 1981. vol. 10. pp. 66–70. (In Russ.).
5. Kosenko P.M., Vavrinchuk S.A. *Electrogastroenterography in patients with complicated peptic ulcer*. Science Book Publishing House. Yelm. USA. 2013. 164 p.

6. Khan S., Ahmad M. A study on B-spline wavelets and wavelet packets. *Applied Mathematics*. 2014. vol. 5. pp. 3001–3010.
7. Svin'in S.F., Popov A.I. [Finite basic functions in the tasks of sampling signals of finite extension]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2015. vol. 6. pp. 50–67. (In Russ.).
8. Riezzo G., Russo F., Indrio F. Electrogastrography in adults and children: the strength, pitfalls, and clinical significance of the cutaneous recording of the gastric electrical activity. *BioMed research international*. 2013. pp. 1–14.
9. Jelektrogastrojenterografija: issledovanie jelektricheskoy aktivnosti zheludka i kishechnika [Analysis of electrical activity of the stomach and intestinal tract]. Available at: <http://www.gastroscan.ru/physician/egg/> (accessed 20.08.2016). (In Russ.).
10. Popov A.I., Tjul'pin A.A., Rudal'jov A.V. [Programme library for digital signal processing in electrogastroenterography]. *Informacionno-izmeritel'nye i upravljajushhie sistemy – Information-measuring and Control Systems*. 2014. vol. 6. pp. 40–45. (In Russ.).
11. Svin'in S.F., Popov A.I., Rudal'jov A.V. [Wavelet analysis and information technologies in problems of electrogastroenterograms processing]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2013. vol. 27. pp. 129–143. (In Russ.).
12. Chen J., Lin Z., McCallum R. W. Noninvasive feature-based detection of delayed gastric emptying in humans using neural networks. *Biomedical Engineering, IEEE Transactions*. 2000. vol. 47. no. 3. pp. 409–412.
13. Liang H. Application of support vector machine to the detection of delayed gastric emptying from electrogastrograms. *Support Vector Machines: Theory and Applications*. 2005. pp. 399–412.
14. Kara S., Dirgenali F., Okkesim Ş. Detection of gastric dysrhythmia using WT and ANN in diabetic gastroparesis patients. *Computers in biology and medicine*. 2006. vol. 36. no. 3. pp. 276–290.
15. Kosenko P.M. et al. [Mathematical modeling of the motor-evacuation disorders of the gastrointestinal tract in patients with ulcerous pyloric duodenal stenosis]. *Novosti hirurgii – Surgery news*. 2014. Issue 22. vol. 2. pp. 224–229. (In Russ.).
16. Popov A.I., Kosenko P.M. *Programma dlja diagnostiki stepeni kompensacii stenoza po dannym jelektrogastrojenterografii* [Software for diagnostics the degree of stenosis compensation by electrogastroenterography data]. Certificate of state registration of computer programs. no. 2016614008. 2016. (In Russ.).
17. Nikol'skij S.M. *Kurs matematicheskogo analiza*. [The course of mathematical analysis.]. M.: Nauka. 1973. Issue 2. 392 p. (In Russ.).
18. Slepjan A.D. [About bandwidth] *TIIER – Institute of engineers in electronics and radiotechnics proceedings*. 1976. vol. 64. no. 3. pp. 4–14. (In Russ.).
19. Yang Z., Chen X., He Z. Wave propagation modeling in one-dimension structures be the B-spline wavelet on interval finite elements. *Applied Mechanics and Materials*. 2012. vol. 105. pp. 3–8.
20. Svin'in S.F. *Teorija i metody formirovanija vyborok signalov s infinيتnymi spektrami* [Theory and methods of sampling signals with infinite spectra]. SPb: Nauka, 2016. 72 p. (In Russ.).
21. Li X. Numerical solution of fractional partial differential equations using cubic B-spline wavelets collocation method. *Australian Communications and Media Authority*. 2012. vol. 1. no. 3. pp. 159–164.
22. Grebennikov A.I. *Metod splajnov i reshenie nekorrektnyh zadach teorii priblizhenij* [Spline method and solving ill-posed problems of approximation theory] M.: Moscow State University. 1983. 208 p. (In Russ.).
23. Rakowski W. Prefiltering in Wavelet Analysis Applying Cubic B-Splines. *Intern. journal of electronics and telecommunications*. 2014. vol. 60. no. 4. pp. 331–340.

А.А. ЗУЕНКО, П.А. ЛОМОВ, А.Г. ОЛЕЙНИК
**ПРИМЕНЕНИЕ МЕТОДОВ РАСПРОСТРАНЕНИЯ
ОГРАНИЧЕНИЙ ДЛЯ УСКОРЕНИЯ ОБРАБОТКИ ЗАПРОСОВ
К ОНТОЛОГИЯМ**

Зуенко А.А., Ломов П.А., Олейник А.Г. **Применение методов распространения ограничений для ускорения обработки запросов к онтологиям.**

Аннотация. Предлагается рассматривать задачу обработки SPARQL-запросов к онтологии как задачу удовлетворения ограничений. Кратко представлен формальный аппарат, служащий для описания задач удовлетворения ограничений с помощью специализированных матрицеподобных структур. Применение данного аппарата позволяет более эффективно по сравнению с табличным представлением описывать, хранить и обрабатывать нечисловые ограничения предметной области. Для ускорения обработки запросов к онтологиям большого объема предлагается применить ранее разработанный авторами метод распространения нечисловых ограничений. По сравнению с традиционным подходом к обработке SPARQL-запросов, базирующимся на динамическом программировании, предлагаемый метод позволяет ускорить их выполнение за счет «компактного» представления онтологии, а также применения оригинальных авторских правил редукции пространства поиска. На конкретном примере показано использование этого математического аппарата для снижения размерности пространства поиска при выполнении запроса к онтологии междисциплинарных знаний.

Ключевые слова: запрос к онтологии, задача удовлетворения ограничений, матричное представление ограничений, распространение ограничений.

1. Введение. Инструменты для работы со структурами онтологических моделей позволяют сопоставлять и связывать различные онтологии, выбирать компоненты существующих онтологий для создания новых, что способствует унификации онтологических моделей и расширяет возможности их совместного использования. Важным аспектом при совместном использовании онтологий различных областей знания является обеспечение их семантической интероперабельности, основанной на одинаковой интерпретации различными онтологиями информационных объектов. Упростить решение проблемы семантической интероперабельности прикладных онтологий, которые предполагается интегрировать, на этапе их разработки позволяет применение шаблонов онтологического проектирования (Ontology Design Patterns, ODP) [1]. ODP положены в основу разрабатываемой авторами онтологии интегрированного пространства знаний (ИПЗ) об Арктической зоне Российской Федерации (АЗРФ). В [2] обоснована актуальность задачи формирования ИПЗ АЗРФ и рассмотрены подходы к ее решению на примере создания единого пространства мультипредметных знаний Кольского научного центра РАН.

Существующие и развиваемые технологии и средства решения проблем технологической и семантической интероперабельности предметных онтологий, объединяемых в ИПЗ АЗРФ, позволяют сформировать комплексную формализованную систему междисциплинарных знаний. Но пополнение этой системы новыми знаниями выводит на передний план еще одну проблему — проблему манипулирования представленными в этой системе знаниями, обусловленную снижением производительности запросов. Значимые задержки отклика системы на запросы проявляются, когда количество триплетов в онтологиях ИПЗ достигает 10 миллионов [3, 4].

В настоящей статье для ускорения обработки запросов к онтологиям большого объема предлагается применить ранее разработанные авторами методы распространения ограничений [5]. Кратко представлен формальный аппарат, служащий для описания с помощью матрицеподобных структур задач удовлетворения ограничений. Применение данного аппарата позволяет более эффективно, по сравнению с табличным представлением, описывать, хранить и обрабатывать нечисловые ограничения предметной области. На конкретном примере показано применение этого математического аппарата для снижения размерности пространства поиска при выполнении запроса к онтологии междисциплинарных знаний.

2. Онтология интегрированного пространства знаний. В качестве основы онтологии ИПЗ использована онтология информационных артефактов (Information Artifact Ontology, IAO) [6], которая, в свою очередь, является расширением одной из распространенных онтологий верхнего уровня — базовой формальной онтологии (Basic Formal Ontology, BFO) [7]. Онтология IAO ориентирована на представление процессов получения информации, их участников, а также информационных источников и носителей.

С целью упрощения применения онтологии ИПЗ было принято решение выделить целостные фрагменты понятийной системы IAO, представляющие отдельные концепты, дополнить их описание с учетом специфики проведения и документирования исследовательской деятельности в РФ и локализовать их в виде отдельных онтологических паттернов содержания (Ontology Content Design Patterns). Каждый такой паттерн реализован в виде мини-онтологии, которая отражает одну точку зрения на несколько связанных понятий предметной области. Использование паттернов для представления знаний в онтологии избавляет пользователя от манипуляции с большим количеством понятий и отношений [8], а также позволяет гарантировать качество такого представления ввиду того, что каждый паттерн представляет

доказавший свою эффективность вариант решения задачи онтологического моделирования. Общая структура онтологии ИПЗ представлена на рисунке 1.

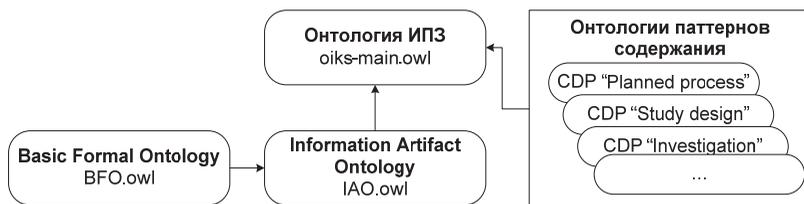


Рис. 1. Общая структура онтологии ИПЗ

В качестве примера рассмотрим один из паттернов разработанной онтологии ИПЗ, используемый для представления процессов исследований и соотносенных с ними объектов. UML-схема данного паттерна приводится на рисунке 2.

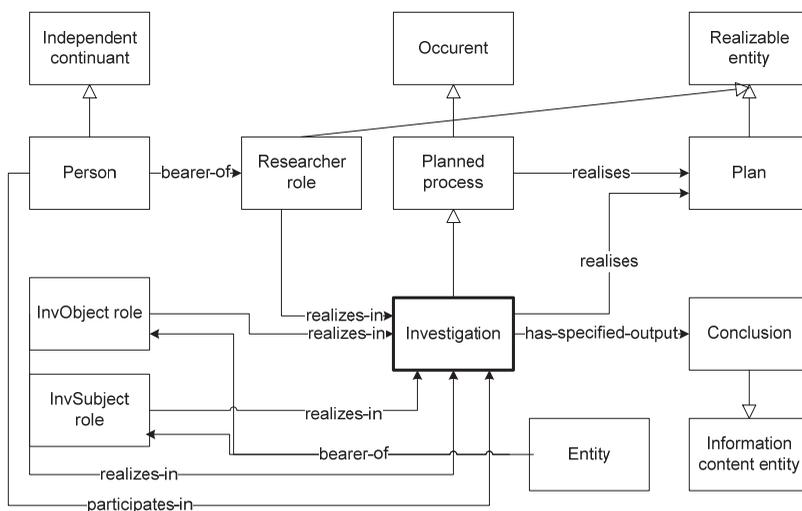


Рис. 2. UML-схема онтологического паттерна содержания «Исследование» (Investigation)

«Процесс исследования» (Investigation) включает в качестве участников (отношение participates-in) персон (Person), реализующих роль «Исследователь» (Researcher), и сущности, реализующие роли «Объект исследования» (Investigation object role) и «Предмет исследования» (Investigation subject role). В результате процесса

исследования продуцируется некоторое заключение (Conclusion), которое является «Информационным объектом» (Information content entity). Понятия «Независимая сущность» (Independent continuant), «Происходящее» (Occurrent), «Реализуемая сущность» (Realizable entity) принадлежат онтологии BFO.

В соответствии с технологией экстремального проектирования онтологий (eXtreme Design methodology, XD) [9, 10] для каждого паттерна содержания задается набор квалификационных вопросов (Competency questions), которые указывают, какую информацию можно получить с помощью онтологии, содержащей данный паттерн. Для рассматриваемого паттерна были определены следующие квалификационные вопросы:

- Кто участвовал в процессе исследования?
- Кто занимался исследованием данной сущности?
- Какие исследования проводились для данной сущности?
- С кем сотрудничал участник в процессе исследования?
- Какую роль выполнял участник исследования?
- Какой предмет в контексте данного объекта исследовался?
- Какие выводы были получены в результате исследования?

Рассмотренный паттерн содержания, а также полная онтология ИПЗ могут быть загружены из GIT-репозитория [11].

При работе с онтологией для каждого квалификационного вопроса формируется соответствующий запрос на специализированном языке SPARQL. Обеспечение приемлемой скорости исполнения таких запросов на больших объемах данных является одной из распространенных проблем [12-14]. В отличие от большинства работ-прототипов, например [4, 15], в настоящей статье предлагается задачу обработки запросов к онтологиям рассматривать не как задачу динамического программирования, а как задачу удовлетворения ограничений (программирования в ограничениях).

3. Задачи удовлетворения ограничений и матричное представление ограничений. Программирование в ограничениях сформировалось на стыке таких направлений, как искусственный интеллект, теория языков программирования, символьные вычисления и вычислительная логика. Технология программирования в ограничениях широко используется при решении многих прикладных задач, таких как календарное планирование, задачи назначения частоты, обработка изображений, тестирование сверхбольших интегральных схем, анализ языков программирования и ЗАДАЧИ понимания естественного языка и т.д. Система программирования в ограничениях стремится сократить перебор вариантов, для ускорения

вычислительных процедур используются методы распространения ограничений, эвристического поиска и т.п.

Дадим определения, необходимые для дальнейшего изложения.

Определение 1. Согласно [16, 17] *сеть ограничений* или *задача удовлетворения ограничений* (Constraint Satisfaction Problem — CSP) определена множеством переменных $V = \{x_1, x_2, \dots, x_n\}$ и множеством ограничений $C = \{C_1, C_2, \dots, C_m\}$. Каждая переменная x_i имеет непустую область определения D_i (домен). Каждое ограничение C_i включает некоторое подмножество переменных (диапазон отношения) и задает допустимые комбинации значений для этого подмножества.

Ограничением называется пара (R, S) , где R — отношение, определенное на диапазоне S . Решение задачи удовлетворения ограничений состоит в присваивании значения каждой переменной так, чтобы одновременно удовлетворялись все ограничения.

Ограничение может быть представлено либо в явной форме путем перечисления кортежей значений, удовлетворяющих заданному условию, либо в виде абстрактного отношения, для которого вводится две операции: проверка кортежа на принадлежность отношению и перечисление элементов отношения [18]. В первом случае ограничения могут задаваться с помощью таблиц, описывающих допустимые сочетания значений переменных [19]. При этом вычисления, необходимые для решения CSP, удобно выполнять с использованием операций в стиле реляционной алгебры. Часть из этих операций предполагают, что отношения сформированы в одной схеме (один диапазон), другие же операции служат для работы с отношениями, сформированными в разных схемах.

Пусть R_1, R_2 — два отношения с одинаковым диапазоном. Тогда к ним можно применить следующие операции: *пересечение* $R_1 \cap R_2$, *объединение* $R_1 \cup R_2$, *разность* $R_1 \setminus R_2$, *дополнение* $U \setminus R_1$ отношения R_1 относительно универсума U (универсального отношения).

Также введем операцию соединения многоместных отношений.

Определение 2. *Соединение* $R_S \bowtie R_T$ двух отношений R_S (с диапазоном S) и R_T (с диапазоном T) является отношением с диапазоном $S \cup T$ и множеством всех кортежей $(a_1, \dots, a_n, b_1, \dots, b_m, c_1, \dots, c_k)$, получаемых слиянием кортежей $(a_1, \dots, a_n, b_1, \dots, b_m) \in R_S$ и $(b_1, \dots, b_m, c_1, \dots, c_k) \in R_T$, где компоненты b_1, \dots, b_m являются значениями переменных из множества $S \cap T$.

Качественные ограничения, такие как правила, логические формулы, многоместные отношения играют важную роль при моделировании многих предметных областей. Существующие на данный момент в теории удовлетворения ограничений методы, основанные на табличном представлении нечисловых ограничений, недостаточно эффективны. В настоящей работе для представления и обработки нечисловых ограничений предлагается использовать два типа матрицеподобных структур: *C*-системы и *D*-системы [20, 21].

Особенностью настоящих исследований является то, что эти структуры предлагается рассматривать как ограничения над конечными доменами, а рассуждения на данных структурах реализовывать в форме процедур удовлетворения ограничений.

С помощью *C*-систем удобно моделировать дизъюнктивные нормальные формы (ДНФ) конечных предикатов. Продемонстрируем это на примере. Пусть задан конечный предикат:

$$\varphi(x, y, z) = (x = a, b) \wedge (y = a, c) \vee (z = d).$$

Для простоты все переменные определены на множестве $\{a, b, c, d\}$. Здесь и далее будем использовать запись вида $(x = a, b)$ для обозначения выражения $(x = a) \vee (x = b)$. Учитывая, что область истинности одноместного предиката $(x = a, b)$ есть $\{a, b\}$, то область истинности предиката $\varphi(x, y, z)$ может быть представлена в виде следующей *C*-системы:

$$R[XYZ] = \begin{bmatrix} \{a, b\} & \{a, c\} & * \\ * & * & \{d\} \end{bmatrix}.$$

Атрибуты X , Y , Z отношения $R[XYZ]$ соответствуют переменным x , y , z формулы $\varphi(x, y, z)$. Заметим, что “*” — сокращенное обозначение всего домена атрибута. *C*-систему $R[XYZ]$ можно преобразовать в многоместное отношение следующим образом:

$$(\{a, b\} \times \{a, c\} \times \{a, b, c, d\}) \cup (\{a, b, c, d\} \times \{a, b, c, d\} \times \{d\}).$$

С помощью *D*-систем моделируются конъюнктивные нормальные формы (КНФ) конечных предикатов. *D*-система записывается как матрица компонент-множеств, которые заключены в перевернутые скобки. *D*-системы позволяют легко вычислять дополнение *C*-систем: берется дополнение каждой компоненты.

Например, $\neg\varphi \equiv (\neg(x = a, b) \vee \neg(y = a, c)) \wedge \neg(z = d)$, что с учетом конечных областей определения переменных равносильно $\neg\varphi \equiv ((x = c, d) \vee (y = b, d)) \wedge (z = a, b, c)$. Предикат $\neg\varphi$ может быть выражен как D -система $\bar{R}[XYZ]$:

$$\bar{R}[XYZ] = \begin{bmatrix} \{c, d\} & \{b, d\} & \emptyset \\ \emptyset & \emptyset & \{a, b, c\} \end{bmatrix}.$$

\emptyset — это фиктивная компонента, не содержащая значений.

В работах [5, 6] приводятся конкретные методы решения задач CSP, основанные на матричном представлении ограничений с конечными доменами в виде D -систем. Данные методы опираются на утверждения, позволяющие зачастую редуцировать пространство поиска без организации ветвления.

С однотипными матрицами ограничений, заданными внутри одного и того же диапазона (в одной схеме отношения), можно выполнять любые операции алгебры множеств. При этом нет необходимости преобразовывать их в совокупность элементарных кортежей — все операции выполняются с компонентами атрибутов или кортежами компонент. Это свойство позволяет существенно уменьшить трудоемкость вычислений. Операции (пересечение, объединение, дополнение) для матрицеподобных структур осуществляются на основании теорем, изложенных в [20]. Ввиду ограничений на объем публикации здесь приводится лишь теорема, непосредственно используемая в дальнейшем.

Теорема 1. Пересечение двух C -систем с одинаковыми схемами равно C -системе, содержащей все непустые пересечения каждой строки первой C -системы с каждой строкой второй C -системы.

Пусть заданы C -системы:

$$R_1[XYZ] = \begin{bmatrix} \{a, b, d\} & \{f, h\} & \{b\} \\ \{b, c\} & * & \{a, c\} \end{bmatrix},$$

$$R_2[XYZ] = \begin{bmatrix} \{a, d\} & * & \{b, c\} \\ \{b, d\} & \{f, h\} & \{a, c\} \\ \{b, c\} & \{g\} & \{b\} \end{bmatrix}.$$

Оценим количество операций сравнения элементов при пересечении отношений в двух случаях: 1) отношения представлены через элементарные кортежи; 2) отношения выражены в виде C -систем.

В первом случае имеем: количество операций = $18 \times 22 \times 3 = 1188$, поскольку отношение R_1 содержит 18 элементарных кортежей, а отношение R_2 — 22 кортежа, причем каждый из элементарных кортежей содержит по 3 элемента. Во втором случае пересечение отношений R_1 и R_2 вычисляется как результат пересечения всех пар C -кортежей, содержащихся в этих C -системах:

$$\begin{aligned} & [\{a, b, d\} \{f, h\} \{b\}] \cap [\{a, d\} * \{b, c\}] = [\{a, d\} \{f, h\} \{b\}]; \\ & [\{a, b, d\} \{f, h\} \{b\}] \cap [\{b, d\} \{f, h\} \{a, c\}] = \emptyset; \\ & [\{a, b, d\} \{f, h\} \{b\}] \cap [\{b, c\} \{g\} \{b\}] = \emptyset; \\ & [\{b, c\} * \{a, c\}] \cap [\{a, d\} * \{b, c\}] = \emptyset; \\ & [\{b, c\} * \{a, c\}] \cap [\{b, d\} \{f, h\} \{a, c\}] = [\{b\} \{f, h\} \{a, c\}]; \\ & [\{b, c\} * \{a, c\}] \cap [\{b, c\} \{g\} \{b\}] = \emptyset. \end{aligned}$$

Затем из непустых C -кортежей формируем итоговую C -систему:

$$R_1 \cap R_2 = \left[\begin{array}{ccc} \{a, d\} & \{f, h\} & \{b\} \\ \{b\} & \{f, h\} & \{a, c\} \end{array} \right].$$

В этом случае получим следующую оценку: количество операций = $(3 \times 2 + 2 \times 3 + 1 \times 2) + (3 \times 2 + 2 \times 2 + 1 \times 2) + (3 \times 2 + 2 \times 1 + 1 \times 1) + (2 \times 2 + 3 \times 3 + 2 \times 2) + (2 \times 2 + 3 \times 2 + 2 \times 2) + (2 \times 2 + 3 \times 1 + 2 \times 1) = 75$. Каждое слагаемое, заключенное в скобки, характеризует пересечение определенной пары C -кортежей и вычисляется как сумма произведений мощностей их соответствующих компонент. Видно, что представление отношений в виде матрицеподобных структур способно существенно ускорить выполнение операций с отношениями.

Как будет показано далее, при использовании матрицеподобных структур операция соединения отношений также упрощается подобно рассмотренной ранее операции пересечения, и ее результат можно вычислить без попарного сравнения всех элементарных кортежей.

Среди разработанных алгоритмов удовлетворения нечисловых ограничений особый интерес представляют алгоритмы распространения ограничений, которые позволяют преобразовать описание задачи CSP за полиномиальное время к более простому виду с сохранением всех решений. Для C - и D -систем это означает, что из доменов переменных удаляются «лишние» значения, могут быть

удалены некоторые значения из компонент, некоторые столбцы и/или строки *C*-систем и *D*-систем, и даже *C*- и *D*-системы целиком.

4. Вывод на совокупности *C*-систем. Приведем утверждения, позволяющие реализовывать эквивалентные преобразования системы ограничений, представленной в виде совокупности *C*-систем [22]. Ранее эти утверждения были сформулированы и применены одним из авторов настоящей статьи для качественного моделирования статических аспектов функционирования технических систем.

Утверждение 1 (У1). Если все строки (кортежи) *C*-системы пусты, то есть содержат хотя бы по одной пустой компоненте каждая, то *C*-система пуста (соответствующая задача CSP несовместна).

Утверждение 2 (У2). Если все компоненты некоторого атрибута (столбца *C*-системы) являются полными, то данный атрибут можно удалить из *C*-системы (удаляются все компоненты, стоящие в соответствующем столбце), а пара «удаляемый атрибут – его домен» сохраняется в векторе частичного решения.

Утверждение 3 (У3). Если домен некоторого атрибута *C*-системы содержит значения, не встречающиеся в соответствующем столбце, то эти значения удаляются из данного домена.

Утверждение 4 (У4). Если строка *C*-системы содержит хотя бы одну пустую компоненту (строка пуста), то строка удаляется.

Утверждение 5 (У5). Если компонента некоторого атрибута содержит значение, не принадлежащее соответствующему домену, то это значение удаляется из компоненты.

Утверждение 6 (У6). Если одна строка *C*-системы полностью доминирует (покомпонентно содержит) другую строку, то доминируемая строка удаляется из *C*-системы.

Рассмотрим применение описанных утверждений при выполнении SPARQL запросов к онтологии ИПЗ, фрагмент которой показан на рисунке 2.

5. Обработка SPARQL-запроса к онтологии с помощью матриц ограничений. При использовании SPARQL онтология рассматривается как RDF-документ, состоящий из набора триплетов вида: «субъект — свойство — объект». Например, «Иванов-В-И — участвует-в — Исследование№213», «Иванов-В-И — выполняет-роль — Координатор». Каждый триплет можно рассматривать как дугу некоторого графа, а весь RDF-документ представлять в виде графа.

Типовой SPARQL-запрос на извлечение данных представляет собой конъюнкцию и/или дизъюнкцию шаблонов триплетов, в соответствии с которыми из онтологии выбираются удовлетворяющие им наборы триплетов, являющиеся решениями (Solutions). Решения

представляют собой подграфы графа онтологии. Например, запрос, состоящий из конъюнкции шаблонов: «?*x* — является-экземпляром — Персона», «?*x* — участвует-в — Исследование № 213» и «?*x* — выполняет-роль — Координатор», позволяет найти подграфы, в которых вершина (?*x*) связана дугами «является-экземпляром», «участвует-в» и «выполняет-роль» с вершинами «Персона», «Исследование № 213» и «Координатор» соответственно. Выполнение запросов осуществляется обработчиком запросов, который обычно является компонентом хранилища триплетов (RDF — triple store). Наличие в запросе длинной связки шаблонов триплетов негативно влияет на скорость его выполнения [23].

Управление большим RDF-хранилищем представляет собой ресурсоемкую задачу с точки зрения физической организации хранилища, проведения индексирования триплетов, а также выполнения запросов. Самым простым решением является хранение всего множества триплетов в одной таблице, но с возрастанием объема данных и сложности запросов, включающих конъюнкцию нескольких шаблонов, появляются проблемы с производительностью, так как для ответа на запрос перебирается большое число связок триплетов. По этой причине обычно используют более сложные способы группировки триплетов в таблицы (cluster-property table), соответствующие:

- 1) одному свойству (предикату);
- 2) свойству (предикату) для сущностей одного класса;
- 3) комбинациям субъектов, предикатов и объектов, которые часто встречаются в большинстве запросов.

Помимо организации структуры RDF-хранилища, возникают сложности с оптимизацией выполнения запросов. При их выполнении приходится оперировать не записями, как в реляционной базе данных, а связками (join) триплетов различной длины, включающих труднопредсказуемые комбинации предикатов. С целью повышения производительности обработчик запроса задействует оптимизатор, который выбирает наиболее дешевый план выполнения запроса на основе функции оценки стоимости. Обычно данная функция учитывает количество промежуточных результатов, порождаемых связкой шаблонов триплетов в том или ином плане выполнения.

Представление RDF-хранилища в рамках предлагаемого подхода способно сократить расход памяти компьютера для хранения RDF-триплетов, а также повысить эффективность обработки SPARQL-запросов за счет структуризации знаний в виде специализированных матрицеподобных структур.

Рассмотрим применение методов распространения ограничений, разработанных ранее одним из авторов статьи, для ускорения обработки

SPARQL-запросов к онтологиям на примере получения ответа на квалификационный вопрос онтологического паттерна содержания для понятия «Процесс исследования» (Investigation): «Кто занимался исследованием магнитного поля (magnetic field)?». В соответствии с UML-схемой (рисунок 2) паттерна «Исследование» (Investigation), необходимо найти персон (экземпляры класса «Person»), реализовавших (отношение «realizes-in») свои роли исследователей (экземпляры класса «Researcher-role») в процессе исследования (экземпляры класса «Investigation»), в котором также была реализована роль объекта исследования (экземпляры класса «InvObjectRole»), носителем которой (отношение «bearer-of») выступил объект «magnetic-field» — экземпляр класса «Entity».

Соответствующий SPARQL-запрос будет иметь вид (листинг 1):

```
PREFIX IKS: < http://www.iimm.ru/ontologies/oiks>
PREFIX INV: < http://www.iimm.ru/ontologies/cdp/investigation-process>
SELECT ?person
WHERE {
  ?person rdf:type IKS:Person.
  ?person IKS:bearer-of ?role.
  ?role rdf:type INV:Researcher-role.
  ?role IKS:realizes-in ?inv.
  ?inv IKS: realizes ?objRole.
  ?objRole IKS:bearer-of "magnetic-field".
}
```

Листинг 1. Пример SPARQL-запроса

Рассмотрим его выполнение на конкретном наборе данных (рисунок 3).

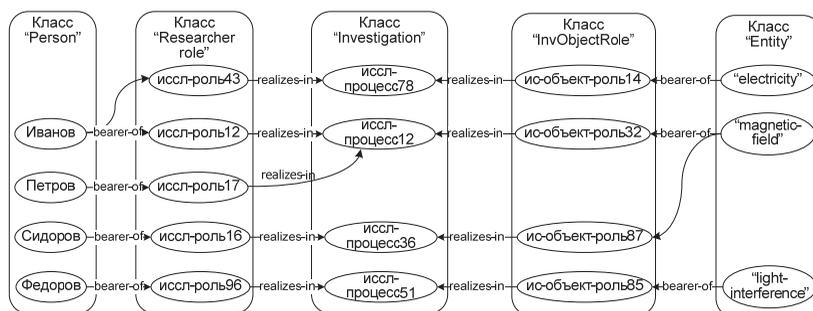


Рис. 3. Примерный фрагмент онтологии ИПЗ

Приведенный фрагмент онтологии включает набор экземпляров, связанных отношениями, используемыми для описания понятия

«Исследование» (Investigation). Для наглядности во фрагменте присутствуют только те классы и отношения, которые представлены в рассматриваемом SPARQL-запросе.

Для классов, экземпляров классов и связей показанного на рисунке 3 фрагмента онтологии введем специальные обозначения. Каждому классу фрагмента онтологии сопоставим определенную переменную (атрибут), а каждому экземпляру класса определенное значение данной переменной так, как это показано в таблице 1.

Таблица 1. Обозначение классов и экземпляров классов

Названия классов и экземпляров классов	Обозначение классов	Обозначение значений переменных (экземпляров классов)
КЛАСС «PERSON» «Иванов» «Петров» «Сидоров» «Федоров»	<i>X</i>	<i>a</i> ₁ <i>a</i> ₂ <i>a</i> ₃ <i>a</i> ₄
КЛАСС «RESEARCHER_ROLE» «иссл-роль43» «иссл-роль12» «иссл-роль17» «иссл-роль16» «иссл-роль96»	<i>Y</i>	<i>b</i> ₁ <i>b</i> ₂ <i>b</i> ₃ <i>b</i> ₄ <i>b</i> ₅
КЛАСС «INVESTIGATION» «иссл-процесс78» «иссл-процесс12» «иссл-процесс36» «иссл-процесс51»	<i>Z</i>	<i>c</i> ₁ <i>c</i> ₂ <i>c</i> ₃ <i>c</i> ₄
КЛАСС «INV_OBJ_ROLE» «исс-объ-роль14» «исс-объ-роль32» «исс-объ-роль87» «исс-объ-роль85»	<i>W</i>	<i>d</i> ₁ <i>d</i> ₂ <i>d</i> ₃ <i>d</i> ₄
КЛАСС «ENTITY» «electricity» «magnetic-field» «light-interference»	<i>L</i>	<i>e</i> ₁ <i>e</i> ₂ <i>e</i> ₃

Тогда в пределах рассматриваемого фрагмента получим следующие соответствия «атрибут — домен атрибута»:

$$X - \{a_1, a_2, a_3, a_4\}, Y - \{b_1, b_2, b_3, b_4, b_5\}, Z - \{c_1, c_2, c_3, c_4\}, W - \{d_1, d_2, d_3, d_4\}, L - \{e_1, e_2, e_3\}.$$

Теперь каждой связи онтологии, которая фигурирует в рассматриваемом SPARQL-запросе, сопоставим ограничение:

$R_1[XY]$ — отношение «bearer-of» между классами X («PERSON») и Y («RESEARCHER_ROLE»);

$R_2[YZ]$ — отношение «realizes-in» между классами Y («RESEARCHER_ROLE») и Z («INVESTIGATION»);

$R_3[WZ]$ — отношение «realizes-in» между классами W («INV_OBJ_ROLE») и Z («INVESTIGATION»);

$R_4[LW]$ — отношение «bearer-of» между классами L («ENTITY») и W («INV_OBJ_ROLE»);

Также для моделирования селектора SPARQL-запроса введем следующее унарное ограничение:

$R_5[L]$ — ограничение на переменную L (КЛАСС «ENTITY»), задающее условие фильтрации (селектор) в SPARQL-запросе.

На рисунке 3 перечисленные отношения $R_1 — R_4$ обозначены стрелками. В соответствии с рисунком 3, учитывая конкретные экземпляры классов, можно записать перечисленные выше отношения в виде предлагаемых матриц ограничений, а именно в виде C -систем:

Отношение $R_1[XY]$:

$$\begin{bmatrix} \{a_1\} & \{b_1, b_2\} \\ \{a_2\} & \{b_3\} \\ \{a_3\} & \{b_4\} \\ \{a_4\} & \{b_5\} \end{bmatrix}.$$

Отношение $R_2[YZ]$:

$$\begin{bmatrix} \{b_1\} & \{c_1\} \\ \{b_2, b_3\} & \{c_2\} \\ \{b_4\} & \{c_3\} \\ \{b_5\} & \{c_4\} \end{bmatrix}.$$

Отношение $R_3[WZ]$:

$$\begin{bmatrix} \{d_1\} & \{c_1\} \\ \{d_2\} & \{c_2\} \\ \{d_3\} & \{c_3\} \\ \{d_4\} & \{c_4\} \end{bmatrix}.$$

Отношение $R_4[LW]$:

$$\begin{bmatrix} \{e_1\} & \{d_1\} \\ \{e_2\} & \{d_2, d_3\} \\ \{e_3\} & \{d_4\} \end{bmatrix}.$$

На примере матрицы R_1 поясним, что обозначает подобная запись. Первая строка данной матрицы соответствует двум стрелкам (дугам) на рисунке 3, проведенным от экземпляра «Иванов» (значение a_1) класса «PERSON» (переменная X) к экземплярам «иссл-роль43» (значение b_1) и «иссл-роль12» (значение b_2) класса «RESEARCHER_ROLE» (переменная Y). По аналогии, каждой строке всех перечисленных C -систем соответствует некоторое множество дуг между экземплярами соответствующих классов онтологии, изображенных на рисунке 3.

Кроме того, известно, что на переменную L в SPARQL-запросе наложено еще одно *дополнительное ограничение* $R_5[L]$: $R_5[L] = [\{e_2\}]$.

За счет представления RDF-триплетов в виде совокупности матрицеподобных структур повышается эффективность обработки запросов к онтологии, как это будет продемонстрировано далее.

В настоящей работе предлагается задачу обработки запроса к онтологии рассматривать как задачу удовлетворения ограничений. Приводимый ранее пример запроса в терминах матриц ограничений будет выглядеть следующим образом:

$$Pr_X (R_1[XY] \bowtie R_2[YZ] \bowtie R_3[WZ] \bowtie R_4[LW] \bowtie R_5[L]).$$

Представим SPARQL-запрос в виде сети ограничений (рисунок 4):

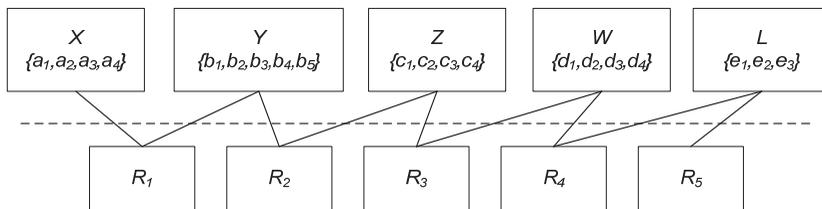


Рис. 4. Исходная задача CSP

Обработка запроса производится на основе приведенных ранее утверждений для совокупности S -систем.

Сеть ограничений является двудольным графом: одна из долей графа — это множество переменных и текущие области определения этих переменных (домены переменных), другая доля — сами ограничения. Ребро связывает две вершины (переменную и ограничение) тогда и только тогда, когда эта переменная формирует диапазон данного ограничения (входит в схему данного отношения).

Далее проиллюстрируем применение авторского метода распространения ограничений к обработке данного запроса, то есть для решения представленной задачи CSP.

При распространении ограничений обычно сначала обрабатывают унарные ограничения, поскольку с их помощью можно непосредственно «сузить» домены переменных. Затем уже активируются бинарные ограничения, поскольку их обработка немного более трудоемка. Алгоритмы удовлетворения унарных и бинарных ограничений выполняются за полиномиальное время, поэтому задачи с множественными отношениями часто сводят к совокупности унарных и бинарных отношений. Активация ограничений осуществляется по событиям. К таким событиям относится «усечение» домена. Вполне типична ситуация, когда на основе анализа определенного ограничения «усекается» домен некоторой переменной, а затем активируется другое ограничение, содержащее в схеме данную переменную. В ходе анализа вновь активированного ограничения может быть «усечен» домен другой переменной, состоящей в том же отношении.

Для нашего примера распишем данную схему вывода по шагам.

Сначала активируется унарное ограничение $R_5[L]$. Удаляем из домена переменной L все значения, кроме значения e_2 (применяем У3). После чего само ограничение $R_5[L]$ исключается из рассмотрения по У2, поскольку единственная компонента этой S -системы становится полной (совпадает с новым доменом переменной L).

Поскольку рассматриваемая задача CSP не содержит больше унарных ограничений (все унарные ограничения уже удовлетворены),

то следующим активируется одно из бинарных ограничений. На текущем шаге выбирается ограничение $R_4[LW]$, так как домен его атрибута L был изменен на предыдущем этапе вывода:

$$\begin{array}{cc} L & W \\ \{e_2\} & \{d_1, d_2, d_3, d_4\} \\ 1 \left[\begin{array}{cc} \{e_1\} & \{d_1\} \\ \{e_2\} & \{d_2, d_3\} \\ \{e_3\} & \{d_4\} \end{array} \right]. \end{array}$$

Для удобства объяснения здесь и далее в верхних двух строках C -системы записываются имена атрибутов и домены атрибутов, а строки C -системы пронумерованы.

Из компонент первого столбца данной C -системы, применяя У5, исключаем значения, не принадлежащие усеченному домену атрибута L . В результате в первом столбце в строках 1 и 3 образуются пустые компоненты, то есть компоненты, не содержащие ни одного значения. Далее, пользуясь У4, удаляем строки 1 и 3. В процессе упрощений ограничение $R_4[LW]$ преобразуется в ограничение:

$$\begin{array}{cc} L & W \\ \{e_2\} & \{d_1, d_2, d_3, d_4\} \\ 2 \left[\begin{array}{cc} * & \{d_2, d_3\} \end{array} \right]. \end{array}$$

Согласно У3, удаляем из домена атрибута W значения, отличные от d_2 и d_3 . После чего обе компоненты второй строки становятся полными и по У2 можно элиминировать оба столбца рассматриваемой C -системы. Значит, данное ограничение исключается из рассмотрения. Сеть ограничений будет выглядеть следующим образом (рисунок 5):

Частичное решение: $L — \{e_2\}$.

Остаток:

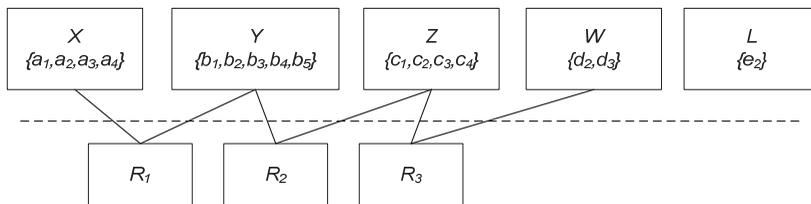


Рис. 5. Редуцированная сеть ограничений

Далее количество вершин и ребер в сети ограничений не изменяется, поэтому дальнейшие шаги процедуры распространения не будем представлять графически.

Следующим активируется ограничение $R_3[WZ]$, поскольку был изменен домен атрибута W , присутствующий в схеме данного отношения. Отношение $R_3[WZ]$ на текущем шаге выглядит так:

$$\begin{array}{cc}
 W & Z \\
 \{d_2, d_3\} & \{c_1, c_2, c_3, c_4\} \\
 1 \left[\begin{array}{cc} \{d_1\} & \{c_1\} \end{array} \right] \\
 2 \left[\begin{array}{cc} \{d_2\} & \{c_2\} \end{array} \right] \\
 3 \left[\begin{array}{cc} \{d_3\} & \{c_3\} \end{array} \right] \\
 4 \left[\begin{array}{cc} \{d_4\} & \{c_4\} \end{array} \right]
 \end{array}$$

Применяем У5 для того, чтобы исключить из первого столбца значения, не принадлежащие домену атрибута W . В первом столбце в строках 1 и 4 образуются пустые компоненты, что позволяет удалить данные строки из рассматриваемой C -системы, пользуясь У4.

C -система преобразуется к виду:

$$\begin{array}{cc}
 W & Z \\
 \{d_2, d_3\} & \{c_1, c_2, c_3, c_4\} \\
 2 \left[\begin{array}{cc} \{d_2\} & \{c_2\} \end{array} \right] \\
 3 \left[\begin{array}{cc} \{d_3\} & \{c_3\} \end{array} \right]
 \end{array}$$

Анализируя второй столбец, можно сузить домен переменной Z до множества $\{c_2, c_3\}$ (утверждение У3). Более никаких операций упрощения к данной C -системе применить нельзя.

Поскольку последние изменения коснулись домена атрибута Z , то следующим активируется ограничение $R_2[YZ]$, содержащее в схеме отношения данный атрибут:

$$\begin{array}{cc}
 Y & Z \\
 \{b_1, b_2, b_3, b_4, b_5\} & \{c_2, c_3\} \\
 1 \left[\begin{array}{cc} \{b_1\} & \{c_1\} \\ \{b_2, b_3\} & \{c_2\} \\ \{b_4\} & \{c_3\} \\ \{b_5\} & \{c_4\} \end{array} \right] .
 \end{array}$$

Пользуясь утверждением У5, исключаем из второго столбца матрицы ограничений значения, которые не принадлежат новому домену переменной Z . В результате строки 1 и 4 матрицы элиминируются. Анализ первых компонент двух оставшихся строк 2 и 3 позволяет на основании У3 сузить домен атрибута Y до множества $\{b_2, b_3, b_4\}$. Тогда ограничение $R_2[YZ]$ примет вид:

$$\begin{array}{cc}
 Y & Z \\
 \{b_2, b_3, b_4\} & \{c_2, c_3\} \\
 2 \left[\begin{array}{cc} \{b_2, b_3\} & \{c_2\} \\ \{b_4\} & \{c_3\} \end{array} \right] .
 \end{array}$$

На заключительном этапе активируется ограничение $R_1[XY]$. После применения правил редукции, ограничение $R_1[XY]$ примет вид:

$$\begin{array}{cc}
 X & Y \\
 \{a_1, a_2, a_3\} & \{b_2, b_3, b_4\} \\
 1 \left[\begin{array}{cc} \{a_1\} & \{b_2\} \\ \{a_2\} & \{b_3\} \\ \{a_3\} & \{b_4\} \end{array} \right] .
 \end{array}$$

В результате получим следующую упрощенную задачу CSP, к которой более не применить ни одного из представленных ранее утверждений (правил редукции), то есть процесс распространения ограничений останавливается, достигнув неподвижной точки:

Области определения переменных $X - \{a_1, a_2, a_3\}$, $Y - \{b_2, b_3, b_4\}$, $Z - \{c_2, c_3\}$, $W - \{d_1, d_2\}$, $L - \{e_2\}$.

Ограничения

Отношение $R_1[XY]$:

$$\begin{bmatrix} \{a_1\} & \{b_2\} \\ \{a_2\} & \{b_3\} \\ \{a_3\} & \{b_4\} \end{bmatrix}.$$

Отношение $R_2[YZ]$:

$$\begin{bmatrix} \{b_2, b_3\} & \{c_2\} \\ \{b_4\} & \{c_3\} \end{bmatrix}.$$

Отношение $R_3[WZ]$:

$$\begin{bmatrix} \{d_2\} & \{c_2\} \\ \{d_3\} & \{c_3\} \end{bmatrix}.$$

Таким образом, процедура распространения ограничений позволила получить более простую задачу CSP по сравнению с исходной: 1) количество ограничений уменьшилось с пяти до трех; 2) суммарное количество значений в доменах переменных уменьшилось с $4 + 5 + 4 + 4 + 3 = 20$ до $3 + 3 + 2 + 2 + 1 = 11$; 3) в каждом из оставшихся ограничений стало меньше строк; 4) некоторые из оставшихся строк содержат редуцированные компоненты, в частности вторая компонента первой строки первого отношения $R_1[XY]$.

При необходимости выполнить соединение упрощенных отношений $R_1[XY]$, $R_2[YZ]$ и $R_3[WZ]$ потребуется проспрягать всего $3 \times 2 \times 2 = 12$ строк C -систем в отличие от первоначальной задачи CSP, где требовалось $4 \times 4 \times 4 \times 3 \times 1 = 192$ подобных операций.

Представленный пример наглядно доказывает, что препроцессинг задачи на основе авторских методов удовлетворения ограничений позволяет существенно снизить размерность решаемой задачи, а иногда и получить ее окончательное решение. В частности, в рассмотренном нами примере SPARQL-запроса в процессе распространения ограничений получено окончательное решение: $X - \{a_1, a_2, a_3\}$. Интересующими нас исследователями магнитного поля являются: Иванов, Петров, Сидоров.

Численные результаты экспериментальной проверки подтвердили полученные в ходе теоретического исследования выводы. В настоящей работе они не приводятся ввиду ограничений на объем публикации.

6. Заключение. Использование онтологических паттернов для формирования онтологий различных областей знания, ориентированных на включение в интегрированное пространство междисциплинарных знаний, существенно облегчает решение задачи обеспечения семантической интероперабельности интегрируемых онтологий. Однако при этом в рамках ИПЗ ожидается быстрый рост числа представленных в онтологиях сущностей и связей, что влечет снижение производительности при обработке запросов к ИПЗ. По сравнению с традиционным подходом к обработке SPARQL-запросов, базирующимся на динамическом программировании, предлагаемый в статье метод обработки запросов позволяет ускорить их выполнение за счет использования специальных матрицеподобных структур для «компактного» представления онтологии, а также применения авторских методов удовлетворения ограничений для организации эффективного вывода на онтологии.

В дальнейшем предполагается модифицировать контекстно-ориентированный подход к управлению ограничениями [24, 25] на задачу интеграции разнородных знаний и данных с целью повышения релевантности извлекаемой информации и скорости обработки запросов конечного пользователя за счет активации только актуальных текущей ситуации компонентов онтологической модели.

Литература

1. *Gangemi A.* Ontology Design Patterns for Semantic Web Content // Proceedings of the Fourth International Semantic Web Conference. Galway. Ireland. Springer. 2005. pp. 262–276.
2. *Олейник А.Г., Ломов П.А.* О формировании единого пространства мультипредметных знаний Кольского научного центра РАН // Труды пятой международной конференции «Системный анализ и информационные технологии» (САИТ-2013). 2013. Т.1. С. 258–265.
3. *Nikolić N. et al.* RDF Stores Performance Test on Servers with Average Specification // Proceedings of the 5th International Conference on Information Science and Technology (ICIST). 2015. pp. 67–72.
4. *Neumann T., Weikum G.* RDF3X: a RISCstyle engine for RDF // Proceedings of the 34th International Conference on Very Large Data Bases (VLDB). Auckland. New Zealand. 2008. pp. 647–659.
5. *Зуенко А.А.* Вывод на ограничениях с применением матричного представления конечных предикатов // Искусственный интеллект и принятие решений. 2014. № 3. С. 21–31.
6. *Smith B. et al.* IAO-Intel: An Ontology of Information Artifacts in the Intelligence Domain // Proceedings of the Eighth International Conference on Semantic Technologies for Intelligence, Defense and Security (STIDS), (CEUR 2013). 2013. vol. 1097. pp. 33–40.
7. Basic Formal Ontology. URL: <http://ifomis.uni-saarland.de/bfo/> (дата обращения: 10.11.2016).
8. *Lomov P., Shishaev M.* Ad-hoc Synthesis of Composite Content Ontology Design Patterns // Proceedings of the 26th International Conference on Information Modelling and Knowledge Bases (EJC 2016). Tampere University of Technology. Pori. 2016. pp. 25–33.

9. *Blomqvist E., Hammar K., Presutti V.* Engineering Ontologies with Patterns: The eXtreme Design Methodology // *Ontology Engineering with Ontology Design Patterns*. IOS Press. 2016. vol. 25. pp. 23–50.
10. *Blomqvist E., Presutti V., Daga E., Gangemi A.* Experimenting with eXtreme Design // *Proceedings of EKAW 2010*. Springer. Berlin-Heidelberg-New York. 2010. LNCS 6317. pp. 120–134.
11. *Ontology of Integrated Knowledge Space*. URL: <https://github.com/palandlom/ontology-of-integrated-knowledge-space>. (дата обращения: 11.11.2016).
12. *Schmidt M., Meier M., Lausen G.* Foundations of SPARQL query optimization // *Proceedings of the 13th International Conference on Database Theory*. 2010. .pp. 4–33.
13. *Tsialiamanis P. et al.* Heuristics-based query optimisation for SPARQL // *Proceedings of the 15th International Conference on Extending Database Technology (EDBT '12)*. ACM. 2012. pp. 324–335.
14. *Aranda C., Arenas M., Corcho M.* Semantics and Optimization of the SPARQL 1.1 Federation Extension The Semantic Web: Research and Applications // *8th Extended Semantic Web Conference*. Springer Berlin Heidelberg. 2011. pp. 1–15.
15. *Gubichev A., Neumann T.* Exploiting the query structure for efficient join ordering in SPARQL queries // *Proceedings of the 17th international conference on Extending Database Technology*. 2014. pp. 439–450.
16. *Bartak R.* Constraint Programming: In Pursuit of the Holy Grail // *Proceedings of the Week of Doctoral Students (WDS99)*. 1999. Part IV. pp. 555–564.
17. *Rutkay Zs.* Constraint satisfaction a survey // *CWI Quarterly*. 1998. vol. 11. pp. 163–214.
18. *Russel S., Norvig P.* Artificial Intelligence: A Modern Approach. 3rd edition // Prentice Hall. 2010. 1132 p.
19. *Щербина О.А.* Удовлетворение ограничений и программирование в ограничениях // *Интеллектуальные системы*. 2011. Т. 15. №. 1–4. С. 54–73.
20. *Kulik B.A., Zuenko A.A., Friedman A.Ya.* Deductive and Defeasible Reasoning on the Basis of a Unified Algebraic Approach // *Scientific and Technical Information Processing*. 2015. vol. 42. no. 6. pp. 8–16.
21. *Zakrevskij A.* Integrated Model of Inductive-Deductive Inference Based on Finite Predicates and Implicative Regularities // *Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems*. IGI Global. 2013. pp. 1–12.
22. *Зуенко А.А.* Качественное моделирование технических систем на основе методов распространения ограничений // *Открытые семантические технологии проектирования интеллектуальных систем: мат. VI Междунар. науч.-техн. конф.* Минск. БГУИР. 2016. С.573–578.
23. *Neumann T., Moerkotte G.* Characteristic sets: Accurate cardinality estimation for RDF queries with multiple joins // *ICDE*. 2011. pp. 984–994.
24. *Smirnov A.V., Levashova T.V., Shilov N.G., Krizhanovskiy A.A.* Knowledge Fusion in Context-Aware Decision Support: Ontology-Based Modeling and Patterns // *Studies in Fuzziness and Soft Computing*. 2014. vol. 314. pp. 35–51.
25. *Smirnov A., Levashova T., Shilov N.* Knowledge Fusion in Context-Aware Decision Support Systems // *KEOD 2014 — Proceedings of the International Conference on Knowledge Engineering and Ontology Development*. 2014. pp. 186–194.

Зуенко Александр Анатольевич — к-т техн. наук, старший научный сотрудник лаборатории информационных технологий управления промышленно-природными системами, Федеральное государственное бюджетное учреждение науки Институт информатики и математического моделирования технологических процессов Кольского научного центра Российской академии наук (ИИММ КНЦ РАН). Область научных интересов: представление и обработка знаний в слабо формализованных предметных областях, программирование в ограничениях. Число научных публикаций — 104.

zuenko@iimm.ru; ул. Ферсмана, 24а, Мурманская обл., Апатиты, 184209; р.т.: +78155574050, Факс: +78155574050.

Ломов Павел Андреевич — к-т техн. наук, научный сотрудник лаборатории региональных информационных систем, Федеральное государственное бюджетное учреждение науки Институт информатики и математического моделирования технологических процессов Кольского научного центра Российской академии наук (ИИММ КНЦ РАН). Область научных интересов: представление знаний, онтологическое моделирование, Semantic web, информационная безопасность. Число научных публикаций — 48. lomov@iimm.ru; ул. Ферсмана, 24а, Мурманская обл., Апатиты, 184209; р.т.: +79522967633, Факс: +7(81555)74050.

Олейник Андрей Григорьевич — д-р техн. наук, заместитель директора по научной работе, Федеральное государственное бюджетное учреждение науки Институт информатики и математического моделирования технологических процессов Кольского научного центра Российской академии наук (ИИММ КНЦ РАН). Область научных интересов: информационные технологии, компьютерное моделирование, базы данных, системы поддержки принятия решений. Число научных публикаций — 166. oleynik@iimm.ru, <http://www.iimm.ru/users/oleynik>; ул. Ферсмана, 24а, Мурманская обл., Апатиты, 184209; р.т.: +7155579220, Факс: +7155574050.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проекты №№ 15-07-03321а, 16-07-00377а, 16-07-00562а).

A.A. ZUENKO, P.A. LOMOV, A.G. OLEYNIK
**APPLICATION OF CONSTRAINT PROPAGATION TECHNIQUES
 TO SPEED UP PROCESSING OF QUERIES TO ONTOLOGIES**

Zuenko A.A., Lomov P.A., Oleynik A.G. Application of Constraint Propagation Techniques to Speed up Processing of Queries to Ontologies.

Abstract. We propose to regard a problem of processing of SPARQL queries to the ontology as a constraint satisfaction problem. The formal apparatus for formalization of constraint satisfaction problems by means of specialized matrix-like structures is briefly given. The use of this apparatus allows one to describe, store and process non-numeric constraints of the subject domain more efficiently compared with the table representation. To speed up processing of queries to big ontologies it is proposed to apply the non-numeric constraint propagation technique previously developed by the authors. Compared with the traditional approach to processing of the SPARQL queries, which is based on dynamic programming, the proposed technique allows one to accelerate their performance by "compact" representation of ontology, as well as by using the original rules of the search space reduction. Application of the mathematical apparatus for reducing the dimension of the search space at processing a query to the ontology of multidisciplinary knowledge is exemplified.

Keywords query to the ontology, constraint satisfaction problem, a matrix representation of constraint, constraint propagation.

Zuenko Alexander Anatolievich — Ph.D., senior researcher of information technologies for nature-industrial systems control laboratory, Institute for Informatics and Mathematical Modelling of Technological Processes of the Kola Science Center of Russian Academy of Sciences (IIMM KSC RAS). Research interests: knowledge representation and processing in poorly formalized subject domains, constraint programming. The number of publications — 104. zuenko@iimm.ru; 24A, Fersman st., 184209, Apatity, Murmansk Region, Russia; office phone: +78155574050, Fax: +78155574050.

Lomov Pavel Andreevich — Ph.D., researcher of regional information systems laboratory, Institute for Informatics and Mathematical Modelling of Technological Processes of the Kola Science Center of Russian Academy of Sciences (IIMM KSC RAS). Research interests: knowledge representation, ontological modeling, Semantic web. The number of publications — 48. lomov@iimm.ru; 24A, Fersman st., 184209, Apatity, Murmansk Region, Russia; office phone: +79522967633, Fax: +7(81555)74050.

Oleynik Andry Gigorievich — Ph.D., Dr. Sci., deputy director, Institute for Informatics and Mathematical Modelling of Technological Processes of the Kola Science Center of Russian Academy of Sciences (IIMM KSC RAS). Research interests: information technologies, computer modeling, databases, systems of decision support. The number of publications — 166. oleynik@iimm.ru, <http://www.iimm.ru/users/oleynik>; 24A, Fersman st., 184209, Apatity, Murmansk Region, Russia; office phone: +7155579220, Fax: +7155574050.

Acknowledgements. This research is supported by RFBR (grants №№ 15-07-03321a, 16-07-00377a, 16-07-00562a).

References

1. Gangemi A. Ontology Design Patterns for Semantic Web Content. Proceedings of the Fourth International Semantic Web Conference. Galway, Ireland. Springer. 2005. pp. 262–276.
2. Olejnik A.G. [About the formation of integrated multi-disciplinary knowledge space]. *Trudy pjatoj mezhdunarodnoj konferencii "Sistemnyj analiz i informacionnyje tehnologii" (SAIT-2013)* [Proceedings of the Fifth International Conference "System Analysis and Information Technologies" (SAIT 2013): Proceedings of the conference]. 2013. Issue 1. pp. 258–265. (In Russ.).
3. Nikolić N. et al. RDF Stores Performance Test on Servers with Average Specification. Proceedings of the 5th International Conference on Information Science and Technology (ICIST). 2015. pp. 67–72.
4. Neumann T., Weikum G. RDF3X: a RISCstyle engine for RDF. Proceedings of the 34th International Conference on Very Large Data Bases (VLDB). Auckland, New Zealand. 2008. pp. 647–659.
5. Zuenko A.A. [Constraint inference based on the matrix representation of finite predicates]. *Iskustvennui intellect I prinyatie reshenii – Artificial intelligence and decision-making*. 2014. vol. 3. pp. 21–31. (In Russ.).
6. Smith B. et al. IAO-Intel: An Ontology of Information Artifacts in the Intelligence Domain. Proceedings of the Eighth International Conference on Semantic Technologies for Intelligence, Defense and Security (STIDS), (CEUR 2013). 2013. vol. 1097. pp. 33–40.
7. Basic Formal Ontology. Available at: <http://ifomis.uni-saarland.de/bfo/> (accessed 10.11.2016).
8. Lomov P., Shishaev M. Ad-hoc Synthesis of Composite Content Ontology Design Patterns. Proceedings of the 26th International Conference on Information Modelling and Knowledge Bases (EJC 2016). Tampere University of Technology. Pori. 2016. pp. 25–33.
9. Blomqvist E., Hammar K., Presutti V. Engineering Ontologies with Patterns: The eXtreme Design Methodology. In: *Ontology Engineering with Ontology Design Patterns*. IOS Press. 2016. vol. 25. pp. 23–50.
10. Blomqvist E., Presutti V., Daga E., Gangemi A. Experimenting with eXtreme Design. Proceedings of EKAW 2010. Springer. Berlin-Heidelberg-New York. 2010. LNCS 6317. pp. 120–134.
11. Ontology of Integrated Knowledge Space. Available at: <https://github.com/palandlom/ontology-of-integrated-knowledge-space>. (accessed 11.11.2016).
12. Schmidt M., Meier M., Lausen G. Foundations of SPARQL query optimization. Proceedings of the 13th International Conference on Database Theory. 2010. pp. 4–33.
13. Tsialiamanis P. et al. Heuristics-based query optimisation for SPARQL. Proceedings of the 15th International Conference on Extending Database Technology (EDBT '12). ACM. 2012. pp. 324–335.
14. Aranda C., Arenas M., Corcho O. Semantics and Optimization of the SPARQL 1.1 Federation Extension The Semantic Web: Research and Applications. 8th Extended Semantic Web Conference. Springer Berlin Heidelberg. 2011. pp. 1–15.
15. Gubichev A., Neumann T. Exploiting the query structure for efficient join ordering in SPARQL queries. Proceedings of the 17th international conference on Extending Database Technology. 2014. pp. 439–450.
16. Bartak R. Constraint Programming: In Pursuit of the Holy Grail. Proceedings of the Week of Doctoral Students (WDS99). 1999. Part IV. pp. 555–564.
17. Ruttkay Zs. Constraint satisfaction a survey. *CWI Quarterly*. 1998. vol. 11. pp. 163–214.
18. Russel S., Norvig P. *Artificial Intelligence: A Modern Approach*. 3rd edition. Prentice Hall. 2010. 1132 p.

19. Shherbina O. A. [Constraint propagation and constraint programming]. *Intellektual'nye sistemy – Intelligent systems*. 2011. vol. 15. no. 1–4. pp. 54–73. (In Russ.).
20. Kulik B.A., Zuenko A.A., Friedman A.Ya. Deductive and Defeasible Reasoning on the Basis of a Unified Algebraic Approach. *Scientific and Technical Information Processing*. 2015. vol. 42. no. 6. 2015. pp. 8–16.
21. Zakrevskij A. Integrated Model of Inductive-Deductive Inference Based on Finite Predicates and Implicative Regularities. In *Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems*. IGI Global. 2013. pp. 1–12.
22. Zuenko A.A. [Qualitative modeling of technical systems based on the constraint propagation techniques] *Otkrytye semanticheskie tehnologii proektirovaniya intellektual'nyh sistem (OSTIS-2016): mat. VI Mezhdunar. nauchn.-tehn. konf.* [Open Semantic Technologies for Intelligent Systems (OSTIS-2016): proceedings of the VI International scientific and technical conference]. Minsk: BGUIR. 2016. pp. 573–578. (In Russ.).
23. Neumann T., Moerkotte G. Characteristic sets: Accurate cardinality estimation for RDF queries with multiple joins. *ICDE*. 2011. pp. 984–994.
24. Smirnov A.V., Levashova T.V., Shilov N.G., Krizhanovsky A.A. Knowledge Fusion in Context-Aware Decision Support: Ontology-Based Modeling and Patterns. *Studies in Fuzziness and Soft Computing*. 2014. vol. 314. pp. 35–51.
25. Smirnov A., Levashova T., Shilov N. Knowledge Fusion in Context-Aware Decision Support Systems. *Proceedings of the International Conference on Knowledge Engineering and Ontology Development – 2014 (KEOD 2014)*. pp. 186–194.

В.В. САПОЖНИКОВ, ВЛ.В. САПОЖНИКОВ, Д.В. ЕФАНОВ, А.Г. КОТЕНКО
**МОДУЛЬНЫЕ КОДЫ С СУММИРОВАНИЕМ
ВЗВЕШЕННЫХ ПЕРЕХОДОВ С ПОСЛЕДОВАТЕЛЬНОСТЬЮ
ВЕСОВЫХ КОЭФФИЦИЕНТОВ, ОБРАЗУЮЩЕЙ
НАТУРАЛЬНЫЙ РЯД ЧИСЕЛ**

Сапожников В.В., Сапожников Вл.В., Ефанов Д.В., Котенко А.Г. Модульные коды с суммированием взвешенных переходов с последовательностью весовых коэффициентов, образующей натуральный ряд чисел.

Аннотация. При синтезе надежных дискретных систем используются разделимые коды с простыми правилами построения и небольшой избыточностью. Часто применяются коды с суммированием с постоянным значением количества контрольных разрядов вне зависимости от длины информационного вектора. В статье показано, что характеристики обнаружения ошибок данными кодами могут быть улучшены за счет модификации правил вычисления значений разрядов контрольных векторов. Предложен алгоритм построения кода с суммированием, основанный на взвешивании переходов между разрядами, занимающими соседние позиции в информационных векторах, и подсчете суммарного веса информационного вектора по модулю, равному степени числа 2. В статье детально анализируются свойства новых кодов с суммированием в сравнении с известными кодами. Установлены условия построения помехоустойчивых модульных взвешенных кодов с суммированием с максимальным количеством обнаруживаемых ошибок в информационных векторах. Улучшение характеристик обнаружения ошибок в новых кодах по сравнению с классическими кодами с суммированием достигается за счет равномерного распределения информационных векторов между всеми контрольными векторами, что, в свою очередь, вносит в класс необнаруживаемых помимо симметричных, некоторую долю монотонных и асимметричных ошибок. В отличие от модульных кодов с суммированием единичных разрядов модульные взвешенные коды обнаруживают большее количество симметричных ошибок, однако доминирующим их видом практически при любой длине информационного вектора являются асимметричные ошибки. Модульные коды с суммированием взвешенных переходов являются перспективным классом разделимых кодов для решения задач технической диагностики дискретных систем.

Ключевые слова: техническая диагностика, функциональный контроль, разделимый код, код Бергера, модульный код с суммированием, код с суммированием взвешенных переходов, необнаруживаемая ошибка, характеристики кода.

1. Введение. При решении задач обеспечения надежности и безопасности дискретных систем, реализующих ответственные технологические процессы, часто используются методы избыточного кодирования [1, 2]. Один из важных классов избыточных кодов составляют разделимые коды: в них выделяется информационный и контрольный векторы, а значения разрядов последних получаются по значениям информационных разрядов. Разделимые коды, или (m,k) -коды (m и k — длины информационных и контрольных векторов соответственно) с простыми правилами построения и небольшим количеством разрядов в контрольных векторах используют при передаче и обработке дан-

ных, а также при решении задач технической диагностики дискретных систем [3, 4]. Например, известным приложением (m,k) -кодов является их использование при организации систем функционального (рабочего, *on-line*) контроля комбинационных логических устройств [5-13].

Одними из наиболее известных (m,k) -кодов, нашедших широкое применение при решении задач построения надежных дискретных устройств, являются классические коды с суммированием, или коды Бергера, предложенные в 1961 г. [14]. Данные коды принадлежат к классу несистематических разделимых кодов и строятся следующим образом: к информационному вектору добавляется контрольный вектор, в разряды которого записывается двоичное число, равное весу информационного вектора. Количество разрядов в контрольных векторах кодов Бергера $k = \lceil \log_2(m+1) \rceil$ (запись $\lceil \dots \rceil$ обозначает целое сверху от вычисляемого значения). Указанные правила построения кодов Бергера позволяют снабдить их свойством обнаружения любых монотонных ошибок в информационных векторах, что, например, используется при организации схем встроенного контроля [15-17]. В [18] установлено, что от свойств обнаружения ошибок в информационных векторах кодами Бергера зависит такой важный показатель схем контроля, как свойства обнаружения неисправностей в контролируемых объектах. При этом сами коды Бергера не обнаруживают достаточно большое количество ошибок в информационных векторах, в особенности в области малой кратности (например, ими не обнаруживается 50% двукратных ошибок в информационных векторах).

Для упрощения контрольного оборудования в дискретных системах применяются модульные коды с суммированием [19, 20]. Они получаются путем незначительного изменения правил построения кодов Бергера: вес информационного вектора подсчитывается по заранее установленному модулю $M \in \{2; 4; \dots; 2^{\lceil \log_2(m+1) \rceil - 1}\}$ (определяется наименьший неотрицательный вычет веса информационного вектора по модулю M), а полученное значение записывается в разряды контрольного вектора. Данные правила построения модульных кодов с суммированием (далее будем обозначать их как $SM(m,k)$ -коды) приводят к увеличению количества не обнаруживаемых кодом ошибок по сравнению с кодом Бергера, но позволяют сохранять их главное свойство — способность обнаружения любых монотонных ошибок, однако до определенной кратности.

Характеристики обнаружения ошибок $SM(m,k)$ -кодами в информационных векторах изучены в большом количестве работ, например, в [21-23]. Отметим наиболее важные их свойства, учет которых

может оказаться полезным при организации систем функционального контроля:

– $SM(m,k)$ -коды не обнаруживают любые симметричные ошибки в информационных векторах;

– $SM(m,k)$ -коды не обнаруживают 100% монотонных ошибок с кратностью $d = jM$, $j = 1, 2, \dots, p$, $p \leq \left\lfloor \frac{m}{M} \right\rfloor$, где запись $\left\lfloor \dots \right\rfloor$ обозначает целое снизу от вычисляемого значения;

– $SM(m,k)$ -коды не обнаруживают некоторую долю асимметричных ошибок с кратностями $d = M + 2j$, где $j = 1, 2, \dots (d \leq m)$;

– доля необнаруживаемых ошибок четной кратностью d от общего количества ошибок данной кратностью является постоянной величиной для $SM(m,k)$ -кода с установленным значением модуля и не зависит от длины информационного вектора.

Часто в качестве модуля при построении $SM(m,k)$ -кодов используют значения $M=4$ и $M=8$. Такие модульные коды называют кодами Боуза-Лина [19]. В [24] приводятся методы синтеза контрольного оборудования для вычисления разрядов контрольных векторов данных кодов.

Следует отметить, что при построении модульных кодов с суммированием в качестве модуля может выступать любое натуральное число $2 \leq M \leq m+1$ [21]. Тем не менее наиболее частого применения среди модульных кодов с суммированием достигли именно коды Боуза-Лина.

Помимо классических и модульных кодов с суммированием при построении надежных дискретных систем также применяются взвешенные коды с суммированием, принцип построения которых впервые описан в работе [14]. Взвешенный код с суммированием строится путем приписывания весовых коэффициентов разрядам, подсчета суммарного значения веса единичных информационных разрядов и последующей записи полученного числа в двоичном виде в разряды контрольного вектора. В качестве весовых коэффициентов могут выступать любые натуральные числа [25].

Принципы взвешивания разрядов и подсчета суммарного веса по заранее установленному модулю могут применяться совместно, что дает целые семейства модульно взвешенных кодов с суммированием с небольшой избыточностью [26].

Классические и модульные коды с суммированием при любых значениях модуля (за исключением $M=2$) имеют крайне неравномерное распределение информационных векторов между всеми контрольными векторами, что является причиной большого общего количества необнаруживаемых ими ошибок. Исследования показывают, что путем

модификации правил вычисления разрядов контрольных векторов можно добиться улучшения характеристик модульных кодов с суммированием. Например, в [27, 28] были получены коды с суммированием с равномерным распределением информационных векторов между всеми контрольными векторами с таким же количеством разрядов в контрольных векторах, как и у кодов Бергера. Такие коды имеют теоретический минимум общего количества необнаруживаемых ошибок в информационных векторах при заданных значениях m и k .

Опишем здесь класс модульных кодов с суммированием с уменьшенным количеством разрядов в контрольных векторах по сравнению с кодами Бергера и его известными модификациями и имеющих уменьшенное количество необнаруживаемых ошибок по сравнению с последними.

2. Модульные коды с суммированием взвешенных переходов. Коды с фиксированным значением количества разрядов в контрольных векторах с улучшенными характеристиками обнаружения ошибок в информационных векторах по сравнению с классическими $SM(m,k)$ -кодами могут быть построены путем установления неравноправия между разрядами информационных векторов. Подобное свойство достигается путем приписывания специальных весовых коэффициентов непосредственно разрядам или же переходам между разрядами, занимающими соседние позиции [29].

В [28] рассмотрен способ построения кода с суммированием взвешенных информационных разрядов с установленной последовательностью весовых коэффициентов, образующей натуральный ряд чисел. При формировании кода используется модульный принцип подсчета суммарного веса информационного вектора, при этом устанавливается модуль $M = 2^{\lceil \log_2(m+1) \rceil}$ (он соответствует модулю классического кода Бергера). Такие коды обозначаются как $WSM(m,k)$ -коды. Они обладают улучшенными характеристиками обнаружения ошибок в информационных векторах по сравнению с классическими кодами Бергера, однако сохраняют и их недостаток — с увеличением длины информационного вектора увеличивается и количество разрядов в контрольных векторах.

Проводя по аналогии с $SM(m,k)$ -кодами модификацию $WSM(m,k)$ -кодов, связанную с уменьшением значения модуля M и выбора его из множества $M \in \{2, 4, \dots, 2^{\lceil \log_2(m+1) \rceil - 1}\}$, в результате анализа характеристик получаемых кодов мы приходим к важному выводу.

Теорема 1. Не существует помехоустойчивых модульно взвешенных кодов с суммированием с последовательностью весовых ко-

эффицентов, образующей ряд натуральных чисел, с модулем $M \in \{2; 4; \dots; 2^{\lceil \log_2(m+1) \rceil - 1}\}$.

Доказательство. Свойство помехоустойчивости (m, k) -кода теряется в том случае, если при его построении не контролируется хотя бы один из разрядов в информационном векторе. При установлении в качестве модуля величины $M = 2^{\lceil \log_2(m+1) \rceil}$ любой наименьший неотрицательный вычет из имеющих для информационного вектора длиной m весовых коэффициентов $[m; m-1; \dots; 2; 1]$ будет принадлежать множеству $\{1; 2; \dots; 2^{\lceil \log_2(m+1) \rceil} - 2; 2^{\lceil \log_2(m+1) \rceil} - 1\}$, а значит, будет влиять на значение суммарного веса информационного вектора. Это следует из того факта, что значение модуля в указанном случае будет всегда больше длины информационного вектора:

$$2^{\lceil \log_2(m+1) \rceil} > m. \quad (1)$$

Уменьшение значения модуля и выбор его из множества $M \in \{2; 4; \dots; 2^{\lceil \log_2(m+1) \rceil - 1}\}$ приводит к тому, что разряд, занимающий в информационном векторе позицию $2^{\lceil \log_2(m+1) \rceil - 1}$, контролироваться не будет. В таком $WSM(m, k)$ -коде всегда будет присутствовать часть одиночных необнаруживаемых ошибок, что и характеризует не помехоустойчивый код. Отсюда следует справедливость утверждения теоремы 1.

Теорема доказана.

Единственным способом построения помехоустойчивого кода при последовательности весовых коэффициентов, образующей натуральный ряд чисел, является выбор значения модуля, исходя из следующего соотношения:

$$M > m. \quad (2)$$

Для значений модулей из множества $M \in \{2; 4; \dots; 2^{\lceil \log_2(m+1) \rceil - 1}\}$ неравенство (2) выполнить нельзя. Отсюда следует, что любой $WSM(m, k)$ -код с модулем, выбранным из указанного множества, помехоустойчивым являться не будет.

Таким образом, на основании принципа взвешивания разрядов информационного вектора весовыми коэффициентами из натурального ряда чисел построить класс кодов с постоянными значениями k вне зависимости от значения m невозможно.

Однако решить поставленную задачу позволяет использование принципа взвешивания не разрядов в информационном векторе, а переходов между ними (рисунок 1).

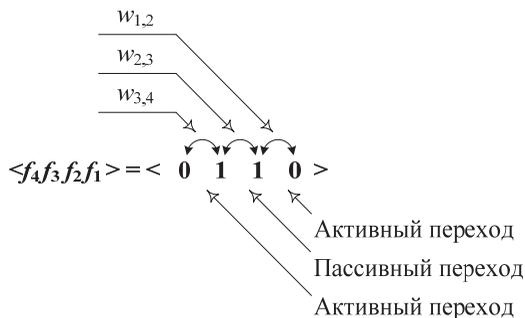


Рис. 1. Элементы информационного вектора

Алгоритм. Получение значений разрядов контрольных векторов модульных кодов с суммированием взвешенных переходов:

1. Устанавливается последовательность весовых коэффициентов переходов между разрядами, занимающими соседние позиции в информационных векторах, образующая натуральный ряд чисел, начиная с младшего разряда информационного вектора:

$$[w] = [w_{m-1,m}; w_{m-2,m-1}; \dots; w_{2,3}; w_{1,2}] = [m-1; m-2; \dots; 2; 1].$$

2. Определяются активные переходы между разрядами в информационных векторах:

$$q_{i,i+1} = f_i \oplus f_{i+1} = 1, \quad i = \overline{1, m-1}. \quad (3)$$

3. Суммируются весовые коэффициенты активных переходов:

$$W = \sum_{i=1}^{m-1} q_{i,i+1} w_{i,i+1} = \sum_{i=1}^{m-1} (f_i \oplus f_{i+1}) w_{i,i+1}. \quad (4)$$

4. Выбирается значение модуля

$$M \in \{2; 4; \dots; 2^{\lceil \log_2 (M_{\max} + 1) \rceil}\} \quad (5)$$

где $M_{\max} = \frac{(m-1)m}{2}$ — сумма максимального количества активных переходов в информационном векторе.

5. Определяется наименьший неотрицательный вычет числа W по модулю M :

$$W_M = W \pmod{M}. \quad (6)$$

6. Число W_M представляется в двоичном виде и записывается в разряды контрольного вектора.

Получаемые по алгоритму коды с суммированием будем обозначать как $WTM(m,k)$ -коды.

Следует отметить, что интерес с практической точки зрения будут представлять такие $WTM(m,k)$ -коды, у которых значение модуля выбирается из множества меньшей мощности, чем множество (5), — $M \in \{2; 4; \dots; 2^{\lceil \log_2(m+1) \rceil}\}$ (значение модуля $M = 2^{\lceil \log_2(m+1) \rceil}$, как отмечалось выше, соответствует модулю классического кода Бергера). Такие $WTM(m,k)$ -коды будут иметь небольшое количество разрядов в контрольных векторах, что будет сказываться на меньшей сложности технической реализации контрольного оборудования в дискретной системе.

Действия алгоритма иллюстрируются в таблице 1, где приводятся все кодовые слова $WT4(4,2)$ -кода.

Таблица 1. Кодовые слова $WT4(4,2)$ -кода

№	Информационный вектор				W	$W \pmod{4}$	Контрольный вектор	
	f_4	f_3	f_2	f_1			g_2	g_1
	$w_{3,4}=3$	$w_{2,3}=2$	$w_{1,2}=1$					
0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	1	0	1
2	0	0	1	0	3	3	1	1
3	0	0	1	1	2	2	1	0
4	0	1	0	0	5	1	0	1
5	0	1	0	1	6	2	1	0
6	0	1	1	0	4	0	0	0
7	0	1	1	1	3	3	1	1
8	1	0	0	0	3	3	1	1
9	1	0	0	1	4	0	0	0
10	1	0	1	0	6	2	1	0
11	1	0	1	1	5	1	0	1
12	1	1	0	0	2	2	1	0
13	1	1	0	1	3	3	1	1
14	1	1	1	0	1	1	0	1
15	1	1	1	1	0	0	0	0

Проанализируем свойства $WT4(4,2)$ -кода. С этой целью представим его в табличной форме, где по столбцам перечислены все контрольные векторы кода (они составляют контрольные группы), каждому из которых сопоставлены все соответствующие информационные векторы [18] (таблица 2). Для рассматриваемого модульного взвешенного кода с суммированием распределение информационных векторов на контрольные группы является равномерным. В [30] показано, что (m,k) -коды, имеющие равномерное распределение информационных векторов между всеми контрольными векторами, имеют минимальное общее количество обнаруживаемых ошибок в информационных векторах при заданных значениях m и k и являются оптимальными по данному критерию:

$$N_{m,k}^{\min} = 2^m (2^{m-k} - 1) \quad (7)$$

Помимо выполнения условия (7) (m,k) -код должен являться помехоустойчивым, чтобы относиться к классу оптимальных кодов с суммированием.

Таблица 2. Распределение информационных векторов $WT4(4,2)$ -кода на контрольные группы

Контрольные группы			
00	01	10	11
Информационные векторы			
0000	0001	0011	0010
0110	0100	0101	0111
1001	1011	1010	1000
1111	1110	1100	1101

Таким образом, $WT4(4,2)$ -код является оптимальным (m,k) -кодом и не обнаруживает 48 ошибок в информационных векторах. К примеру, классический модульный код с суммированием $S4(4,2)$ -код таковым не является и не обнаруживает 56 ошибок в информационных векторах.

Табличная форма задания (m,k) -кода удобна при анализе характеристик обнаруживаемых им ошибок и установления их количества по

каждой кратности $d \in \{1; 2; \dots; m\}$ и виду (одионочная, мононононная, симметричная или асимметричная [21]). На рисунке 2 показан такой анализ.

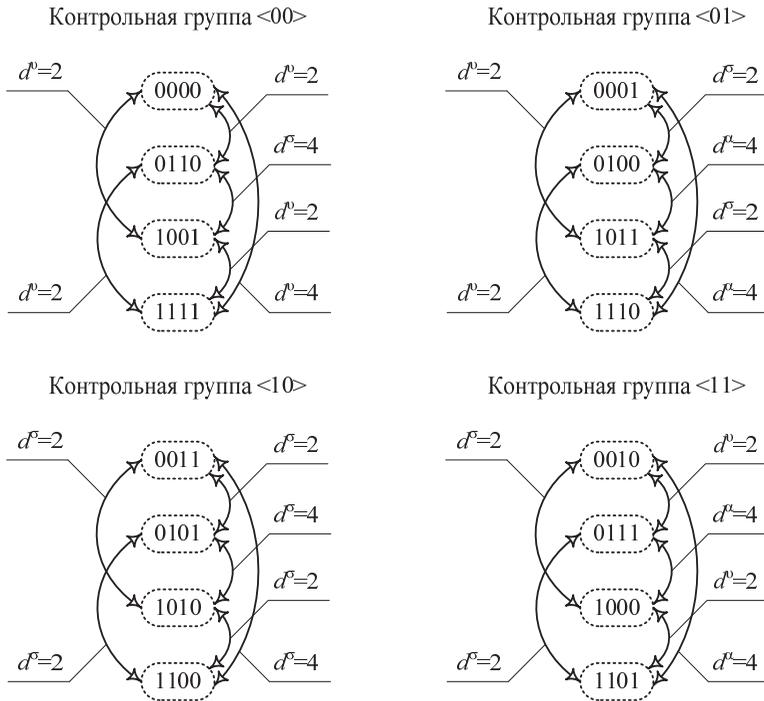


Рис. 2. Характеристики необнаруживаемых переходов внутри каждой контрольной группы

Все возможные переходы внутри каждой контрольной группы соответствуют искажениям, приводящим к необнаруживаемым ошибкам в информационных векторах. Для $WT_4(4,2)$ -кода каждая контрольная группа дает по 12 необнаруживаемых ошибок (количество необнаруживаемых ошибок внутри группы определяется как удвоенное число всех возможных парных переходов; для оптимального кода оно равно $2C_2^{2-n}$). На рисунке 2 на каждом переходе указана кратность d и в виде верхнего индекса — вид (v — мононононная, σ — симметричная или α — асимметричная) необнаруживаемой ошибки. Рассматриваемым кодом не обнаруживается 18 мононононных ошибок (16 двукратных и 2 четырехкратных), 22 симметричных ошибки (16 двукратных и 6 четырехкратных) и 8 асимметричных четырехкратных оши-

бок. Наличие необнаруживаемых ошибок различных видов характерно для любых $WTM(m,k)$ -кодов, при этом их кратности $d \in \{1; 2; \dots; m\}$.

Отметим следующую важную особенность $WTM(m,k)$ -кодов.

Теорема 2. Модульно взвешенный код с суммированием переходов между разрядами, занимающими соседние позиции в информационных векторах, с модулем $M \in \{2; 4; \dots; 2^{\lceil \log_2(m+1) \rceil - 1}\}$ будет помехоустойчивым тогда и только тогда, когда значение длины информационного вектора будет удовлетворять соотношению:

$$m \neq jM + 1, \quad j \in \{1, 2, \dots, p\}, \quad p = \left\lfloor \frac{m}{M} \right\rfloor. \quad (8)$$

Доказательство. Рассмотрим $WTM(m,k)$ -код с последовательностью весовых коэффициентов, образующей натуральный ряд чисел, начиная с младшего разряда информационного вектора $[m-1; m-2; \dots; 2; 1]$, с длиной информационного вектора $m=M+1$. Весовые коэффициенты для такой длины информационного вектора образуют ряд $[M; M-1; \dots; 2; 1]$. Наименьшие неотрицательные вычеты каждого из весовых коэффициентов, за исключением $w_{m-1,m}$, будут являться натуральными числами. Поскольку $w_{m-1,m}=M$ значение наименьшего неотрицательного вычета для данного весового коэффициента будет равно $w_{m-1,m} \pmod{M} = M \pmod{M} = 0$. Другими словами, вне зависимости от значения активационной функции $q_{m-1,m} = f_{m-1} \oplus f_m = f_M \oplus f_{M+1}$ в суммарный вес информационного вектора не будет добавляться никакого значения ни при $f_m=0$, ни при $f_m=1$. Таким образом, информационные векторы вида $\langle 0 f_{m-1} f_{m-2} \dots f_2 f_1 \rangle$ и $\langle 1 f_{m-1} f_{m-2} \dots f_2 f_1 \rangle$ при одинаковых значениях $m-1$ младших разрядов будут иметь одинаковый суммарный вес и одинаковые контрольные векторы. Расстояние Хэмминга между такими векторами будет равно $d=1$. Поскольку существует ровно 2^{m-1} векторов вида $\langle 0 f_{m-1} f_{m-2} \dots f_2 f_1 \rangle$, то взаимные переходы их в векторы $\langle 1 f_{m-1} f_{m-2} \dots f_2 f_1 \rangle$ при наличии искажения старшего разряда приводят к возникновению 2^m необнаруживаемых одиночных ошибок. Таким образом, $WTM(m,k)$ -код при $m=M+1$ не будет являться помехоустойчивым. При длине информационного вектора $m=M+2$ весовые коэффициенты $WTM(m,k)$ -кода образуют ряд $[M+1; M; M-1; \dots; 2; 1]$, а их наименьшие неотрицательные вычеты — $[1; 0; M-1; \dots; 2; 1]$. В таком случае разряд f_{M+1} будет влиять на значение активационных функций $q_{M,M+1} = f_M \oplus f_{M+1}$ и $q_{M+1,M+2} = f_{M+1} \oplus f_{M+2}$. Для функции $q_{M,M+1}$ значение разряда f_{M+1} не будет существенным, так как вне зависимости от ее значения будет умножено на нулевой весовой коэффициент.

ент. Данная функция при формировании суммарного веса информационного вектора при $f_m=0, f_{m-1}=0$ и $f_m=1, f_{m-1}=1$ будет вносить в значение суммарного веса информационного вектора значение 0, а в случае $f_m=1, f_{m-1}=0$ или $f_m=0, f_{m-1}=1$ — значение $w_{m,m-1}=1$. Таким образом, разряд f_{M+1} будет контролироваться хотя бы одной активационной функцией, а его значение будет влиять на суммарный вес информационного вектора.

На рисунке 3 дается иллюстрация описываемых случаев для примера $WT4(5,2)$ и $WT4(6,2)$ кодов.

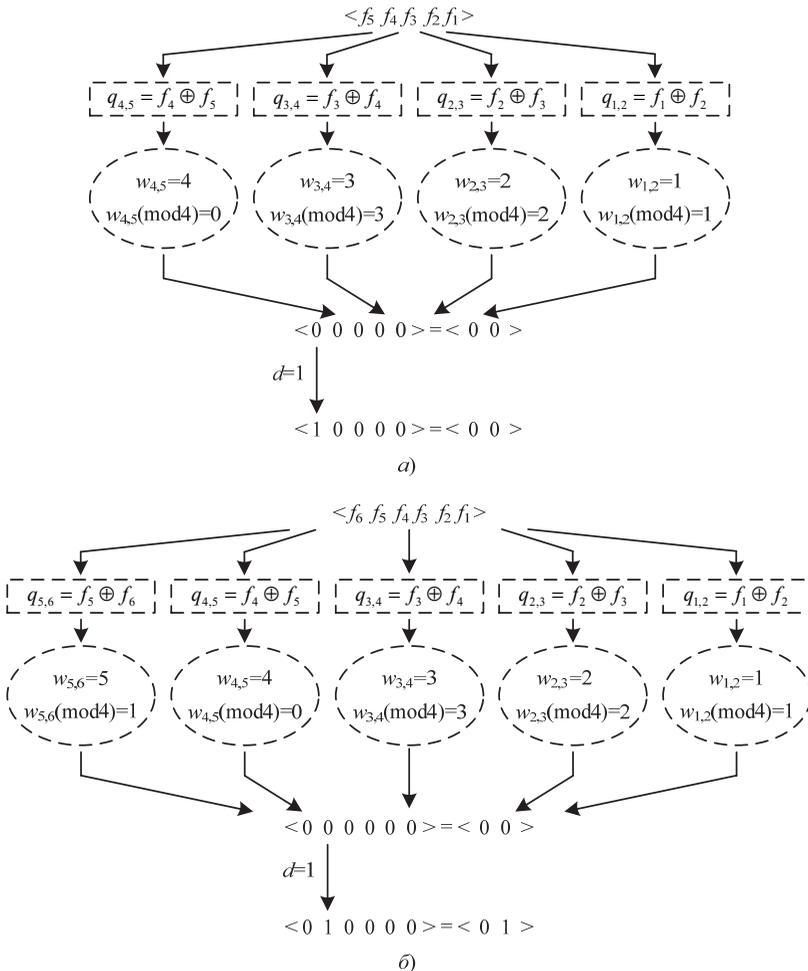


Рис. 3. Модульный взвешенный код: а) не помехоустойчивый; б) помехоустойчивый

Исходя из вышеизложенного, не помехоустойчивым будет только такой $WTM(m,k)$ -код, у которого длина информационного вектора $m=M+1$.

Уменьшение значения модуля и выбор его из множества $M \in \{2;4;\dots;2^{\lceil \log_2 (m+1) \rceil - 1}\}$ приводит к тому, что описанная выше ситуация с информационными векторами и разрядами, занимающими в них позиции jM , $j \in \{1,2,\dots,p\}$, $p = \left\lfloor \frac{m}{M} \right\rfloor$, повторяется. Неконтролируемым в

любом случае при $m \neq jM + 1$, $j \in \{1,2,\dots,p\}$, $p = \left\lfloor \frac{m}{M} \right\rfloor$ остается старший разряд информационного вектора. Остальные $WTM(m,k)$ -коды являются помехоустойчивыми.

Теорема доказана.

Теорема 2 определяет условия построения помехоустойчивого $WTM(m,k)$ -кода.

К примеру, помехоустойчивый $WT4(5,2)$ -код построен быть не может. В табл. 3 для $WTM(m,k)$ -кодов с небольшими значениями модулей $M=2, 4, 8$ указаны длины информационных векторов, при которых возможно построение помехоустойчивых кодов.

Исследования $WTM(m,k)$ -кодов показали справедливость следующих положений.

Свойство 1. При длинах информационных векторов $m \neq jM + 1$,

$j \in \{1,2,\dots,p\}$, $p = \left\lfloor \frac{m}{M} \right\rfloor$ $WTM(m,k)$ -код со значением модуля

$M \in \{2;4;\dots;2^{\lceil \log_2 (m+1) \rceil - 1}\}$ не является помехоустойчивым и имеет в классе необнаруживаемых следующее количество одиночных ошибок:

$$N_{m,1} = 2^{jM+1}, \quad j \in \{1,2,\dots,p\}, \quad p = \left\lfloor \frac{m}{M} \right\rfloor. \quad (9)$$

Свойство 2. $WTM(m,k)$ -коды со значением модуля $M \in \{2;4;\dots;2^{\lceil \log_2 (m+1) \rceil - 1}\}$ являются оптимальными в том случае, если длина информационного вектора удовлетворяет выражению (8) и следующему соотношению:

$$m \geq \frac{M}{2} + 1. \quad (10)$$

Таблица 3. Значения m при которых могут быть построены $WTM(m,k)$ -коды

Код	Длина информационного вектора, m															
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
$WT2(m,k)$	×	–	×	–	×	–	×	–	×	–	×	–	×	–	×	
$WT4(m,k)$	×	×	×	–	×	×	×	–	×	×	×	–	×	×	×	
$WT8(m,k)$	–	–	×	×	×	×	×	–	×	×	×	×	×	×	×	

Свойство 3. Помехоустойчивые $WT2(m,k)$ -коды являются классическими кодами паритета и не обнаруживают 100% ошибок с четными кратностями.

Свойство 4. Любой $WTM(m,k)$ -код при четном значении длины информационного вектора обнаруживает любые ошибки в информационных векторах с нечетными кратностями.

Свойство 5. Любой $WTM(m,k)$ -код при нечетном значении длины информационного вектора в классе необнаруживаемых имеет ошибки с кратностями $d \in \{2; 3; \dots; m - 2; m\}$, причем, не обнаруживается 100% ошибок с максимальной кратностью $d=m$:

$$N_{m,m} = 2^m. \tag{11}$$

3. Сравнительные характеристики классических и взвешенных модульных кодов с суммированием. $SM(m,k)$ -коды получили широкое распространение за счет того, что при фиксированной длине контрольных векторов их использование при синтезе дискретных устройств дает небольшую структурную избыточность и приемлемые характеристики обнаружения ошибок [20, 21]. Сравним $SM(m,k)$ -коды с соответствующими $WTM(m,k)$ -кодами на примере кодов с количеством контрольных разрядов $k=2$ и $k=3$.

В таблицах 4 и 5 приводятся рассчитанные значения числа не обнаруживаемых описываемыми кодами ошибок в информационных векторах. В последних графах таблиц приводится значение доли необнаруживаемых ошибок в $WTM(m,k)$ -кодах по сравнению с $SM(m,k)$ -кодами (значение величины ζ_m , %). $WTM(m,k)$ -коды имеют минимально возможное общее количество необнаруживаемых ошибок в информационных векторах, тогда как $SM(m,k)$ -коды не обнаруживают большее количество ошибок.

С увеличением длины информационного вектора при фиксированном значении M распределения информационных векторов по контрольным группам в классических модульных кодах приближаются к равномерному распределению. По этой причине с увеличением длины информационного вектора значение величины ζ_m приближается к 100%. При этом скорость приближения величины ζ_m к 100% увеличивается с уменьшением значения M , а наиболее существенный эффект по сравнению с $SM(m,k)$ -кодами достигается на малых длинах информационных векторов.

Таблица 4. Характеристики необнаруживаемых ошибок для кодов с двумя контрольными разрядами

m	k	$S4(m,k)$	$WT4(m,k)$	$\zeta_m, \%$
4	2	56	48	85,7143
5	2	240	224	93,3333
6	2	992	960	96,7742
7	2	4032	3968	98,4127
8	2	16256	16128	99,2126
9	2	65280	65024	99,6078
10	2	261632	261120	99,8043
11	2	1047552	1046528	99,9022
12	2	4192180	4190208	99,9530
13	2	16773120	16769024	99,9756
14	2	67100672	67092480	99,9878
15	2	268419072	268402688	99,9939
16	2	1073709056	1073676288	99,9969
17	2	4294901760	4294836224	99,9985
18	2	17179738112	17179607040	99,9992
19	2	68719214592	68718952448	99,9996
20	2	274877382656	274876858368	99,9998

При решении задач технической диагностики определяющее значение имеют характеристики обнаружения (m,k) -кодом ошибок различных видов (монотонных (*unidirectional*), симметричных (*symmetrical*) и асимметричных (*asymmetrical*)). Анализ распределений необнаруживаемых ошибок в информационных векторах $WTM(m,k)$ -кодов (рисунок 4 и рисунок 5) показывает, что доминирующим видом

необнаруживаемых ошибок в них являются асимметричные ошибки. Их доля в общем количестве необнаруживаемых ошибок с увеличением длины информационного вектора возрастает, тогда как доля монотонных и симметричных ошибок уменьшается. При этом доля монотонных ошибок от общего количества необнаруживаемых ошибок (величина ν_m , %) с увеличением длины информационного вектора убывает постепенно. Характер же изменения долей симметричных и асимметричных ошибок от общего количества необнаруживаемых ошибок (величины σ_m и α_m соответственно, %), напротив, характеризуется некоторым разбросом относительно аппроксимирующей кривой. При четных значениях m происходит увеличение величины σ_m относительно предыдущего значения длины информационного вектора $m-1$, при нечетных значениях m — наоборот, уменьшение относительно предыдущего значения. Однако сохраняется общая тенденция уменьшения величины σ_m . Прямо противоположный характер изменения зависимости имеет величина α_m .

Таблица 5. Характеристики необнаруживаемых ошибок для кодов с тремя контрольными разрядами

m	k	$S8(m,k)$	$WT8(m,k)$	ζ_m , %
8	3	12616	7936	62,9042
9	3	48144	32256	66,9990
10	3	184112	130048	70,6353
11	3	706464	522240	73,9231
12	3	2721236	2093056	76,9156
13	3	10523968	8380416	79,6317
14	3	40853696	33538048	82,0931
15	3	159156352	134184960	84,3102
16	3	622051456	536805376	86,2960
17	3	2438377728	2147352576	88,0648
18	3	9583248128	8589672448	89,6322
19	3	37751400960	34359214080	91,0144
20	3	149019369984	137437904896	92,2282

Например, для $WT4(m,k)$ и $WT8(m,k)$ кодов уже при длине информационных векторов $m=8$ свыше 50% необнаруживаемых ошибок являются асимметричными и менее 40% необнаруживаемых ошибок – симметричными.

Совершенно иной характер изменения величин σ_m , ν_m и α_m имеют классические $SM(m,k)$ -коды (рисунок 6 и рисунок 7). При малых длинах информационных векторов ($m < 20$) в классе необнаруживаемых у данных кодов доминируют симметричные ошибки, однако происходит постепенное уменьшение значения σ_m с увеличением длины информационного вектора. Скорость такого уменьшения различна для $SM(m,k)$ -кодов с различными значениями модулей: чем меньше значение M , тем быстрее изменяется процентное соотношение между симметричными ошибками и другими их видами. При увеличении длины информационного вектора постепенно увеличивается значение величины α_m , а значение величины ν_m уменьшается.

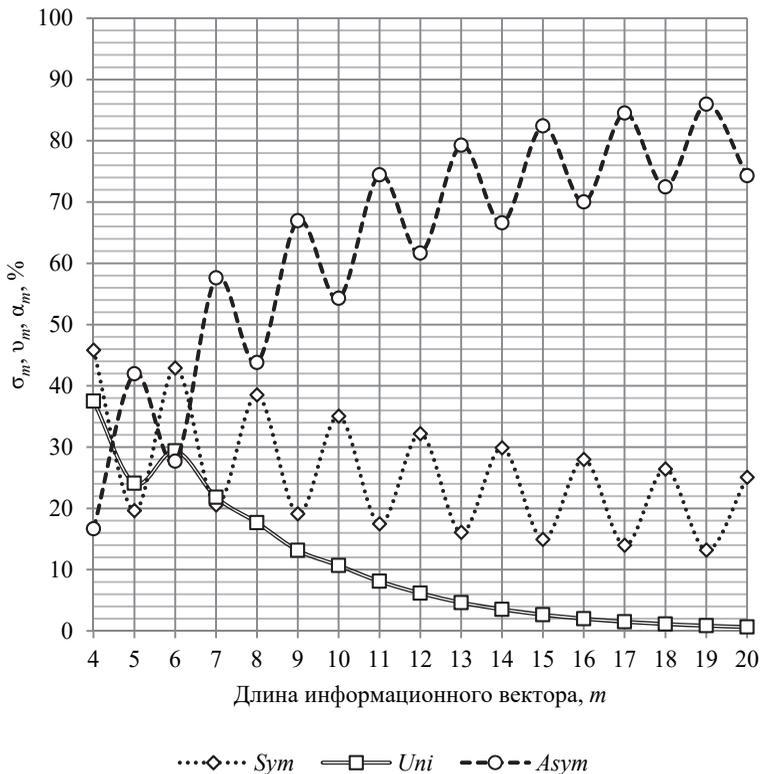


Рис. 4. Распределение необнаруживаемых ошибок по видам в $WT4(m,k)$ -кодах

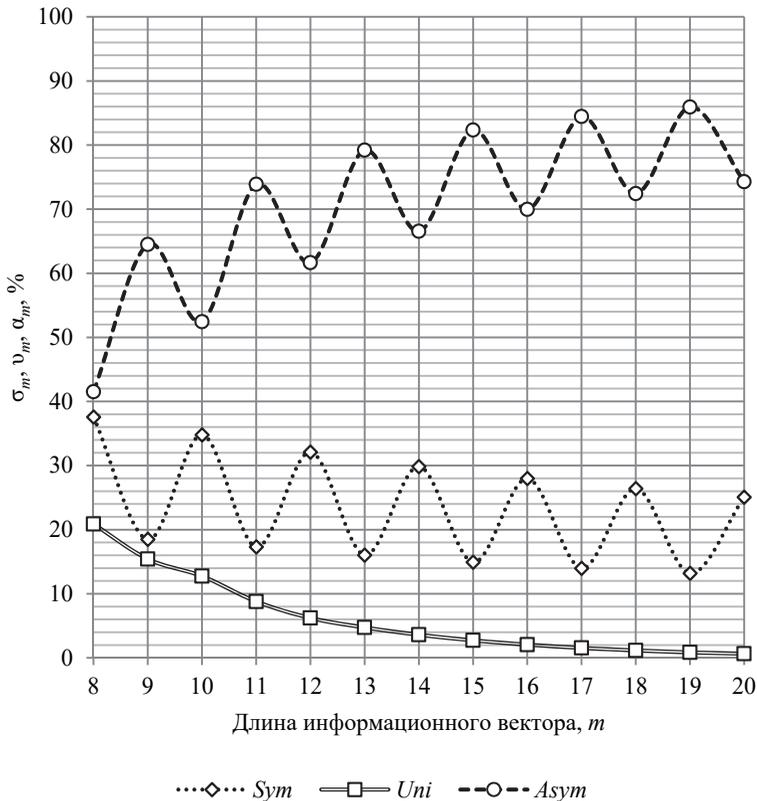


Рис. 5. Распределение необнаруживаемых ошибок по видам в $WT8(m,k)$ -кодах

Следует отметить, что некоторым предельным случаем для модульных кодов является выбор значения модуля $M = 2^{\lceil \log_2(m+1) \rceil}$. Такой модуль соответствует модулю кода Бергера, для которого в классе необнаруживаемых присутствуют только симметричные ошибки, а значение величины $\sigma_m = 100\%$ вне зависимости от длины информационного вектора [21]. Таким образом, чем ближе для данной длины информационного вектора значение M к величине $M = 2^{\lceil \log_2(m+1) \rceil}$, тем ближе и значение величины σ_m к 100% (сравните графики на рисунке 6 и рисунке 7 при одинаковом значении длины информационного вектора).



Рис. 6. Распределение необнаруживаемых ошибок по видам в $S4(m,k)$ -кодах

Одной из важных характеристик (m,k) -кодов, которые следует учитывать при организации систем функционального контроля, является обнаруживающая способность в области ошибок малых кратностей, и прежде всего — в области двукратных ошибок. На рисунке 8 и рисунке 9 даются графики долей необнаруживаемых двукратных ошибок от общего количества двукратных ошибок (величин β_2 , %) для кодов с суммированием с двумя и тремя контрольными разрядами. Известно [21], что для $SM(m,k)$ -кодов значение величины β_2 не зависит от длины информационного вектора и является постоянным. Для любого $SM(m,k)$ -кода с $M > 2$ значение $\beta_2 = 50\%$. Значение же величины β_2 для модульных кодов со взвешенными переходами является меньшим. Например, для $WT4(m,k)$ -кодов с длинами информационных векторов $m = 4 \div 20$ среднее значение величины $\beta_2 = 40,2347\%$, а для $WT8(m,k)$ -кодов с длинами информационных векторов $m = 8 \div 20$ — $\beta_2 = 25,4582\%$.

Чем выше значение M для $WTM(m,k)$ -кода, тем меньше значение величины β_2 .

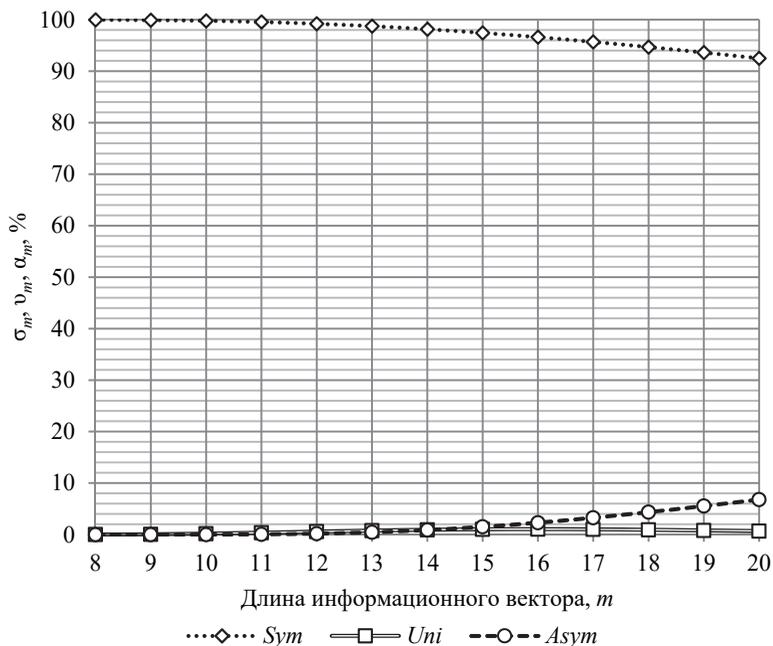


Рис. 7. Распределение необнаруживаемых ошибок по видам в $S8(m,k)$ -кодах

В таблицах 6 и 7 даются более детальные характеристики рассматриваемых кодов с суммированием: указаны доли не обнаруживаемых кодами ошибок различных кратностей $d \in \{1; 2; \dots; 8\}$ от общего количества ошибок данных кратностей для $SM(m,k)$ и $WTM(m,k)$ кодов с модулями $M=4$ и $M=8$.

При четных значениях m значения долей не обнаруживаемых $WT4(m,k)$ -кодами ошибок кратностью d от общего количества ошибок данной кратностью при $m \geq 6$ соизмеримы с данным показателем для $S4(m,k)$ -кода.

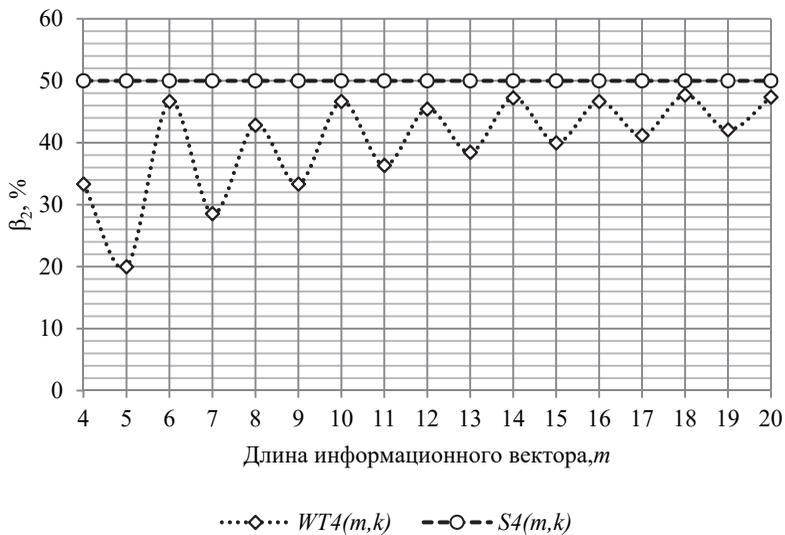


Рис. 8. Доля необнаруживаемых двукратных ошибок в $WT4(m,k)$ и $S4(m,k)$ кодах

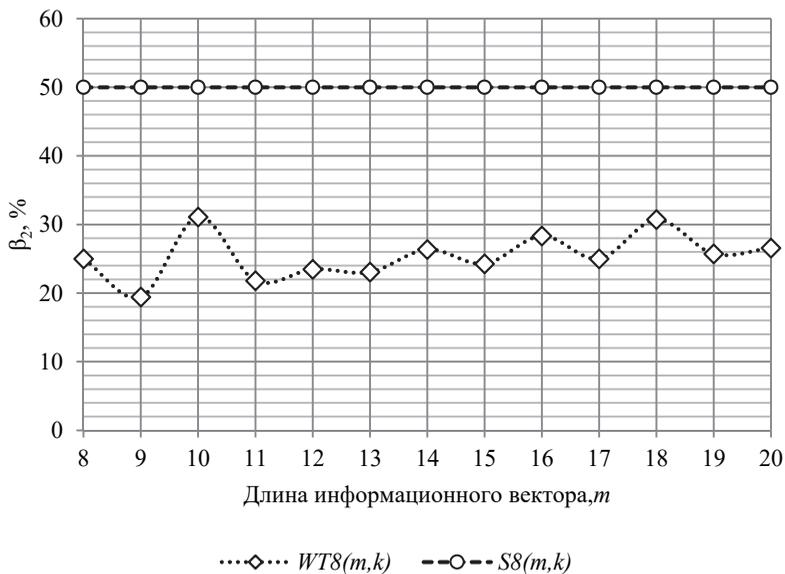


Рис. 9. Доля необнаруживаемых двукратных ошибок в $WT8(m,k)$ и $S8(m,k)$ кодах

Таблица 6. Доли не обнаруживаемых ошибок $S4(m,k)$ и $WT4(m,k)$ кодами по каждой кратности

m	Значение доли необнаруживаемых ошибок кратностью d от общего количества ошибок данной кратностью, %							
	1	2	3	4	5	6	7	8
$S4(m,k)$ -коды								
$\forall m$	0	50	0	50	0	50	0	50
$WT4(m,k)$ -коды								
4	0	33,3333	0	100	–	–	–	–
5	20	20	20	20	100	–	–	–
6	0	46,6667	0	46,6667	0	100	–	–
7	0	28,5714	25,7143	25,7143	28,5714	0	100	–
8	0	42,8571	0	54,2857	0	42,8571	0	100
9	11,1111	33,3333	14,2857	30,1587	30,1587	14,2857	33,3333	11,1111
10	0	46,6667	0	50,4762	0	50,4762	0	46,6667
11	0	36,3636	15,1515	33,3333	21,645	21,645	33,3333	15,1515
12	0	45,4545	0	51,5152	0	48,9177	0	51,5152
13	7,6923	38,4615	10,4895	35,6643	19,8135	26,3403	26,3403	19,8135
14	0	47,2527	0	50,4496	0	49,9167	0	49,9167
15	0	40	10,7692	37,4359	16,317	29,6503	23,6053	23,6053
16	0	46,6667	0	50,7692	0	49,6503	0	50,272
17	5,8824	41,1765	8,2353	38,8235	14,9321	32,1267	20,4443	26,6146
18	0	47,7124	0	50,3268	0	49,9246	0	50,016
19	0	42,1053	8,3591	39,9381	13,0031	34,0557	18,5044	29,0307
20	0	47,3684	0	50,4644	0	49,8452	0	50,0834

При нечетных значениях m значения долей необнаруживаемых $WT4(m,k)$ -кодами ошибок кратностью d от общего количества ошибок данной кратностью существенно меньше указанного показателя для $S4(m,k)$ -кода, что связано с присутствием для взвешенных кодов с суммированием в классе необнаруживаемых ошибок с нечетными кратностями.

Следует отметить, что $WT4(m,k)$ -коды с четными значениями m даже несколько «проигрывают» $S4(m,k)$ -кодам по характеристикам обнаружения ошибок с кратностями $d=4$ и $d=8$. Такая закономерность характерна и для ошибок с большими кратностями (в таблице 6 они не представлены).

Более заметны преимущества модульных взвешенных кодов с суммированием при сравнении между собой $WT8(m,k)$ и $S8(m,k)$ кодов (таблица 7).

Таблица 7. Доли не обнаруживаемых ошибок $S8(m,k)$ и $WT8(m,k)$ кодами по каждой кратности

m	Значение доли необнаруживаемых ошибок кратностью d от общего количества ошибок данной кратностью, %							
	1	2	3	4	5	6	7	8
$S8(m,k)$ -коды								
$\forall m$	0	50	0	37,5	0	31,25	0	28,125
$WT8(m,k)$ -коды								
8	0	25	0	22,8571	0	25	0	100
9	11,1111	19,4444	8,3333	12,6984	12,6984	8,3333	19,4444	11,1111
10	0	31,1111	0	23,3333	0	23,3333	0	31,1111
11	0	21,8182	8,4848	15,4545	10,8225	10,8225	15,4545	8,4848
12	0	23,4848	0	25,0505	0	25	0	25,0505
13	0	23,0769	4,1958	16,7832	10,4895	13,1702	13,1702	10,4895
14	0	26,3736	0	24,1758	0	25,2081	0	25,2081
15	0	24,2857	4,9451	18,1319	8,0753	14,7053	12,0124	12,0124
16	0	28,3333	0	24,7253	0	24,6503	0	25,439
17	5,8824	25	5	18,9076	7,2721	15,9502	10,1501	13,4677
18	0	30,719	0	25,098	0	24,6283	0	25,1291
19	0	25,731	4,4376	19,711	6,5703	16,9099	9,1927	14,5855
20	0	26,5789	0	24,871	0	24,9329	0	25,0742

$WT8(m,k)$ -коды обнаруживают большее количество ошибок как кратностью $d=2$, так и ошибок других представленных в таблице 7 четных кратностей, по сравнению с $S8(m,k)$ -кодами. Тем не менее $WT8(m,k)$ -коды с нечетными значениями длин информационных векторов в классе необнаруживаемых имеют некоторую долю ошибок с нечетными кратностями. В среднем их доля от общего количества ошибок соответствующей кратности для диапазона нечетных значений длин информационных векторов 8÷20 равна: $d=3$ — 5,8994%; $d=5$ — 9,3214%; $d=7$ — 13,2374%. С увеличением значения d (с учетом только нечетных d) доля необнаруживаемых ошибок от общего количества ошибок данной кратностью увеличивается.

Следует отметить, что в целом улучшение обнаруживающей способности в $WTM(m,k)$ -кодах по сравнению с $SM(m,k)$ -кодами достигается не только за счет более сбалансированного распределения ин-

формационных векторов между всеми контрольными векторами, но и для нечетных значений m — за счет добавления в класс необнаруживаемых ошибок с нечетными кратностями.

4. Заключение. Характеристики обнаружения ошибок в информационных векторах $SM(m,k)$ -кодов улучшают за счет простых правил модификации, связанных с установлением неравноправия между разрядами в информационных векторах. Описанный в статье алгоритм модификации, основанный на операции взвешивания переходов между разрядами, занимающими соседние позиции в информационных векторах, а также вычислении суммарного веса информационного вектора по модулю $M \in \{2; 4; \dots; 2^{\lceil \log_2(m+1) \rceil - 1}\}$, позволяет уменьшить число необнаруживаемых $SM(m,k)$ -кодами ошибок.

В статье показано, что модульные взвешенные коды с суммированием, в отличие от классических модульных кодов с суммированием единичных разрядов, не обнаруживают минимальное общее количество ошибок в информационных векторах при своих значениях m и k . Улучшение свойств модульных кодов с суммированием достигается за счет перераспределения информационных векторов между всеми контрольными векторами. При этом в класс необнаруживаемых добавляется существенное количество (по сравнению с количеством таких ошибок в $SM(m,k)$ -кодах) так называемых асимметричных ошибок в информационных векторах.

Предложенные в работе семейства $WTM(m,k)$ -кодов расширяют множество разделимых кодов с простыми правилами построения и небольшой избыточностью, пригодных для решения задач технической диагностики дискретных систем автоматики и вычислительной техники.

Литература

1. *Кравцов Ю.А., Архипов Е.В., Бакин М.Е.* Перспективные способы кодирования рельсовых цепей тональной частоты // Автоматика на транспорте. 2015. Том 1. № 2. С. 119–126.
2. *Шаманов В.И.* Обобщенная математическая модель процесса эксплуатации систем автоматики и телемеханики // Автоматика на транспорте. 2016. Том 2. № 2. С. 163–179.
3. *Ubar R., Raik J., Vierhaus H.-T.* Design and Test Technology for Dependable Systems-on-Chip (Premier Reference Source) // Information Science Reference. Hershey—New York. IGI Global. 2011. 578 p.
4. *Fujiwara E.* Code Design for Dependable Systems: Theory and Practical Applications // New Jersey: John Wiley & Sons. 2006. 720 p.
5. *Nicolaidis M., Zorian Y.* On-Line Testing for VLSI – A Compendium of Approaches // Journal of Electronic Testing: Theory and Application. 1998. vol. 12. Issue 1-2. pp. 7–20.
6. *Дрозд А.В. и др.* Рабочее диагностирование безопасных информационно-управляющих систем / Под ред. А.В. Дрозда и В.С. Харченко // Харьков: Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ». 2012. 614 с.

7. *Drozdz A. et al.* Objects and Methods of On-Line Testing: Main Requirements and Perspectives of Development // Proceedings of 14th IEEE East-West Design & Test Symposium (EWDTS'2016). 2016. pp. 72–76.
8. *Kharchenko V., Kondratenko Yu., Kasprzyk J.* Green IT Engineering: Concepts, Models, Complex Systems Architectures // Springer Book series "Studies in Systems, Decision and Control". 2017. vol. 74. 305 p.
9. *Mitra S., McCluskey E.J.* Which Concurrent Error Detection Scheme to Choose? // Proceedings of International Test Conference. USA. 2000. pp. 985–994.
10. *Blyudov A.A.* On the Synthesis of Test Equipment for Modulo Codes with Summation // Proceedings of Petersburg Transport University. 2013. vol. 1. pp. 53–58.
11. *Пархоменко П.П., Согомонян Е.С.* Основы технической диагностики (оптимизация алгоритмов диагностирования, аппаратурные средства) // М.: Энергоатомиздат. 1981. 320 с.
12. *Matrosova A.Yu., Ostantin S.A., Kirienko I.E.* All Stuck-at Fault Test Patterns and Incompletely Specified Boolean Functions // Proceeding of the 11th International Workshop on Boolean Problems. Freiberg. Germany. 2014. pp. 165–170.
13. *Черкасова Т.Х.* Об обнаружении ошибок в системах автоматики и вычислительной техники с помощью кодов Бергера и его модификаций // Сборник трудов научно-практической конференции Проблемы безопасности и надежности микропроцессорных комплексов. СПб: ИГУПС. 2015. С. 167–172.
14. *Berger J.M.* A Note on Error Detecting Codes for Asymmetric Channels // Information and Control. 1961. vol. 4. Issue 1. pp. 68–73.
15. *Согомонян Е.С., Слабаков Е.В.* Самопроверяемые устройства и отказоустойчивые системы // М.: Радио и связь. 1989. 207 с.
16. *Busaba F.Y., Lala P.K.* Self-Checking Combinational Circuit Design for Single and Unidirectional Multibit Errors // Journal of Electronic Testing: Theory and Applications. 1994. vol. 5. no. 1. pp. 19–28.
17. *Göessel M., Ocheretny V., Sogomonyan E., Marienfeld D.* New Methods of Concurrent Checking: Edition 1 // Dordrecht: Springer Science+Business Media B.V. 2008. 184 p.
18. *Ефанов Д.В., Сапожников В.В., Сапожников Вл.В.* О свойствах кода с суммированием в схемах функционального контроля // Автоматика и телемеханика. 2010. № 6. С. 155–162.
19. *Bose B., Lin D.J.* Systematic Unidirectional Error-Detection Codes // IEEE Transactions on Computers. 1985. vol. C-34. pp. 1026–1032.
20. *Das D., Toubia N.A.* Synthesis of Circuits with Low-Cost Concurrent Error Detection Based on Bose-Lin Codes // Journal of Electronic Testing: Theory and Applications. 1999. vol. 15. no. 1-2. pp. 145–155.
21. *Sapozhnikov V., Sapozhnikov V.I., Efanov D.* Modular Sum Code in Building Testable Discrete Systems // Proceedings of 13th IEEE East-West Design & Test Symposium (EWDTS'2015). 2015. pp. 181–187.
22. *Ефанов Д.В., Сапожников В.В., Сапожников Вл.В.* Применение модульных кодов с суммированием для построения систем функционального контроля комбинационных логических схем // Автоматика и телемеханика. 2015. № 10. С. 152–169.
23. *Черпанова М.Р.* Исследование влияния значения модуля кода с суммированием на структурную избыточность систем функционального контроля // Известия Петербургского университета путей сообщения. 2016. № 2. С. 279–288.
24. *Piestrak S.J.* Design of Self-Testing Checkers for Unidirectional Error Detecting Codes // Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 1995. 111 p.
25. *Das D., Toubia N.A.* Weight-Based Codes and Their Application to Concurrent Error Detection of Multilevel Circuits // Proceedings of 17th IEEE Test Symposium. 1999. pp. 370–376.

26. *Das D., Touba N.A., Seuring M., Gossel M.* Low Cost Concurrent Error Detection Based on Modulo Weight-Based Codes // Proceedings of the 6th IEEE International On-Line Testing Workshop (IOLTW). Spain. 2000. pp. 171–176.
27. *Sapozhnikov V. et al.* Optimum Sum Codes, that Effectively Detect the Errors of Low Multiplicities // Radioelectronics & Informatics. 2015. Issue 1. pp. 17–22.
28. *Сапожников В.В., Сапожников Вл.В., Ефанов Д.В., Никитин Д.А.* Модульно взвешенный код с суммированием для систем технического диагностирования // Информатика и системы управления. 2015. № 3. С. 53–62.
29. *Дмитриев В.В.* О двух способах взвешивания и их влиянии на свойства кодов с суммированием взвешенных переходов в системах функционального контроля логических схем // Известия Петербургского университета путей сообщения. 2015. № 3. С. 119–129.
30. *Блюдов А.А., Ефанов Д.В., Сапожников В.В., Сапожников Вл.В.* Построение модифицированного кода Бергера с минимальным числом необнаруживаемых ошибок информационных разрядов // Электронное моделирование. 2012. Том 34. № 6. С. 17–29.

Сапожников Валерий Владимирович — д-р техн. наук, профессор, профессор кафедры автоматки и телемеханики на железных дорогах, Петербургский государственный университет путей сообщения Императора Александра I (ПГУПС). Область научных интересов: надежностный синтез дискретных устройств, синтез безопасных систем, синтез самопроверяемых схем, техническая диагностика дискретных систем. Число научных публикаций — 600. port.at.pgups@gmail.com; Московский пр., 9, Санкт-Петербург, 190031; р.т.: +7(812)310-0788, Факс: +7(812)457-8579.

Сапожников Владимир Владимирович — д-р техн. наук, профессор, профессор кафедры автоматки и телемеханики на железных дорогах, Петербургский государственный университет путей сообщения Императора Александра I (ПГУПС). Область научных интересов: надежностный синтез дискретных устройств, синтез безопасных систем, синтез самопроверяемых схем, техническая диагностика дискретных систем. Число научных публикаций — 600. at.pgups@gmail.com; Московский пр., 9, Санкт-Петербург, 190031; р.т.: +7(812)310-0788, Факс: +7(812)457-8579.

Ефанов Дмитрий Викторович — к-т техн. наук, доцент кафедры автоматки и телемеханики на железных дорогах, Петербургский государственный университет путей сообщения Императора Александра I (ПГУПС). Область научных интересов: дискретная математика, надежность и техническая диагностика дискретных систем. Число научных публикаций — 200. TrES-4b@yandex.ru; Московский пр., 9, Санкт-Петербург, 190031; р.т.: +7(911)709-2164, Факс: +7(812)457-8579.

Котенко Алексей Геннадьевич — д-р техн. наук, доцент, заведующий кафедрой управления эксплуатационной работой, Петербургский государственный университет путей сообщения Императора Александра I (ПГУПС). Область научных интересов: организация движения на транспорте, риск-ориентированные методы управления, обработка неточных данных и представление знаний с неопределенностью. Число научных публикаций — 117. algenko@gmail.com; Московский пр., 9, Санкт-Петербург, 190031; р.т.: +7(812)457-8992.

V.V. SAPOZHNIKOV, VL.V. SAPOZHNIKOV, D.V. EFANOV, A.G. KOTENKO
**MODULO CODES WITH SUMMATION
OF WEIGHTED TRANSITIONS WITH NATURAL NUMBER
SEQUENCE OF WEIGHTS**

Sapozhnikov V.V., Sapozhnikov V.V., Efanov D.V., Kotenko A.G. Modulo Codes with Summation of Weighted Transitions with Natural Number Sequence of Weights.

Abstract. Systematic codes with simple formation rules and low redundancy are used in the reliable digital systems design. Codes with summation with a constant number of check bits, which does not depend on data vector length, are often applied. The paper shows that error detection properties of these codes could be improved by a modification of check bits calculation rules. The authors offer the algorithm of the formation of a code with summation based on weighing the transitions between bits on neighboring positions of data vectors and calculation of the total weight of the data vector by the modulo, which is degree of two. The paper contains a detailed analysis of properties of new codes compared with known ones. The authors state the conditions of the formation of error-tolerant modulo weight-based codes with summation with the maximum number of detected errors in data vectors. Improving of error detection characteristics in new codes, compared with classic ones, is achieved by the uniform distribution of data vectors among check vectors. This, in its turn, brings some part of monotonous and asymmetric errors to the class of undetectable in addition to symmetric ones. Modulo weight-based codes detect more symmetric errors than modulo codes with summation of ones, but asymmetric errors are the dominant sort of them for almost any length of data vector. Modulo codes with summation of weighted transitions are the perspective class of systematic codes to solve the tasks of digital systems technical diagnostics.

Keywords: technical diagnostics, concurrent error detection, systematic code, Berger code, modulo codes with sum, weight-transition code with sum, undetectable error, code characteristics.

Sapozhnikov Valery Vladimirovich — Dr. Sci., professor, professor of automation and remote control on railways department, Emperor Alexander I St. Petersburg state transport university. Research interests: reliable synthesis of discrete devices, safe systems synthesis, self-checking systems synthesis, discrete systems technical diagnostics. The number of publications — 600. port.at.pgups@gmail.com; 9, Moskovsky ave., St. Petersburg, 190031; office phone: +7(812)310-0788, Fax: +7(812)457-8579.

Sapozhnikov Vladimir Vladimirovich — Dr. Sci., professor, professor of automation and remote control on railways department, Emperor Alexander I St. Petersburg state transport university. Research interests: reliable synthesis of discrete devices, safe systems synthesis, self-checking systems synthesis, discrete systems technical diagnostics. The number of publications — 600. at.pgups@gmail.com; 9, Moskovsky ave., St. Petersburg, 190031; office phone: +7(812)310-0788, Fax: +7(812)457-8579.

Efanov Dmitry Viktorovich — Ph.D., associate professor of automation and remote control on railways department, Emperor Alexander I St. Petersburg state transport university. Research interests: discrete mathematics, reliability and technical diagnostics of discrete devices. The number of publications — 200. TrES-4b@yandex.ru; 9, Moskovsky ave., St. Petersburg, 190031; office phone: +7(911)709-2164, Fax: +7(812)457-8579.

Kotenko Alexey Gennad'evich — Dr. Sci., associate professor, head of operational work department, Emperor Alexander I St. Petersburg state transport university. Research interests: transport traffic organization, risk-oriented control methods, inaccurate data processing and uncertain presentation of data. The number of publications — 117. algenko@gmail.com; 9, Moskovsky ave., St. Petersburg, 190031; office phone: +7(812)457-8992.

References

1. Kravtsov Yu.A., Arhipov E.V., Bakin M.E. [Advanced coding schemes of voice-frequency track circuits]. *Avtomatika na transporte – Automation on Transport*. 2015. vol. 1. Issue 2. pp. 119–126. (In Russ.).
2. Shamanov V.I. [Generalised mathematical body of the operational process of automation and remote control system]. *Avtomatika na transporte – Automation on Transport*. 2016. vol. 2. Issue 2. pp. 163–179. (In Russ.).
3. Ubar R., Raik J., Vierhaus H.-T. Design and Test Technology for Dependable Systems-on-Chip (Premier Reference Source). Information Science Reference. Hershey – New York. IGI Global. 2011. 578 p.
4. Fujiwara E. Code Design for Dependable Systems: Theory and Practical Applications. New Jersey: John Wiley & Sons. 2006. 720 p.
5. Nicolaidis M., Zorian Y. On-Line Testing for VLSI – A Compendium of Approaches. *Journal of Electronic Testing: Theory and Application*. 1998. vol. 12. Issue 1-2. pp. 7–20.
6. Drozd A.V. et al. *Rabochee diagnostirovanie bezopasnyh informacionno-upravlyayushchih system. Pod red. A.V. Drozda i V.S. Harchenko* [Objects and methods of on-line testing for safe instrumentation and control systems. Edited by A.V. Drozd, and V.S. Harchenko]. Kharkov. National Aerospace University "KhAI". 2012. 614 p. (In Russ.).
7. Drozd A. et al. Objects and Methods of On-Line Testing: Main Requirements and Perspectives of Development. Proceedings of 14th IEEE East-West Design & Test Symposium (EWDTS 2016). 2016. pp. 72–76.
8. Kharchenko V., Kondratenko Yu., Kacprzyk J. Green IT Engineering: Concepts, Models, Complex Systems Architectures. Springer Book series "Studies in Systems, Decision and Control". 2017. vol. 74. 305 p.
9. Mitra S., McCluskey E.J. Which Concurrent Error Detection Scheme to Choose? Proceedings of International Test Conference. USA. 2000. pp. 985–994.
10. Blyudov A.A. On the Synthesis of Test Equipment for Modulo Codes with Summation. Proceedings of Petersburg Transport University. 2013. vol. 1. pp. 53–58.
11. Parhomenko P.P., Sogomonyan E.S. *Osnovy tekhnicheskoy diagnostiki (optimizaciya algoritmov diagnostirovaniya, apparaturnye sredstva)* [Technical diagnosis fundamentals (Diagnostic algorithm optimization, apparatus means)]. Moscow: Energoatomizdat. 1981. 320 p. (In Russ.).
12. Matrosova A.Yu., Ostanin S.A., Kirienko I.E. All Stuck-at Fault Test Patterns and Incompletely Specified Boolean Functions. Proceeding of the 11th International Workshop on Boolean Problems. Freiberg, Germany. 2014. pp. 165–170.
13. Cherkasova T.H. [About error detection in automation and computer engineering with Berger code and its modification]. *Sbornik trudov nauchno-prakticheskoy konferencii Problemy bezopasnosti i nadezhnosti mikroprocessornykh kompleksov* [Proceedings of scientific-practical conference "Safety and reliability of microprocessors problems"]. St. Petersburg. Emperor Alexander I St. Petersburg state transport university. 2015. pp. 167–172. (In Russ.).
14. Berger J.M. A Note on Error Detecting Codes for Asymmetric Channels. *Information and Control*. 1961. vol. 4. Issue 1. pp. 68–73.

15. Sogomonyan E.S., Slabakov E.V. *Samoproveryaemye ustroystva i otkazoustoychivyye sistemy* [Self-Checking and Fail-Safety Systems]. M.: Radio and svjaz'. 1989. 207 p. (In Russ.).
16. Busaba F.Y., Lala P.K. Self-Checking Combinational Circuit Design for Single and Unidirectional Multibit Errors. *Journal of Electronic Testing: Theory and Applications*. 1994. vol. 5. no. 5. pp. 19–28.
17. Göessel M., Ocheretny V., Sogomonyan E., Marienfeld D. *New Methods of Concurrent Checking: Edition 1*. Dordrecht: Springer Science+Business Media B.V. 2008. 184 p.
18. Efanov D.V., Sapozhnikov V.V., Sapozhnikov V.I.V. [On summation code properties in functional control circuits]. *Avtomatika i telemekhanika – Automation and remote control*. 2010. vol. 6. pp. 155–162. (In Russ.).
19. Bose B., Lin D.J. Systematic Unidirectional Error-Detection Codes. *IEEE Transactions on Computers*. 1985. vol. C-34. pp. 1026–1032.
20. Das D., Toubia N.A. Synthesis of Circuits with Low-Cost Concurrent Error Detection Based on Bose-Lin Codes. *Journal of Electronic Testing: Theory and Applications*. 1999. vol. 15. no. 1-2. pp. 145–155.
21. Sapozhnikov V., Sapozhnikov V.I., Efanov D. Modular Sum Code in Building Testable Discrete Systems. Proceedings of 13th IEEE East-West Design & Test Symposium (EWDTS'2015). 2015. pp. 181–187.
22. Efanov D.V., Sapozhnikov V.V., Sapozhnikov V.I.V. [Application of modular summation codes to concurrent error detection systems for combinational Boolean circuits]. *Avtomatika i telemekhanika – Automation and remote control*. 2015. vol. 10. pp. 152–169. (In Russ.).
23. Cherepanova M.R. [A study into influence of sum code module value on structure redundancy of concurrent error detection systems]. *Izvestija Peterburgskogo universiteta putej soobshhenija – Proceedings of Petersburg Transport University*. 2016. vol. 2. pp. 279–288.
24. Piestrak S.J. *Design of Self-Testing Checkers for Unidirectional Error Detecting Codes*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej. 1995. 111 p.
25. Das D., Toubia N.A. Weight-Based Codes and Their Application to Concurrent Error Detection of Multilevel Circuits. Proceedings of 17th IEEE Test Symposium. 1999. pp. 370–376.
26. Das D., Toubia N.A., Seuring M., Gossel M. Low Cost Concurrent Error Detection Based on Modulo Weight-Based Codes. Proceedings of the 6th IEEE International On-Line Testing Workshop (IOLTW). Spain. 2000. pp. 171–176.
27. Sapozhnikov V. et al. Optimum Sum Codes, that Effectively Detect the Errors of Low Multiplicities. *Radioelectronics & Informatics*. 2015. vol. 1. pp. 17–22.
28. Sapozhnikov V.V., Sapozhnikov V.I.V., Efanov D.V., Nikitin D.A. [Modulo weighted code with summation for technical diagnostics systems]. *Informatika i sistemy upravlenija – Information science and control systems*. 2015. vol. 3. pp. 53–62. (In Russ.).
29. Dmitriev V.V. [On the two weighing methods and their influence on properties of sum codes of weighted transitions in functional logical scheme control systems]. *Izvestija Peterburgskogo universiteta putej soobshhenija – Proceedings of Petersburg Transport University*. 2015. vol. 3. pp. 119–129. (In Russ.).
30. Blyudov A.A., Efanov D.V., Sapozhnikov V.V., Sapozhnikov V.I.V. [Formation of the Berger modified code with minimum number of undetectable errors of data bits]. *Jelektronoe modelirovanie – Electronic Modeling*. 2012. vol. 34. no. 6. pp. 17–29. (In Russ.).

С.А. МАРЧЕНКОВ, А.С. ВДОВЕНКО, Д.Ж. КОРЗУН
**РАСШИРЕНИЕ ВОЗМОЖНОСТЕЙ СОВМЕСТНОЙ
ДЕЯТЕЛЬНОСТИ В ИНТЕЛЛЕКТУАЛЬНОМ ЗАЛЕ НА
ОСНОВЕ СЕРВИСОВ ЭЛЕКТРОННОГО ТУРИЗМА**

Марченков С.А., Вдовенко А.С., Корзун Д.Ж. Расширение возможностей совместной деятельности в интеллектуальном зале на основе сервисов электронного туризма.

Аннотация. Система интеллектуального зала реализует цифровое сервисно-ориентированное окружение совместной деятельности участников. В статье исследуется задача по расширению таких систем сервисами электронного туризма. Предлагается сценарий, расширяющий возможности совместной деятельности в интеллектуальном зале за счет вовлечения самих участников в составление культурной программы для изучения региона, в котором проходит мероприятие. В качестве конкретного программного приложения используется система SmartRoom, предназначенная для информационно-организационной поддержки проведения совместной деятельности участников в помещении в рамках таких мероприятий как конференции и совещания. На основе подхода интеллектуальных пространств для системы SmartRoom разработан и исследован набор сервисов, реализующий предлагаемый сценарий составления культурной программы. Предложены онтологическая модель и способы организации взаимодействия, обеспечивающие построение и доставку сервисов на основе накопления и обработки информации множеством программных агентов в интеллектуальном пространстве системы SmartRoom. Выполнено экспериментальное исследование производительности полученных программных реализаций.

Ключевые слова: интеллектуальные пространства, информационные сервисы, электронный туризм, интеллектуальный зал, Smart-M3, онтологическая модель.

1. Введение. Туризм является динамически развивающимся сектором использования разнообразных цифровых услуг для поддержки социально-экономической и культурной деятельности человека [1], определяя такую прикладную область информационных технологий, как электронный туризм (е-туризм). Развитие информационных технологий и их широкое применение в сфере туризма определяют е-туризм [2]. Число Интернет-пользователей в мире растет, достигая к настоящему времени около 3,3 миллиардов [3], что формирует существенное множество потенциальных потребителей сервисов е-туризма. Такие пользователи начинают понимать преимущества информационных сервисов при планировании и осуществлении своей туристской деятельности [2, 4, 5].

Распространенный вариант информационных сервисов е-туризма реализуется на основе веб-подхода, когда пользователь использует свое персональное мобильное устройство (например, смартфон или планшет) в качестве средства доступа к различным веб-сервисам [5]. Таким образом, при условии наличия доступа к глобальной сети Интернет, пользователь получает (как сервис) необходимую информацию для возникающей у него задачи в текущем месте пребывания. В статье рассматривается более общий случай, когда построение сервисов выполняется текущей вычислительной средой, в которой

в данный момент находится пользователь. В этом случае, во-первых, построение и доставка сервиса происходит не только на удаленной системе, а можно использовать и доступные цифровые устройства вычислительной среды. Во-вторых, исходные сервисы вычислительной среды могут быть дополнены за счет сервисов e-туризма, повышая эффективность информационной поддержки пользователей при решении ими тех задач, для которых эта среда создавалась.

В качестве референтного класса вычислительных сред в статье рассматриваются такие цифровые окружения совместной деятельности человека как интеллектуальный зал [6, 7, 8], предназначенные для информационно-организационной поддержки людей-участников при проведении ими мероприятий в помещении. Исследуется задача по расширению интеллектуального зала за счет сервисов e-туризма. В частности, предлагается использовать сервисы составления культурной программы, которые не только автоматизируют для организаторов составление такой программы с учетом имеющихся возможностей, но и позволяют участникам самим участвовать в ее составлении в ходе выполнения основной деятельности с учетом имеющихся интересов и предпочтений.

Для исследуемой задачи разработка сервисов e-туризма может использовать подход интеллектуальных пространств, который определяет способы создания сервисно-ориентированного окружения в заданной вычислительной среде с разделением ресурсов между разнородными цифровыми устройствами [9, 10]. В частности, система SmartRoom [11] предоставляет набор сервисов для поддержки проведения таких видов совместной деятельности, как конференции и совещания. Персональные мобильные устройства являются для пользователей точками доступа к соответствующим сервисам и точками управления ими [12]. Исходные сервисы системы SmartRoom направлены на информационно-организационную поддержку совместной деятельности в помещении, оборудованном мультимедийной аппаратурой для отображения информации и имеющим локальную беспроводную сеть для взаимодействия внутри помещения и доступа к ресурсам глобальной сети Интернет. Другие варианты применения интеллектуальных пространств для задач e-туризма представлены, например, в работах [4, 13, 14].

В статье предлагается сценарий расширения интеллектуального зала сервисами e-туризма, повышающий эффективность совместной деятельности участников за счет привлечения их к составлению культурной программы для изучения региона, в котором проходит мероприятие [15]. Система SmartRoom рассматривается как конкретное программное приложение для изучения способов расширения дополняющими сервисами. Представлены решения проектирования, включающие архитектуру системы для создания расширенного интеллектуального пространства SmartRoom и онтологическую модель для накопления, представления и использования туристской информации

при построении сервисов. Проведено экспериментальное исследование с реализованным программными прототипом, демонстрирующее практическую осуществимость предлагаемого способа расширения окружений совместной деятельности сервисами e-туризма.

2. Сервисы e-туризма в интеллектуальном зале. В качестве одного из классов вычислительных сред для сервисов электронного туризма рассмотрим такие цифровые окружения поддержки совместной деятельности людей как интеллектуальный зал (ИЗ). Последний может быть реализован в виде «интеллектуального пространства» (ИП) в помещении, оборудованном мультимедийной аппаратурой для восприятия пользователями совместно создаваемой информации [11].

В общем случае под ИП понимается вычислительная среда, обеспечивающая динамическое множество участников контекстно-зависимыми и персонализированными сервисами [16, 17]. Участники представлены программными агентами, выполняющимися на окружающих вычислительных устройствах и удаленных ЭВМ, включая персональные мобильные устройства пользователей. Построение и доставка пользователям информационных сервисов выполняется множеством взаимодействующих агентов. Взаимодействие является косвенным — через общее информационное хранилище, использующее известные методы Семантического Веба для интероперабельного представления и доступа к информации [18].

В качестве конкретного варианта ИЗ рассмотрим систему SmartRoom [11]. Она реализована на программной платформе SmartM3 [19] следующим образом. Доступ к ИП выполняется с помощью брокера семантической информации (брокер SIB, от англ. semantic information broker), который позволяет представить информационное содержимое не только непосредственно с помощью данных, описывающих объекты, но и их семантики посредством моделей представления знаний Семантического Веба. Программные агенты действуют как процессоры знаний (агент КР, от англ. knowledge processor), взаимодействие которых друг с другом происходит через брокера SIB на основе совместного формирования информационного содержимого ИП и выполняет построение и доставку сервисов. Сервисы ИЗ разделяются на прикладные и опорные.

Прикладные сервисы. Непосредственно используемые пользователями сервисами являются сервис презентации и сервис повестки мероприятия [11], отвечающие за визуализацию информации на двух общественных экранах в помещении: экран презентации и экран повестки мероприятия. Экраны представляют собой широкоформатные дисплеи, видимые для всех участников в помещении. Экран презентации предназначен для управляемого показа слайдов текущим докладчиком во время выступления. Одновременно на экране повестки показывается порядок выступлений, докладчики и названия выступлений,

выделенного времени и др. Аналогичные сервисы доступны участнику на его персональном мобильном устройстве с помощью агента КР, реализующего мобильного клиента [12].

Оторные сервисы. Отвечают за автоматизацию обработки информации во время мероприятия для визуализации (доставки) на общественных экранах или экранах персональных мобильных устройств. Например, сервис управления конференцией составляет и по ходу мероприятия поддерживает программу конференции.

При проведении мероприятия в ИЗ, его участники в дополнение к основной деятельности могут быть заинтересованы в участии в культурной программе и, более того, готовы сами способствовать ее составлению и проведению. В то же время, известные ИЗ не предоставляют сервисов для автоматизации составления культурной программы с учетом интересов и предпочтений участников, так как ориентированы только на информационную поддержку основной деятельности при проведении мероприятия. Возникает задача по расширению ИЗ сервисами е-туризма с целью повышения эффективности предоставляемой информационной поддержки для учета других видов совместной деятельности, связанных с основной. В исследуемом случае решение этой задачи требует как исследования и разработки сценариев расширения ИЗ сервисами е-туризма, так и развития имеющихся архитектурных и онтологических моделей для программной реализации такого ИЗ в виде расширенного ИП.

Рассмотрим предлагаемый нами способ расширения системы SmartRoom дополнительными сервисами [11, 20, 21]. Концептуальная модель сервисов системы представлена на рисунке 1. Один из возможных вариантов для е-туризма — это сценарий составления культурной программы для участников конференции.

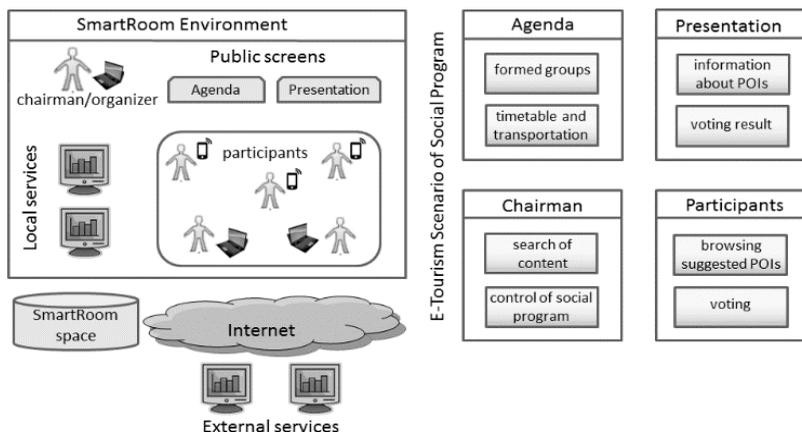


Рис. 1. Архитектурная схема расширения системы интеллектуального зала для поддержки электронного туризма

Культурная программа определяет точки интереса, которые можно посетить в свободное от основной деятельности время. Организаторы предоставляют предопределенный набор точек интереса в регионе проведения конференции. На основе этой информации участник может строить свои планы и принимать решения, связанные с интересующими его местами, временем их посещения и составом экскурсионных групп.

Процесс принятия решения является итеративным: участник обновляет свое решение в зависимости от решений других участников, которые можно увидеть на экранах прикладных сервисов. В результате, организаторы фиксируют культурную программу, определяя: (1) сформированные группы посещения точек интереса, (2) расписание посещения и (3) транспортную поддержку. Сценарий, в первую очередь, ориентирован на применение внутри помещения, где проходит конференция. Процесс построения и результат динамически отслеживается участниками на экранах презентации и повестки мероприятия, а также на экранах персональных мобильных устройств.

Для реализации описанного сценария предлагается набор дополнительных сервисов (сервис культурной программы, сервис поиска и сервис веб-страниц), приведенный в таблице 1. Соответствующая архитектурная схема приведена на рисунке 2. Каждый сервис реализуется в виде отдельного агента КР.

Таблица 1. Сервисы е-туризма для сценария составления культурной программы участниками конференции

Название	Описание
Сервис культурной программы	Управление составлением культурной программы, основанной как на доступных точках интереса, так и на интересах и предпочтениях участников.
Сервис поиска	Поиск информации для точек интереса во внешних источниках данных (напр., фотографии с веб-сервиса Flickr).
Сервис веб-страниц	Генерация веб-страниц на основе шаблонов и размещение соответствующих ссылок для доступа в пространство ИЗ.

Сервис культурной программы предоставляет информацию о возможных точках интереса для культурной программы каждому участнику системы ИЗ. Соответствующий агент КР запускается на компьютере организатора с графическим интерфейсом пользователя для просмотра информации о возможных точках интереса, наблюдения предлагаемых участниками решений и отслеживания процесса

составления культурной программы. Сервис культурной программы взаимодействует с сервисом поиска и сервисом веб-страниц.

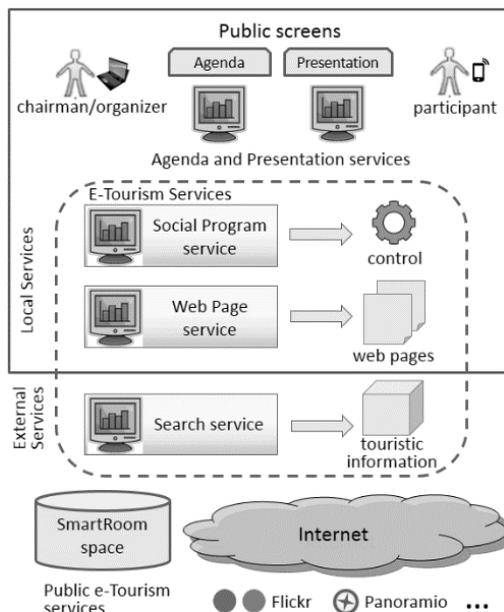


Рис. 2. Интеграция сервисов е-туризма в систему SmartRoom

Сервис поиска находит фотографии и другую информацию о соответствующих точка интереса. Поиск выполняется в ограниченной географической зоне вокруг места проведения конференции. Результат поиска размещается в пространстве ИЗ для последующего использования сервисом культурной программы.

Сервис веб-страниц отвечает за генерацию и хранение веб-страниц, в виде которых информация доставляется участникам. Сервис использует известные веб-технологии для упрощения процесса доставки с визуализацией на стороне мобильных клиентов (и других прикладных сервисов). В пространстве ИЗ размещаются только ссылки (URL) на соответствующие веб-страницы. Сервис реализуется как веб-сервер с доступом в Интернет.

Исходные сервисы системы SmartRoom — сервис презентации и сервис повестки мероприятия — также задействованы для визуализации туристской информации (в виде веб-страниц). Первый отображает страницу с доступными точками интереса и процесс голосования. Второй показывает текущий вариант культурной программы, которая составляется совместными усилиями участников.

3. Онтологическая модель. Для представления необходимой информации в сценарии построения культурной программы используется собственная онтологическая модель, дополняющая основную онтологическую модель системы SmartRoom. Под онтологической моделью понимается иерархическая структура понятий, объектов, определений, свойств и отношений определенной предметной области [22]. Онтологическая модель позволяет эффективно представлять информацию для обработки в отличие от способа, когда информация представлена в виде совокупности или коллекции по общей тематике, объединяемой общим понятием антологии [23]. Основными составными частями предлагаемой онтологической модели являются онтологическая модель сервиса и онтологическая модель пользовательского профиля.

Онтологическая модель сервиса культурной программы (класс SocialProgramService) представлена на рисунке 3. Ключевыми являются свойства hasClientUrl, hasAgendaServiceUrl и hasPresentationServiceUrl. Они используются для доступа к графическому интерфейсу пользователя, если он реализован как веб-приложение (рисунок 2). Значения этих свойств публикуются сервисом веб-страниц. Схожим образом свойства могут быть использованы и для других прикладных сервисов. Свойства serviceName, serviceDescription и serviceStatus являются общими для всех сервисов системы SmartRoom.

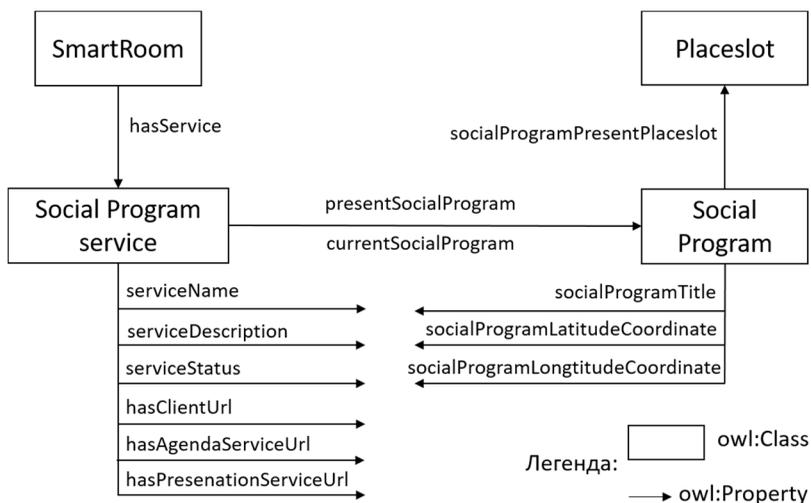


Рис. 3. Онтологическая модель сервиса культурной программы

Когда организаторы добавляют новый экземпляр культурной программы, соответствующий индивид класса SocialProgram публику-

ется в пространстве ИЗ. Этот индивид связан с индивидом сервиса, что позволяет идентифицировать индивиды между экземплярами сервисов в рамках одного пространства. Свойства данных, соответствующие названию и координатам, задаются при создании программы. Координаты места проведения программы используются для автоматического поиска точек интереса. Каждый вариант культурной программы представлен множеством индивидов класса Placeslot, доступных через объектное свойство socialProgramPresentPlaceslot.

Онтологическая модель для представления точек интереса показана на рисунке 4. Основными индивидами являются слот места (класс Placeslot), место (класс Place) и фото (класс Photo).

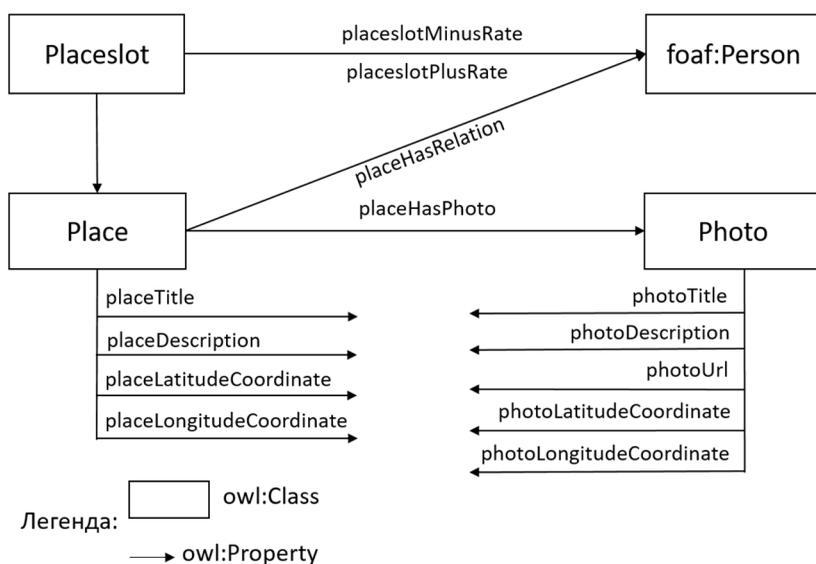


Рис. 4. Онтологическая модель точек интереса

Индивид класса Placeslot имеет объектное свойство placeslotPresentPlace, связывающее с индивидом класса Place. Базовый индивид места может быть расширен дополнительными свойствами, относящимися только к сервису культурной программы. Подобный способ используется для инкапсуляции индивидов, чтобы не влиять на базовую онтологическую модель и взаимодействие других сервисов, работающих с этими же индивидами. Другими объектными свойствами являются placeslotPlusRate и placeslotMinusRate для хранения результатов голосования участников с указанием связи на индивида участника. Данные свойства очищаются, когда начинается новое голосование.

Индивиды места (класс Place) и фото (класс Photo) являются общими для различных предметно-ориентированных сценариев [24]. Таким образом, онтологическая модель поддерживает вывод новых знаний относительно индивидов класса Person и Place. Поиск новых точек интереса может выполняться автоматически, основываясь на имеющихся индивидах классов Place и Photo.

Представленная онтологическая модель семантически связывает участников ИЗ с точками интереса, что сходно с результатами в [25, 26]. В общем случае такой тип связей в ИП позволяет не привлекать внешние источники данных. Примером такой связи является индивид класса Place с объектным свойством placeHasRelation. С точкой интереса может быть дополнительно связана историческая информация. В частности, может быть установлено отношение между отдельно взятой точкой интереса и регионом участника (напр., архитектор здания из этой же страны).

4. Построение сервисов. Рассмотрим способы построения введенных дополнительных сервисов на основе взаимодействия агентов КР в ИП системы SmartRoom. Сервис культурной программы является прикладным (ориентирован на конечного пользователя) — программа отображается на общественных экранах в помещении ИЗ или непосредственно на персональных мобильных устройствах участников. Сервисы поиска и веб-страниц являются опорными, обеспечивая информационную поддержку сервиса культурной программы.

А. Сервис культурной программы. Предоставляет организаторам управление и реализует составление культурной программы самими участниками. Функции управления включают в себя участие в составлении культурной программы, добавление новых точек интереса в программу, задание начальной информации, инициацию поиска информации в медиа-источниках.

Основными индивидами для онтологического построения культурной программы является класс SocialProgram и Placeslot. Индивиды классов Place и Photo содержат информацию о точках интереса и их фотографиях. Данная информация предоставляется другими сервисами. Взаимодействие с сервисами поиска и веб-страниц организовано через использование уведомлений по операции подписки в ИП.

При построении культурной программы участник предоставляет свои индивидуальные предпочтения и голосует схожим образом, как описано в [27]. Когда участник голосует за точку интереса, сервис веб-страниц посылает уведомление сервису культурной программы. Последний, в свою очередь, обрабатывает уведомление и обновляет программу. Голоса хранятся в индивидах класса Placeslot в соответствующую

щих свойствах `placeslotPlusRating` и `placeslotMinusRating` с заданием нужного индивида класса `Person`. Затем культурная программа формирует новый объект `JSON` и отправляет его сервису веб-страниц. Данный объект используется при построении содержимого страницы и ее визуализации для участников.

Б. Сервис поиска. Доступно множество открытых источников данных в сети Интернет с туристской информацией: описание точек интереса и событий, доступные туры, прогнозы погоды, транспорт, размещение и прочее. В частности, можно использовать такие хостинг-сервисы фотографий, как веб-сервисы `Flickr` и `Google Places`. Сервис поиска извлекает эту информацию. Описательная информация получается, учитывая местоположение точки интереса. Каждое найденное фото получает представление в виде индивида `Photo` со связью с соответствующим местом через объектное свойство `placeHasPhoto`.

В. Сервис веб-страниц. Доставка информации пользователям выполняется с помощью сервиса веб-страниц. На стороне пользователя достаточно иметь веб-обозреватель. Построение страницы происходит динамически с использованием параметров запроса. Исходное содержимое получается из ИП.

Каждый сервис, которому необходимо построение веб-страниц, должен определить шаблоны для этих страниц. В базовом варианте шаблон создается с помощью языка разметки `HTML` и языка программирования `Javascript`. Также шаблон может использовать стили `CSS` для визуальных эффектов. Ссылки на сгенерированные страницы публикуются в ИП и доступны через индивида класса `Service`: свойства `hasClientUrl`, `hasAgendaServiceUrl`, и `hasPresentationServiceUrl`.

Имеется набор глобальных объектов `JSON`, каждый из них хранит информацию сервиса. Эти объекты используются как содержимое для визуализации на стороне пользователя. Они также передаются как параметры для генерации соответствующих страниц.

Г. Взаимодействие. Взаимодействие между сервисами организовано на основе модели уведомлений [28]. Когда сервис публикует данные, другой сервис получает уведомление об изменениях в рамках подписки. Модель уведомлений описывает набор требуемых вариантов взаимодействия сервисов друг с другом. Отдельный сервис описывает собственный тип уведомлений, где каждый тип отвечает за вариант взаимодействия и представляется в ИП в виде набора `RDF`-троек. Необходимый для разрабатываемых сервисов набор уведомлений представлен в таблице 2. Используется тип запросов один-к-одному, за исключением уведомления `updateServiceInformation`, который используется действие один-ко-многим.

Таблица 2. Модель уведомлений для поддержки сервисов электронного туризма

Сервис	Уведомление	Параметр	Описание
Веб-страниц	startUpdateServicePage	Service	Вызывается для обновления содержимого сервиса класса Service
Культурная программа	voteUpdate	Placeslot	Активирует пересчет голосов для слота места
Презентации / повестки мероприятия	startServiceMode	Service	Показывает на экране сервис класса Service
Презентации / повестки мероприятия	stopServiceMode	Service	Прекращения показа сервиса класса Service, переключение в режим конференции
Презентации / повестки мероприятия	updateServicePage	Service	Обновление сервиса класса Service на экране

Используется операция подписки в ИП при представлении запроса в виде индивида для определения изменений конкретных свойств. После инициализации подписки все происходящие изменения, которые относятся к подписанному индивиду, отправляются всем подписчикам. В результате возрастает нагрузка на стороне брокера SIB и подписанных агентов КР, реализующих сервисы. Для уменьшения нагрузки используется «индивид запроса», который представляет запрос к ИП. Когда сервис публикует индивид запроса, этот сервис также подписывается на свойство, отвечающее за индикацию завершения обработки. Когда отвечающий сервис заканчивает выполнять запрошенные действия, этот сервис устанавливает статус обработки в соответствующее свойство. Отправитель получает только одно событие по подписке относительно данного индивида, и после проверки статуса запроса сервис удаляет индивид. Схожие идеи взаимодействия можно найти в [29].

Пример индивида запроса показан на рисунке 5. Взаимодействие между сервисом культурной программы и сервисом поиска с использованием этого индивида запроса представлено на рисунке 6. Сервис культурной программы и сервис поиска взаимодействуют, используя подписку на индивида класса SearchActivity. Индивид содержит параметры поиска (напр., местоположение точки интереса, тип поиска). Сервис поиска извлекает параметры и обращается к соответствующим веб-сервисам.

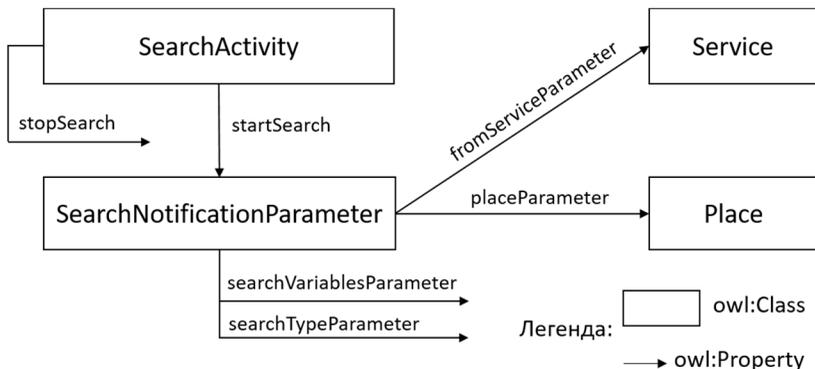


Рис. 5. Онтологическая модель индивида запроса поиска

При взаимодействии с сервисом веб-страниц использование операции подписки затруднительно, так как сервис не обладает информацией о всех индивидах. Сервис веб-страниц выступает в роли посредника между ИП и клиентской стороной (веб-приложением). Требуемое взаимодействие организовано с использованием параметризованных REST-запросов [30, 31]. Например, когда участник голосует, то клиентское веб-приложение посылает POST-запрос по специальному адресу. Запрос содержит следующие параметры: идентификатор слота места индивида класса Placeslot, идентификатор участника индивида класса Person и значение решения.



Рис. 6. Взаимодействие между сервисом культурной программы и сервисом поиска, используя индивид класса SearchActivity

Частным случаем является взаимодействие сервисов культурной программы и веб-страниц. Сервис культурной программы использует

подписку на уведомление VoteUpdate. Когда уведомление достигает сервиса культурной программы, последний создает новый объект JSON. Содержимое этого объекта определяется онтологической моделью культурной программы. Индивиды и их свойства хранятся в этом объекте.

После формирования объект JSON отправляется сервису веб-страниц по специальному адресу в POST-запросе. Сервис заменяет текущий экземпляр новым объектом JSON. Новый экземпляр используется для фоновых обновлений с использованием AJAX-вызовов и для построения страницы модулем шаблонов. Отметим, что данный механизм взаимодействия поддерживает расширение системы SmartRoom другими сервисами, взаимодействующими с сервисом веб-страниц.

Процесс поиска и инициализации построения культурной программы с использованием предложенных способов взаимодействия изображен на рисунке 7. После определения доступных точек интереса в программе организаторы запускают поиск фотографий и другой медиаинформации. Затем организаторы запускают процесс построения выбранной программы. Это включает в себя преобразование программы в объект JSON и отправку его сервису веб-страниц. Сервис обновляет параметры для построения шаблона и предоставляет обновленные страницы.

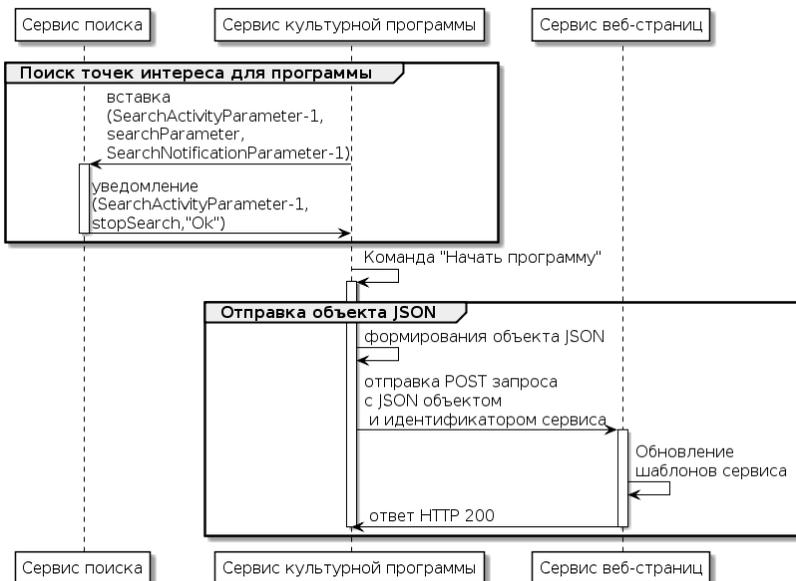


Рис. 7. Последовательность процессов поиска и начала построения культурной программы

Процесс доступа клиентов к страницам сервиса и процесс голосования представлены на рисунке 8. Каждый клиент может получить адрес страницы сервиса из свойства `hasClientUrl` индивида класса `SocialProgramService`. Участники голосуют за некоторые точки интереса, и соответствующие свойства публикуются в ИП. Сервис веб-страниц публикует уведомление для сервиса культурной программы, который пересчитывает голоса и формирует новый объект JSON. Страница на стороне клиента обновляется в фоновом режиме с использованием технологии AJAX-вызовов.

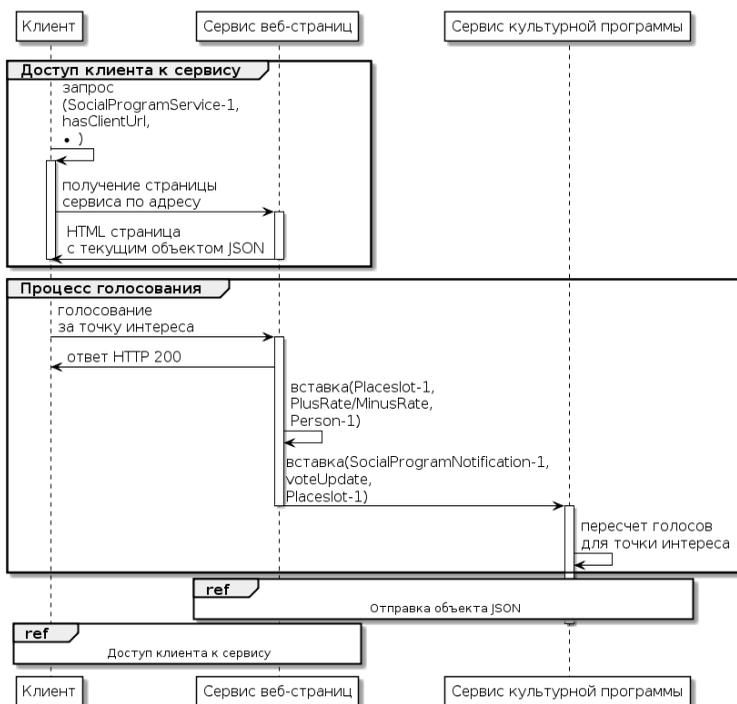


Рис. 8. Последовательность для доступа клиентов и процесса голосования

Процесс доступа к страницам сервиса для сервисов общественных экранов повестки мероприятия и презентации представлен на рисунке 9. Когда организаторам необходимо показать культурную программу, они переключают режимы экранов на сервисах повестки мероприятия и презентации. Процесс доступа аналогичен случаю с мобильными клиентами участников, используя свойства `hasAgendaServiceUrl` для сервиса повестки мероприятия и `hasPresentationServiceUrl` для сервиса презентации.

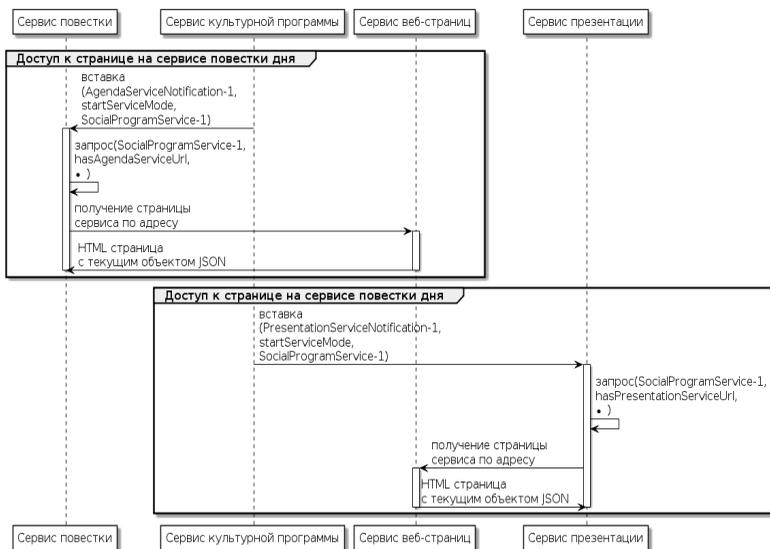


Рис. 9. Последовательность для доступа сервисов повестки мероприятия и презентации

5. Экспериментальная реализация. Исходный код программной реализации разработанных сервисов е-туризма для расширения системы SmartRoom является открытым и распространяется вместе с основной системой (см. <https://sourceforge.net/projects/smartroom/>). Таблица 3 показывает технологии, использованные при реализации.

Таблица 3. Используемые технологии

Технология	Сервис	Описание
Visual C# и .Net Framework	Культурной программы, поиска	Язык программирования
Python	Веб-страниц	
Windows Presentation Foundation (WPF)	Культурной программы	Графический интерфейс пользователя для организаторов
SmartSlog SDK	Культурной программы, поиска	
Python KPI	Веб-страниц	Взаимодействие с пространством ИЗ
World Around Me search engine	Поиска	
CherryPy	Веб-страниц	Поиск медиа-информации, включая фото и описание точки интереса
Мако	Веб-страниц	
HTML, JavaScript, CSS, jQuery	Веб-страниц	Веб-каркас для обработки REST-запросов и распространения веб-страниц с содержимым сервисов
	Веб-страниц	Шаблонизатор для построения веб-страниц
	Веб-страниц	Программные инструменты веб-страниц

На рисунке 10 представлены примеры экрана сервиса культурной программы.

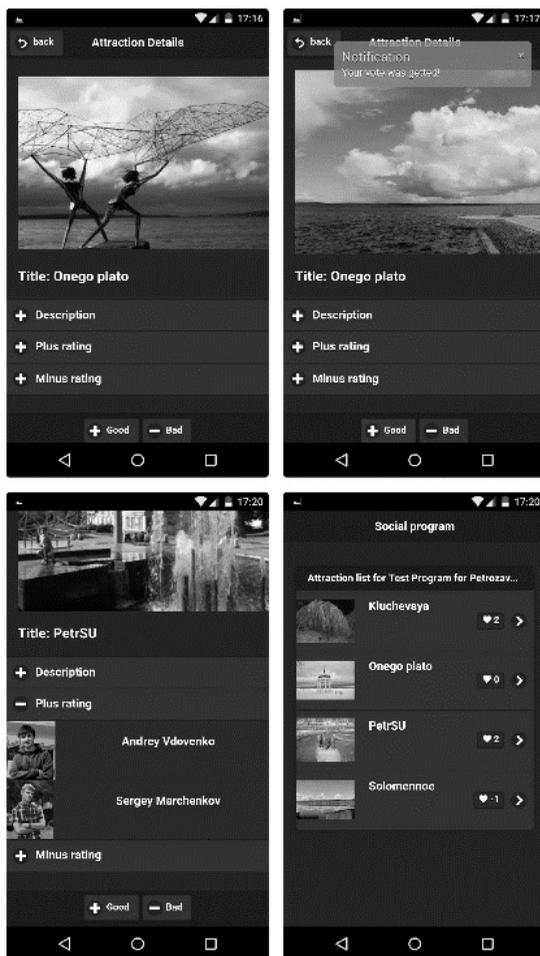
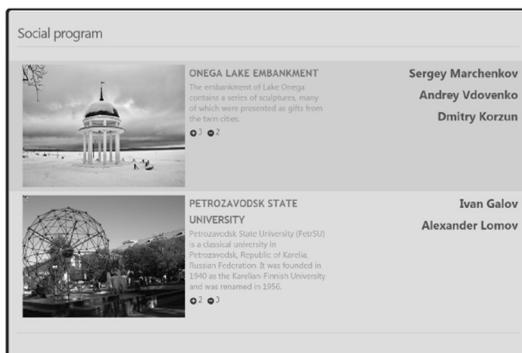


Рис. 10. Примеры снимков экрана для сервиса культурной программы на персональном мобильном устройстве

Первые три снимка демонстрируют представление сервиса на персональном мобильном устройстве участника: 1) страница информации о точке интереса, включающая слайд-шоу из найденных фотографий; 2) страница с примером уведомления о процессе голосования за данную точку интереса; 3) пример страницы с результатами голосо-

вания участников за точку интереса. Последний снимок экрана представляет список точек интереса и результатов голосования за каждый из них. На рисунке 11 показаны соответствующие снимки экранов для сервисов повестки мероприятия и презентации.



a)



b)

Рис. 11. Примеры снимков общественных экранов при составлении культурной программы: а) экран сервиса повестки мероприятия; б) экран сервиса презентации

Сервис поиска использует язык программирования Visual C# и библиотеку .Net Framework. Механизм сериализации использован для обработки ответов веб-сервисов. Сервис поиска использует библиотеку

ку поиска из приложения WorldAroundMe [32]. Она состоит из ядра и модулей драйверов, позволяющих расширять набор доступных веб-сервисов и функций. Модуль ядра перенаправляет поисковые запросы от сервисов электронного туризма соответствующим драйверам. Драйвер, в свою очередь, обрабатывает запрос сервисов и добавляет данные к точкам интереса (напр., ссылки на изображения).

Сервис веб-страниц использует язык программирования Python. Используется библиотека CherryPy для построения веб-каркаса и шаблонизатор Мако для построения веб-страниц. Основным модулем сервиса веб-страниц является диспетчер запросов, который используется для перенаправления запросов соответствующим модулям и вызова действий на основе заданного адреса. Вторым является модуль процессора знаний, использующий библиотеку Python KPI. Модуль процессора знаний отвечает за операции взаимодействия с пространством ИЗ. Третий модуль отвечает за работу с шаблонами, написанными с использованием языков разметки HTML и CSS, а также языка программирования JavaScript. Дополнительно используются библиотеки jQuery и jQuery Mobile для упрощения кода. Модуль шаблонов строит веб-страницы с информацией для пользователей и предоставляет им возможность взаимодействовать с сервисами.

Для изучения производительности реализованных сервисов были проведены экспериментальные исследования. Для экспериментов использовались два персональных компьютера (Linux/Windows) и пять смартфонов (Windows Phone и Android). Каждый компьютер оборудован двухъядерным процессором с тактовой частотой 2,4 Гц, 4 Гб оперативной памяти, жестким диском объемом 2,5 Тб со стандартной скоростью чтения и записи. Тестовая культурная программа состояла из 10 различных точек интереса. Данное количество точек выбрано из расчета среднего количества мест, которые заявляются на конференциях как наиболее популярные. Измерялось время выполнения для следующих операций: поиск фотографий для каждого места, отправка объекта JSON сервисом культурной программы сервису веб-страниц, передача обновленных голосов сервисом веб-страниц сервису культурной программы.

Для каждой точки интереса производился поиск 20 фотографий. Измеренное время составило около 10 сек. Данное время допустимо, так как запрос фотографий выполняется заранее, до начала использования участниками. Основную часть поиска занимает взаимодействие с внешними веб-сервисами (сервис Flickr), так как публикация индивида запроса и результата поиска в ИП предполагает передачу небольшого количество информации в виде индивидов и свойств, и зави-

сит только от пропускной способности сети между сервисами и брокером SIB. Каждый запрос к веб-сервисам выполняется последовательно. Скорость передачи данных при организации веб-взаимодействий сервисов является допустимой для их практического применения.

Операция отправки сервисом культурной программы объекта JSON сервису веб-страниц включает в себя выполнение нескольких шагов. На стороне сервиса культурной программы происходит получение локальной структуры списка точек интереса, затем вся структура преобразовывается в предопределенный формат JSON, и после происходит отправка полученного объекта на сервис веб-страниц по протоколу HTTP. Полученное время для данной операции составило в среднем около 0,5 сек для списка из 10 точек интереса. Наиболее узким среди всех шагов является преобразование локальной структуры к формату JSON, но, используя оптимизированные библиотеки, получается решить эту проблему, что отражено в результатах эксперимента.

Передача обновленных голосов сервисом веб-страниц сервису культурной программы заключается в работе с интеллектуальными пространствами. Когда участник голосует, сервис веб-страниц производит публикацию значений пользователя в интеллектуальное пространство и после публикует уведомление для сервиса культурной программы, указывающее на обновление информации о голосах для точки интереса. В свою очередь, сервис культурной программы получает уведомление и производит обновление локальной информации о результатах голосования. Измеренное время выполнения операции составило в среднем приблизительно 0,6 сек. Данный показатель считается подходящим для рассматриваемой области применения.

6. Заключение. В статье выполнено исследование по разработке цифровых окружений совместной деятельности людей на основе подхода интеллектуальных пространств с целью расширения таких окружений информационными сервисами e-туризма. В качестве референтного класса рассмотрены такие окружения как интеллектуальный зал, а в качестве конкретного примера используется система SmartRoom, реализованная на платформе Smart-M3. Для расширения этой системы предложен сценарий построения культурной программы участниками мероприятия. Предложены необходимые решения проектирования для реализации этого сценария на основе набора дополнительных сервисов в составе: сервис культурной программы, сервис поиска и сервис веб-страниц. Построение и доставка этих сервисов выполняется в рамках взаимодействия программных агентов в интеллектуальном пространстве системы SmartRoom, создаваемом в оборудованном помещении. Проведено экспериментальное исследование с

реализованным программным прототипом. Так, при задании для культурной программы 10 возможных точек интереса с 20 фотографиями каждая, операция поиска занимает до 10 с, а операция обновления данных сервиса — около 0,5 с, что является достаточным при практическом использовании для исследуемого класса цифровых окружений совместной деятельности людей.

Литература

1. Statistics and facts on the global tourism industry. URL: <http://www.statista.com/topics/962/global-tourism> (дата обращения: 18.03.2016).
2. *Sebastia L. et al.* e-Tourism: a tourist recommendation and planning application // International Journal on Artificial Intelligence Tools. 2009. vol. 55. no. 05. pp. 717–738.
3. World Internet users and 2015 population stats. URL: <http://www.internetworldstats.com/stats.htm> (дата обращения: 18.03.2016).
4. *Smirnov A. et al.* Intelligent mobile tourist guide // Internet of Things, Smart Spaces, and Next Generation Networking. 2013. pp. 94–106.
5. *Borràs J., Moreno A., Valls A.* Intelligent tourism recommender systems: A survey // Expert Systems with Applications. 2014. vol. 41. no. 16. pp. 7370–7389.
6. *Muntean M.I.* Some Collaborative Systems Approaches in Knowledge-Based Environments // New Research on Knowledge Management Models and Methods. 2012. pp. 379–394.
7. *Ронжин А.Л., Карпов А.А.* Сравнительный анализ функциональности прототипов интеллектуальных пространств // Труды СПИИРАН. 2013. Т. 1. №. 24. С. 277–290.
8. *Кашевник А.М. и др.* Интеллектуальная система автоматизированного проведения конференций // Труды СПИИРАН. 2010. Т. 3. №. 14. С. 228–245.
9. *Balandin S., Waris H.* Key properties in the development of smart spaces // Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments. 2009. pp. 3–12.
10. *Ovaska E., Cinotti T.S., Toninelli A.* The design principles and practices of interoperable smart spaces // Advanced Design Approaches to Emerging Software Systems: Principles, Methodology and Tools. 2012. pp. 18–47.
11. *Korzun D., Galov I., Kashevnik A., Balandin S.* Virtual shared workspace for smart spaces and M3-based case study // Proceedings of 15th Conference Open Innovations Association FRUCT. 2014. pp. 60–68.
12. *Vdovenko A., Marchenkov S., Korzun D.* Mobile multi-service smart room client: Initial study for multi-platform development // Proceedings of 13th Conference of Open Innovations Association FRUCT and 2nd Seminar on e-Tourism for Karelia and Oulu Region. 2013. pp. 143–152.
13. *Kulakov K., Shabaev A.* An approach to creation of smart space-based trip planning service // Proceedings of 16th Conference Open Innovations Association FRUCT. 2014. pp. 38–44.
14. *Varfolomeyev A., Korzun D., Ivanovs A., Petrina O.* Smart personal assistant for historical tourism // Proceedings of 2nd Int'l Conf. on Environment, Energy, Ecosystems and Development (EEEAD'2014). 2014. pp. 9–15.
15. *Vdovenko A.S., Marchenkov S.A., Korzun D.G.* Enhancing the smartroom system with e-tourism services // Proceedings of 17th Conference Open Innovations Association FRUCT. 2015. pp. 237–246.
16. *Городецкий В.И.* Агенты и извлечение знаний из данных в интеллектуальном пространстве // Автоматизация управления и интеллектуальные системы и

- среды: Сб. научн. тр. Первой Международной конференции. 2010. С. 24–36.
17. *Korzun D.Zh.* Формализм сервисов и архитектурные абстракции для программных приложений интеллектуальных пространств // Программная инженерия. 2015. С. 3–12.
 18. *Kiljander J. et al.* Enabling Semantic Technology Empowered Smart Spaces // Journal of Computer Networks and Communications. 2012. vol. 2012. pp. 1–14.
 19. *Honkola J. et al.* Smart-M3 information sharing platform // The IEEE symposium on Computers and Communications. 2010. pp. 1041–1046.
 20. *Korzun D.G. et al.* Integration of Smart-M3 applications: Blogging in smart conference // Smart Spaces and Next Generation Wired/Wireless Networking. 2011. pp. 51–62.
 21. *Marchenkov S., Korzun D.* User presence detection based on tracking network activity in smartroom // Proceedings of 16th Conference Open Innovations Association FRUCT. 2014. pp. 45–50.
 22. *Wache H. et al.* Ontology-based integration of information – a survey of existing approaches // IJCAI-01 workshop: ontologies and information sharing. 2001. vol. 2001. pp. 108–117.
 23. *Александров В.В., Андреева Н.А., Кулешов С.В.* Методы построения информационно-логистических систем // СПб.: Изд-во Политехнического университета. 2005. 109 с.
 24. *Nguyen T. et al.* On the practicalities of place-based virtual communities: Ontology-based querying, application architecture, and performance // Expert Systems with Applications. 2014. vol. 41. no. 6. pp. 2859–2873.
 25. *Skillen K. L. u et al.* Ontological user modelling and semantic rule-based reasoning for personalisation of Help-On-Demand services in pervasive environments // Future Generation Computer Systems. 2014. vol. 34. pp. 97–109.
 26. *Martin-Vicente M.I. et al.* A semantic approach to improve neighborhood formation in collaborative recommender systems // Expert Systems with Applications. 2014. vol. 41. no. 17. pp. 7776–7788.
 27. *Evers C. et al.* The user in the loop: Enabling user participation for self-adaptive applications // Future Generation Computer Systems. 2014. vol. 34. pp. 110–123.
 28. *Galov I.V., Korzun D.G.* A Notification Model for Smart-M3 Applications // Internet of Things, Smart Spaces, and Next Generation Networks and Systems. 2014. pp. 121–132.
 29. *Bernini D. et al.* Architectural abstractions for spaces-based communication in Smart Environments // Journal of Ambient Intelligence and Smart Environments. 2012. vol. 4. no. 3. pp. 253–277.
 30. *Riva C., Laitkorpi M.* Designing web-based mobile services with REST // Service-Oriented Computing – ICSOC 2007 Workshops. 2009. pp. 439–450.
 31. *Ivan C., Popa R.* Cloud based Cross Platform Mobile Applications Building and integrating cloud services with mobile client applications // Advances in Computer Science: an International Journal. 2014. vol. 3. no. 2. pp. 69–77.
 32. *Вдовенко А.С., Корзун Д.Ж.* Программный клиент World Around Me для мобильных устройств ОС Windows Phone // Всероссийский конкурс научно-исследовательских работ студентов и аспирантов в области информатики и информационных технологий. 2012. С. 555–559.

Марченков Сергей Александрович — аспирант кафедры информатики и математического обеспечения, институт математики и информационных технологий, Петрозаводский государственный университет (ПетрГУ), младший научный сотрудник кафедры информатики и математического обеспечения института математики и информационных технологий, Петрозаводский государственный университет (ПетрГУ). Область научных интересов: интеллектуальные пространства, Интернет вещей, семантический веб, онто-

логическое моделирование. Число научных публикаций — 16. marchenk@cs.karelia.ru; пр. Ленина, 33, Петрозаводск, 185910; р.т.: +7(8142)711015.

Вдовенко Андрей Сергеевич — аспирант кафедры информатики и математического обеспечения, институт математики и информационных технологий, Петрозаводский государственный университет (ПетрГУ), младший научный сотрудник кафедры информатики и математического обеспечения института математики и информационных технологий, Петрозаводский государственный университет (ПетрГУ). Область научных интересов: интеллектуальные пространства, интеллектуальный зал, онтологические модели. Число научных публикаций — 23. vdovenko@cs.petrstu.ru, <http://www.cs.petrstu.ru>; пр. Ленина, 33, Петрозаводск, 185910; р.т.: +7(8142)711015.

Корзун Дмитрий Жоржевич — доцент, ведущий научный сотрудник кафедры информатики и математического обеспечения института математики и информационных технологий, Петрозаводский государственный университет (ПетрГУ). Область научных интересов: анализ распределенных систем, дискретное моделирование, повсеместные вычисления и интеллектуальные пространства, Интернет вещей, технологии разработки ПО, проектирование алгоритмов и вычислительная сложность, линейный диофантов анализ и его приложения, теория формальных языков и методы трансляции. Число научных публикаций — 152. dkorzun@cs.karelia.ru; пр. Ленина, 33, Петрозаводск, 185910; р.т.: +79095689571, Факс: +7(8142)711000.

Поддержка исследований. Статья подготовлена при финансовой поддержке отделения гуманитарных и общественных наук РФФИ, проект № 16-01-12033. Исследование поддержано Минобрнауки России, инициативный проект № 2.5124.2017 в рамках базовой части государственного задания на 2017-2019 гг.

S.A.MARCHENKOV, A.S.VDOVENKO, D.G.KORZUN
**ENHANCING THE OPPORTUNITIES OF COLLABORATIVE
 WORK IN AN INTELLIGENT ROOM USING E-TOURISM
 SERVICES**

Marchenkov S.A., Vdovenko A.S., Korzun D.G. Enhancing the Opportunities of Collaborative Work in an Intelligent Room using e-Tourism Services.

Abstract. An intelligent room system creates a digital service-oriented environment for collaborative work. This paper studies the problem of enhancing such a system with services of electronic tourism (e-Tourism). We define a scenario to enhance the opportunities of collaborative work through involving the participants in the making of a social program to explore the region where the collaborative activity takes place. As a specific software application, we consider the SmartRoom system, which supports collaborative work in a room during such activities as conferences and meetings. Using the smart spaces approach, we develop a set of services to implement the scenario of social program creation. We propose an ontological model and interaction mechanisms for service construction and delivery based on shared information collection and processing by multiple software agents in the smart space created by the SmartRoom system. Our experimental performance study illustrates the proposed service enhancement.

Keywords: smart spaces, information service, e-Tourism, intelligent room, Smart-M3, ontological model.

Marchenkov Sergey Aleksandrovich — Ph.D. student of computer science department of mathematics and information technology institute , Petrozavodsk State University (PetrSU), junior researcher of computer science department of mathematics and information technology institute , Petrozavodsk State University (PetrSU). Research interests: smart spaces, Internet of Things, semantic web, ontological modeling. The number of publications — 16. marchenk@cs.karelia.ru; 33, Lenin Str., Petrozavodsk, 185910, Russia; office phone: +7(8142)711015.

Vdovenko Andrey Sergeevich — Ph.D. student of computer science department of mathematics and information technology institute , Petrozavodsk State University (PetrSU), junior researcher of computer science department of mathematics and information technology institute , Petrozavodsk State University (PetrSU). Research interests: smart spaces, intelligent room, ontological model. The number of publications — 23. vdovenko@cs.petsu.ru, http://www.cs.petsu.ru; 33, Lenin Str., Petrozavodsk, 185910, Russia; office phone: +7(8142)711015.

Korzun Dmitry Georzhevich — Ph.D., associate professor, leading researcher of computer science department of mathematics and information technology institute , Petrozavodsk State University (PetrSU). Research interests: analysis and evaluation of distributed systems, discrete modeling, ubiquitous computing in smart spaces, Internet of Things, software engineering, algorithm design and complexity, linear Diophantine analysis and its applications, theory of formal languages and parsing. The number of publications — 152. dkorzun@cs.karelia.ru; 33, Lenin Str., Petrozavodsk, 185910, Russia; office phone: +79095689571, Fax: +7(8142)711000.

Acknowledgements. The article was published with financial support from Department for Humanities of Russian Fund for Basic Research according to project № 16-01-12033. This

research is supported by the Ministry of Education and Science of Russia within project no. 2.5124.2017 of the basic part of state research assignment for 2017–2019.

References

1. Statistics and facts on the global tourism industry. Available at: <http://www.statista.com/topics/962/global-tourism> (accessed 18.03.2016).
2. World Internet users and 2015 population stats. Available at: <http://www.internetworldstats.com/stats.htm> (accessed 18.03.2016).
3. Sebastia L. et al. e-Tourism: a tourist recommendation and planning application. *International Journal on Artificial Intelligence Tools*. 2009. vol. 55. no. 05. pp. 717–738.
4. Smirnov A. et al. Intelligent mobile tourist guide. Internet of Things, Smart Spaces, and Next Generation Networking. 2013. pp. 94–106.
5. Borràs J., Moreno A., Valls A. Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*. 2014. vol. 41. no. 16. pp. 7370–7389.
6. Muntean M.I. Some Collaborative Systems Approaches in Knowledge-Based Environments. *New Research on Knowledge Management Models and Methods*. 2012. pp. 379–394.
7. Ronzhin A.L., Karpov A.A. [A comparative analysis of smart space prototypes functionality]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2013. vol. 1. no. 24. pp. 277–290. (In Russ.).
8. Kashevnik A. M. et al. [Intelligent system for conference management automation]. *Trudy SPIIRAN– SPIIRAS Proceedings*. 2010. vol. 3. no. 14. pp. 228–245. (In Russ.).
9. Balandin S., Waris H. Key properties in the development of smart spaces. *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*. 2009. pp. 3–12.
10. Ovaska E., Cinotti T.S., Toninelli A. The design principles and practices of interoperable smart spaces. *Advanced Design Approaches to Emerging Software Systems: Principles, Methodology and Tools*. 2012. pp. 18–47.
11. Korzun D., Galov I., Kashevnik A., Balandin S. Virtual shared workspace for smart spaces and M3-based case study. *Proceedings of 15th Conference Open Innovations Association FRUCT*. 2014. pp. 60–68.
12. Vdovenko A., Marchenkov S., Korzun D. Mobile multi-service smart room client: Initial study for multi-platform development. *Proceedings of 13th Conference of Open Innovations Association FRUCT and 2nd Seminar on e-Tourism for Karelia and Oulu Region*. 2013. pp. 143–152.
13. Kulakov K., Shabaev A. An approach to creation of smart space-based trip planning service. *Proceedings of 16th Conference Open Innovations Association FRUCT*. 2014. pp. 38–44.
14. Varfolomeyev A., Korzun D., Ivanovs A., Petrina O. Smart personal assistant for historical tourism. *Proceedings of 2nd Int'l Conf. on Environment, Energy, Ecosystems and Development (EEEAD'2014)*. 2014. pp. 9–15.
15. Vdovenko A.S., Marchenkov S.A., Korzun D.G. Enhancing the smartroom system with e-tourism services. *Proceedings of 17th Conference Open Innovations Association FRUCT*. 2015. pp. 237–246.
16. Gorodetskii V. I. [Agents and knowledge extraction from data in the smart space] *Avtomatizacija upravlenija i intellektual'nye sistemy i sredy: Sb. nauchn. tr. Pervoj Mezhdunarodnoj konferencii. [Automation of management and intelligent systems and environments: Proceedings of First International Conference]*. 2010. pp. 24–36. (In Russ.).
17. Korzun D.G. [Formalism of services and architectural abstraction for software applications of smart spaces]. *Programmnyaya inzheneriya. – Software engineering*. 2015. pp. 3–12. (In Russ.).

18. Kiljander J. et al. Enabling Semantic Technology Empowered Smart Spaces. *Journal of Computer Networks and Communications*. 2012. vol. 2012. pp. 1–14.
19. Honkola J. et al. Smart-M3 information sharing platform. The IEEE symposium on Computers and Communications. 2010. pp. 1041–1046.
20. Korzun D.G. et al. Integration of Smart-M3 applications: Blogging in smart conference. Smart Spaces and Next Generation Wired/Wireless Networking. 2011. pp. 51–62.
21. Marchenkov S., Korzun D. User presence detection based on tracking network activity in smartroom. Proceedings of 16th Conference Open Innovations Association FRUCT. 2014. pp. 45–50.
22. Wache H. et al. Ontology-based integration of information – a survey of existing approaches. IJCAI-01 workshop: ontologies and information sharing. 2001. vol. 2001. pp. 108–117.
23. Aleksandrov V.V., Andreeva N.A., Kuleshov S.V. *Metody postroenija informacionno-logisticheskikh sistem* [Methods of construction information and logistics systems]. SPb.: Izd-vo Politehnicheskogo universiteta, 2005. 109 p. (In Russ.).
24. Nguyen T. et al. On the practicalities of place-based virtual communities: Ontology-based querying, application architecture, and performance. *Expert Systems with Applications*. 2014. vol. 41. no. 6. pp. 2859–2873.
25. Skillen K. L. et al. Ontological user modelling and semantic rule-based reasoning for personalisation of Help-On-Demand services in pervasive environments. *Future Generation Computer Systems*. 2014. vol. 34. pp. 97–109.
26. Martín-Vicente M.I. et al. A semantic approach to improve neighborhood formation in collaborative recommender systems. *Expert Systems with Applications*. 2014. vol. 41. no. 17. pp. 7776–7788.
27. Evers C. et al. The user in the loop: Enabling user participation for self-adaptive applications. *Future Generation Computer Systems*. 2014. vol. 34. pp. 110–123.
28. Galov I.V., Korzun D.G. A Notification Model for Smart-M3 Applications. Internet of Things, Smart Spaces, and Next Generation Networks and Systems. 2014. pp. 121–132.
29. Bernini D. et al. Architectural abstractions for spaces-based communication in Smart Environments. *Journal of Ambient Intelligence and Smart Environments*. 2012. vol. 4. no. 3. pp. 253–277.
30. Riva C., Laitkorpi M. Designing web-based mobile services with REST. Service-Oriented Computing – ICSOC 2007 Workshops. 2009. pp. 439–450.
31. Ivan C., Popa R. Cloud based Cross Platform Mobile Applications Building and integrating cloud services with mobile client applications. *Advances in Computer Science: an International Journal*. 2014. vol. 3. no. 2. pp. 69–77.
32. Vdovenko A.S., Korzun D.G. [The software client World Around Me for mobile devices OS Windows Phone]. *Vserossijskij konkurs nauchno-issledovatel'skikh rabot studentov i aspirantov v oblasti informatiki i informacionnyh tehnologij* [All-Russian contest of research works of students and graduate students in computer science and information technologies]. 2012. pp. 555–559. (In Russ.).

Ю.А. КОТОВ
**АППРОКСИМАЦИЯ РАСПРЕДЕЛЕНИЙ ЧАСТОТ БУКВЕННЫХ
БИГРАММ ТЕКСТА ДЛЯ ИДЕНТИФИКАЦИИ БУКВ**

Котов Ю.А. Аппроксимация распределений частот буквенных биграмм текста для идентификации букв.

Аннотация. В статье рассмотрены особенности применения методов частотного упорядочивания и аппроксимации для решения задачи идентификации знаков текста. Определены условия реализации метода Якобсена для получения наименьшей погрешности идентификации. Предложен метод аппроксимации одномерных и двумерных распределений частот знаковых биграмм текста и буквенных биграмм эталона языка текста. Приведены экспериментальные данные о погрешностях метода Якобсена и предложенного метода аппроксимации для русскоязычных текстов.

Погрешность предложенного метода меньше, чем у метода Якобсена. Метод может быть использован для идентификации знаков текста любого языка, для которого существует эталонное распределение частот буквенных биграмм.

Ключевые слова: аппроксимация, идентификация, буква, биграмма, простая замена, шифр.

1. Введение. Алфавит любого языка представляет собой множество упорядоченных кодовых знаков, обозначающих буквы этого языка. Буквы языка однозначно связаны с их порядковым номером в исходном алфавите, но могут быть представлены разными — в том числе неизвестными — знаками, и в то же время одинаковые знаки в разных текстах могут обозначать одну и ту же, но, возможно, неизвестную букву. Будем считать, что в каждом конкретном тексте для обозначения буквы языка используется только один знак. Тогда идентифицировать знак в произвольном тексте на некотором языке означает приписать ему такие числовые характеристики, получаемые из данного текста, которые позволяют определить номер знака в исходном алфавите этого языка и соответственно букву, которую данный знак представляет [1].

Сложность решения данной задачи заключается в том, что каждый текст имеет собственное упорядочение множества используемых в нем знаков. Математически это означает, что каждый текст определен в своей собственной системе координат. Для формального анализа дискретных множеств, которые представляют тексты и их элементы, необходимо преобразование индивидуальных систем координат каждого текста к общей системе. В большинстве случаев такое преобразование осуществляется отождествлением множеств используемых в тексте знаков, «букв», с некоторым известным алфавитом. Если же отождествление невозможно, тексты признаются несравнимыми, хотя, возможно, отличаются только множествами используемых знаков.

Традиционно задача идентификации знаков таких текстов трактуется как криптографическая, связанная с шифром простой замены [2]. Для решения данной задачи могут применяться методы перебора, частотного упорядочивания [3–7], генетические алгоритмы [9–11], марковские цепи [12–13], методы, основанные на вычислении энтропии [2, 14–15] и статистической проверке гипотез [16], а также другие методы и их комбинации, например, [1, 10, 13, 17]. В основе большинства этих методов лежат методы частотного упорядочивания, представляющие частный случай методов аппроксимации.

2. Методы упорядочивания и аппроксимации в задаче идентификации знаков. Идея использования метода частотного упорядочивания для решения задачи идентификации знаков основана на следующих математических принципах [3–6]. Пусть есть множество пар знаков эталона и частоты появления этих знаков $\mathbf{E}=\{<e_i, h_i>\}$ и аналогичное множество $\mathbf{T}=\{<z_j, k_j>\}$ для знаков анализируемого текста T . Соответствие между знаками \mathbf{E} и \mathbf{T} заранее неизвестно. Так как знаки имеют нечисловую природу, построить функции для множеств \mathbf{E} и \mathbf{T} можно только на основании значений частот знаков, то есть значений из области изменения функций.

Для решения этой задачи предположим, что знаки множеств \mathbf{E} и \mathbf{T} подчиняются одинаковым функциональным закономерностям и в любой системе координат имеют одинаковый знак приращения функции в каждой точке области определения, а в области изменения функции нет совпадающих значений. Тогда в системе координат, в которой значения частот знаков эталона \mathbf{E} упорядочены (по убыванию или возрастанию), частоты знаков \mathbf{T} будут упорядочены таким же образом. Построим в этой системе функции $y=f_E(x)$ и $y=f_T(x)$, применив одинаковый способ упорядочивания к элементам множеств \mathbf{E} и \mathbf{T} и приписав в результате знакам из \mathbf{E} и \mathbf{T} значения из множества $\mathbf{N}=\{0,1,2,\dots,n-1\}$, $x \in \mathbf{N}$, где n — мощность множеств \mathbf{E} и \mathbf{T} .

Суммарная разность (1):

$$W(\mathbf{T}) = \sum_{i=0}^{n-1} |f_E(i) - f_T(i)|, \quad (1)$$

значений $f_E(x)$ и $f_T(x)$ будет минимальной по построению. Предполагается, что $W(T)$ стремится к нулю при росте объема анализируемого текста, то есть в пределе функции $f_T(x)$ и $f_E(x)$ совпадают. По результатам одинакового упорядочивания знаки анализируемого текста z_i отождествляются с соответствующими знаками эталона e_i , $z_i \equiv e_i$.

Погрешность такого решения определяется количеством знаков из \mathbf{T} , правильно отождествленных со знаками эталона \mathbf{E} , и зависит от соответствия эталона и анализируемого текста исходным посылкам метода, способа формирования эталона и его шкалы измерения, а также от объема анализируемого текста.

Значения эталона зависят не только от общего объема текстов, из которого он сформирован, но также от количества и разнообразия этих текстов, то есть от того, насколько в эталоне учитывается влияние стилистики тематической области текстов и их авторов. Очевидно следующее замечание.

Замечание 1. Наименьшую погрешность методов идентификации знаков текста, использующих сравнение с эталоном, можно ожидать в случаях, когда анализируемый текст является фрагментом текста, из которого сформирован эталон.

Таким образом, наиболее объективную информацию о методах идентификации знаков текста, использующих эталон, можно получить, когда анализируемый текст и эталон сформированы независимо друг от друга и представляют разные тексты. Замечание 1 необходимо, потому что в ряде работ, посвященных подобным методам, в частности [9-13], результаты работы методов представляются на фрагментах текста и эталонах, полученных из одного текста.

Группа методов частотного упорядочивания тесно связана с методами аппроксимации, а точнее — *поиска наилучшей* аппроксимации, применяемыми для решения задачи идентификации знаков текста. Эта связь определяется функцией $W(\mathbf{T})$. В терминах методов аппроксимации данная функция может рассматриваться как погрешность аппроксимации функции $f_E(x)$ функцией $f_T(x)$. Так как функция $W(\mathbf{T})$ — экстремальная функция, то ее можно взять в качестве целевой функции и осуществить поиск минимального значения $W(\mathbf{T})$, то есть решить оптимизационную задачу.

Тогда, казалось бы, можно отказаться от упорядочивания вообще, и сформулировать задачу идентификации знаков текста как задачу поиска наилучшей в смысле минимума функции $W(\mathbf{T})$ аппроксимации произвольной эталонной функции $f(E)$.

Однако справедливо не только то, что для одномерного случая подобная сложность решения избыточна, но и следующее замечание.

Замечание 2. Минимальное значение $W(\mathbf{T})$ не является единственным для различных упорядочений \mathbf{T} .

Это легко показать для монотонных функций $f_E(x)$ и $f_T(x)$.

Отсюда следует, что в общем случае погрешность решения задачи идентификации знаков текста на основе метода аппроксимации будет больше при использовании неупорядоченных множеств \mathbf{E} и \mathbf{T} .

Справедливо также и то, что любой метод частотного упорядочивания является частным случаем метода аппроксимации.

3. Метод Якобсена. Непосредственно для решения задачи идентификации знаков текста (в формулировке раскрытия шифра простой замены) известен метод Якобсена [7]. Метод основан на минимизации значения целевой функции:

$$W(\mathbf{T}) = \sum_{ij} |t_{ij} - b_{ij}|, \quad (2)$$

где t_{ij} , b_{ij} — значения частот знаковых (буквенных) биграмм анализируемого и эталонного текстов.

В одномерном случае по формуле (1) $W(\mathbf{T})$ может быть интерпретирована как площадь, ограниченная кривыми, аппроксимирующими значения частот множеств \mathbf{E} и \mathbf{T} , а поиск наилучшей аппроксимации — как поиск наилучшего совпадения форм этих кривых. В отличие от этого в двумерном случае, описываемом формулой (2), ведется поиск наилучшей аппроксимации поверхности $z_B = f_B(x, y)$, аппроксимирующей множество частот биграмм эталона \mathbf{B} , поверхностью $z_T = f_T(x, y)$, аппроксимирующей множество частот биграмм \mathbf{T} анализируемого текста. В этом случае $W(\mathbf{T})$ можно интерпретировать как объем, заключенный между соответствующими поверхностями.

Множества \mathbf{B} и \mathbf{T} образованы тройками $\mathbf{B} = \{ \langle \langle e_i, e_j \rangle, b_{ij} \rangle \}$, $\mathbf{T} = \{ \langle \langle w_i, w_j \rangle, t_{ij} \rangle \}$, где e_i , w_i — знаки, b_{ij} , t_{ij} — частоты биграмм эталона и анализируемого текста соответственно. Они могут быть представлены симметрично упорядоченными по знакам таблицами (матрицами), по строке которых идентифицируется первый знак биграммы, по столбцу — второй. Значениями таблицы являются значения соответствующих частот b_{ij} или t_{ij} .

Поиск минимального значения $W(\mathbf{T})$ в методе Якобсена осуществляется следующим образом [7]. По таблице биграмм \mathbf{T} последовательно просматриваются пары знаков — сначала соседние, затем — отстоящие на один знак и так до тех пор, пока не будет рассмотрена пара, состоящая из первого и последнего знака таблицы. Назовем просмотр одинаково отстоящих пар *уровнем* просмотра, индекс пары на уровне — ее *местом*. Для каждого уровня просмотр ведется с начала таблицы.

При каждом просмотре пары проверяется необходимость ее перестановки в смысле минимизации значения $W(\mathbf{T})$. Для этого по формуле (2) вычисляется $W_0(kl)$ для исходной пары (k — уровень пары,

l — место), и $W_1(kl)$ — для условно переставленной. Если выполняется условие (3):

$$W_0(kl) > W_1(kl), \quad (3)$$

то пара переставляется, и просмотр начинается с начала с нулевого уровня.

Так продолжается до тех пор, пока при очередном просмотре пар с нулевого уровня не будет найдено ни одной перестановки, включая последнюю пару.

Метод Якобсена представляет собой обыкновенную сортировку с переменным шагом, адаптированную к решению задачи идентификации знаков текста на основе двумерного распределения частот знаковых биграмм и функции (3). Этот метод является методом двумерного упорядочивания, аналогичным одномерному методу упорядочивания частот знаков. Основным его недостатком, как и других методов упорядочивания, является возможность остановки в локальном минимуме целевой функции [8].

Метод Якобсена является классическим и эффективным представителем группы методов аппроксимации, применяемых для решения задачи идентификации знаков текста. Представляется очевидным его применение в качестве эталонного метода для сравнения с другими методами этой группы.

В работе [7] проверка метода была проведена для англоязычного текста «Alice in Wonderland» с использованием эталона биграмм, сформированного по тексту «Moby Dick». В данной работе была проведена экспериментальная проверка метода Якобсена на представительных выборках текстовых фрагментов русскоязычных текстов (раздел 5) с использованием независимого эталона биграмм из [16]. По результатам эксперимента можно сделать следующие замечания.

Метод Якобсена имеет низкую погрешность идентификации знаков, но значения ее существенно выше, чем это заявлено автором — например, не 2% при объеме текста в 1000 знаков [7], а 22% — 38% (таблицы 1 и 2, рисунки 1 и 2).

При этом наименьшую погрешность удается достичь только при замене условия (3) условием (4):

$$W_0(kl) - W_1(kl) > \varepsilon, \quad \varepsilon = 10^{-8}, \quad (4)$$

то есть только при учете погрешности вычислений. При других значениях ε , в том числе равном нулю, как в условии (3), погрешность метода значительно возрастает, иногда в несколько раз.

И наконец погрешность метода также возрастает, если матрицы **В** и **Т** не упорядочены по частоте появления знаков эталона и анализируемого текста соответственно (объяснение см. в замечании 2 раздела 2) (необходимость начальной упорядоченности указана в [7]).

Данные о погрешности метода Якобсена, реализованного с учетом указанных замечаний, применительно к русскоязычным текстам приведены в таблицах 1 и 2 и на рисунках 1 и 2 раздела 5.

4. Описание метода аппроксимации. При использовании для идентификации знаков текста двумерных распределений знаковых биграмм, как в методе Якобсена [7], следует заметить, что из общего распределения биграмм для каждого знака эталона и анализируемого текста можно выделить два индивидуальных *условных* распределения биграмм, начинающихся и заканчивающихся данным знаком. Это одномерные распределения, получаемые в своеобразных «сечениях» общего двумерного распределения.

Тогда для идентификации знаков можно было бы применить метод Пирсона проверки однородности индивидуальных распределений [18], если бы не ограничения, связанные с тем, что такие распределения определены в различных системах координат для эталона и анализируемого текста.

Возьмем из метода Пирсона способ аппроксимации распределений, но модифицируем сам метод следующим образом:

1) оценку аппроксимации по распределению хи-квадрат Пирсона заменим прямой оценкой погрешности аппроксимации, учитывающей пропущенные значения и влияние «хвостов» сравниваемых распределений;

2) аппроксимацию распределений для каждого знака анализируемого текста будем проводить на всем множестве индивидуальных распределений частот биграмм знаков («букв») эталона;

3) идентификацию знаков анализируемого текста и образца будем осуществлять на основе минимальной оценки погрешности аппроксимации индивидуальных распределений;

4) введем специальную процедуру разрешения коллизий идентификации;

5) общую погрешность идентификации знаков анализируемого текста и образца будем оценивать на основе формулы (2).

Эти модификации приводят к следующему методу аппроксимации.

Пусть **AB** и **AT** — множества знаков («букв»), встречающихся в эталонном и анализируемом текстах соответственно, упорядоченные по частоте появления этих знаков. Назовем номер знака в этой упорядоченности его идентификатором. $V_B = |\mathbf{AB}|$ и $V_T = |\mathbf{AT}|$ — количества знаков

для множеств \mathbf{AB} и \mathbf{AT} соответственно, $VT \leq VB$. Множества \mathbf{AB} и \mathbf{AT} одновременно либо содержат, либо не содержат знак «пробел».

Пусть \mathbf{B} — таблица значений частот буквенных биграмм эталонного текста (образец), симметрично упорядоченная в соответствии с \mathbf{AB} , \mathbf{T} — таблица значений частот знаковых биграмм анализируемого текста (тест), симметрично упорядоченная в соответствии с \mathbf{AT} . Диагональные элементы в обеих таблицах обнулены. С учетом этих соглашений по упорядоченности, будем далее под \mathbf{B} и \mathbf{T} понимать соответствующие квадратные матрицы значений частот биграмм $b_{ij} \in \mathbf{B}$, $t_{ij} \in \mathbf{T}$.

1. Для каждого элемента $a_m^T \in \mathbf{AT}$ последовательно вычислим по формулам (5) два вектора \mathbf{S}_m^1 и \mathbf{S}_m^2 по строкам и столбцам матриц \mathbf{B} и \mathbf{T} соответственно:

$$s_{mj}^1 = \sum_{i=1}^{VB} \frac{(t_{mi} - b_{ji})^2}{t_{mi} + b_{ji}}, \quad s_{mj}^2 = \sum_{i=1}^{VB} \frac{(t_{im} - b_{ij})^2}{t_{im} + b_{ij}} \quad (5)$$

При этом если знаменатель в формулах (5) равен нулю, то значение соответствующей дроби в сумме принимается равным нулю.

Из (5) следует, что каждая строка (столбец) теста \mathbf{T} сравниваются со всеми строками (столбцами) эталона \mathbf{B} .

2. Для каждого элемента $a_m^T \in \mathbf{AT}$ по формулам (6) вычислим нормированную обратную ошибку аппроксимации для векторов \mathbf{S}_m^1 и \mathbf{S}_m^2 — векторы \mathbf{R}_m^1 и \mathbf{R}_m^2 соответственно:

$$r_{mj}^1 = 1 - \frac{s_{mj}^1}{\sum_{i=1}^{VB} b_{ji} + \sum_{i=1}^{VT} t_{mi}}, \quad r_{mj}^2 = 1 - \frac{s_{mj}^2}{\sum_{i=1}^{VB} b_{ij} + \sum_{i=1}^{VT} t_{im}} \quad (6)$$

и вычислим общую ошибку $\mathbf{R}_m = \mathbf{R}_m^1 + \mathbf{R}_m^2$.

Критерий (6) сформирован на основе простых геометрических представлений о близости форм подобных кривых и линейной интерполяции пропущенных значений, и на основе формул (1) и (5).

3. Упорядочим векторы \mathbf{R}_m по убыванию значений r_{mk} , с сохранением значения первоначального упорядочивания k . Значение k является идентификатором знака («буквы») эталона $a_k^B \in \mathbf{AB}$, и таким образом по значению r_{mk} устанавливается упорядоченная пара $\langle m, k \rangle$,

идентифицирующая знак теста $a_m^T \in \mathbf{AT}$ на множестве знаков эталона \mathbf{AB} . В качестве такой пары выберем первую пару из упорядоченного вектора \mathbf{R}_m .

4. Так как каждый знак $a_m^T \in \mathbf{AT}$ идентифицируется независимо от других знаков \mathbf{AT} на всем множестве знаков эталона \mathbf{AB} , то при выполнении шага 3 могут возникать коллизии, когда для одного и того же знака эталона выбираются два и более знака теста.

Для разрешения коллизий осуществляется следующая процедура «выталкивания». Если пара $\langle m, k \rangle$ уже идентифицирована, но появляется «претендент» j , то выбирается пара $\langle m, k \rangle$ или $\langle j, k \rangle$, у которой больше значение r_{mk} или r_{jk} и происходит новая идентификация. Выбывший знак переходит на место, определяемое в порядке убывания ошибки из собственного упорядоченного вектора \mathbf{R}_m или \mathbf{R}_j («выталкивается»). Образуется тройка $\langle m, i, k \rangle$ означающая, что знак $a_m^T \in \mathbf{AT}$ идентифицирован как знак $a_k^B \in \mathbf{AB}$ по значению $r_{mk}^i \in \mathbf{R}_m$, $i = \overline{1, VB}$. Процедура продолжается, пока все коллизии не будут разрешены.

Заметим, что процедура выталкивания не эквивалентна простому упорядочиванию максимальных значений.

5. Выбираем следующий элемент $a_{m+1}^T \in \mathbf{AT}$ и переходим к шагу 3 и так до полной идентификации всех знаков анализируемого текста.

6. Шаги 1-5 дают *частное* решение \mathbf{U}_n исходной задачи, $n \geq 1$, в соответствии с которым таблица \mathbf{T} переупорядочивается и образуется таблица \mathbf{T}_n .

Погрешность полученного решения $W(\mathbf{T}_n)$ вычисляется по формуле (2) для матриц \mathbf{B} и \mathbf{T}_n . Шаги 1-5 повторяются до тех пор, пока выполняется условие (7):

$$W(\mathbf{T}_{n-1}) - W(\mathbf{T}_n) > \varepsilon, \quad \varepsilon = -0.2, \quad \mathbf{T}_0 = \mathbf{T}, \quad n < 20, \quad (7)$$

аналогичное условию (4). В отличие от условия (4), условием (7) допускается увеличение $W(\mathbf{T}_n)$ относительно $W(\mathbf{T}_{n-1})$ и совпадающие значения, поэтому вводится ограничение на количество итераций n . Это позволяет пропускать локальные минимумы целевой функции при поиске глобального экстремума и ограничивать поиск в случаях неудачи. Значения ε и ограничения для n получены в результате вычислительного эксперимента. Если при получении очередного решения \mathbf{U}_n не

было осуществлено ни одной перестановки по сравнению с U_{n-1} , то итерационный процесс (7) прекращается независимо от значения n .

В качестве окончательного результата работы метода принимается результат предпоследней итерации — решение U_{n-1} .

Возможность с помощью итерационного процесса (7) управлять получаемой погрешностью наряду с использованием аппроксимации одномерных условных распределений биграмм для идентификации знаков текста является основной особенностью рассмотренного метода аппроксимации, отличающей его от метода Якобсена.

Так же как и метод Якобсена, рассмотренный метод аппроксимации позволяет идентифицировать знаки в текстах на любых языках, для которых существуют эталонные распределения частот буквенных биграмм, то есть является универсальным.

5. Результаты вычислительного эксперимента. Экспериментальная проверка погрешности метода Якобсена и предложенного метода аппроксимации была проведена на двух выборках фрагментов русскоязычных текстов.

Выборка 1 была сформирована из 100 научно-популярных и художественных текстов разных жанров и авторов; выборка 2 — из 100 текстов учебных пособий для вузов разных авторов из различных областей знаний: математика, химия, физика, машиностроение и т.д.

Из текстов 1-го и 2-го типов случайным образом были выделены последовательные фрагменты различной длины.

Выделение фрагментов для выборки 1 происходило после удаления из текста всех пробелов. Выделение происходило по принципу вложенности: сначала выделялись фрагменты большей длины, затем из них выделялись меньшие фрагменты последовательным удалением постоянного объема знаков. При этом начала фрагментов для одного текста совпадали. Это означает, что фрагменты разной длины для одного текста в выборке 1 включены друг в друга, а длина фрагментов совпадает со шкалой длин, представленной в эксперименте.

Фрагменты для выборки 2 выделялись со случайного знака текста с учетом пробелов, которые после этого удалялись. Таким образом, в отличие от выборки 1, выборка 2 сформирована из фрагментов случайной длины, начинающихся со случайного знака текста. Пересечения фрагментов для одного текста в выборке 2 возможны только случайно. Полученные для данных фрагментов значения ошибок отнесены к ближайшим значениям *сверху* шкалы длин фрагментов, одинаковой со шкалой выборки 1.

Выборки фрагментов для разных групп производились независимо друг от друга. Фрагменты обеих выборок распределены по 4

группам: группа 1 — фрагменты длиной от 400 до 1800 знаков, с шагом 200, группа 2 — фрагменты от 2000 до 10000 знаков, с шагом 2000, группа 3 — фрагменты от 30000 до 90000 знаков, с шагом 20000, группа 4 — фрагменты от 100000 до 350000 знаков, с шагом 50000.

Таблица 1. Результаты вычислительного эксперимента 1

<i>N</i>	<i>K</i>	<i>O1m</i>	<i>O1j</i>	<i>O2m</i>	<i>O2j</i>	<i>max</i> <i>O2m</i>	<i>max</i> <i>O2j</i>	<i>SD O2m</i>	<i>SD O2j</i>
Группа 1									
400	20	20	20	12.30	16.10	24	27	4.95	5.47
600	49	45	47	9.44	9.74	22	24	5.46	5.51
800	64	57	57	8.47	7.25	25	26	5.77	5.18
1000	76	64	62	6.30	6.97	23	24	4.46	6.34
1200	82	56	49	4.50	4.31	19	22	3.26	4.10
1400	86	53	50	3.83	4.30	17	20	3.23	3.34
1600	90	41	41	3.00	4.17	13	15	1.95	3.12
1800	93	35	26	2.86	3.88	9	15	1.68	2.86
Группа 2									
2000	93	32	26	2.59	3.12	6	12	1.17	2.06
4000	99	12	13	2.08	2.38	3	4	0.28	0.62
6000	100	7	9	2.43	3.00	3	5	0.49	1.15
8000	100	3	5	2.67	3.40	3	5	0.47	0.80
10000	100	3	6	3.33	2.83	5	3	1.25	0.37
Группа 3									
30000	50	2	1	2.00	3.00	2	3	0.00	0.00
50000	50	1	1	2.00	2.00	2	2	0.00	0.00
70000	50	1	1	2.00	2.00	2	2	0.00	0.00
90000	50	0	1	0.00	2.00	0	2	0.00	0.00
Группа 4									
100000	20	0	0	0.00	0.00	0	0	0.00	0.00
150000	20	0	0	0.00	0.00	0	0	0.00	0.00
200000	20	0	1	0.00	3.00	0	3	0.00	0.00
250000	20	0	0	0.00	0.00	0	0	0.00	0.00
300000	20	0	1	0.00	3.00	0	3	0.00	0.00
350000	20	0	0	0.00	0.00	0	0	0.00	0.00

Всего в эксперименте использовался 2631 фрагмент, из которых выборка 1 содержит 1372 фрагмента (группа 1 — 560, группа 2 — 492, группа 3 — 200, группа 4 — 120), а выборка 2 — 1259 фрагментов (группа 1 — 547, группа 2 — 481, группа 3 — 183, группа 4 — 48). При этом в каждом из фрагментов использовались все буквы русского алфавита (31, «Е» — «Е,Ё», «Ь» — «Ь,Ъ»).

В эксперименте определялось 2 вида ошибок: $O1$ — ошибка полной идентификации, как количество ошибочных фрагментов, содержащих хотя бы один неправильно идентифицированный знак; $O2$ — ошибка идентификации знаков, как среднее количество неправильно идентифицированных знаков для всех ошибочных фрагментов данной длины (по шкале длин). Для ошибки $O2$ дополнительно определялись минимальное и максимальное значение и стандартное отклонение. Результаты тестов для выборок 1 и 2 приведены в таблицах 1 и 2.

Таблица 2. Результаты вычислительного эксперимента 2

N	K	$O1m$	$O1j$	$O2m$	$O2j$	max $O2m$	max $O2j$	SD $O2m$	SD $O2j$
Группа 1									
400	10	10	10	17.60	20.00	24	29	5.22	7.05
600	30	30	30	12.47	14.30	26	26	7.26	5.76
800	42	38	41	9.37	8.12	22	21	5.92	5.33
1000	69	65	68	10.94	10.97	28	26	6.71	5.90
1200	67	64	66	7.25	9.01	22	26	5.22	5.81
1400	126	111	120	6.50	7.70	29	25	5.11	5.78
1600	118	103	109	6.51	6.74	28	27	6.01	5.41
1800	85	70	76	7.43	7.93	30	28	6.86	7.34
Группа 2									
2000	87	70	78	5.66	6.06	21	22	4.90	5.05
4000	97	71	81	4.55	4.60	18	19	3.82	3.40
6000	99	62	73	4.45	4.71	25	24	4.70	4.72
8000	99	50	70	3.66	3.70	14	25	3.03	3.90
10000	99	47	63	3.09	3.22	18	18	2.67	2.39
Группа 3									
30000	46	18	26	2.17	2.69	3	6	0.37	0.82
50000	46	13	21	2.85	3.38	7	12	1.46	2.30
70000	45	19	22	2.11	2.54	3	4	0.31	0.58
90000	46	18	24	2.44	2.71	6	8	0.96	1.21
Группа 4									
100000	8	4	5	3.50	3.60	6	8	1.50	2.24
150000	8	4	5	3.25	3.60	6	8	1.64	2.24
200000	8	3	6	3.67	3.33	6	7	1.70	1.70
250000	8	4	6	2.25	2.67	3	3	0.43	0.47
300000	8	3	5	3.00	4.00	4	9	0.82	2.53
350000	8	3	7	3.67	3.28	6	7	1.70	1.58

В таблицах 1 и 2 столбец N содержит значения длин текстов в знаках (без учета пробела), K — количество текстов; столбцы $O1$, $O2$ — абсолютные значения ошибок; столбцы max , SD — максимальное значение и стандартное отклонение ошибки $O2$. Минимальное значение $O2$ в таблицах 1 и 2 пропущено, так как для всех фрагментов текстов, содержащих больше четырехсот знаков, оно равно двум.

Для наглядного представления о динамике ошибок кусочно-линейная аппроксимация их нормированного значения представлена на соответствующих графиках рисунка 1 и рисунка 2.

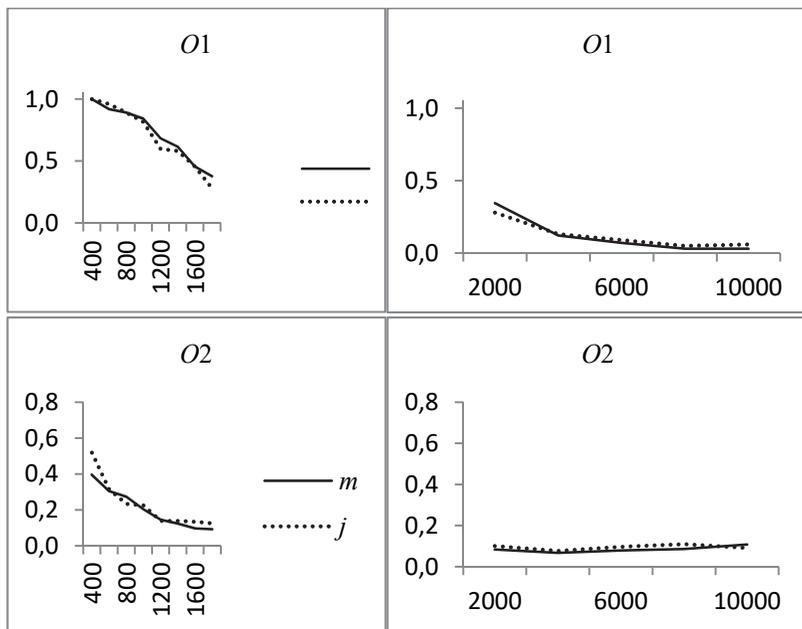


Рис. 1. Графики распределения нормированных ошибок $O1$ и $O2$ методов MAP – m и JAC – j для групп 1 и 2 фрагментов выборки 1

Индекс j в таблицах и на рисунках означает метод Якобсена, который сокращенно обозначим как JAC, индекс m — рассмотренный в разделе 4 метод аппроксимации, который сокращенно обозначим как MAP. Оба метода используют одинаковый эталон биграмм [16], независимый от выборок 1 и 2.

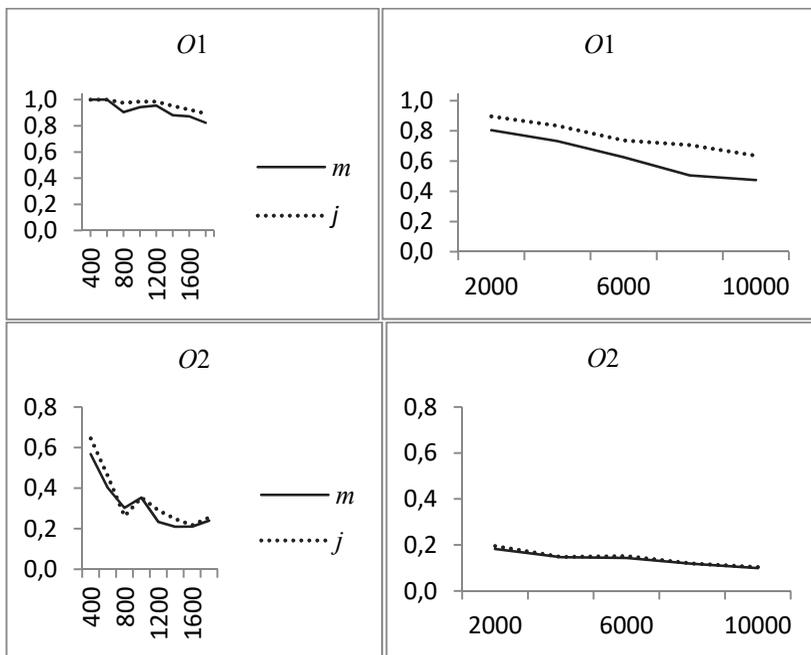


Рис. 2 Графики распределения нормированных ошибок $O1$ и $O2$ методов MAP – m и JAC – j для групп 1 и 2 фрагментов выборки 2.

Таблица 3. Методы JAC

G	P	$P min$	$P max$	$SD P$	S	$S min$	$S max$	$SD S$
Выборка 1								
1	34.71	7	82	10.93	1857.38	805	4307	545.76
2	25.09	7	63	8.73	1404.75	678	3060	415.17
3	18.99	8	53	6.91	1135.20	652	2440	294.04
4	16.19	6	41	6.18	1005.70	580	1946	243.17
Выборка 2								
1	43.13	14	97	11.84	2224.71	845	5522	632.14
2	36.55	14	84	11.59	1885.98	836	4017	524.52
3	27.92	10	59	8.33	1504.19	711	2793	404.99
4	23.56	13	35	5.30	1301.48	826	2273	354.28

В таблицах 3 и 4 представлены данные по основным вычислительным характеристикам методов JAC и MAP соответственно. Столбец G содержит номер группы текстов, P — среднее число выполненных перестановок, столбцы $x min$, $x max$, $SD x$ — минимальное, макси-

мальное значение величины x (P , S или U) и стандартное отклонение соответственно.

Таблица 4. Методы MAP

G	P	P_{min}	P_{max}	$SD P$	U	U_{min}	U_{max}	$SD U$
Выборка 1								
1	78.24	15	493	77.22	5.81	1	20	5.23
2	30.59	12	124	11.66	2.28	1	7	0.90
3	22.44	11	48	7.13	1.62	1	4	0.69
4	20.63	11	54	8.12	1.33	1	4	0.72
Выборка 2								
1	119.95	21	572	101.12	8.32	1	20	6.19
2	59.93	19	429	47.57	4.45	1	20	3.50
3	38.31	19	269	19.37	2.90	1	20	1.76
4	36.19	23	63	10.26	2.77	2	4	0.71

Столбец S таблицы 3 содержит среднее общего числа выполненных просмотров в методе JAC, столбец U таблицы 4 — среднее общего числа итераций (частных решений) в методе MAP. Значения P в таблице 4 представляют сумму перестановок по всем выполненным итерациям.

6. Обсуждение результатов. Из данных эксперимента можно сделать следующие выводы.

1) Выборки 1 и 2 представляют разные модели текстов. Выборка 1 представляет модель семантически связного, последовательно развивающегося (с точки зрения изложения) текста, словарь которого является наиболее общим для всех носителей языка. Такая модель дает возможность оценить в первую очередь зависимость погрешности метода от объема (длины) одного «усредненного» текста.

Как видно из данных таблицы 1 и графиков рисунка 1, в диапазоне объемов текстов от 400 до 4000 знаков погрешность $O1$ в методе MAP в среднем на 7% больше, чем в методе JAC.

В диапазоне объемов текстов от 4000 знаков и выше погрешность $O1$ в методе MAP в среднем на 26% меньше, чем в методе JAC.

Погрешность $O2$ во всем диапазоне анализируемых текстов в среднем меньше на 17% для метода MAP, чем для метода JAC. Она будет меньше в 20 точках шкалы измерений из 23, чем в методе JAC. То есть при *наличии* ошибки число неправильно идентифицированных знаков для метода MAP будет в среднем меньше.

С учетом указанных различий в целом погрешности методов MAP и JAC зависят от объема «усредненного» текста для текстов выборки 1 приблизительно одинаково.

2) Выборку 2 можно интерпретировать как модель «произвольного» текста. Семантика учебных пособий едина в рамках учебной дисциплины, но различается, иногда значительно, в разных разделах одного пособия. Изложение содержания, как правило, лаконично и ведется с использованием большого числа локальных сокращений. Текст перемежается большим количеством чисел, формул, таблиц и графиков, при формальном исключении которых возникают синтаксические «разрывы». Используются специальные терминологические словари. Каждый случайный фрагмент текста из учебного пособия, в котором оставлены только буквы языка, представляет собой семантический и синтаксический кластер, не всегда и не полностью грамматически правильный и понятный произвольному носителю языка. Усреднение данных по выборке 2 дает возможность оценить зависимость погрешности метода от объема (длины) одного «произвольного» текста.

Данные таблицы 2 и графиков рисунка 2 показывают, что для выборки 2 погрешности O_1 и O_2 в методе MAP в среднем на 7% меньше, чем в методе JAC во всем диапазоне шкалы объемов текстов. При этом погрешность O_1 меньше в 22 точках шкалы из 23, чем в методе JAC. Погрешность O_2 меньше в 20 точках шкалы из 23, чем в методе JAC.

В целом метод MAP на текстах выборки 2 имеет меньшие погрешности O_1 и O_2 , чем метод JAC, а их зависимость от объема «произвольного» текста для метода MAP является более стабильной, чем для метода JAC.

3) Метод MAP медленнее, чем метод JAC, что показывают данные таблиц 3 и 4. Сравнительную оценку можно получить исходя из того, что в методе MAP для получения одного частного решения U требуется *не менее* 961 просмотра, аналогичного просмотрам S метода JAC.

Однако из опыта следует, что такое замедление малозаметно и не существенно для решения задачи идентификации знаков текста, особенно с учетом уменьшения погрешности идентификации. При этом из данных таблиц 3 и 4 видно, что уменьшение погрешности в методе MAP обусловлено поиском решения в большем пространстве реальных перестановок, чем в методе JAC.

4) Суммируя данные по выборкам 1 и 2, можно сделать вывод, что погрешность метода MAP в целом меньше, чем метода JAC.

Следует отметить определенную относительность значения ошибки при решении задачи идентификации знаков текста. Например, для одного и того же фрагмента текста получено значение O_2 , равное двум по методу MAP, и четырем — по методу JAC. Однако в первом случае неправильно идентифицированы часто встречающиеся гласные (А, И), тогда как во втором — редко встречающиеся согласные (Г, Б, Ш, Ж).

Какую из двух полученных ошибок — меньшую или большую — будет в данном случае легче исправить, зависит от конкретного текста.

Данные о подобном качественном различии решений, получаемых по методам МАР и JAC, представлены на примере текстов (общее количество — 547) группы 1 выборки 2 в таблице 5. Здесь столбец М представляет метод идентификации, столбцы «ОЕАИ», «НТС» и т.д. — группы идентифицируемых знаков текста, значения столбцов — суммарное количество текстов, в которых встретилась хотя бы одна ошибка при идентификации знаков данной группы.

Таблица 5. Сравнение методов МАР, JAC

М	ОЕАИ	НТС	ВРЛП	КМД	ЙЬЬ	УЯЮЭ	ЗГЬЧХ	ШЖЩЦФ
МАР	158	111	178	242	106	179	327	459
JAC	127	94	197	301	147	188	361	500

Из таблицы 5 можно увидеть, что при решении задачи по методу МАР происходит уменьшение ошибки идентификации по сравнению с методом JAC в шести группах знаков из восьми, начиная с третьей, но в первых двух группах погрешность идентификации может возрасти.

7. Заключение. В статье рассмотрены особенности применения методов упорядочения и аппроксимации для решения задачи идентификации знаков текста, и взаимосвязь данных методов. Показано, что метод Якобсена [7] является методом двумерного упорядочивания частот знаковых биграмм, и определены условия, при выполнении которых данный метод имеет наименьшую погрешность при использовании независимого эталона частот буквенных биграмм.

Предложен метод аппроксимации, основанный на сравнении одномерных и двумерных распределений частот знаковых биграмм. При сравнении одномерных условных распределений частот знаковых биграмм используется аппроксимация формы распределений, аналогичная аппроксимации, применяемой в методе Пирсона для проверки однородности двух распределений. Выбор наилучшей аппроксимации осуществляется на основе предложенной оценки погрешности. Частные решения по идентификации знаков текста на основе одномерных условных распределений последовательно улучшаются в ходе итерационного процесса аппроксимации двумерных распределений частот знаковых биграмм.

Это позволяет управлять погрешностью идентификации, которая в целом меньше, чем у метода Якобсена на 12- 17%. Приведены результаты экспериментальной проверки погрешностей предложенного метода аппроксимации и метода Якобсена на двух представительных выборках фрагментов из русскоязычных текстов.

Предложенный в статье метод аппроксимации прост в реализации, является универсальным и может быть использован для идентификации знаков текста любого языка, для которого существует эталонное распределение частот буквенных биграмм.

Литература

1. *Котов Ю.А.* Детерминированная идентификация буквенных биграмм в русскоязычных текстах // Труды СПИИРАН. 2016. Вып. 1(44). С. 181–197.
2. *Шеннон К.* Теория связи в секретных системах // Работы по теории информации и кибернетике // М.: ИЛ. 1963. С. 333–369.
3. *Бабенко Л.К. и др.* Развитие криптографических методов и средств защиты информации // Известия ЮФУ. Технические науки. 2012. № 4. С. 40–50.
4. *Бабенко Л.К., Ицуква Е.А.* Анализ симметричных криптосистем // Известия ЮФУ. Технические науки. 2012. № 12. С. 136–147.
5. *Глухов М.М., Круглов И.А., Пичкур А.Б., Черёмушкин А.В.* Введение в теоретико-числовые методы криптографии // СПб.: Лань. 2011. 400 с.
6. *Минеев М. П., Чубариков В. Н.* Лекции по арифметическим вопросам криптографии // М.: Изд-во «Попечительский совет Механико-математического факультета МГУ им. М.В. Ломоносова». 2010. 186 с.
7. *Jakobsen T.* A fast Method for Cryptanalysis of Substitution Ciphers // Cryptologia. 1995. vol. 19. no. 3. pp. 265–274.
8. *Corlett E.* An Exact A* Method for Solving Letter Substitution Ciphers //University of Toronto. 2011. URL: <ftp://ftp.cs.toronto.edu/pub/gh/Corlett-MSc-2011.pdf> (дата обращения 16.10.2016).
9. *Varagada S.R., Reddy P.S.* A Survey of Cryptanalytic Works Based on Genetic Algorithms // International Journal of Emerging Trends & Technology in Computer Science (IJETTCSS). 2013. vol. 2. no. 5. pp. 18–22.
10. *Singh A.P., Pal S.K., Bhatia M.P.S.* The Firefly Algorithm and Application in Cryptanalysis of Monoalphabetic Substitution Ciphers // American Journal of Computer Science and Engineering Survey. 2013. vol. 1. no. 1. pp. 33–52.
11. *Морозенко В.В., Плешкова И.Ю.* О применении генетического алгоритма для криптоанализа шифра Тритемия-Белазо-Виженера // Современные проблемы науки и образования: электронный научный журнал. 2014. № 2. С. 1–11.
12. *Chen J., Rosenthal J.S.* Decrypting classical cipher text using Markov chain Monte Carlo // Statistics and Computing. 2011. vol. 22. no. 2. pp. 397–413.
13. *Bhateja A., Kumar S., Bhateja A.K.* Cryptanalysis of Vigenere Cipher using Particle Swarm Optimization with Markov chain random walk // International Journal on Computer Science and Engineering (IJCSE). 2013. vol. 5. no. 5. pp. 422–429.
14. *Васильев Е.М., Жданова Д.В.* Диахроническое исследование энтропии графем русского письма // Вестник Воронежского государственного технического университета. 2010. № 4. С.1–3.
15. *Васильев Е.М., Гусев К.Ю.* Анализ избыточности русскоязычного текста // Вестник Воронежского государственного технического университета. 2010. № 8. С.1–4.
16. *Жданов О. Н., Куденкова И. А.* Криптоанализ классических шифров // Красноярск: Изд-во Сиб. гос. аэрокосм. ун-та им. акад. М.Ф. Решетнева. 2008. 107 с.
17. *Mohan M., Devi M.K.K., Prakash V.J.* Security Analysis and Modification of Classical Encryption Scheme // Indian Journal of Science and Technology. 2015. vol. 8 no. 8. pp. 542–548.
18. *Губарев В.В.* Введение в теоретическую информатику // Новосибирск: Изд-во НГТУ. 2014. 420 с.

Котов Юрий Алексеевич — к-т физ.-мат. наук, доцент кафедры защиты информации факультета автоматики и вычислительной техники, Новосибирский государственный технический университет (НГТУ). Область научных интересов: информационная и компьютерная безопасность, криптография и криптоанализ, математическое обеспечение вычислительных систем. Число научных публикаций — 25. kotov@corp.nstu.ru; пр. К. Маркса, 20, Новосибирск, 630073; р.т.: +7(383)346-58-03, Факс: +7(383)346-58-03.

YU.A. KOTOV
**APPROXIMATION OF DISTRIBUTIONS OF TEXT CHARACTERS
 BIGRAMS FREQUENCIES FOR ALPHABETIC CHARACTERS
 IDENTIFICATION**

Kotov Yu.A. Approximation of Distributions of Text Characters Bigrams Frequencies for Alphabetic Characters Identification.

Abstract. The article discusses the application features of methods of the frequencies ordering and approximation to solve the problem of text characters identification. The conditions for realization of Jacobsen's method for receiving the least error of identification are defined. The method of approximation of one- and two-dimensional distributions of the frequencies of characters bigrams of the text and the language is offered. The experimental data about errors of Jacobsen's method and the offered approximation method for Russian language texts are provided.

The error of the offered method is less than that of Jacobsen's method. This method can be used for identification of text characters for any language that has a reference distribution of the alphabetic characters bigrams frequencies.

Keywords: approximation, identification, character, bigram, one-to-one substitution, cypher.

Kotov Yuri Alexeevich — Ph.D., associate professor of information protection department of faculty of automation and computer engineering, Novosibirsk State Technical University (NSTU). Research interests: information and computer security, cryptography, software technologies and development of information systems. The number of publications — 25. kotov@corp.nstu.ru; 20, pr. K. Marksa, Novosibirsk, 630073; office phone: +7(383)346-58-03, Fax: +7(383)346-58-03.

References

1. Kotov Yu.A. [Determinate Identification of Russian Text Letter Bigrams]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2016. vol. 1(44). pp. 181–197. (In Russ.).
2. Shannon K. *Raboty po teorii informacii i kibernetike* [Works on the theory of information and cybernetics]. M.: IL. 1963. 832 p. (In Russ.).
3. Babenko L.K. et al. [Development of cryptographic methods and information security tools]. *Izvestija JuFU. Tehniceskie nauki – Izvestiya SFedU. Engineering sciences*. 2012. vol. 4. pp. 40–50. (In Russ.).
4. Babenko L.K., Ishchukova E.A. [Analysis of symmetric cryptosystems]. *Izvestija JuFU. Tehniceskie nauki – Izvestiya SFedU. Engineering sciences*. 2012. vol. 12. pp. 136–147. (In Russ.).
5. Gluhov M.M., Kruglov I.A., Pichkur A.B., Cheryomushkin A.V. *Vvedenie v teoretiko-chislovye metody kriptografii* [Introduction to number-theoretic methods in cryptography]. SPb.: Lan'. 2011. 400 p. (In Russ.).
6. Mineev M.P., Chubarikov V.N. *Lekcii po arifmeticheskim voprosam kriptografii* [Lectures on arithmetic cryptography]. M.: Izd-vo «Popechitel'skij soviet Mehaniko-matematicheskogo fakul'teta MGU im. M. V. Lomonosova». 2010. 186 p. (In Russ.).
7. Jakobsen T. A fast Method for Cryptanalysis of Substitution Ciphers. *Cryptologia*. 1995. vol. 19. no. 3. pp. 265–274.
8. Corlett E. An Exact A* Method for Solving Letter Substitution Ciphers //University of Toronto, 2011. Available at: <ftp://ftp.cs.toronto.edu/pub/gh/Corlett-MSc-2011.pdf> (accessed 16.10.2016).

9. Baragada S.R., Reddy P.S. A Survey of Cryptanalytic Works Based on Genetic Algorithms. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*. 2013. vol. 2. no. 5. pp. 18–22.
10. Singh A.P., Pal S.K., Bhatia M.P.S. The Firefly Algorithm and Application in Cryptanalysis of Monoalphabetic Substitution Ciphers. *American Journal of Computer Science and Engineering Survey*. 2013. vol. 1. no. 1. pp. 33–52.
11. Morozenko V.V., Pleshkova I.Yu. [On the application of a genetic algorithm for cryptanalysis of the cipher Triteria-Belazo-Vigenère]. *Sovremennyye problemy nauki i obrazovaniya: jelektronnyj nauchnyj zhurnal – Modern problems of science and education*. 2014. vol. 2. pp. 1–11. (In Russ.).
12. Chen J., Rosenthal J.S. Decrypting classical cipher text using Markov chain Monte Carlo. *Statistics and Computing*. 2011. vol. 22. no. 2. pp. 397–413.
13. Bhateja A., Kumar S., Bhateja A.K. Cryptanalysis of Vigenere Cipher using Particle Swarm Optimization with Markov chain random walk. *International Journal on Computer Science and Engineering (IJCSE)*. 2013. vol. 5. no. 5. pp. 422–429.
14. Vasil'ev E.M., Zhdanova D.V. [Diachronic study of the entropy of the graphemes of the Russian writing]. *Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta – Bulletin of Voronezh State Technical University*. 2010. vol. 4. pp. 1–3.
15. Vasil'ev E.M., Gusev K.Yu. [Redundancy analysis of Russian text]. *Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta – Bulletin of Voronezh State Technical University*. 2010. vol. 8. pp. 1–4.
16. Zhdanov O.N., Kudenkova I.A. *Kriptoanaliz klassicheskikh shifrov* [Cryptanalysis of classical ciphers]. Krasnojarsk: Izd-vo Sib. gos. ajerokosm. un-ta im. akad. M.F. Reshetneva. 2008. 107 p. (In Russ.).
17. Mohan M., Devi M.K.K., Prakash V.J. Security Analysis and Modification of Classical Encryption Scheme. *Indian Journal of Science and Technology*. 2015. vol. 8 no. 8. pp. 542–548.
18. Gubarev V.V. *Vvedenie v teoreticheskuyu informatiku* [Introduction to theoretical informatics]. Novosibirsk: Izd-vo NGTU. 2014. 420 p. (In Russ.).

Ю.Н. БАЛОНИН, А.А. ВОСТРИКОВ, А.М. СЕРГЕЕВ, И.С. ЕГОРОВА
**О ВЗАИМОСВЯЗЯХ КВАЗИОРТОГОНАЛЬНЫХ МАТРИЦ,
ПОСТРОЕННЫХ НА ИЗВЕСТНЫХ
ПОСЛЕДОВАТЕЛЬНОСТЯХ ЧИСЕЛ**

Балонин Ю.Н., Востриков А.А., Сергеев А.М., Егорова И.С. О взаимосвязях квазиортогональных матриц, построенных на известных последовательностях чисел.

Аннотация. Цель работы: показать связь чисел, принадлежащих известным последовательностям, и квазиортогональных матриц, существующих на порядках, равных этим числам, а также взаимосвязь таких матриц через алгоритмы вычисления. Методы: анализ последовательностей квазиортогональных матриц абсолютного и локального максимумов детерминанта, выделение в матрицах структурных инвариантов, сопоставление алгоритмов вычисления таких матриц. Результаты: рассмотрены известные последовательности натуральных чисел, сформулировано определение матрицы, ассоциированной с натуральным числом. Приведены последовательности чисел, для которых доказано существование ассоциированных с ними квазиортогональных матриц. Высказано предположение, что ассоциированные матрицы существуют для всех натуральных чисел. Рассмотрены свойства типов таких матриц, их взаимосвязи через алгоритмы вычисления. Приведены модифицированные алгоритмы и основные цепочки матриц Эйлера и Мерсенна, последовательности порядков которых являются системообразующими. Практическая значимость: квазиортогональные матрицы абсолютного и локального максимумов детерминанта имеют непосредственное практическое значение для задач помехоустойчивого кодирования, сжатия и маскирования видеoinформации. Их разнообразие позволяет разработчикам технических систем значительно облегчить выбор матрицы, оптимальной для конкретной задачи.

Ключевые слова: числовые последовательности, числа Мерсенна, числа Ферма, ортогональные матрицы, квазиортогональные матрицы, матрицы Адамара, матрицы Мерсенна, матрицы Эйлера, матрицы Ферма, цепочки матриц, алгоритмы вычисления квазиортогональных матриц.

1. Введение. Практический интерес к ортогональным (квазиортогональным) матрицам обусловлен основными свойствами таких матриц, обеспечивающими им широкое применение в цифровых системах обработки и преобразования информации [1]. Ортогональность матриц позволяет реализовывать конгруэнтные преобразования, а существующие ортогональные базисы, включая симметричные, циклические, двуциклические и другие конструкции значительно расширяют возможности выбора оптимальной матрицы для решения конкретной задачи преобразования информации.

В теории кодирования, например, столбцы ортогональных матриц Адамара используются для построения кодов с большим кодовым расстоянием [2]. Специальный порядок нумерации столбцов таких матриц в цифровой обработке сигналов, сжатии и маскировании изоб-

ражений интерпретируется как двухуровневое представление широко используемых функций Уолша [3].

Первый алгоритм вычисления ортогональных по столбцам (квазиортогональных) матриц с элементами $\{1, -1\}$ четных порядков $n = 2^k$, где k — целое, был предложен Сильвестром (Sylvester) [4].

Адамар (Hadamard) сформулировал гипотезу о существовании таких матриц на порядках 1, 2 и $4t$, где t — натуральное число, включающих порядки 2^k , заложив основы исследований [1, 5], во-первых, соответствия последовательностей целых чисел и квазиортогональных матриц на порядках, равных этим числам; во-вторых, вложенности числовых последовательностей при сохранении качества соответствия чисел и матриц.

С последовательностью $4t$ соседствуют известные последовательности нечетных чисел $4t+1$ и $4t-1$ ($4t+3$), на обособленные числовые свойства которых обратили внимание еще Ферма и Эйлер.

Последовательность нечетных чисел вида $4t-1$ включает последовательность чисел Мерсенна, задаваемую как $n = 2^k - 1$ и начинающуюся с чисел 1, 3, 7, 15, 31, ... Последовательность чисел Ферма, определяемая формулой $n = 2^{2^k} + 1$, начинается с чисел 3, 5, 17, 257, 65537, 4294967297, 18446744073709551617, ... и принадлежит последовательности чисел вида $4t+1$.

Отличительной особенностью матриц Адамара, поставленных в соответствие последовательности чисел Сильвестра $n = 2^k$, является их оптимальность по детерминанту. В общем случае квазиортогональные матрицы могут иметь как глобальный, так и локальный максимум детерминанта [4, 6].

Определение 1. Субоптимальную (строго оптимальную) по детерминанту квазиортогональную (или соответствующую ей ортогональную) матрицу порядка, равного некоторому числу числовой последовательности, будем называть матрицей, ассоциированной с этим числом.

Цель настоящей статьи — показать, что последовательностям натуральных чисел соответствуют последовательности матриц, которые будем называть ассоциированными с ними. На четных порядках это могут быть оптимальные матрицы Адамара [5], Белевича [7] и тому подобное, на нечетных — субоптимальные матрицы, ассоциированные с последовательностями Мерсенна и Ферма. Указанное обстоятельство дает хорошую ориентацию в разнообразии базисов и позволяет разработчикам технических систем значительно облегчить выбор матрицы, оптимальной для конкретной задачи обработки или преобразования информации.

2. Таблица квазиортогональных матриц. Для большей определенности в типах встречаемых квазиортогональных матриц мы приведем таблицу 1 матриц порядков $4t \pm k$, здесь $k \leq 3$, классифицируемых по принадлежности их элементов заданным константам (как у матриц Адамара) или функциям уровня, зависящим от порядка n . Порядки $4t-3$ (или $4t+1$) сложнее прочих тем, что ортогональность столбцов матриц достижима при введении дополнительного уровня d для элементов диагонали (матрицы Зейделя) или каймы s (матрицы Ферма).

Таблица 1. Значения уровней семейств матриц

Символ	Порядок n	Матрица	Значения элементов
H	$4t$	Адамара	$1, -1$
C	$2t, 4t$	Белевича	$1, -1, 0$
W	$t, 2t, 3t, 4t$	Себерри (взвешенная)	$1, -1, 0$
M	$4t-1$	Мерсенна	$1, -b$, где $b = \frac{t}{t + \sqrt{t}}$
E	$4t-2$	Эйлера	$1, -b$, где $b = \frac{t}{t + \sqrt{2t}}$
S	$4t-3$	Зейделя	$1, -b, d$, где $b = 1 - 2d$, $d = \frac{1}{1 + \sqrt{n}}$
F	$4t+1$	Ферма	$1, -b, s$, где $q = n - 1 = 4u^2$, $p = q + \sqrt{q}$, $b = \frac{2n - p}{p} = 1 - \frac{2u - 1}{2u + 1} \times \frac{1}{u}$ $s = \frac{\sqrt{nq} - 2\sqrt{q}}{p} = \frac{\sqrt{nu} - 1}{2u + 1} \times \frac{1}{\sqrt{u}}$

3. Квазиортогональные матрицы Мерсенна. В работе [8] предложены квазиортогональные матрицы, существующие на нечетных порядках $n = 2^k - 1$, которые могут быть вложены в последовательность $4t-1$ при сохранении значения модулей элементов матрицы.

Определение 2. Матрица Мерсенна (**M**) — квадратная матрица порядка $n = 4t - 1$, состоящая из чисел $\{a = 1, -b\}$, столбцы которой ортогональны $\mathbf{M}_n^T \mathbf{M}_n = \mu \mathbf{I}$, $b = \frac{t}{t + \sqrt{t}}$, $\mu = \frac{p + qb^2}{2}$, $p = n - 1$, $q = n + 1$ (порядок матрицы Адамара).

4. Вычисление матриц Мерсенна. Для вычисления матриц Мерсенна повышенного порядка может использоваться *модифицированная последовательность Сильвестра* [8], представляемая в виде:

$$S_{2n} = \begin{pmatrix} M_n & M_n \\ M_n & M_n^* \end{pmatrix}$$

и отличающаяся от классической тем, что матрица M_n^* образована перестановкой местами элементов $a=1$ и $-b$. У матриц Адамара это приводит к смене знака. Матрицы порядков, равных числам последовательности Мерсенна, образованы дополнением указанной основы строкой и столбцом в виде:

$$M_{2n+1} = \begin{pmatrix} -\lambda & e^T \\ e & S_{2n} \end{pmatrix},$$

где $\lambda = -a$ — собственное число, а e — собственный вектор матрицы S_{2n} .

5. Прикладная сторона, построение фильтров Мерсенна. В работе с участием авторов подробно раскрывается построение на основе матриц Мерсенна фильтров Мерсенна. К удобству читателя приведем краткую выдержку из этой работы, поясняющую ее суть. Элементарными преобразованиями (умножениями на -1 , перестановками строк и столбцов) матрицы Мерсенна порядков $n = 2^k - 1$ переупорядочиваются к форме Мерсенна-Уолша (рисунок 1).

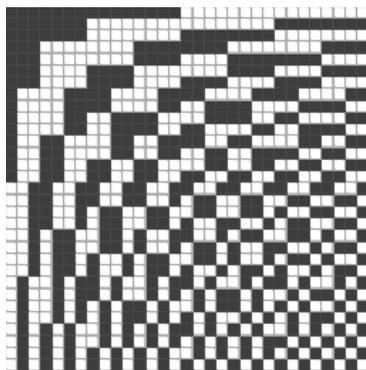


Рис. 1. Квазиортогональная матрица Мерсенна-Уолша M_{31}

Фильтр Мерсенна оперирует двумерным спектром изображения P , который вычисляется по формуле $S = M^T P M$ (столбцы M нормализованы до единичных норм). Высокочастотная часть спектра со-

держится в нижнем правом углу элементов S , которые в целях фильтрации либо обнуляются, либо умножаются на понижающие элементы. После передачи по каналу связи исходное изображение восстанавливается $P = MSM^T$, не меняя обозначений для большей простоты. Восстановленные изображения будут отличаться от исходного на величину ошибки сжатия изображения, которая равна 0, если не вносить изменения. Напомним, что в задачах маскирования сжатие не производится или производится вынужденно с целью скрытия факта маскирования. Для маскирования большое значение играет не столько частота, сколько иррациональность уровней. В отличие от матриц Адамара перебор перестает играть роль ключа для скрытия изображения, нужно знать еще и уровень, а при иррациональном значении его не подобрать рациональными приближениями.

Система ортогональных функций, порождаемых упорядоченными по частоте столбцами матриц Мерсенна, принимает значения $\{1, -b\}$. Перечислим ее отличительные особенности и назовем причины, по которым такой базис сигналов может быть интересным. Система функций Мерсенна — Уолша — двухуровневая, она такая же, как и классическая система функций Уолша. Она отличается от функций Уолша пониженным по амплитуде нижним значением $-b$, которое с ростом размерности системы стремится к -1 . В этом смысле она отличается от системы функций Уолша, но является достаточно близкой аппроксимацией ее на нечетных значениях порядка. Систему функций Мерсенна — Уолша отличает также пониженное на единицу количество порождающих ее элементов, она более проста для вычисления.

Любой базис отличает предпочтительная область его применения. Система функций Мерсенна — Уолша более высокочастотная, чем система функций Уолша, в ее составе нет функции нулевой частоты (константы). Таким образом, она более удобна для построения полосовых фильтров изображений. Первые единичные столбец и строка нормализованных матриц Адамара представляют собой ненужную составляющую, которая у полосовых фильтров никакой нагрузки не несет, поскольку отвечает частоте, которую они фильтруют, соответственно, она означает лишние затраты процессорного времени. Однако заметим, что простое удаление канвы матрицы Адамара отбрасыванием ее первых строки и столбца нарушает ортогональность столбцов усеченной матрицы.

6. Взаимосвязи квазиортогональных матриц. В работе [9] раскрывается гипотеза о существовании аналогичных матриц на всех нечетных порядках $n = 4t - 1$. Она имеет для матриц Мерсенна системное значение и отвечает гипотезе Адамара, расширяющей трактовку матриц порядков Сильвестра на порядки $n = 4t$. Покажем, каким образом можно находить дополнительные к основной последовательности

матрицы Мерсенна. Согласно работе [8], ассоциированные с числами последовательностей Сильвестра и Мерсенна матрицы встречаются чаще, чем относительно более редкие матрицы, ассоциированные с числами последовательности Ферма [10]. Причем матрицы Адамара порядков $4t$ можно получить на основе матриц Мерсенна порядков $4t-1$, и наоборот. Последовательности чисел Сильвестра и Мерсенна связаны между собой однозначно, столь же однозначно связаны между собой и ассоциированные с ними матрицы.

Взаимосвязь 1. Матрица \mathbf{H}_{4t} вычисляется путем окаймления округленной до целых матрицы Мерсенна в виде:

$$\mathbf{H}_{4t} = \begin{pmatrix} -\lambda & e^T \\ e & \mathbf{M}_{4t-1} \end{pmatrix},$$

с заменой элементов $-b$ образующей ее матрицы на -1 . Здесь λ , e — собственное число и собственный вектор округленной целочисленной матрицы \mathbf{M}_{4t-1} соответственно.

Возможен и обратный ход этого алгоритма, когда усечением нормированной матрицы Адамара (с изменением по знаку) и изменением отрицательных значений ее элементов до расчетного значения уровня $-b$, вычисляемого по формулам, приведенным в определении 2, формируется матрица Мерсенна. Такая взаимосвязь показывает, что с точностью до знака округленная матрица Мерсенна является ядром (core) нормализованной матрицы Адамара.

В качестве примера матриц, отличных от матриц основной последовательности, на рисунке 2 и на рисунке 3 приведены портреты матриц Адамара \mathbf{H}_{12} с инвертированным по знаку, чтобы обеспечить должное значение $-\lambda$, ядром и матрицы Мерсенна \mathbf{M}_{11} соответственно.

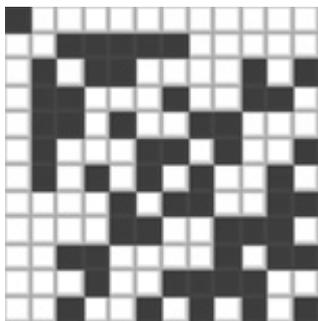


Рис. 2. Портрет матрицы \mathbf{H}_{12}

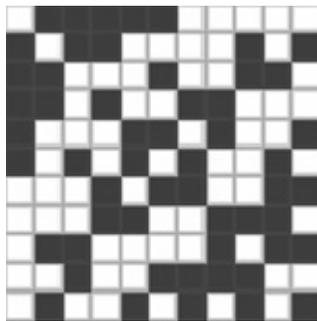


Рис. 3. Портрет матрицы \mathbf{M}_{11}

Следовательно, матрица Мерсенна, включая порядки, отличные от чисел Мерсенна, может быть найдена посредством матрицы Адамара

ра, и наоборот. Более того, каждая из таких матриц Мерсенна является исходной для получения ветви матриц применением описанного выше и модифицированного для нечетных порядков алгоритма Сильвестра.

В работах [8, 10] матрицы Мерсенна и Ферма представлены самостоятельно, отдельно друг от друга. Однако существуют некоторые общие структуры квазиортогональных матриц четного и нечетного порядков.

Взаимосвязь 2. Матрица Ферма F_{4u^2+1} , где u — целое число, может быть получена путем окаймления *регулярной* матрицы Адамара в виде:

$$F_{4u^2+1} = \begin{pmatrix} -\lambda & e^T \\ e & H_{4u^2} \end{pmatrix},$$

где λ , e — собственное число и собственный вектор *регулярной* матрицы Адамара H_{4u^2} (суммы строк и столбцов одинаковы), у которой отрицательные элементы заменены значениями, характерными для матриц Ферма.

Элемент -1 матрицы Адамара следует принять равным значению $-b$, $b = \frac{2n-p}{p} = 1 - \frac{2u-1}{2u+1} \times \frac{1}{u}$, элементы $s = \frac{\sqrt{ng-2\sqrt{q}}}{p}$ собственного вектора являются элементами ($a < s \leq b$) каймы матриц Ферма, $q = n-1 = 4u^2$ (порядок регулярной матрицы Адамара), $p = q + \sqrt{q}$ [10].

Возможен обратный ход алгоритма вычисления вложенной матрицы Адамара по матрице Ферма. Отличие матриц Ферма от матриц Мерсенна состоит в том, что первые существуют только для значений порядков, равных числам Ферма, а также расширенных порядков, равных числам вида $n = 4u^2 + 1$, где u — натуральное число, имеем: 3, 5, 17, 37, 65, 101, 145, 197, 257, ...

Матрица Адамара после нормализации (операции, при которой знаки элементов ее первого столбца и строки совпадают) дальнейшим усечением ее с пересчетом уровня переводится сначала в матрицу Мерсенна, а потом усечением — в матрицу Эйлера [6] четного порядка с модулем уровня $b = \frac{q - \sqrt{8q}}{q-8}$, $q = n+2$, $n = 4t-2$.

Матрицы Эйлера сходны с матрицами Адамара в том, что они четных порядков, для них справедливо следующее разложение на квадратные блоки:

$$E_{2n} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & -\mathbf{A}^T \end{pmatrix},$$

верхние блоки образуются из матриц Мерсенна $\mathbf{A} = \mathbf{B} = \mathbf{M}_n$. Расширение этого простого частного случая на матрицы с не совпадающими блоками даст весь набор порядков $n = 4t - 2$ матриц с указанными уровнями.

Матрицы Мерсенна есть для всех значений $n = 4t - 1$ без исключения. Матрицы Эйлера удобны для технических приложений тем, что они существуют, соответственно, для всех $n = 4t - 2$. Они удачным образом заполняют известные пробелы порядков 22, 34, 58 и т. п. среди квазиортогональных матриц Белевича [7]. Объясняется это тем, что матрицы Эйлера могут быть иррациональны, тогда как целочисленные матрицы Белевича даны только для порядков, на которых $n - 1$ разложимо в сумму квадратов целых чисел.

7. Другие взаимосвязи квазиортогональных матриц. В работе [4] вводится определение, согласно которому порядок матриц Эйлера четен и равен $n = 4t - 2$, где t — натуральное число.

На основе матриц Мерсенна порядков $n = 4t - 1$, принимающих значения 3, 7, 11, 15..., вычисляются матрицы Эйлера порядков 6, 14, 30.... Так как каждая матрица Эйлера расширением ее каймой дает матрицу Мерсенна, серия матриц складывается в *основную* цепочку $\mathbf{M}_3 - \mathbf{E}_6 - \mathbf{M}_7 - \mathbf{E}_{14} - \mathbf{M}_{15} - \dots$

Рассмотрим алгоритмы построения квазиортогональных матриц заданных порядков, а также произведем оценку сложности таких процедур. Заметим, что в итерационных алгоритмах вычисления матриц Мерсенна (см. раздел 4 настоящей статьи) нет необходимости вычислять собственные числа и собственные векторы \mathbf{S}_{2n} .

Для уменьшения трудоемкости решения этой задачи, служащее источником определения уровней, ищется аналитически [8]: половину элементов собственного вектора e матрицы \mathbf{S}_{2n} составляют элементы $-b$, остальную — элементы a , что дает возможность быстрого вычисления матриц требуемых базисов в виде:

$$\mathbf{M}_{2n+1} = \begin{pmatrix} 1 & e^T \\ e & \mathbf{S}_{2n} \end{pmatrix}.$$

Для большего понимания материала приведем числовой пример. Итерации вычисления матриц Мерсенна основной последовательности $n = 2^k - 1$ начинаются с матрицы:

$$\mathbf{M}_3 = \begin{pmatrix} a & -b & a \\ -b & a & a \\ a & a & -b \end{pmatrix}, \text{ причем } \mathbf{M}_3^* = \begin{pmatrix} -b & a & -b \\ a & -b & -b \\ -b & -b & a \end{pmatrix},$$

где $a=1$, модуль второго элемента $b = \frac{t}{t+\sqrt{t}}$ принимает значение 0,5 при $t=1$ (порядок $n=4t-1=3$) и растет до $b=2-\sqrt{2}=0,5858\dots$ при $t=1$ (порядок $n=7$). Учитывая выписанное выше выражение, расчетная матрица Мерсенна $\mathbf{M}_7 = \begin{pmatrix} 1 & e^T \\ e & \mathbf{S}_6 \end{pmatrix}$ при $e = (-b, -b, -b, 1, 1, 1)^T$ также

квазиортогональна, итерации можно продолжить до бесконечности.

На рисунке 4 изображены первые три матрицы этого бесконечного семейства $\mathbf{M}_3, \mathbf{M}_7, \mathbf{M}_{15}$.

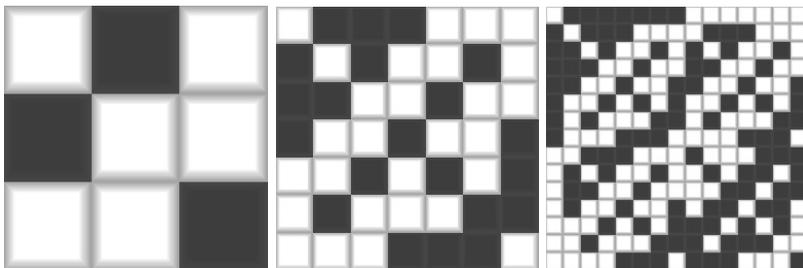


Рис. 4. Квазиортогональные матрицы Мерсенна $\mathbf{M}_3, \mathbf{M}_7, \mathbf{M}_{15}$

Приведенные выше формулы, как видно, несложно компьютеризировать, получая цепочки ортогональных по столбцам матриц Мерсенна *нечетных порядков* $2^k - 1$ — столь же легко, как и цепочки матриц Адамара *четных порядков* 2^k , содержащих элементы 1, -1.

Число уровней матриц Мерсенна ровно такое же, причем с ростом порядка значение модуля уровня второго элемента $b = \frac{t}{t+\sqrt{t}}$ стремится к 1, то есть эти две цепочки матриц сближаются между собой по значениям их элементов.

Алгоритм вычисления матриц Эйлера четных порядков использует алгоритм вычисления матриц Мерсенна, то есть не превосходит его по трудоемкости:

$$\mathbf{E}_{2n} = \begin{pmatrix} \mathbf{M}_n & \mathbf{M}_n \\ \mathbf{M}_n^T & -\mathbf{M}_n^T \end{pmatrix}.$$

В работе [11] показано, что цепочка последовательно вычисляемых матриц может начинаться непосредственно с матрицы Эйлера, если соответствующей ей по порядку матрицы Белевича нет.

Эти матрицы обладают сходными свойствами, вместе они образуют расширенное семейство матриц типа Адамара, и могут использоваться в прикладных алгоритмах, расширяя их возможности.

8. Рекомендации по вычислению цепочек матриц. Дополнительный анализ [6] показывает, что вместо стартовых матриц Мерсена (когда их нет) можно использовать отличающиеся от них диагональными элементами матрицы Зейделя S_n , в таком случае цепочка обретает другой предиктор: $S_5 - E_{10} - M_{11} - E_{22} - M_{23} - \dots$ или $S_9 - E_{18} - M_{19} - E_{38} - M_{39} - \dots$ и т.п. Случай подробно рассмотрен с участием соавтора этой статьи в [11] и не комментируется подробно для полезной краткости. В работе [12] также показывается, что помимо универсальных матриц Эйлера существует сходная с ней частная структура, связанная с *золотым сечением*, а пропуск в последовательности матриц Белевича на порядке 22 [13] восполняется не только матрицей Эйлера E_{22} , но и квазиортогональными матрицами с тремя [14] или шестью значениями элементов [15].

Тем не менее благодаря отмеченным взаимосвязям основное множество квазиортогональных матриц может быть вычислено на основе бициклических матриц Эйлера, когда блоки A, B — циклические.

Для квазиортогональных матриц произведение $A^T B + B^T A$ диагонально по определению. Несложно показать, что округление элементов матрицы Эйлера до 1 и -1 не сказывается на преобладании диагонали, причем в силу сохранения структуры блоков внедиагональные элементы хотя и теряют свое нулевое значение, но не перестают быть равными друг другу. Общее число элементов на диагонали и вне ее пропорционально квадрату размера блока, то есть растет очень быстро, что ограничивает грубый поиск матриц Эйлера тривиальным перебором $n!$ бинарных последовательностей блоков.

Тем не менее для матриц значений порядков до 100 и выше случайный поиск вполне возможен, если принять во внимание два существенных обстоятельства.

Первое из них состоит в том, что если блоки A, B — циклические матрицы, то их произведение и сумма $A^T B + B^T A$ (матричная невязка) тоже циклическая. То есть для них имеет место не квадратичный, а линейный рост сложности задачи. Элементы первой строки матрицы невязок для округленных матриц Эйлера, за исключением первого элемента диагонали, равны -2. Из них можно контролировать половину, поскольку вторая половина строки зависима.

Второе обстоятельство состоит в ассоциативной связи матриц Эйлера с числами. Размер первой матрицы Эйлера равен 6, из двух блоков один может быть симметричным, второй — кососимметричным (с точностью до элементов диагонали). Соответственно, все мат-

рицы размеров $q \times 6$, q — простое нечетное число, сохраняют это качество. Учет симметрии позволяет экономить при генерации случайных последовательностей ресурсы процессора. Изучение семейств матриц Эйлера по признакам их симметрии резко упрощает матричные генераторы. С точки зрения теории сигналов, контролируемая квадратичная невязка называется спектральной плотностью сигнальной последовательности PSD (power spectral density) [16].

На рисунке 5 в качестве примера приведена найденная перебором симметричная матрица Эйлера E_{66} размера $n=11 \times 6=66$, а на рисунке 6 матрица квадратичной невязки половинного размера 33 ее PSD-теста, причем для полного контроля ортогональности хватает вычисления и проверки на значение -2 половины (без диагонального элемента) ее верхней строки.

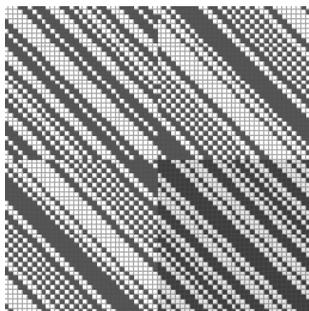


Рис. 5. Портрет матрицы E_{66}

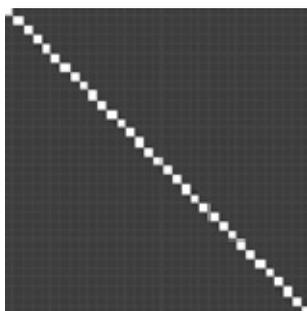


Рис. 6. Портрет PSD-теста для E_{66}

Если генератор случайной последовательности распараллелен и работает быстрее технического устройства, на которое возложена проверка ортогональности (чаще всего это компьютер), можно контролировать не все, а несколько выборочно взятых значений квадратичной невязки. Это составляет основу так называемых PSD-тестов, фильтрующих миллионы ненужных для синтеза матриц Эйлера последовательностей.

9. Заключение. Ассоциация чисел с матрицами позволяет навести порядок среди множества матриц разнообразных базисов и свести проблему их поиска к нахождению бициклических матриц Эйлера. Как показано выше, сложности с их вычислением обходимы учетом специфики их структуры и присущих матрицам Эйлера симметрий. Выделение различных типов симметричных семейств значительно облегчает конструирование бинарных последовательностей для построения этих универсальных матриц и представляет непосредствен-

ный практический интерес для постановки как научных, так и учебных заданий в рамках целевых соревнований среди студентов в системе высшего образования по количеству и качеству находимых матриц (programming contests).

В процессе поиска квазиортогональных матриц четных и нечетных порядков, близких по своим свойствам к матрицам Адамара, удалось выделить классы матриц Мерсенна, Эйлера, Ферма. Матрицы Мерсенна порядков $4t-1$ регламентируют структуру матриц слева и справа от них, включая матрицы Адамара и Эйлера четных порядков $4t$ и $4t-2$ для t равных натуральным числам, а также матриц Ферма нечетных порядков $4t+1$ для t равных квадратам натуральных чисел u^2 (квадратичная ветвь квазиортогональных матриц). При отсутствии пропусков среди матриц Ферма, элементарным наращиванием каймы и пересчетом отрицательных элементов, периодически равных -1 и $-b$, можно было бы получить все квазиортогональные матрицы вдоль всей числовой оси, начиная со стартовой матрицы $\mathbf{H}_1 = 1$.

Пропуски матриц Ферма усложняют эту картину, лишая цепочки надстраиваемых матриц преемственности, что связано с разнообразием ассоциируемой числовой системы. Вместе с тем, появляются специфические матрицы-артефакты, наиболее полно отражающие особенность числа. Например, среди модификаций матриц Эйлера можно отметить, например, матрицу \mathbf{G}_{10} со значениями элементов b , равными золотому сечению $b = \frac{\sqrt{5}-1}{2} \approx 0,618$.

Литература

1. Algebraic Design Theory and Hadamard Matrices / Edited by C.J. Colbourn // Springer Proceedings in Mathematics & Statistics. 2014. vol. 133. 260 p.
2. *Shalom E.* La conjecture de Hadamard (I) – Images des Mathématiques, CNRS. 2012. URL: <http://images.math.cnrs.fr/La-conjecture-de-Hadamard-I.html> (дата обращения: 15.03.2016).
3. *Балонин Н.А., Балонин Ю.Н., Востриков А.А., Сергеев М.Б.* Вычисление матриц Мерсенна-Уолша // Вестник компьютерных и информационных технологий. 2014. № 11. С. 51–55.
4. *Sylvester J.J.* Thoughts on inverse orthogonal matrices, simultaneous sign successions, and tessellated pavements in two or more colours, with applications to Newton's rule, ornamental tile-work, and the theory of numbers // Philosophical Magazine. 1867. vol. 34. no. 232. pp. 461–475.
5. *Hadamard J.* Résolution d'une question relative aux determinants // Bulletin des Sciences Mathématiques. 1893. vol. 17. pp. 240–246.
6. *Balonin N.A., Sergeev M.B.* Quasi-Orthogonal Local Maximum Determinant Matrices // Applied Mathematical Sciences. 2015. vol. 9. no. 6. pp. 285–293.
7. *Belevitch V.* Theorem of 2n-terminal networks with application to conference telephony // Electr. Commun. 1950. vol. 26. pp. 231–244.
8. *Балонин Н.А., Сергеев М.Б., Мироновский Л.А.* Вычисление матриц Адамара-Мерсенна // Информационно-управляющие системы. 2012. № 5. С. 92–94.

9. *Sergeev A. Generalized Mersenne Matrices and Balonin's Conjecture // Automatic Control and Computer Sciences. 2014. vol. 4. pp. 35–43.*
10. *Балонин Н.А., Сергеев М.Б., Мироновский Л.А. Вычисление матриц Адамара-Ферма // Информационно-управляющие системы. 2012. № 6. С. 90–93.*
11. *Balonin N.A., Vostrikov A.A., and Sergeev M.B. On Two Predictors of Calculable Chains of Quasi-Orthogonal Matrices // Automatic Control and Computer Sciences. 2015. vol. 49. no. 3. pp. 153–158.*
12. *Balonin N.A., Sergeev M.B. Quasi-Orthogonal Matrices with Level Based on Ratio of Fibonacci Numbers // Applied Mathematical Sciences. 2015. vol. 9. no. 86. pp. 4261–4268.*
13. *Balonin N.A., Seberry J. A Review and New Symmetric Conference Matrices // Информационно-управляющие системы. 2014. № 4(71). pp. 2–7.*
14. *Балонин Н.А., Сергеев М.Б. Взвешенная конференц-матрица, обобщающая матрицу Белевича на 22-м порядке // Информационно-управляющие системы. 2013. № 5. С. 97–98.*
15. *Балонин Ю.Н., Сергеев М.Б. М-матрица 22-го порядка // Информационно-управляющие системы. 2011. № 5. С. 87–90.*
16. *Djokovic D.Z., Kotsireas I. S. Compression of periodic complementary sequences and applications // Des. Codes Cryptogr. 2015. vol. 74. pp. 365–377.*

Балонин Юрий Николаевич — инженер кафедры вычислительных систем и сетей, Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП). Область научных интересов: вычислительные методы, теория чисел, компьютерное моделирование. Число научных публикаций — 13. yuraball@mail.ru; ул. Большая Морская, 67, Санкт-Петербург, 190000; р.т.: +7(812)355-17-07.

Востриков Антон Александрович — к-т техн. наук, доцент, доцент кафедры вычислительных систем и сетей, Санкт-Петербургский государственный университет аэрокосмического приборостроения (ГУАП). Область научных интересов: обработка видеoinформации, IP-сети, распределенные информационно-управляющие системы, встраиваемые системы управления, системы обеспечения безопасности. Число научных публикаций — 38. vostricov@mail.ru; ул. Большая Морская, 67, Санкт-Петербург, 190000; р.т.: +79219430422.

Сергеев Александр Михайлович — старший преподаватель кафедры вычислительных систем и сетей, Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП). Область научных интересов: обработка изображений, вычислительная математика. Число научных публикаций — 7. mbse@mail.ru; ул. Большая Морская, 67, Санкт-Петербург, 190000; р.т.: +7(812) 494-70-44.

Егорова Ирина Сергеевна — аспирант кафедры вычислительных систем и сетей, Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП). Область научных интересов: специализированные системы обработки информации, проектирование цифровых устройств на базе ПЛИС, автоматизация проектирования вычислительных систем. Число научных публикаций — 4. irina.egorova@ask-lab.com; ул. Большая Морская, 67, Санкт-Петербург, 190000; р.т.: +7(921)090-55-32

Ju.N. BALONIN, A.A. VOSTRIKOV, A.M. SERGEEV, I.S. EGOROVA
**ON RELATIONSHIPS AMONG QUASI-ORTHOGONAL
 MATRICES CONSTRUCTED ON THE KNOWN SEQUENCES OF
 PRIME NUMBERS**

Balonin Ju.N., Vostrikov A.A., Sergeev A.M., Egorova I.S. On Relationships among Quasi-orthogonal Matrices Constructed on the Known Sequences of Prime Numbers.

Abstract. The objective of the paper is to show the relationship among numbers, belonging to known sequences of prime numbers, and quasi-orthogonal matrices, existing for orders equal to these numbers, and the relationship among such matrices based on calculating algorithms. Methods: analysis of the sequences of quasi-orthogonal matrices with absolute and local maximum of its determinant, detecting of the structural invariants in the matrices, matching algorithms for calculating these matrices. Results: known sequences of natural numbers are considered, the definition of a matrix, associated with a natural number, is formulated. Sequences of numbers with proved existence of quasi-orthogonal matrices associated with them are presented. It is suggested that associated matrices exist for all positive natural numbers. Properties of these types of matrices, their relationships based on calculating algorithms are considered. There are modified algorithms and general key chains of Euler and Mersenne matrices presented, a sequence of orders of which are systemically important. Practical value: quasi-orthogonal matrices of absolute and local maximum of determinant have immediate practical value for error-correcting coding tasks, video compression and masking. Their diversity allows developers of technical systems greatly facilitate a matrix selection, optimal one for a particular task.

Keywords: numerical sequences, Mersenne numbers, Fermat numbers, orthogonal sequences, quasiorthogonal matrices, Hadamard matrices, Mersenne matrices, Euler matrices, Fermat Matrices, matrix chains, quasiorthogonal matrix algorithms.

References

1. Algebraic Design Theory and Hadamard Matrices. Edited by C.J. Colbourn. Springer Proceedings in Mathematics & Statistics. 2014. vol. 133. 260 p.
2. Shalom E. La conjecture de Hadamard (I) – *Images des Mathématiques*, CNRS. 2012. <http://images.math.cnrs.fr/La-conjecture-de-Hadamard-I.html> (accessed: 15.03.2016).
3. Balonin N.A., Balonin Yu.N., Vostrikov A.A., Sergeev M.B. [Computation of Mersenne-Walsh Matrices]. *Vestnik komp'yuternykh i informatsionnykh tekhnologii – Herald of computer and information technologies*. 2014. vol. 11. pp. 51–55. (In Russ.).
4. Sylvester J.J. Thoughts on inverse orthogonal matrices, simultaneous sign successions, and tessellated pavements in two or more colours, with applications to Newton's rule, ornamental tile-work, and the theory of numbers. *Philosophical Magazine*. 1867. vol. 34. no. 232. pp. 461–475.
5. Hadamard J. Résolution d'une question relative aux déterminants. *Bulletin des Sciences Mathématiques*. 1893. vol. 17. pp. 240–246.
6. Balonin N.A., Sergeev M.B. Quasi-Orthogonal Local Maximum Determinant Matrices. *Applied Mathematical Sciences*. 2015. vol. 9. no. 6. pp. 285–293.
7. Belevitch, V. Theorem of 2n-terminal networks with application to conference telephony, *Electr. Commun.* 1950. vol. 26. pp. 231–244.
8. Balonin N.A., Sergeev M.B., Mironovsky L.A. [Calculation of Hadamard-Mersenne Matrices]. *Informatsionno-upravliaiushchie sistemy – Information and Control Systems*. 2012. vol. 5(60). pp. 92–94 (In Russ.).

9. Sergeev A. Generalized Mersenne Matrices and Balonin's Conjecture, *Automatic Control and Computer Sciences*. 2014. no. 4. pp. 35–43.
10. Balonin N.A., Sergeev M.B. Mironovsky L.A. [Calculation of Hadamard-Fermat Matrices]. *Informatsionno-upravliaiushchie sistemy – Information and Control Systems*. 2012. vol. 6(61). pp. 90–93 (In Russ.).
11. Balonin N.A., Vostrikov A.A., Sergeev M.B. On Two Predictors of Calculable Chains of Quasi-Orthogonal Matrices. *Automatic Control and Computer Sciences*. 2015. vol. 49. no. 3. pp. 153–158.
12. Balonin N.A., Sergeev M.B. Quasi-Orthogonal Matrices with Level Based on Ratio of Fibonacci Numbers. *Applied Mathematical Sciences*. 2015. vol. 9. no. 86. pp. 4261–4268.
13. Balonin N.A., Seberry, Jennifer. A Review and New Symmetric Conference Matrices. *Informatsionno-upravliaiushchie sistemy – Information and Control Systems*. 2014. no. 4 (71). pp. 2–7. (In English).
14. Balonin N.A., Sergeev M.B. [Weighted Conference Matrix Generalizing Belevich Matrix at the 22nd Order]. *Informatsionno-upravliaiushchie sistemy – Information and Control Systems*. 2013. vol. 5(66). pp. 97–98 (In Russ.).
15. Balonin Yu.N., Sergeev M.B. [M-matrix of 22nd order]. *Informatsionno-upravliaiushchie sistemy – Information and Control Systems*. 2011. vol. 5(54). pp. 87–90 (In Russ.).
16. Djokovic D.Z., Kotsireas I. S. Compression of periodic complementary sequences and applications. *Des. Codes Cryptogr*. 2015. vol. 74. pp. 365–377.

Balonin Yuriy Nikolaevich — engineer of computer systems and networks department, Saint Petersburg State University of Aerospace Instrumentation (SUAI). Research interests: numerical methods, number theory, computer simulation. The number of publications — 13. yuraball@mail.ru; 67, B. Morskaiia St., 190000, Saint-Petersburg; office phone: +7(812)355-17-07.

Vostrikov Anton Aleksandrovich — Ph.D., associate professor, associate professor of computer systems and networks department, Saint Petersburg State University of Aerospace Instrumentation (SUAI). Research interests: videoimage compression, IP-nets, distributed information and control, board control systems, safety systems. The number of publications — 38. vostricov@mail.ru; 67, B. Morskaiia St., 190000, Saint-Petersburg; office phone: +79219430422.

Sergeev Alexander Mikhailovich — senior lecturer of computer systems and networks department, Saint Petersburg State University of Aerospace Instrumentation (SUAI). Research interests: videoimage compression, numerical methods. The number of publications — 7. mbse@mail.ru; 67, B. Morskaiia St., 190000, Saint-Petersburg; office phone: +7(812) 494-70-44.

Egorova Irina Sergeevna — Ph.D. student of computer systems and networks department, Saint Petersburg State University of Aerospace Instrumentation (SUAI). Research interests: special information systems, PLIS digital system design, numerical system automatic design. The number of publications — 4. irina.egorova@ask-lab.com; 67, B. Morskaiia St., 190000, Saint-Petersburg; office phone: +7(921)090-55-32.

Signed to print 01.02.2017

Printed in Publishing center GUAP, 67, B. Morskaya, St. Petersburg, 190000, Russia

The journal is registered in Russian Federal Agency for Communications
and Mass-Media Supervision, certificate ПИ № ФС77-41695 dated August 19, 2010 г.
Subscription Index 29393, Russian Post Catalog

Подписано к печати 01.02.2017. Формат 60×90 1/16. Усл. печ. л. 15,0. Заказ № 15.

Тираж 150 экз., цена свободная

Отпечатано в Редакционно-издательском центре ГУАП, 190000, Санкт-Петербург, Б. Морская, д. 67

Журнал зарегистрирован Федеральной службой по надзору в сфере связи
и массовых коммуникаций,
свидетельство ПИ № ФС77-41695 от 19 августа 2010 г.

Подписной индекс 29393 по каталогу «Почта России»

РУКОВОДСТВО ДЛЯ АВТОРОВ

Взаимодействие автора с редакцией осуществляется через личный кабинет на сайте журнала «Труды СПИИРАН» <http://www.proceedings.spiiras.nw.ru>. При регистрации авторам рекомендуется заполнить все предложенные поля данных.

Подготовка статьи ведется с помощью текстовых редакторов MS Word 2007 и выше. Объем основного текста – от 15 до 25 страниц включительно. Формат страницы документа – А5 (148 мм ширина, 210 мм высота); ориентация – портретная; все поля – 20 мм. Верхний и нижний колонтитулы страницы – пустые. Основной шрифт документа – Times New Roman, основной кегль (размер) шрифта – 10 pt. Переносы разрешены. Абзацный отступ устанавливается размером в 10 мм. Межстрочный интервал – одинарный. Номера страниц не проставляются.

В основную часть допускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно рассмотрены на нашем сайте в разделе «Руководство для авторов».

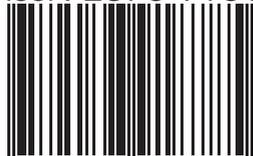
AUTHOR GUIDELINES

Interaction between each potential author and the Editorial board is realized through the personal account on the website of the journal "Proceedings of SPIIRAS" <http://www.proceedings.spiiras.nw.ru>. At the registration the authors are requested to fill out all data fields in the proposed form.

The submissions should be prepared using MS Word 2007 text editor or higher versions, at that, only manuscripts in *.docx format will be considered. The text of the paper in the main part of it should be from 15 – 25 pages of A5 size that is 210 X 148 mm; orientation – portrait; all margins – 20 mm. The font of the main paper text is Times New Roman of 10 pt font size. The pages' headers and footers should be empty; indentation – 10 mm; line spacing – single; pages are not numbered; hyphenations are allowed.

Certain figures, tables, listings and formulas are allowed in the main section, and their typography is considered by the paper template in more detail in journal web.

ISSN 2078-9181



9 772078 918785 >

