

ОБЗОРНАЯ СТАТЬЯ

Учёт многочастичных взаимодействий в молекулярных J-агрегатах и нелинейные оптические эффекты в этих системах

ФОТОНИКА И ОПТОИНФОРМАТИКА

ГКР спектроскопия молекул бактериородопсина, адсорбированных на серебряные наноструктурные пленки

ОПТИЧЕСКИЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Аблиция твердого тела под действием сверхкоротких лазерных импульсов: исследование методами молекулярной динамики

АВТОМАТИЧЕСКОЕ УПРАВЛЕНИЕ И РОБОТОТЕХНИКА

Фундаментальная матрица линейной непрерывной системы в задаче оценки ее транспортного запаздывания

НОВЫЕ МАТЕРИАЛЫ И НАНОТЕХНОЛОГИИ

Тепловые и электрические поля при искровом плазменном спекании термоэлектрических материалов
Влияние одностадийной обработки кварцевой керамики гелеобразующими водными растворами на прочностные характеристики

Исследование сорбционных характеристик полимерных минерал-наполненных композитов для медицины

Кинетика кристаллизации полимерных нанокомпозитов на основе полиамида 12, модифицированного наночастицами Cr_2O_3

Неорганические люминофоры в стекле на основе свинцово-силикатных стекол

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Улучшение качества распознавания в сетях глубокого обучения с помощью метода имитации отжига

Экспериментальные исследования программно-аппаратных средств ввода и выделения речевого сигнала пользователя систем голосовой аутентификации

Создание систем поддержки совместных научных исследований в науках о земле в среде COLLA на основе метода куста событий

Двуязычная многомодальная система для аудиовизуального синтеза речи и жестового языка по тексту

Совершенствование Police Office Model для обеспечения безопасности роевых робототехнических систем

Использование реляционной теории при оптимальном проектировании интегральных схем

Оптимизация интервалов проверки информационной безопасности систем

Алгоритм семантического анализа текста, основанный на базовых семантических шаблонах с удалением

Архитектура веб-ориентированной автоматизированной системы технологической подготовки производства

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

Моделирование импульсного пробоя в гелии с использованием адаптивных методов

О построении алгоритмов работы бескарданного гирогоризонткомаса на электростатическом гироскопе

Моделирование и экспериментальное исследование чувствительного элемента волоконно-оптического гидрофона

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ

К вопросу повышения результативности массового онлайн-курса

Информационная инфраструктура образовательной среды с использованием технологии виртуальных машин

КРАТКИЕ СООБЩЕНИЯ

Двухслойные фазо-компенсирующие интерференционные системы

Прототип профессиональной социальной сети VareNtsnet для информационной поддержки управления развитием Баренцева Евро-Арктического региона

Веретенев Н.А., Левинский Б.Н., Нестеров Л.А., Розанов Н.Н., Файнберг Б.Д., Фёдоров С.В. 1

Хайслер Ф., Пилогина Е.С., Червинский С.Д., Самусев А.К., Липовский А.А. 18

Иванов Д.С., Вейко В.П., Яковлев Е.Б., Ретфельд Б., Гарсия М.Э. 23

Дударенко Н.А., Полинова Н.А., Ушаков А.В. 32

Булат Л.П., Пшеница-Северин Д.А., 38

Нефедова И.А., Новотельнова А.В., Гуревич Ю.Г.

Евстропьев С.К., Волькин В.М., Шашкин А.В., 46

Дукельский К.В., Коробейников А.Г.,

Гатчин Ю.А., Поляков В.И.

Игнатъева Ю.А., Успенская М.В., Борисов О.В., 52

Олехнович Р.О., Евсеев Р.А., Касанов К.Н.

Шаповал Е.С., Зуев В.В. 57

Асеев В.А., Бирик А.Ю., Колобкова Е.В., 64

Некрасова Я.А., Никоноров Н.В., Романов А.Е.,

Тузова Ю.В., Швалева М.А.

Потапов А.С., Батищева В.В., Пан Ш. 71

Файзулаева О.Н., Невлюдов И.Ш. 77

Дивьякко П., Пшеничный К.А., Бенке Б., 83

Канжелева О.М.

Карпов А.А., Железны М. 92

Зикратов И.А., Гуртов А.В., Зикратова Т.В., 99

Козлова Е.В.

Демидов Д.В. 110

Богатырев В.А., Богатырев А.В., Богатырев С.В. 119

Мочалова А.В. 126

Филоков Н.Е. 133

Елисеев С.И., Демидов В.И., Чирцов А.С., 139

Колобов В.И., Кудрявцев А.А., Богданов Е.А.

Емельянцева Г.И., Медведков А.А., Цай Тицзин 147

Ефимов М.Е., Плотников М.Ю., Куликов А.В. 158

Лисицына Л.С., Перишин А.А., Усков В.Л. 164

Береснев А.Д., Гусарова Н.Ф., Иванов Р.В., 172

Федотов Д.Н.

Никандров Г.В., Путилин Э.С., Губанова Л.А., 183

Стародубов Д.

Маслобоев А.В., Странд Р.М. 186

Главный редактор – В.О. Никифоров¹, д.т.н., профессор

МЕЖДУНАРОДНЫЙ РЕДАКЦИОННЫЙ СОВЕТ

Elias Aifantis, PhD, Professor, Aristotle University of Thessaloniki, Thessaloniki, Greece;
Emeritus Professor, Michigan Technological University, Michigan, USA
Denis Efimov, HDR, charge de recherche 1ere grade, Institut national de recherche en informatique
et en automatique, Villeneuve-d'Ascq, France
Yurii Gun'ko, PhD, Professor, School of Chemistry, Trinity College Dublin, Dublin, Ireland
Yuri Kivshar, FAA, Distinguished Professor, Australian National University, Canberra, Australia
Ilya Kolmanovsky, PhD, Full Professor, University of Michigan, Ann Arbor, USA
Konstantin R. Simovski, Dr. Sci. Full Professor, Aalto University, Aalto, Finland
Peter Sloot, PhD, Professor, University of Amsterdam, Amsterdam, Netherlands
Mart Tamre, Doctor's Degree, Professor, Tallinn University of Technology, Tallinn, Estonia
Tatiana Timofeeva, PhD, Associate Professor, New Mexico Highlands University, Las Vegas, USA

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

А.А. Бобцов ¹ , д.т.н., профессор	В.Г. Мельников ¹ , д.т.н., доцент
А.В. Бухановский ¹ , д.т.н.	Н.В. Никоноров ¹ , д.т.н., профессор
Ю.А. Гатчин ¹ , д.т.н., профессор	А.А. Ожиганов ¹ , д.т.н., профессор
Н.Ф. Гусарова ¹ , к.т.н., с.н.с. (редактор)	Н.Н. Розанов ³ , д.ф.-м.н., профессор, чл.-корр. РАН
Н.С. Кармановский ¹ , к.т.н., доцент (заместитель главного редактора)	А.Е. Романов ^{1,7} , д.ф.-м.н., профессор, вед.н.с.
Ю.Л. Колесников ¹ , д. ф.-м.н., профессор	И.Г. Сидоркина ⁴ , д.т.н., профессор
С.А. Козлов ¹ , д.ф.-м.н., профессор	О.А. Степанов ⁶ , д.т.н., профессор
А.Г. Коробейников ² , д.т.н., профессор	А.А. Шальто ¹ , д.т.н., профессор
Л.С. Лисицына ¹ , д.т.н., профессор	Ю.Г. Якушенков ⁵ , д.т.н., профессор

Ответственный секретарь – Г.О. Артемова¹, к.т.н.

¹ Университет ИТМО, Санкт-Петербург, Россия

² Санкт-Петербургский филиал Федерального государственного бюджетного учреждения науки Института
земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова Российской академии наук
(СПбФ ИЗМИРАН), Санкт-Петербург, Россия

³ ОАО «Государственный оптический институт (ГОИ) им. С.И. Вавилова», Санкт-Петербург, Россия

⁴ Поволжский государственный технологический университет, Йошкар-Ола, Россия

⁵ Московский государственный университет геодезии и картографии (МИИГАиК), Москва, Россия

⁶ ОАО «Концерн «ЦНИИ «Электроприбор», Санкт-Петербург, Россия

⁷ Физико-технический институт имени А.Ф. Иоффе, Санкт-Петербург, Россия

Журнал издается Университетом ИТМО с 2001 года как периодическое научное и научно-образовательное издание.

Прежние названия журнала: «Научно-технический вестник Санкт-Петербургского государственного института точной механики и оптики (технического университета)» – до 11 выпуска 2003 года; «Научно-технический вестник Санкт-Петербургского государственного университета информационных технологий, механики и оптики» – с 12 выпуска 2004 года по 76 выпуск 2011 года.

Издание зарегистрировано в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (свидетельство ПИ ФС77-47243 от 10 ноября 2011 года).

ISSN 2226-1494

Английское название: «Scientific and Technical Journal of Information Technologies, Mechanics and Optics».

Транслитерированное название: «Nauchno-Tekhnicheskii Vestnik Informatsionnykh Tekhnologii, Mekhaniki i Optiki».

Журнал включен в каталог периодических изданий Ulrich's Periodical Directory.

Журнал входит в утвержденный Высшей Аттестационной комиссией «Перечень ведущих рецензируемых научных журналов и изданий, выпускаемых в Российской Федерации, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата наук».

Периодичность издания – 6 выпусков в год.

Плата за публикации не взимается.

Адрес: 197101, Россия, Санкт-Петербург, Кронверкский пр., д. 49, Университет ИТМО, оф. 330
Телефон/факс (812) 233 45 51 <http://ntv.ifmo.ru> e-mail: karmanov@mail.ifmo.ru

Подписано к печати 01.09.2014 Тираж 350 экз. Заказ № 41(93)

Отпечатано в учреждении «Университетские телекоммуникации»
Адрес: 197101, Санкт-Петербург, Кронверкский пр., 49

ПРАВИЛА ПРЕДСТАВЛЕНИЯ РУКОПИСЕЙ

ОБЩИЕ ПОЛОЖЕНИЯ

Редакция принимает к рассмотрению оригинальные (нигде ранее не опубликованные, за исключением материалов научных конференций) работы по тематике журнала, представляющие новые научные результаты, полученные лично авторами статьи. Публикуются также обзоры, характеризующие современное состояние актуальных направлений развития науки и техники, сообщения о научных конференциях, материалы научных дискуссий и рецензии на новые книги.

Тематика журнала включает разделы, отражающие достижения науки, техники и технологии в области современного приборостроения по направлениям: оптика, оптические системы и технологии, оптическое материаловедение и нанотехнологии, информационные и компьютерные системы и технологии, методы и системы защиты информации, системный анализ, системы автоматического управления, мехатроника, системы автоматизированного проектирования, микроэлектроника, биомедицинские технологии. В качестве дополнительных разделов журнал также публикует работы по экономике и финансам.

Материал статьи должен быть изложен в соответствии со следующей структурой.

В аннотации, рассчитанной на самый широкий круг читателей, необходимо кратко (объемом до 100 слов) изложить научное содержание статьи.

Во введении необходимо представить содержательную постановку рассматриваемого вопроса, краткий анализ известных из научной литературы решений, критику их недостатков и преимущества (особенности) предлагаемого подхода.

В основном тексте статьи должна быть представлена строгая постановка решаемой задачи, изложены и обстоятельно разъяснены (доказаны) полученные утверждения и выводы, приведены результаты экспериментальных исследований или математического моделирования, иллюстрирующие сделанные утверждения. Основной текст статьи может быть разбит на отдельные разделы: «Постановка задачи», «Основной результат», «Результаты моделирования» и т.п.

В заключении необходимо кратко сформулировать основные результаты, прокомментировать их и, если возможно, указать направления дальнейших исследований.

Все статьи проходят процедуру рецензирования. После ее завершения автору направляется экземпляр рукописи с замечаниями рецензента и научного редактора для доработки. При опубликовании статьи указываются даты ее поступления и последующих переработок.

ОСНОВНЫЕ ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ

Объем статьи, включая иллюстрации, таблицы и список литературы, не должен превышать 6 страниц.

Оригинал-макет статьи выполняется в редакторе Word for Windows. Шрифт Times New Roman, размер шрифта 12 pt, межстрочный интервал одинарный, поля сверху 20 мм, остальные поля 25 мм. Абзацный отступ 10 мм.

Аннотация печатается шрифтом Times New Roman 10 обычный, межстрочный интервал одинарный, выравнивается по ширине страницы. В аннотации не допускаются ссылки на цитируемую литературу и громоздкие формулы.

Название раздела печатается шрифтом Times New Roman 12 полужирный, межстрочный интервал одинарный, отделяется от предыдущего раздела полutorной пустой строкой, а от последующего текста одной пустой строкой.

Название подраздела печатается шрифтом Times New Roman 12 полужирный, межстрочный интервал одинарный, пустыми строками не отделяется.

Формулы набираются в редакторе Microsoft Equation Ed. 3.0. Параметры стиля: текста — Times New Roman, прямой; функции — Times New Roman, прямой; переменной — Times New Roman, наклонный; греческих букв — Symbol, прямой; символов — Symbol, прямой; матрицы-вектора — Times New Roman, прямой, полужирный; чисел — Times New Roman, прямой. Размеры: обычный — 12 pt, крупный индекс — 10 pt, мелкий индекс — 8 pt, крупный символ — 16 pt, мелкий символ — 12 pt. Отдельные переменные в тексте допускается набирать в текстовом режиме с соблюдением требований, указанных для формул. Разделительный знак десятичных дробей — запятая.

При оформлении рисунков и графиков используется шрифт Times New Roman. Стиль написания переменных должен соответствовать требованиям к основному тексту.

Используются только единицы измерения, соответствующие системе СИ.

Список использованной литературы оформляется в соответствии с ГОСТ 7.0.5-2008. Нумерация позиций — согласно очередности ссылок в тексте.

Плата с аспирантов за публикацию рукописей не взимается.

Более подробную информацию смотрите на сайте <http://ntv.ifmo.ru>

SCIENTIFIC AND TECHNICAL JOURNAL

OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS

SEPTEMBER-OCTOBER 2014

NUMBER 5 (93)

ISSN 2226-1494

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ



REVIEW ARTICLE

Accounting of many-particle interactions in molecular J-aggregates and nonlinear optical effects in these systems

N.A. Veretenov, B.N. Levinsky, L.A. Nesterov, N.N. Rosanov, B.D. Fainberg, S.V. Fedorov 1

PHOTONICS AND OPTOINFORMATICS

SERS of bacteriorhodopsin with out-diffused silver nanoislands

F. Heisler, E.S. Piliugina, S.D. Chervinskii, A.K. Samusev, A.A. Lipovskii 18

OPTICAL ENGINEERING

Solid body ablation under exposure to ultra short laser pulses: study by molecular dynamics methods

D.S. Ivanov, V.P. Veiko, E. B. Yakovlev, B. Rethfeld, M.E. Garcia 23

AUTOMATIC CONTROL AND ROBOTICS

Fundamental matrix of linear continuous system in the problem of estimating its transport delay

N.A. Dudarenko, N.A. Polinova, A.V. Ushakov 32

MATERIAL SCIENCE AND NANOTECHNOLOGIES

Thermal and electric fields at spark plasma sintering of thermoelectric materials

L.P. Bulat, D.A. Pshenai-Severin, I.A. Nefedova, A.V. Novotelnova, Yu.G. Gurevich 38

Influence of quartz ceramics single-stage processing by gel-forming water solutions on its strength characteristics

S.K. Evstropiev, V.M. Volynkin, A.V. Shashkin, K.B. Dukel'sky, A.G. Korobeynikov, Yu.A. Gatchin, V.I. Polyakov 46

Investigation of sorption characteristics of polymeric mineral-filled composites for medicine

Yu. A. Ignatieva, M.V. Uspenskaya, O.V. Borisov, R.O. Olekhnovich, R.A. Evseev, K.N. Kasanov 52

Crystallization kinetics of polymeric nanocomposites based on polyamide12 modified by Cr₂O₃ nanoparticles

E.S. Shapoval, V.V. Zuev 57

Inorganic phosphors in glass based on lead silicate glasses

V.A. Aseev, A.Yu. Bibik, E.V. Kolobkova, Ya. A. Nekrasova, N.V. Nikonorov, A.E. Romanov, Ju.V. Tuzova, M.A. Shvaleva 64

COMPUTER SCIENCE

Improvement of recognition quality in deep learning networks by simulated annealing method

A.S. Potapov, V.V. Batishcheva, Sh. Pang 71

Experimental study of firmware for input and extraction of user's voice signal in voice authentication systems

O.N. Faizulaieva, I. Sh. Nevlyudov 77

Systems for support of collaborative studies in the COLLA environment based on the event bush method

P. Diviacco, C. A. Pshenichny, B. Behncke, O.M. Kanzheleva 83

Bilingual multimodal system for text-to-audiovisual speech and sign language synthesis

A.A. Karpov, M. Zelezny 92

Police Office Model improvement for security of swarm robotic systems

I.A. Zikratov, A.V. Gurtov, T.V. Zikratova, E.V. Kozlova 99

Relational theory application for optimal design of integrated circuits

D.V. Demidov 110

Intervals optimization of systems information security inspection

V.A. Bogatyrev, A.V. Bogatyrev, S.V. Bogatyrev 119

Algorithm for semantic text analysis by means of basic semantic templates with deletion

A.V. Mochalova 126

Architecture of web based computer-aided manufacturing system

N.E. Filyukov 133

MODELING AND SIMULATION

Simulation of pulsed breakdown in helium by adaptive methods

S.I. Eliseev, V.I. Demidov, A.S. Chirtsov, A.A. Kudryavtsev, V.I. Kolobov, E.A. Bogdanov 139

On algorithms creation for strapdown stabilized gyrocompass operation based on electrically suspended gyroscope

G.I. Emelyantsev, A.A. Medvedkov, C. Tijing 147

Modeling and experimental study of a fiber optic hydrophone sensing element

M.E. Efimov, M.Yu. Plotnikov, A.V. Kulikov 158

INFORMATION TECHNOLOGIES IN EDUCATION

New approaches to efficiency of massive online course

L.S. Lisitsyna, A.A. Pershin, V.L. Uskov 164

Information infrastructure of the educational environment with virtual machine technology

A.D. Beresnev, N.F. Gusarova, R.V. Ivanov, D.N. Fedotov 172

BRIEF PAPERS

Two-layer phase compensating interference systems

G.N. Nikandrov, E.S. Putilin, L.A. Gubanova, D. Starodubov 183

A prototype of BarentsNet professional social network for information support of development management for Barents Euro-Arctic region

A.V. Masloboev, R.M. Strand 186

ОБЗОРНАЯ СТАТЬЯ REVIEW ARTICLE

УДК 535.2

УЧЁТ МНОГОЧАСТИЧНЫХ ВЗАИМОДЕЙСТВИЙ В МОЛЕКУЛЯРНЫХ *J*-АГРЕГАТАХ И НЕЛИНЕЙНЫЕ ОПТИЧЕСКИЕ ЭФФЕКТЫ В ЭТИХ СИСТЕМАХ

Н.А. Веретенев^{a,b}, Б.Н. Левинский^c, Л.А. Нестеров^{a,b}, Н.Н. Розанов^{a,b,d}, Б.Д. Файнберг^{c,e},
С.В. Фёдоров^{a,b}

^a Государственный оптический институт им. С.И. Вавилова, 199034, Санкт-Петербург, Россия, torrek@gmail.com

^b Национальный исследовательский университет ИТМО, 197101, Санкт-Петербург, Россия

^c Holon Institute of Technology, 58102, Holon, Israel

^d Физико-технический институт им. А.Ф. Иоффе, 194021, Санкт-Петербург, Россия

^e School of chemistry, Tel Aviv University, 69978, Tel Aviv, Israel

Аннотация. В работе обобщаются материалы исследований, проведенных авторами за последние годы, и анализируются полученные результаты. Предметом статьи является учет многочастичных взаимодействий в резонансно-возбуждаемых лазерным излучением *J*-агрегатах. При таком рассмотрении принимаются во внимание не только парные взаимодействия, но и взаимодействия данной частицы с тремя и большим числом частиц одновременно. Среди проведенных исследований можно выделить три основных направления.

В рамках первого направления решалась задача вывода из первых принципов системы уравнений движения для молекул *J*-агрегатов с учетом многочастичных взаимодействий, а также парных корреляций между частицами. Вывод уравнений из первых принципов приводит в общем случае к системе взаимозацепляющихся уравнений для средних от произведения *n* операторов, относящихся к *n* разным молекулам системы. Поскольку от уравнения к уравнению *n* возрастает, то возникают проблемы, связанные с расщеплением этой системы, а также факторизацией средних высших порядков. Отдельную и наиболее сложную задачу при этом представляет корректное вычисление релаксационных членов, возникающих при учете механизма экситон-экситонной аннигиляции. Первое направление конкретно связано с рассмотрением и решением всех вышеперечисленных задач.

В рамках второго направления на основе выведенных уравнений проводилось исследование бистабильности в рассматриваемых системах при конкретном учете трехчастичных взаимодействий. При этом основное внимание было уделено анализу однородных режимов в *J*-агрегатах. В частности, показано, что учет многочастичных вкладов сдвигает границу существования бистабильности в область меньших констант экситон-экситонной аннигиляции.

Наконец, третье направление исследований связано с анализом модуляционной неустойчивости для стационарных состояний *J*-агрегатов, рассмотренных и изученных при исследовании бистабильности в рамках второго направления. Проведено изучение границ областей устойчивости и неустойчивости и их сопоставление с границами существования бистабильности. Полученные результаты позволяют более надежно выделить область параметров, где могут наблюдаться существенно нелинейные эффекты, которые можно было бы использовать для создания схем молекулярной памяти и, в более широком контексте, для создания устройств оптической логики.

Ключевые слова: *J*-агрегаты, экситон-экситонная аннигиляция, диполь-дипольное взаимодействие, многочастичные взаимодействия, бистабильность, диссипативные солитоны, модуляционная неустойчивость.

Благодарности. Представленные исследования поддержаны грантом РФФИ 13-02-00527-а, грантом Университета ИТМО 414656, а также при финансовой поддержке ведущих университетов Российской Федерации (субсидия 074-U01).

ACCOUNTING OF MANY-PARTICLE INTERACTIONS IN MOLECULAR *J*-AGGREGATES AND NONLINEAR OPTICAL EFFECTS IN THESE SYSTEMS

N.A. Veretenov^{a,b}, B.N. Levinsky^c, L.A. Nesterov^{a,b},
N.N. Rosanov^{a,b,d}, B.D. Fainberg^{c,e}, S.V. Fedorov^{a,b}

^a Vavilov State Optical Institute, Institute for Laser Physics, 199034, St. Petersburg, Russia, torrek@gmail.com

^b University ITMO, 197101, St. Petersburg, Russia

^c Holon Institute of Technology, 58102 Holon, Israel

^d Ioffe Physical Technical Institute, 194021, St. Petersburg, Russia

^e School of chemistry, Tel Aviv University, 69978, Tel Aviv, Israel

Abstract. The paper deals with generalization of investigation materials done by the authors in recent years and analysis of obtained results. The subject of the paper is accounting of many-particle interactions in molecular *J*-aggregates at their resonance excitation by laser radiation. In this case, not only twin interactions are taken into considerations, but also the interactions of a given particle with three and more particles simultaneously. Three basic directions can be denoted among carried out investigations. The first direction is connected with derivation (from the first principles) of motion equations for molecular of *J*-aggregates in view of many-particle interactions, and also twin correlations between particles. The derivation of equations from the first principles leads in general to the system of coupled equations for the means of products of *n* operators relating to *n* different molecules. Since *n* increases in every following equation, the problems arise, connected with uncoupling of this system and also factorization of the means with the highest *n*. The most difficult and complicated problem in this process is correct calculation of relaxed terms, arising due to exciton-exciton annihilation. The first direction is connected concretely with solution of all above mentioned problems. Within the second direction the study of bistability has

been carried out on the basis of obtained equations, in view of three-particle interactions. Meanwhile primary attention has been concentrated on analysis of homogeneous regimes in J -aggregates. It has been shown, in particular, that accounting of many-particle contributions leads to the shift of bistability boundary into region of smaller constants of exciton-exciton annihilation. And, at last, the third direction of investigations is connected with analysis of modulational instability for stationary states of J -aggregates considered earlier at bistability study. The study of stability region boundaries and their correlation with boundaries of bistability regions has been carried out. Obtained results give the possibility to determine more rigorously the parameters region, where nonlinear effects can be observed. They can be used for development of optical logic devices and, in particular, molecular memory circuits.

Keywords: J -aggregates, exciton-exciton annihilation, dipole-dipole interaction, many-particle interactions, bistability, dissipative solitons, modulational instability.

Acknowledgements. These researches are supported by the Russian Foundation for Basic Research, grant 13-02-00527-a, ITMO University grant 414656, and partially financially supported by the Government of the Russian Federation (grant 074-U01).

Введение

Ориентированные J -агрегаты цианиновых красителей обладают коллективным (экситонным) механизмом их резонансного возбуждения, что проявляется в крайне неординарном поведении их фотоотклика (см. [1–6], а также обзоры [7, 8]). Одна из наиболее важных особенностей этих систем – это гигантские значения нелинейных восприимчивостей, что, в свою очередь, при субпикосекундных временах релаксации делает эти наноструктуры весьма перспективными для разнообразных приложений [9, 10]. Значительный интерес представляет перспектива использования J -агрегатов в схемах молекулярной памяти. Необходимая для запоминания информации бистабильность в одиночных J -агрегатах была предсказана и изучена в работах [11, 12], а в ансамбле агрегатов (в тонкой пленке) – в [13–15]. В [16] для J -агрегатов были предсказаны «наносолитоны» – диссипативные молекулярные солитоны нанометровых размеров. Дальнейший анализ этих структур содержится в работах [17–19], результаты суммированы в монографии [20]. Все перечисленные работы были выполнены в приближении одночастичной матрицы плотности без учета трехчастичных и более высокого порядка взаимодействий, а также корреляций между молекулами. Следующий принципиальный шаг – учет многочастичных взаимодействий молекул в цепочке и двухчастичных корреляций между ними – выполнен в работе [21], в которой построена последовательная теория резонансного возбуждения молекулярных цепочек когерентным поддерживающим излучением при учете указанных факторов. В работе [22] было проанализировано влияние эффекта трехчастичных взаимодействий на характеристики стационарных однородных режимов идеальной бесконечной цепочки и условия бистабильности для этих режимов. Наконец, в работе [23] проведено детальное исследование модуляционной неустойчивости стационарных состояний однородных режимов, рассмотренных в [22] при учете многочастичных взаимодействий. Это исследование позволило определить границы областей устойчивости и неустойчивости и соотнести их с границами существования бистабильности.

Настоящая работа существенно использует результаты работ [21–23], в которых проводился учет вкладов от многочастичных взаимодействий, и состоит из трех разделов. Первый раздел обзора посвящен выводу из первых принципов уравнений движения для молекул цепочек при последовательном учете многочастичных вкладов, связанных с экситон-экситонной аннигиляцией и диполь-дипольным взаимодействием. В следующем разделе проанализировано влияние трехчастичных взаимодействий на бистабильность стационарных состояний однородных режимов в J -агрегатах. В заключительном разделе рассматривается модуляционная неустойчивость стационарных состояний, исследованных в предыдущем разделе.

Вывод уравнений движения для резонансно возбуждаемых J -агрегатов с учетом многочастичных взаимодействий и парных корреляций между молекулами

Модель J -агрегата и аппарат для описания системы молекул. Отдельный J -агрегат моделируется в виде цепочки, состоящей из трехуровневых молекул, первый и второй уровни которых связаны с аналогичными уровнями других молекул посредством диполь-дипольного взаимодействия, а взаимодействие с третьим уровнем осуществляется с помощью механизма экситон-экситонной аннигиляции. При этом внешнее монохроматическое поддерживающее излучение может напрямую взаимодействовать лишь с переходом $1 \rightarrow 2$. Считается также, что частота перехода $2 \rightarrow 3$ близка к частоте перехода $1 \rightarrow 2$. В отсутствие радиационных переходов с первого и второго уровней на третий накачка этого уровня осуществляется за счет механизма непосредственной передачи возбуждения от одной молекулы другой, находящейся в соседнем узле цепочки. В этом процессе одна из молекул, находящаяся на втором возбужденном уровне, взаимодействуя с другой молекулой в аналогичном состоянии, отдает ей свою энергию, переходя в основное состояние, в то время как вторая молекула переходит на третий уровень. Предполагается, что третий уровень по своей природе является электронно-колебательным и крайне быстро распадается с последующей передачей энергии на второй и первый уровни.

Все описанные выше процессы можно рассмотреть также в рамках теории экситонов. Это рассмотрение показывает, что в цепочке молекул существуют одноэкситонные состояния с энергиями, близкими ко второму возбужденному уровню, а также двухэкситонные состояния с энергиями, близкими к

энергии третьего уровня. В результате механизм передачи возбуждения между соседними молекулами приводит к возбуждению двухэкситонного состояния, которое, взаимодействуя с близким по энергии электронно-колебательным уровнем молекулы, очень быстро аннигилирует. Было обнаружено, что с ростом интенсивности накачки роль экситон-экситонной аннигиляции возрастает, и она начинает существенно влиять на оптический отклик экситонной системы [11, 24–34]. Наиболее часто для анализа J -агрегатов используется подход, при котором цепочка молекул описывается уравнениями типа уравнений Блоха для одночастичных матриц плотности. При этом взаимодействие между молекулами выводится на основе классического выражения для запаздывающего взаимодействия между системой диполей, которыми моделируются молекулы. Кроме этого, в систему обычно феноменологически вводится также упомянутое выше взаимодействие, приводящее к экситон-экситонной аннигиляции. В получающейся таким образом системе уравнений, как правило, учитываются только двухчастичные взаимодействия, которые представлены в факторизованной форме, т.е. без учета корреляций между молекулами.

Однако, как будет показано далее, систему уравнений для J -агрегатов можно вывести также из первых принципов. В этом случае возникает иерархия взаимосцепляющихся уравнений для средних от произведений операторов, относящихся к разным молекулам цепочки. Такая система содержит средние, начиная от одночастичных и кончая N -частичными средними (N – число молекул в цепочке), причем $N \gg 1$.

Важным аспектом данной задачи является то, что третий уровень молекул представляет собой систему большого числа колебательных подуровней, взаимодействие с которой приводит к диссипации энергии и необратимости процесса экситон-экситонной аннигиляции. Если, исходя из первых принципов, произвести корректный учет такого взаимодействия, то в уравнениях движения добавится ряд многочастичных вкладов, описывающих релаксацию системы, связанную с экситон-экситонной аннигиляцией, но отсутствующих в рамках чисто феноменологического подхода.

При этом, если ограничиться рассмотрением уравнений только для одночастичных средних, а все многочастичные средние в этих уравнениях факторизовать, то мы приходим к традиционным уравнениям [11, 17], в которых теперь учтены поправки, связанные с трехчастичными взаимодействиями. Если также принять во внимание систему уравнений и для двухчастичных средних, то тем самым можно учесть парные корреляции между молекулами. В настоящей работе рассмотрен вывод как уточненных уравнений Блоха, так и системы уравнений, учитывающей парные корреляции между молекулами [21, 22].

Рассмотрим линейную цепочку, состоящую из N трехуровневых молекул. Как было упомянуто выше, такая цепочка может быть использована для моделирования J -агрегатов. Будем полагать, что наименьшее состояние каждой молекулы определяется вектором состояния $|g\rangle$ или $|1\rangle$, а энергия этого состояния равна E_1 . Соответственно второе состояние будет определяться вектором состояния $|e\rangle$ или $|2\rangle$ с энергией E_2 . Третьему состоянию сопоставим вектор $|f\rangle$ или $|3\rangle$ с энергией E_3 . Очевидно, что $E_3 > E_2 > E_1$. Для уровней энергии существует и другой набор обозначений, используемый в тексте, а именно: $E_1 = E_g$, $E_2 = E_e$ и $E_3 = E_f$. Молекуле, расположенной в узле цепочки с номером m , будут соответствовать вектора состояний $|mg\rangle$, $|me\rangle$ и $|mf\rangle$. С помощью этих векторов для каждой молекулы можно построить следующие операторы рождения и уничтожения, которые являются операторами проецирования на соответствующие состояния молекулы: $B_m = |mg\rangle\langle me|$ – оператор, описывающий уничтожение возбуждения в молекуле m на уровне « e » и ее переход в основное состояние $|mg\rangle$, а также $B_m^+ = |me\rangle\langle mg|$ – оператор, описывающий рождение возбуждения в молекуле m на уровне « e ».

По этому же принципу можно определить операторы $C_m = |mg\rangle\langle mf|$ и $C_m^+ = |mf\rangle\langle mg|$, а также $D_m = |me\rangle\langle mf|$ и $D_m^+ = |mf\rangle\langle me|$.

Определим также операторы числа молекул N_{mg} , N_{me} и N_{mf} в состояниях $|mg\rangle$, $|me\rangle$ и $|mf\rangle$ соответственно:

$$\begin{cases} N_{mg} = B_m B_m^+ = |mg\rangle\langle me|me\rangle\langle mg| = |mg\rangle\langle mg|, \\ N_{me} = B_m^+ B_m = |me\rangle\langle mg|mg\rangle\langle me| = |me\rangle\langle me|, \\ N_{mf} = C_m^+ C_m = |mf\rangle\langle mg|mg\rangle\langle mf| = |mf\rangle\langle mf|, \\ N_{mg} + N_{me} + N_{mf} = 1. \end{cases}$$

Все операторы, относящиеся к разным молекулам, очевидно, коммутируют между собой (по поводу определений и обозначений см. также [35, 36]).

Гамильтониан системы. Общий гамильтониан системы складывается из гамильтониана свободных молекул, а также гамильтонианов взаимодействия молекул с внешним полем и между собой. В частности, этот гамильтониан включает в себя также взаимодействие, приводящее к экситон-экситонной ан-

нигиляции. Гамильтониан взаимодействия с внешним электромагнитным полем с частотой ν_{12} можно представить в форме

$$H_{el}(t) = -\frac{1}{2} \sum_m (\boldsymbol{\mu}^{12} \mathbf{e}_1) (B_m^+ E_1 \exp(-i\nu_{12}t) + \text{э.с.}) . \quad (1)$$

Здесь внешнее поле определяется формулой $\mathbf{E} = \frac{1}{2} (\mathbf{e}_1 E_1 \exp(-i\nu_{12}t) + \text{к.с.})$, а сам гамильтониан (1) представляет взаимодействие указанного поля с вектором поляризации \mathbf{P} системы молекул: $H_{el}(t) = -\mathbf{PE}$. Это взаимодействие учитывается в приближении вращающегося поля. При этом $\boldsymbol{\mu}^{12}$ является дипольным моментом молекулы для перехода $1 \rightarrow 2$.

Таким образом, гамильтониан системы оказывается зависящим от времени. На практике более удобно иметь дело со стационарным гамильтонианом. Нетрудно показать, что существует элементарное унитарное преобразование операторов системы, применение которого позволяет устранить зависимость от времени в (1). Это преобразование [13] приводит к замене старых операторов B_m и D_m новыми b_m и d_m по правилу

$$\begin{cases} b_m = B_m \exp(i\nu_{12}t), \\ d_m = D_m \exp(i\nu_{12}t), \end{cases}$$

а также к сдвигу энергий уровней в гамильтониане свободных молекул H_0 (см. ниже). Оператор C_m , как показывает анализ, не входит ни в один из составляющих гамильтонианов взаимодействия, и поэтому в данном контексте он не используется для описания системы.

Общий гамильтониан H системы можно представить как сумму составляющих его гамильтонианов:

$$H = H_0 + H_{el} + H_{int} + H_{annih},$$

где H_0 – гамильтониан свободных молекул,

$$H_0 = \sum_m \{ \hbar(\omega_{m12} - \nu_{12}) b_m^+ b_m + \sum_v \hbar(\omega_{m13v} - 2\nu_{12}) d_{mv}^+ d_{mv} \} . \quad (2)$$

При этом в (2), а также всюду ниже мы теперь учитываем, что реально третий уровень молекул является электронно-колебательным и расщепляется на систему колебательных подуровней, которым соответствует индекс v . Здесь также предполагается, что $E_{1m} = E_{mg} = 0$, и тогда $\hbar\omega_{m12}$ и $\hbar\omega_{m13v}$ являются энергиями второго и третьего уровня.

Гамильтонианы H_{el} , H_{int} и H_{annih} последовательно описывают взаимодействие с полем накачки, диполь-дипольное взаимодействие (см. также [12, 35–38]), а также взаимодействие, обусловленное экситон-экситонной аннигиляцией. Эти гамильтонианы имеют следующий вид:

$$H_{el} = -\frac{1}{2} \sum_m (\boldsymbol{\mu}^{12} \mathbf{e}_1) (b_m^+ E_1 + \text{э.с.}),$$

$$H_{int} = \frac{\hbar}{2} \sum_{k \neq l} \Delta_{lk} (b_l^+ b_k + b_l b_k^+),$$

$$H_{annih} = \sum_{k \neq l} (V_{kl} b_k d_{lv}^+ + V_{lk} d_{lv} b_k^+).$$

Константы взаимодействия Δ_{lk} [12] равны

$$\Delta_{lk} = \frac{\mu^2}{\hbar a^3} \left\{ \left[\frac{\cos(k_0 a |l-k|)}{|l-k|^3} + k_0 a \frac{\sin(k_0 a |l-k|)}{|l-k|^2} \right] \times (1 - 3 \cos^2 \theta) - (k_0 a)^2 \frac{\cos(k_0 a |l-k|)}{|l-k|} \sin^2 \theta \right\}.$$

Здесь μ – матричный элемент дипольного момента для перехода $1 \rightarrow 2$; k_0 – волновой вектор падающего излучения; a – постоянная решетки; θ – угол между направлением дипольного момента $\boldsymbol{\mu}$ и осью цепочки; $l, k (l \neq k) : 1, 2, \dots, N$; N – число молекул в цепочке.

Вывод уравнений движения. Зная гамильтониан системы, можно получить уравнение движения для произвольного оператора A системы. Такое уравнение имеет вид

$$\frac{dA}{dt} = \frac{i}{\hbar} [H, A]. \quad (3)$$

Следует отметить, что полученное таким образом уравнение движения не содержит взаимодействия через поперечное поле излучения, которым обмениваются молекулы. В силу того, что такое взаимо-

действие содержит затухание, связанное с излучением, оно обычно учитывается непосредственно в уравнениях движения путем добавления в правую часть (3) оператора определенного типа [39]. В результате уравнение (3) переходит в уравнение вида

$$\frac{dA}{dt} = \frac{i}{\hbar} [H, A] + 2 \sum_{l,k} \gamma_{lk} [b_l^+ A b_k - \frac{1}{2} (b_l^+ b_k A + A b_l^+ b_k)]. \quad (4)$$

Здесь γ_{lk} задается формулой (см. [12])

$$\gamma_{lk} = \frac{\mu^2}{\hbar a^3} \left\{ \left[k_0 a \frac{\cos(k_0 a |l-k|)}{|l-k|^2} - \frac{\sin(k_0 a |l-k|)}{|l-k|^3} \right] \times (1 - 3 \cos^2 \theta) + (k_0 a)^2 \frac{\sin(k_0 a |l-k|)}{|l-k|} \sin^2 \theta \right\}.$$

В качестве примера приведем полученное на основе (4) уравнение для оператора числа молекул на втором уровне N_{me} (m – номер молекулы):

$$\frac{dN_{me}}{dt} = i \left\{ -\frac{1}{2\hbar} (\mu^2 \mathbf{e}_i) E_i (b_m - b_m^+) + \frac{1}{\hbar} \sum_{n \neq m} [(\Delta_{mn} + i\gamma_{mn}) b_n^+ b_m - (\Delta_{mn} - i\gamma_{mn}) b_m^+ b_n] \right\} + \frac{i}{\hbar} [H_{annih}, N_{me}]. \quad (5)$$

Скобки Пуассона с H_{annih} мы пока не раскрываем, поскольку этот член требует особого анализа. Нетрудно видеть, что в полученные уравнения, кроме одночастичных операторов b_m , входят также двухчастичные с индексами m и n . Очевидно, необходимо получить уравнения и для этих операторов. В свою очередь, в выведенных уравнениях появляются новые операторы, в том числе трехчастичные, которые представляют произведения одночастичных операторов, относящихся к трем разным молекулам. В конечном счете мы приходим к бесконечной иерархии сцепляющихся уравнений, содержащих многочастичные операторы все более высокого порядка.

Физический интерес представляют не сами операторы, а средние, получаемые их усреднением по матрице плотности системы. В результате бесконечная система уравнений для операторов переходит в иерархическую систему для средних от этих операторов. При этом возникают следующие вопросы: как сделать такую систему конечной и пригодной для решения, каким минимальным набором средних от операторов можно ограничиться для адекватного описания рассматриваемой системы. Если мы хотим ограничиться описанием системы с помощью средних лишь от одночастичных операторов, то для замыкания системы уравнений нужно представить все многочастичные средние, входящие в уравнения для одночастичных средних, в виде произведений от последних. Очевидно, при таком описании мы пренебрегаем корреляциями между молекулами. Для учета парных корреляций в описании системы следует добавлять, кроме одночастичных, также и двухчастичные средние. В результате возникает добавочная система уравнений и для этих средних. А чтобы замкнуть такую систему, многочастичные (число частиц больше двух) средние необходимо представить в виде произведений одночастичных и двухчастичных средних. При учете трехчастичных корреляций весь описанный выше процесс повторяется. Таким образом, минимальный набор операторов и средних от них, необходимый для описания системы, определяется той точностью, с какой мы хотим описать взаимодействия между молекулами цепочки. Следует отметить, что здесь представлена общая концепция выбора базисного набора средних от операторов системы. Реально на практике возможны определенные отклонения от этой схемы (не принципиального характера), связанные с конкретными свойствами системы (см. [21]).

Целью настоящей работы являлся учет парных корреляций между молекулами, а также многочастичных вкладов в уравнениях, которые возникают при анализе системы из первых принципов и отсутствуют при феноменологическом рассмотрении.

С учетом сказанного выше в данной работе в качестве базисного набора операторов, а следовательно и средних от них, выбраны следующие операторы: b_m , N_{me} , N_{mf} , $N_{me} N_{ne}$, $N_m b_n$ и $b_m b_n$ ($m \neq n$). Сюда следует включить также набор эрмитовски-сопряженных операторов. Физический смысл средних от этих операторов будет пояснен ниже.

Последовательно подставляя указанные операторы в качестве A в уравнение (4), а затем усредняя его, получаем полную систему уравнений для описания молекул цепочки вида

$$\frac{d\langle A \rangle}{dt} = \frac{i}{\hbar} \langle [H, A] \rangle + 2 \sum_{l \neq k} \gamma_{lk} \left[\langle b_l^+ A b_k \rangle - \frac{1}{2} (\langle b_l^+ b_k A \rangle + \langle A b_l^+ b_k \rangle) \right].$$

Следующий этап работы заключается в вычислении релаксационных членов, связанных с экситон-экситонной аннигиляцией, и замыкании полученной системы уравнений путем факторизации многочастичных средних, представляя их в виде произведений средних от операторов, входящих в базисный набор. Мы не будем останавливаться на правилах факторизации, поскольку они достаточно подробно изложены в работах [21], а также [35, 38, 40]. Укажем лишь на интерпретацию средних от операторов базисного набора, физический смысл которых достаточно ясен. Так, величина $\langle b_m \rangle$ из этого набора пропор-

циональна дипольному моменту молекулы m при переходе $1 \rightarrow 2$. $\langle N_{me} \rangle$ и $\langle N_{mf} \rangle$ определяют населенности второго и третьего уровней молекулы соответственно, а $\langle N_{me} N_{ne} \rangle$ ($m \neq n$) можно связать с населенностью двухэкситонных состояний.

Расчет релаксационных членов, связанных с механизмом экситон-экситонной аннигиляции. Вычисление в уравнениях движения членов, связанных с коммутатором, в который входит H_{annih} , требует особого анализа. Как упоминалось ранее, мы рассматриваем третий уровень молекул f , состоящий из ряда колебательных подуровней, отмечаемых индексом v . Эти подуровни можно охарактеризовать плотностью состояний $\rho(E) = \sum_v \delta(E - E_{fv})$, необходимой для вычисления вероятностей перехода. Фактически система электронно-колебательных подуровней представляет собой резервуар, взаимодействуя с которым, возбужденная молекула необратимо релаксирует в более низкое энергетическое состояние.

Рассмотрим основную идею расчета такой релаксации для населенности второго уровня $\langle N_{me} \rangle$. Нетрудно показать, что коммутатор операторов N_{me} и H_{annih} можно привести к виду

$$\frac{i}{\hbar} [H_{annih}, N_{me}] = \frac{i}{\hbar} \sum_{p \neq m} [(V_{pm} b_p d_{mv}^+ - h.c.) + (V_{mp} b_m d_{pv}^+ - h.c.)]. \quad (6)$$

Далее необходимо исследовать эволюцию во времени операторов, входящих в выражение (6). Мы будем полагать, что их эволюцию, связанную с механизмом экситон-экситонной аннигиляции, можно описать независимо от влияния всех других взаимодействий в системе [41]. В результате для оператора $b_p d_{mv}^+$ получаем уравнение вида

$$\frac{d(b_p d_{mv}^+)}{dt} = \frac{i}{\hbar} [(H_0 + H_{annih}), b_p d_{mv}^+]. \quad (7)$$

Усредняя (7) по матрице плотности системы и используя приближение марковского процесса, это уравнение можно формально решить [21], а результат подставить в усредненное уравнение (6). В итоге получаем выражение для вклада в релаксацию системы, связанного с механизмом экситон-экситонной аннигиляции:

$$\frac{d\langle N_{me} \rangle}{dt} = - \sum_{p \neq m} \{w_{mp} \langle N_{me} N_{pe} \rangle + \text{Re} \sum_{\substack{k \neq p \\ k \neq m}} [(\Gamma_{pmk} + i2\Delta_{pmk}) \langle N_{me} b_k^+ b_p \rangle + (\Gamma_{mppk} + i2\Delta_{mppk}) \langle N_{pe} b_k^+ b_m \rangle]\}, \quad (8)$$

в котором константы Δ_{pmk} , Γ_{pmk} и w_{mp} имеют следующий вид:

$$\begin{cases} \Delta_{pmk} = \frac{1}{\hbar^2} P \sum_v \frac{V_{pm} V_{mk}}{\omega_{fv} - 2\omega_e}, \\ \Gamma_{pmk} = \frac{2\pi}{\hbar^2} \sum_v V_{pm} V_{mk} \delta(\omega_{fv} - 2\omega_e), \\ w_{mp} = 2\Gamma_{mpm}. \end{cases}$$

Константа w_{mp} совпадает с такой же константой, фигурирующей, в частности, в [33].

Усредняя уравнение (5) и подставляя в него вместо члена $\frac{i}{\hbar} [H_{annih}, N_{me}]$ правую часть уравнения (8), получаем уравнение для эволюции $\langle N_{me} \rangle$. Для замыкания системы уравнений, в которую входит $\langle N_{me} \rangle$, в этих уравнениях необходимо, как об этом упоминалось выше, провести факторизацию многочастичных членов [21]. В результате для $\langle N_{me} \rangle$ окончательно получаем следующее уравнение:

$$\begin{aligned} \frac{d\langle N_{me} \rangle}{dt} = & -\gamma_2 \langle N_{me} \rangle + i \left\{ -\frac{1}{2\hbar} (\mathbf{u}^{12} \mathbf{e}_1) E_1 (\langle b_m \rangle - \langle b_m^+ \rangle) + \right. \\ & + \frac{1}{\hbar} \sum_{n \neq m} [(\Delta_{mn} + i\gamma_{mn}) \langle b_n^+ \rangle \langle b_m \rangle - (\Delta_{mn} - i\gamma_{mn}) \langle b_m^+ \rangle \langle b_n \rangle] \left. \right\} - \sum_{l \neq m} \{w_{ml} \langle N_{me} N_{le} \rangle + \\ & + \text{Re} \sum_{\substack{k \neq l \\ k \neq m}} [(\Gamma_{lmk} + i2\Delta_{lmk}) (\langle b_k^+ \rangle \langle N_{me} b_l \rangle + \langle b_l \rangle \langle N_{me} b_k^+ \rangle - \langle b_k^+ \rangle \langle b_l \rangle \langle N_{me} \rangle) + \\ & + (\Gamma_{mlk} + i2\Delta_{mlk}) (\langle b_k^+ \rangle \langle N_{le} b_m \rangle + \langle b_m \rangle \langle N_{le} b_k^+ \rangle - \langle b_k^+ \rangle \langle b_m \rangle \langle N_{le} \rangle)] \left. \right\} + \Gamma_{32} \langle N_{mf} \rangle. \end{aligned} \quad (9)$$

Аналогичный подход к вычислению релаксационных членов, связанных с экситон-экситонной аннигиляцией, применяется при выводе уравнений движения и для всех остальных средних, входящих в базисный набор. В результате возникает замкнутая система уравнений, в которой учтены двухчастичные корреляции между молекулами. Ввиду крайней громоздкости этих уравнений мы их не приводим (см. [21]). Отметим также, что в (9) феноменологически учтены вклад $\Gamma_{32} \langle N_{mf} \rangle$, связанный с переходом с третьего уровня на второй, а также скорость поперечной релаксации γ_2 .

Учет вкладов от трехчастичных взаимодействий в традиционных уравнениях для одночастичных матриц плотности. При обычном феноменологическом подходе система молекул в цепочках отдельных J -агрегатов описывается уравнениями типа уравнений Блоха для одночастичных матриц плотности. В этих уравнениях учитываются лишь двухчастичные взаимодействия, которые представлены в факторизованной форме. В рамках подхода, изложенного выше, этим уравнениям соответствует система уравнений для одночастичных средних. Однако анализ такой системы из первых принципов показал, что в ней, кроме двухчастичных вкладов, возникают также вклады, учитывающие трехчастичные взаимодействия. Если полученные уравнения факторизовать, то мы приходим к традиционным уравнениям [11, 17, 22], в которых, однако, учтены поправки, связанные с трехчастичными взаимодействиями. Эти поправки возникают в рамках изложенного выше подхода к расчету релаксационных членов, обусловленных экситон-экситонной аннигиляцией. Ниже приведена полученная таким образом система уравнений (10)–(13), в которой поправки подчеркнуты прямыми линиями. Нетрудно показать (см. также [11]), что

$$\langle N_{kf} \rangle = \rho_{33}^{(k)}, \quad \langle N_{ke} \rangle = \rho_{22}^{(k)}, \quad \langle N_{kg} \rangle = \rho_{11}^{(k)}, \quad \langle b_k \rangle = \frac{1}{2} R_k, \quad w = 2\alpha_s.$$

Эти формулы связывают средние от одночастичных операторов с элементами одночастичных матриц плотности. При этом $\rho_{11}^{(k)}$, $\rho_{22}^{(k)}$ и $\rho_{33}^{(k)}$ представляют диагональные элементы, соответствующие уровням 1, 2 и 3 молекулы с номером k ; R_k – недиагональный элемент, соответствующий переходу $1 \rightarrow 2$. Тогда:

$$\dot{\rho}_{22}^{(k)} = -\frac{1}{2} \operatorname{Re} \left[\sum_{l=1, l \neq k}^N (\gamma_{lk} + i\Delta_{lk}) R_l R_k^* - i\Omega R_k^* \right] + \Gamma_{32} \rho_{33}^{(k)} - \Gamma_{21} \rho_{22}^{(k)} - 2\alpha_s \rho_{22}^{(k)} [\rho_{22}^{(k-1)} + \rho_{22}^{(k+1)}] - \frac{1}{4} \alpha_s \operatorname{Re} \left[R_k \left(R_{k-2}^* \rho_{22}^{(k-1)} + R_{k+2}^* \rho_{22}^{(k+1)} \right) \right] - \frac{1}{4} \alpha_s \rho_{22}^{(k)} \left(R_{k-1}^* R_{k+1} + R_{k+1}^* R_{k-1} \right), \quad (10)$$

$$\dot{\rho}_{33}^{(k)} = -(\Gamma_{31} + \Gamma_{32}) \rho_{33}^{(k)} + \alpha_s \rho_{22}^{(k)} [\rho_{22}^{(k-1)} + \rho_{22}^{(k+1)}] + \frac{1}{4} \alpha_s \rho_{22}^{(k)} \left(R_{k-1}^* R_{k+1} + R_{k+1}^* R_{k-1} \right), \quad (11)$$

$$\dot{R}_k = -(\Gamma_{\perp} + i\Delta_k) R_k + \sum_{l=1, l \neq k}^N (\gamma_{lk} + i\Delta_{lk}) R_l [\rho_{22}^{(k)} - \rho_{11}^{(k)}] - i\Omega [\rho_{22}^{(k)} - \rho_{11}^{(k)}] - \alpha_s R_k [\rho_{22}^{(k-1)} + \rho_{22}^{(k+1)}] - \frac{1}{2} \alpha_s \left(R_{k-2} \rho_{22}^{(k-1)} + R_{k+2} \rho_{22}^{(k+1)} \right) \left(1 - 2\rho_{22}^{(k)} - \rho_{33}^{(k)} \right) - \frac{1}{8} \alpha_s R_k \left(R_{k-1}^* R_{k+1} + R_{k+1}^* R_{k-1} \right), \quad (12)$$

$$\rho_{11}^{(k)} = 1 - \rho_{22}^{(k)} - \rho_{33}^{(k)}. \quad (13)$$

Здесь α_s – скорость экситон-экситонной аннигиляции; Ω – частота Раби; Γ_{mn} – скорость релаксации с уровня m на уровень n ; Γ_{\perp} – скорость поперечной релаксации молекулы; $\Delta_k \equiv \Delta$ – отстройка от резонанса для изолированной молекулы.

В следующей главе эта система уравнений будет использована для анализа бистабильности в однородных цепочках.

Исследование бистабильности в молекулярных J -агрегатах при резонансном оптическом возбуждении с учетом трехчастичных взаимодействий

Анализ показывает, что факторизованные трехчастичные вклады могут включать в качестве множителя среднее от оператора числа молекул на нижнем уровне. Особенностью рассматриваемых систем (см. [11, 17]) является то, что в равновесном состоянии такое среднее близко к 1, ввиду малой вероятности заселения верхних уровней молекул. Это напрямую следует из условия нормировки, принятого в настоящей работе (см. [21]): $N_{mg} + N_{me} + N_{mf} = 1$, где N_{mg} , N_{me} и N_{mf} являются операторами числа молекул в узле m на первом, втором и третьем уровнях соответственно. В связи с этим указанные трехчастичные вклады являются фактически двухчастичными и совпадают по порядку величин с двухчастичными вкладами, присутствующими в традиционной системе уравнений. Таким образом, уже только одно это обстоятельство может как количественно, так и качественно повлиять на свойства исследуемых объектов.

Ниже с учетом новых вкладов в уравнения движения проводится анализ бистабильности в J -агрегатах в специальном случае, соответствующем однородной цепочке молекул. В такой цепочке считается, что индивидуальные характеристики всех молекул не зависят от их расположения, а сама цепочка предполагается бесконечной.

Таким образом, если в системе уравнений (10)–(13) полагать, что все элементы матрицы плотности не зависят от номера молекулы, то мы получим однородную цепочку. Если теперь, в свою очередь, в полученной системе уравнений положить равными нулю производные по времени, то возникает система уравнений для стационарных состояний однородных цепочек. Как показал анализ, выполненный в работах [11, 17] без учета многочастичных членов, в таких состояниях возникает бистабильность. Исходя из этого, представляет интерес выяснить, каково влияние многочастичных эффектов на проявление бистабильности в этих состояниях. Такие исследования удобно проводить для уравнений в безразмерной форме. В нашем случае для этого в исходной системе (10)–(13) производится нормировка времени на величину γ_R :

$$t \rightarrow \gamma_R t,$$

где

$$\gamma_R = 2 \sum_{l=k=1}^{\infty} \gamma_{lk}.$$

В соответствии с этим параметры уравнений также нормируются на γ_R . Эти нормированные параметры будут обозначаться следующим образом: α – нормированная скорость экситон-экситонной аннигиляции; $\bar{\Delta}$ – отстройка от частоты перехода $1 \rightarrow 2$, учитывающая влияние окружающих молекул; $\bar{\Omega}$ – частота Раби.

Анализ уравнений без учета многочастичных эффектов показал, что бистабильность существует в достаточно широкой области значений $\alpha \geq 0$. Более детальные расчеты показывают, что область существования гистерезиса по этому параметру и при $\bar{\Delta} = -10$ достигает значения $\alpha \approx 40$. На рис. 1 изображены кривые гистерезисной зависимости населенности второго уровня ρ_{22}^0 от $\bar{\Omega}^2$ при $\bar{\Delta} = -10$ и $\alpha = 0, 1, 5, 10, 15$ и 25. Индекс «0» означает стационарное состояние.

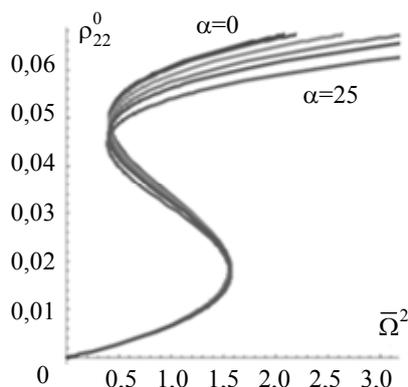


Рис. 1. Бистабильность населенности второго возбужденного уровня при изменении частоты Раби без учета многочастичных поправок. Коэффициент экситон-экситонной аннигиляции принимает по порядку значения: $\alpha = 0, 1, 5, 10, 15$ и 25; $\bar{\Delta} = -10$

Однако, как следует из дальнейшего анализа, учет многочастичных вкладов приводит к уменьшению области существования бистабильности. На рис. 2 изображены бистабильные зависимости населенности второго уровня ρ_{22}^0 от $\bar{\Omega}^2$ при учете трехчастичных вкладов в случае $\bar{\Delta} = -10$ и различных значениях параметра α . Из расчетов следует, что с увеличением α области бистабильности сужаются, и она исчезает при конечном значении $\alpha \approx 22,7$. Таким образом, учет многочастичных вкладов в целом приводит к заметному ограничению области существования бистабильности в зависимости от постоянной экситон-экситонной аннигиляции.

На рис. 3 изображена зависимость ширины области бистабильности $\delta\rho_{22}^0$ (разность значений правой и левой границ бистабильности на рис. 2) для населенности второго уровня от величин расстройки и постоянной экситон-экситонной аннигиляции. Полученная поверхность позволяет определить не только область существования самой бистабильности, но в принципе также и наиболее вероятную область параметров, в которой может быть сформирован диссипативный солитон. Пересечение этой поверхности с плоскостью $\delta\rho_{22}^0 = 0$, очевидно, определяет границу существования бистабильности. Эта граница изображена на рис. 4 в плоскости параметров α и $\bar{\Delta}$. При этом кривая 1 ограничивает сверху область бистабиль-

ности, полученную без учета многочастичных вкладов, а кривая 2 – полученную при учете таких вкладов. Вертикальные пунктирные линии соответствуют значениям α , при которых $\left| \frac{d\bar{\Delta}}{d\alpha} \right| \rightarrow \infty$. Согласно расчетам, эти значения равны $\alpha \cong 63,4$ и $\alpha \cong 101,5$. Таким образом, графики наглядно показывают, что при учете многочастичных вкладов область существования бистабильности по параметру α в целом сужается.

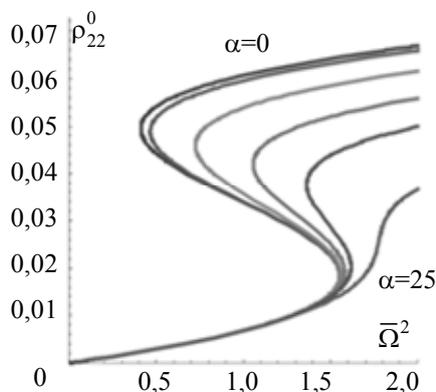


Рис. 2. Бистабильная зависимость населенности второго уровня ρ_{22}^0 от частоты Раби для стационарных состояний. Бистабильность исчезает при значениях скорости экситон-экситонной аннигиляции $\alpha \cong 22,7$

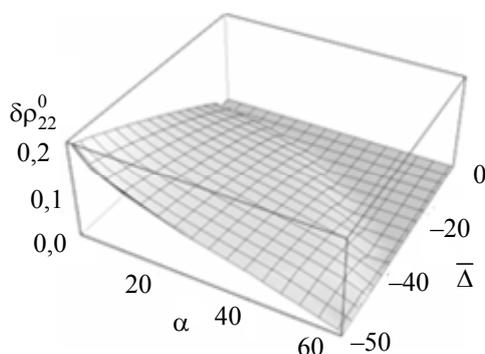


Рис. 3. Зависимость ширины области бистабильности населенности второго уровня $\delta\rho_{22}^0 = \rho_{22,up}^0 - \rho_{22,dn}^0$, где $\rho_{22,up}^0, \rho_{22,dn}^0$ – значения для левой и правой границ бистабильности на рис. 2, от коэффициента экситон-экситонной аннигиляции α и расстройки $\bar{\Delta}$

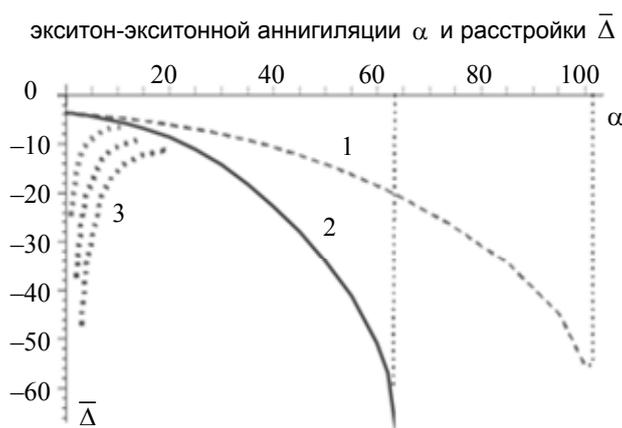


Рис. 4. Границы областей существования бистабильности в плоскости параметров α и $\bar{\Delta}$. Кривая 1 (штриховая линия) ограничивает сверху область бистабильности без учета поправок. Кривая 2 (сплошная линия) соответствует случаю учета многочастичных поправок. Этому случаю соответствует также набор кривых (слева направо) $\rho_{33}^0 / \rho_{22}^0 = 0,1; 0,2$ и $0,3$ под номером 3 (ρ_{33}^0 – населенность третьего уровня).

Вертикальные пунктирные линии – это границы по параметру α , при стремлении к которым $\left| \frac{d\bar{\Delta}}{d\alpha} \right| \rightarrow \infty$

Фактически при учете трехчастичных взаимодействий граница существования бистабильности проявляет тенденцию к смещению в сторону меньших констант взаимодействия в механизме экситон-экситонной аннигиляции. Этот эффект можно трактовать как увеличение эффективности данного механизма при учете трехчастичных вкладов, что уменьшает требование к величине константы взаимодействия.

Модуляционная неустойчивость однородных режимов резонансного возбуждения молекулярных J -агрегатов

В данном, заключительном разделе работы анализируется модуляционная неустойчивость стационарных состояний однородных цепочек, исследованных в предыдущем разделе при учете трехчастичных вкладов. Целью такого анализа является определение границ областей устойчивости и неустойчивости и соотношение их с границами существования бистабильности.

Исследование модуляционной неустойчивости проводится в рамках полученной ранее системы уравнений (10)–(13). Для однородных цепочек, как упоминалось в предыдущем разделе, все элементы одночастичных матриц плотности не зависят от номера молекулы. Следовательно, можно положить: $\rho_{mn}^{(k)} \equiv \rho_{mn}$ ($n = 1, 2, 3$) и $R_k \equiv R$. Для стационарных состояний, как это делалось и ранее, используем обозначения ρ_{22}^0 и R_0 , причем $R_0 = R_0^R + i R_0^J$, где R_0^R и R_0^J – вещественная и мнимая части R_0 .

Наличие бистабильности свидетельствует о сильной нелинейности системы по отношению к подерживающему излучению и служит признаком возможности формирования в ней диссипативных солитонов [16–20]. Именно в связи с этим необходимо проведение анализа устойчивости и эволюции малых возмущений стационарных состояний однородных цепочек. Этот анализ проводится в линейном приближении.

В общем виде решение, близкое к стационарному однородному состоянию, можно представить в следующей форме:

$$\begin{cases} \rho_{33}(k) = \rho_{33}^0 + \frac{1}{2} \left(x_1 e^{\lambda t + ikq} + \text{к.с.} \right), \\ \rho_{22}(k) = \rho_{22}^0 + \frac{1}{2} \left(x_2 e^{\lambda t + ikq} + \text{к.с.} \right), \\ \text{Re } R_k = R_0^R + \frac{1}{2} \left(x_3 e^{\lambda t + ikq} + \text{к.с.} \right), \\ \text{Im } R_k = R_0^J + \frac{1}{2} \left(x_4 e^{\lambda t + ikq} + \text{к.с.} \right). \end{cases} \quad (14)$$

Здесь x_i являются малыми возмущениями стационарных однородных состояний, которые описывают модуляционную неустойчивость системы в случае $\text{Re } \lambda > 0$. При этом параметр q можно рассматривать как нормированное волновое число возмущения. Из свойств симметрии следует, что, если положить $q = \pi s$, то изменение параметра s достаточно рассматривать лишь в интервале $0 \leq s < 1$.

Подстановка (14) в систему (10)–(13) и ее линеаризация по x_i приводят к системе линейных уравнений вида

$$\begin{cases} (a_{11} - \lambda)x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 = 0, \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + a_{23}x_3 + a_{24}x_4 = 0, \\ a_{31}x_1 + a_{32}x_2 + (a_{33} - \lambda)x_3 + a_{34}x_4 = 0, \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + (a_{44} - \lambda)x_4 = 0. \end{cases} \quad (15)$$

Коэффициенты a_{mn} являются функциями параметров системы, в том числе параметров α , $\bar{\Omega}$, $\bar{\Delta}$ и s . Требование обращения в нуль определителя системы уравнений (15) позволяет найти собственные значения λ , вещественные части которых определяют нарастание или убывание возмущений, а следовательно, и устойчивость состояний. Указанное требование приводит к алгебраическому уравнению 4-го порядка по λ с вещественными коэффициентами, так что имеется четыре ветви корней. При этом, если при заданных фиксированных значениях параметров стационарных состояний вещественная часть хотя бы одного из значений λ при каком-либо s в интервале $0 \leq s < 1$ является положительной, то такой режим является неустойчивым. Ввиду громоздкости вида коэффициентов a_{mn} реален только численный анализ (15).

Проведенный численный анализ устойчивости населенности второго уровня ρ_{22}^0 в зависимости от изменения параметров α и $\bar{\Omega}$ и фиксированных значениях других параметров показал, что стационарные состояния, отвечающие верхней ветви бистабильной зависимости, всегда устойчивы. В то же время состояния, соответствующие средней ветви, как обычно, неустойчивы, и область неустойчивости существ-

вует также на нижней ветви бистабильной зависимости. При этом найденная область неустойчивости выходит за пределы области бистабильности.

Результаты данного анализа продемонстрированы на рис. 5, на котором приведено изменение населенности второго уровня в зависимости от частоты Раби при $\alpha = 5$ (внутри области бистабильности) и $\alpha = 25$ (вне этой области). На графиках отмечены границы областей неустойчивости.

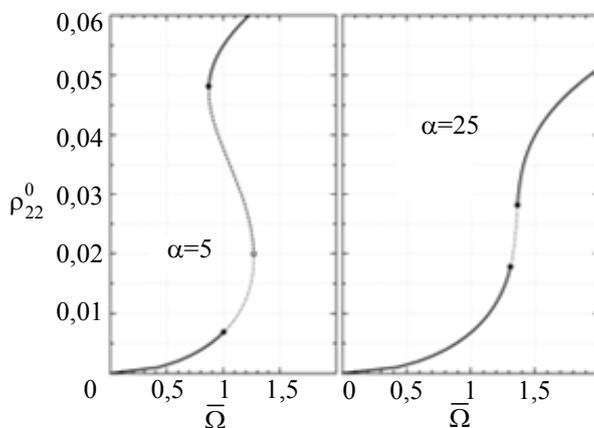


Рис. 5. Изменение населенности второго возбужденного уровня в зависимости от частоты Раби при двух значениях параметра α ; $\bar{\Delta} = -10$. В первом случае ($\alpha = 5$) наблюдается бистабильность, которая исчезает с увеличением α (кривая при $\alpha = 25$). Сплошная линия соответствует области устойчивости стационарных состояний, пунктирная – неустойчивой промежуточной ветви бистабильности, штрихпунктирная – области неустойчивости на нижней ветви или же вне области бистабильности. Эти области на рисунке разделены кружками

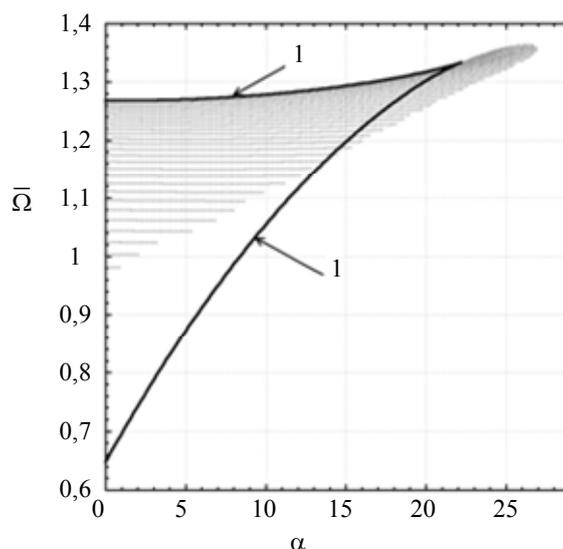


Рис. 6. Область неустойчивости (заштрихованная часть графика), соответствующая нижней ветви бистабильной зависимости функции $\rho_{22}^0(\alpha, \bar{\Omega})$. 1 – границы области существования бистабильности

Наиболее общее представление о границах области неустойчивых стационарных состояний, соответствующих нижней ветви, можно получить из рис. 6. Этой области соответствует заштрихованный участок, который располагается не только внутри границ бистабильности, но и выходит за ее пределы.

Описанное выше явление неустойчивости, а также наблюдавшаяся бистабильность были проанализированы в области параметров α и $\bar{\Omega}$ при выбранном фиксированном значении отстройки $\bar{\Delta}$. Однако эти явления могут наблюдаться и в области других параметров, в частности, в области параметров α и $\bar{\Delta}$ при фиксированном значении частоты Раби.

На рис. 7 приведена зависимость населенности второго молекулярного уровня ρ_{22}^0 от параметра $\bar{\Delta}$, как и ранее, для двух значений α : для $\alpha = 5$ (внутри области бистабильности) и $\alpha = 20$ (вне этой области). На этих графиках границы, связанные с проявлениями неустойчивости, помечены кружками. Анализ показал, что и в данном случае состояния, соответствующие верхней ветви бистабильной зависимости, являются устойчивыми, а средней – неустойчивыми.

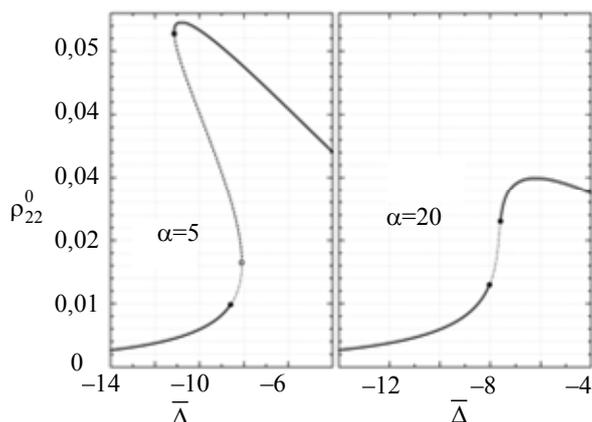


Рис. 7. Изменение населенности второго возбужденного уровня в зависимости от отстройки $\bar{\Delta}$ при фиксированных частных значениях параметра α ; $\bar{\Omega} = 1$. В первом случае ($\alpha = 5$) наблюдается бистабильность, которая исчезает с увеличением α (кривая при $\alpha = 20$). Сплошная линия соответствует области устойчивости стационарных состояний, пунктирная – неустойчивой промежуточной ветви бистабильности, штрихпунктирная линия – области неустойчивости на нижней ветви или же неустойчивым состояниям вне области бистабильности. Эти области на рисунке разделены кружками

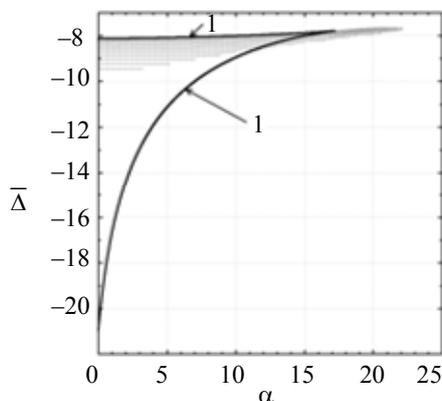


Рис. 8. Область неустойчивости (заштрихованная часть графика), соответствующая нижней ветви бистабильной зависимости функции $\rho_{22}^0(\alpha, \bar{\Delta})$. 1 – границы области существования бистабильности

Область неустойчивости существует также для состояний, соответствующих нижней ветви бистабильности, и отчасти для состояний вне границ бистабильной зависимости. Этой области на рис. 8 соответствует заштрихованный участок поверхности.

Таким образом, в данном разделе были рассмотрены результаты линейного анализа устойчивости стационарных состояний однородных режимов резонансно возбуждаемых молекулярных J -агрегатов с учетом трехчастичных взаимодействий молекул. Проведено изучение границ областей устойчивости и неустойчивости и их сопоставление с границами существования бистабильности. Эти результаты существенны для последующего исследования пространственно неоднородных структур в молекулярных цепочках конечных длин и наносолитонов.

Заключение

В обзоре обобщены результаты исследований, выполненных авторами в рамках трех связанных между собой направлений. Впервые получена система уравнений для описания отдельных J -агрегатов с учетом многочастичных взаимодействий и парных корреляций между молекулами цепочек. Это позволяет более корректно исследовать свойства J -агрегатов и нелинейные оптические эффекты в этих системах при их резонансном возбуждении. Проанализировано влияние трехчастичных взаимодействий на бистабильность стационарных состояний однородных цепочек. Показано, что учет трехчастичных взаимодействий заметно сужает область существования бистабильности и, в частности, сдвигает ее в сторону меньших скоростей экситон-экситонной аннигиляции. Проведено исследование модуляционной неустойчивости определенных выше стационарных состояний относительно пространственных возмущений в J -агрегатах. Найдены области устойчивости и неустойчивости. Проведено сопоставление границ этих областей с границами области существования бистабильности. Полученные результаты важны как для определения областей параметров, где наиболее вероятно существование устойчивых стационарных со-

стояний однородных режимов в J -агрегатах, так и для исследования возможности возбуждения диссипативных солитонов.

Проведенные исследования позволяют более надежно выделить область параметров, где могут наблюдаться существенно нелинейные эффекты, которые можно было бы использовать для создания схем молекулярной памяти и, в более широком контексте, для создания устройств оптической логики.

Литература

1. De Boer S., Wiersma D.A. Dephasing-induced damping of superradiant emission in J -aggregates // *Chemical Physics Letters*. 1990. V. 165. N 1. P. 45–53.
2. Fidler H., Terpstra J., Wiersma D.A. Dynamics of Frenkel excitons in disordered molecular aggregates // *Journal of Chemical Physics*. 1991. V. 94. N 10. P. 6895–6907.
3. Fidler H., Knoester J., Wiersma D.A. Optical properties of disordered molecular aggregates: a numerical study // *Journal of Chemical Physics*. 1991. V. 95. N 11. P. 7880–7890.
4. Богданов В.А., Викторова Е.Н., Куля С.В., Спино А.С. Нелинейная кубическая восприимчивость и дефазировка экситонных переходов в молекулярных агрегатах // *Письма в ЖЭТФ*. 1991. Т. 53. № 2. С. 100–103.
5. Wang Y. Resonant third-order optical nonlinearity of molecular aggregates with low-dimensional excitons // *Journal of the Optical Society of America B*. 1991. V. 8. N 5. P. 981–985.
6. Gadonas R., Feller K.-H., Pugzlys A. Wavelength dependent nonlinear optical properties of pseudoisocyanine J -aggregates // *Optics Communications*. 1994. V. 112. N 3–4. P. 157–162.
7. Spano F.C., Knoester J. Fermions on a Frenkel chain: nonlinear optical response of linear molecular aggregates / In: *Advances in Magnetic and Optical Resonance*. Ed. W.S. Warren. Academic Press, 1994. P. 117.
8. Knoester J., Spano F.C. Theory of pump-probe spectroscopy of molecular J -aggregates / In: *J-Aggregates*. Ed. T. Kobayashi. Singapore: World Scientific, 1996. P. 111–160.
9. Furuki M., Tian M., Sato Y., Pu L.S., Tatsuura S., Wada O. Terahertz demultiplexing by a single shot time to space conversion using a film of squarylium dye J -aggregates // *Applied Physics Letters*. 2000. V. 77. N 4. P. 472–474.
10. Авдеева В.И., Кучьянов А.С., Плеханов А.И., Соломинский Ю.Л., Толмачев А.И., Шапиро Б.И. Пассивная синхронизация мод Nd:YAG-лазера с насыщающимся поглотителем в виде тонкой желатиновой пленки с J -агрегатами органического красителя // *Квантовая электроника*. 2003. Т. 33. № 6. С. 539–541.
11. Malyshev V.A., Glaeske H., Feller K.-H. Optical bistable response of an open Frenkel chain: exciton-exciton annihilation and boundary effects // *Physical Review A – Atomic, Molecular, and Optical Physics*. 1998. V. 58. N 1. P. 670–678.
12. Malyshev V.A., Moreno P. Mirrorless optical bistability of linear molecular aggregates // *Physical Review A – Atomic, Molecular, and Optical Physics*. 1996. V. 53. N 1. P. 416–423.
13. Glaeske H., Malyshev V.A., Feller K.-H. Effects of higher exciton manifolds and exciton-exciton annihilation on optical bistable response of an ultrathin glassy film comprised of oriented linear Frenkel chains // *Physical Review A – Atomic, Molecular, and Optical Physics*. 2002. V. 65. N 3. P. 033821/1–033821/10.
14. Klugkist J.A., Malyshev V., Knoester J. Intrinsic optical bistability of thin films of linear molecular aggregates: the one-exciton approximation // *Journal of Chemical Physics*. 2007. V. 127. N 16. Art. 164705.
15. Klugkist J.A., Malyshev V., Knoester J. Intrinsic optical bistability of thin films of linear molecular aggregates: the two-exciton approximation // *Journal of Chemical Physics*. 2008. V. 128. N 8. Art. 084706.
16. Киселев Ал.С., Киселев Ан.С., Розанов Н.Н. Наноразмерные диссипативные дискретные солитоны в резонансно возбуждаемых молекулярных J -агрегатах // *Письма в ЖЭТФ*. 2008. Т. 87. № 11–12. С. 763–766.
17. Высотина Н.В., Малышев В.А., Маслов В.Г., Нестеров Л.А., Розанов Н.Н., Федоров С.В., Шацев А.Н. Моделирование взаимодействия ориентированных J -агрегатов с резонансным лазерным излучением // *Оптика и спектроскопия*. 2010. Т. 109. № 1. С. 117–124.
18. Высотина Н.В., Розанов Н.Н., Федоров С.В., Шацев А.Н. Движение молекулярных диссипативных солитонов в ориентированных линейных J -агрегатах при наклонном падении возбуждающего лазерного излучения // *Оптика и спектроскопия*. 2010. Т. 109. № 1. С. 125–127.
19. Rosanov N.N., Fedorov S.V., Shatsev A.N., Vysotina N.V. Dissipative molecular solitons // *European Physical Journal D*. 2010. V. 59. N 1. P. 3–12.
20. Розанов Н.Н. Диссипативные оптические солитоны. От микро- к нано- и атто-. М.: Физматлит, 2011. 536 с.
21. Левинский Б.Н., Нестеров Л.А., Файнберг Б.Д., Розанов Н.Н. Вывод уравнений движения для резонансно возбуждаемых молекулярных J -агрегатов при учете многочастичных эффектов // *Оптика и спектроскопия*. 2013. Т. 115. № 3. С. 464–479.

22. Нестеров Л.А., Федоров С.В., Розанов Н.Н., Левинский Б.Н., Файнберг Б.Д. Анализ бистабильности в молекулярных J-агрегатах при резонансном оптическом возбуждении с учетом многочастичных эффектов // *Оптика и спектроскопия*. 2013. Т. 115. № 4. С. 572–580.
23. Веретеннов Н.А., Нестеров Л.А., Розанов Н.Н., Федоров С.В. Модуляционная неустойчивость одно-родных режимов резонансного возбуждения молекулярных J-агрегатов // *Оптика и спектроскопия*. 2014. Т. 117. № 2. С. 267–274.
24. Sundström V., Gillbro T., Gadonas R.A., Piskarskas A. Annihilation of singlet excitons in J aggregates of pseudoisocyanine (PIC) studied by pico- and subpicosecond spectroscopy // *The Journal of Chemical Physics*. 1988. V. 89. N 5. P. 2754–2762.
25. Stiel H., Daehne S., Teuchner K. J-aggregates of pseudoisocyanine in solution: new data from nonlinear spectroscopy // *Journal of Luminescence*. 1988. V. 39. N 6. P. 351–357.
26. Gagel R., Gadonas R., Laubereau A. Evidence for biexcitons and dynamic Stark effect in J-aggregates from femtosecond spectroscopy // *Chemical Physics Letters*. 1994. V. 217. N 3. P. 228–233.
27. Minoshima K., Taiji M., Misawa K., Kobayashi T. Femtosecond nonlinear optical dynamics of excitons in J-aggregates // *Chemical Physics Letters*. 1994. V. 218. N 1–2. P. 67–72.
28. Van Burgel M., Wiersma D.A., K. Duppen. The dynamics of one-dimensional excitons in liquids // *The Journal of Chemical Physics*. 1995. V. 102. N 1. P. 20–33.
29. Gadonas R., Feller K.-H., Pugzlys A., Jonusauskas Q., Oberlé J., Rullière C. Wavelength and intensity-dependent transient degenerate four-wave mixing in pseudoisocyanine J-aggregates // *Journal of Chemical Physics*. 1997. V. 106. N 20. P. 8374–8383.
30. Gaižauskas E., Feller K.-H., Gadonas R. Annihilation enhanced four-wave mixing in molecular aggregates // *Optics Communications*. 1995. V. 118. N 3–4. P. 360–366.
31. Malyshev V.A., Glaeske H., Feller K.-H. Effect of exciton-exciton annihilation on optical bistability of a linear molecular aggregate // *Optics Communications*. 1997. V. 140. N 1–3. P. 83–88.
32. Malyshev V.A., Glaeske H., Feller K.-H. Effect of exciton-exciton annihilation on optical bistability of one-dimensional Frenkel excitons // *Journal of Luminescence*. 1998. V. 76–77. P. 455–459.
33. Malyshev V.A., Glaeske H., Feller K.-H. Exciton-exciton annihilation in linear molecular aggregates at low temperature // *Chemical Physical Letters*. 1999. V. 305. N 1–2. P. 117–122.
34. Malyshev V.A., Kozlov G.G., Glaeske H., Feller K.-H. Channels of the exciton-exciton annihilation in one-dimensional aggregates at low temperature // *Chemical Physical Letters*. 2000. V. 254. N 1. P. 31–38.
35. Mukamel S., Abramavicius D. Many-body approaches for simulating coherent nonlinear spectroscopies of electronic and vibrational excitons // *Chemical Reviews*. 2004. V. 104. N 4. P. 2073–2098.
36. Renger T., May V., Kuhn O. Ultrafast excitation energy transfer dynamics in photosynthetic pigment-protein complexes // *Physics Report*. 2001. V. 343. N 3. P. 137–254.
37. Spano F., Mukamel S. Nonlinear susceptibilities of molecular aggregates: enhancement of (3) by size // *Physical Review A*. 1989. V. 40. N 10. P. 5783–5801.
38. Spano F.C., Mukamel S. Excitons in confined geometries: size scaling of nonlinear susceptibilities // *The Journal of Chemical Physics*. 1991. V. 95. N 10. P. 7526–7540.
39. Lemberg R.H. Radiation from an N-atom system. I. General formalism // *Physical Review A*. 1970. V. 2. N 3. P. 883–888.
40. Mukamel S. *Principles of Nonlinear Optical Spectroscopy*. NY: Oxford University Press, 1995. 543 p.
41. Fainberg B.D., Jouravlev M., Nitzan A. Light-induced current in molecular tunneling junctions excited with intense shaped pulses // *Physical Review B – Condensed Matter and Materials Physics*. 2007. V. 76. N 24. Art. 245329.



Веретеннов Николай Александрович закончил Физический факультет СПбГУ в 2002 году. В настоящее время – младший научный сотрудник ГОИ им. С.И. Вавилова, ведущий инженер кафедры Оптики лазеров Университета ИТМО. С 2004 по 2006 г.г. работал в Свободном университете Брюсселя, Бельгия. Область научных интересов – нелинейная оптика, лазеры, оптические солитоны, компьютерное моделирование.

Nikolay A. Veretenov graduated from the Physical Faculty of Saint Petersburg State University in 2002. Now he is a junior scientific researcher in Vavilov Optical State Institute, a leading engineer at the Laser Optics Department of ITMO University. From 2004 to 2006 he was working in Universite Libre de Bruxelles, Belgium. His areas of scientific interests are: nonlinear optics, lasers, optical solitons and computer simulation.



Левинский Борис Нахимович – кандидат физико-математических наук, преподаватель в физической лаборатории, физик-исследователь, Технологический институт, Холон, Израиль. Области научных интересов – квантовая химия, теоретическая физика, прикладная электродинамика, оптика, физика и оптика больших молекул, теория экситонов, теория полупроводников, коллективные эффекты в физике, нанопизика. С отличием закончил физический факультет Московского государственного университета в 1967 г. (специальность – «Физика»). В 1992 г. защитил кандидатскую диссертацию (специальность – «Математика и физика»). С 1967 по 1975 г.г. работал в Всесоюзном научно-исследовательском институте химических средств защиты растений, а с 1976 по 1994 г.г. во Всесоюзном научно-исследовательском институте оптико-физических измерений. С 1994 по 1996 г.г.

работал экспертом в фирме по технологическому трансферу технологий в страны Европы. С 1997 по 2001 г.г. работал как алгоритмист в частной фирме в Иерусалиме. С 2001 г. по настоящее время работает в Технологическом институте гор. Холон в качестве преподавателя в физической лаборатории и одновременно как физик-исследователь. Имеет свыше 50 публикаций в советских и иностранных журналах и докладов на конференциях. Принимал участие как исполнитель ряда фундаментальных и прикладных проектов.

Levinsky Boris has got PhD degree in the field of Physics and Mathematics He is a tutor in Physical Laboratory, scientist-researcher at Holon Institute of Technology, Israel. His areas of scientific interests are: Quantum Chemistry, Theoretical Physics, Applied Electrodynamics, Optics, Physics and Optics of large molecules, theory of excitons, theory of semiconductors, collective effects in Physics and Nanophysics. In 1967 he graduated with honours from the Physical Faculty of Moscow University ("Physics" specialty). In 1992 he defended his PhD thesis ("Mathematics and Physics" specialty). From 1967 to 1975 he was working at National Research Institute of Chemical Science. From 1976 to 1994 he was working at National Research Institute of Optical & Physical Measurements, Moscow, Russia. From 1994 to 1996 he was an expert for transfer of advanced Russian technologies to European countries. From 1997 to 2001 he was working as an algorithm researcher at the private company in Jerusalem. From 2001 to nowadays, he is working as a tutor and scientist-researcher in the Physical Laboratory at Holon Institute of Technology, Israel. He has got more than 50 publications in the Soviet and foreign editions and conference reports. He is a co-investigator of a number of fundamental and applied projects.

Boris N. Levinsky – PhD, scientist-researcher, tutor in Physical Laboratory, Holon Institute of Technology, Israel, borisl@hit.ac.il



Нестеров Леонид Андреевич – ведущий инженер, Университет ИТМО. Окончил физический факультет Ленинградского государственного университета в 1963 г. После этого работал в Государственном оптическом институте им. С.И. Вавилова (сейчас это второе место работы), а позднее в Научно-исследовательском центре «Corning» (США) в Санкт-Петербурге. Сотрудник Университета ИТМО с 2006 г. Область исследований: нелинейная и квантовая оптика, вопросы оптической связи.

Leonid A. Nesterov is a leading engineer in ITMO University, Saint Petersburg, Russia. He graduated from the Physical Faculty of Leningrad State University in 1963. Thereafter he has been working in Vavilov State Optical Institute (it is the second place of work now), and further in Scientific&Research Center "Corning" (USA) in Saint Petersburg. In 2006 he began to work in ITMO University. His areas of researches are: Nonlinear and Quantum Optics, problems of optical communications.

Leonid A. Nesterov – leading engineer, Saint Petersburg, Russia, nesterovla@mail.ru



Розанов Николай Николаевич – член-корреспондент Российской академии наук, доктор физико-математических наук, профессор, начальник отдела ОАО «ГОИ им. С.И. Вавилова», заведующий кафедрой Оптики лазеров Университета ИТМО. Лауреат премии им. Д.С. Рождественского РАН (2007 г.). Область научных интересов – физическая и нелинейная оптика, лазерная физика, физика солитонов.

Розанов Николай Николаевич – член-корр. РАН, доктор физ.-мат.наук, начальник отдела ОАО «ГОИ им. С.И. Вавилова»; заведующий кафедрой «Оптики лазеров» Университета ИТМО, Санкт-Петербург, Россия, nrosanov@yahoo.com

Nikolay N. Rosanov is a corresponding member of the Russian Academy of Sciences (RAS), doctor of phys.&math. sciences, Professor, section head of Vavilov State Institute, head of Laser Optics Department of ITMO University. Laureate of the RAS

prize named after D.S. Rozhdestvenskii (2007). The areas of scientific interests are: Physical and Nonlinear Optics, Laser Physics, Physics of solitons.

Nicolay N. Rosanov – RAS Corresponding member, D.Sc., Section head of Vavilov State Institute; Head of Laser Optics Department, ITMO University, Saint Petersburg, Russia, nrosanov@yahoo.com



Борис Д. Файнберг окончил ЛИТМО (ныне Университет ИТМО) в 1969 г. Он получил степень Ph.D. (1977 г.) в Государственном оптическом институте имени С.И. Вавилова (Санкт-Петербург, Россия), где он работал как младший научный сотрудник, старший научный сотрудник и старший научный сотрудник с дипломом в отделе молекулярной спектроскопии (до 1990 г.). В 1991 г. он начал работать в Школе Химии Тель-Авивского Университета в качестве научного сотрудника (до 1998 г.) и затем как приглашенный профессор, начиная с 2000 г. В 1998 г. он был принят в штат Холонского Технологического Института (Холон, Израиль) как доцент и затем (2003 г.) как полный профессор химической физики. Он был приглашенным профессором в Институте атомных и молекулярных наук, Тайпэй, Тайвань (2000 г.); Институте физики и химии материалов, Страсбург, Франция (2001 и 2002 г.г.); в Университете Гумбольдта в Берлине (2003 г.); в Университете Аугсбурга, Германия (2007 г.); в Северо-Западном Университете, Эванстон, Иллинойс, США (2011 и 2012 г.г.). Он является автором более чем 120 статей. Его научные интересы включают немарковские эффекты в четырехфотонной спектроскопии, когерентный оптический контроль, молекулярную электронику, нанофотонику и плазмонику. Борис Файнберг – приглашенный профессор в Школе Химии Тель-Авивского Университета, Тель-Авив, Израиль; полный профессор химической физики Холонского технологического института, Холон, Израиль, fainberg@hit.ac.il

Boris D. Fainberg graduated from LITMO (now ITMO University) in 1969. He received his PhD degree in 1977 at Vavilov State Optical Institute (St.-Petersburg, Russia) where he was employed as a junior research scientist, senior research scientist and associate professor at the department of molecular spectroscopy (till 1990). In 1991 he moved to the School of Chemistry at Tel-Aviv University as a scientific researcher till 1998 and then as a visiting professor beginning from 2000. In 1998 he joined the staff of the Faculty of Sciences at Holon Institute of Technology (Holon, Israel) as an associate professor and then (in 2003) as a full professor of Chemical Physics. He was a visiting professor at Institute of Atomic and Molecular Science, Taipei, Taiwan (2000), Institute of Physics and Chemistry of Materials, CNRS, Strasbourg (in 2001 and 2002), Humboldt University of Berlin (in 2003), University of Augsburg (in 2007), Northwestern University, Evanston, IL (in 2011 and 2012). He is an author of more than 120 papers. His research interests include non-Markovian effects in four-photon spectroscopy, coherent optical control, molecular electronics, nanophotonics and plasmonics. Boris D. Fainberg – visiting professor in the School of Chemistry at Tel-Aviv University, Tel-Aviv, Israel; professor of Chemical Physics in Holon Institute of Technology, Holon, Israel, fainberg@hit.ac.il



Федоров Сергей Васильевич – кандидат физ.-мат. наук, доцент кафедры Оптики лазеров Университета ИТМО, ведущий научный сотрудник ОАО «ГОИ им. С.И. Вавилова». Окончил физический факультет Ленинградского государственного университета в 1979 г. В 1993 г. работал в Техническом университете Кайзерслаутерна, Германия. В 2000 г. работал в Институте твердого тела и теоретической оптики Йенского университета имени Фридриха Шиллера, Германия. Область научных интересов – физическая и нелинейная оптика, лазерная физика, физика солитонов, нанооптика, спазеры.

Sergey V. Fedorov is a candidate of phys.&math. sciences, a leading researcher in Vavilov State Optical Institute and associate professor in ITMO University, Saint Petersburg, Russia. He graduated from the Physical Faculty of Leningrad State University in 1979. He was a guest researcher in University of Kaiserslautern, Germany, Physics department in 1993. He worked as a guest researcher in Friedrich-Shiller-University of Jena, Germany, Institute of Theory of Solid State and Theoretical Optics in 2000. The areas of scientific interests are: Physical and Nonlinear Optics, Laser Physics, Physics of solitons, Nanooptics, spasers. Sergey V. Fedorov – associate professor at the Department of Laser Optics, ITMO University, Saint Petersburg, Russia; leading researcher at Vavilov State Optical Institute, Saint Petersburg, Russia, sfedorov2006@bk.ru

- | | |
|--|--|
| <i>Веретенков Николай Александрович</i> | – младший научный сотрудник, ОАО ГОИ им. С.И. Вавилова, 199034, Санкт-Петербург, Россия; ведущий инженер, Университет ИТМО, 197101, Санкт-Петербург, Россия, torrek@gmail.com |
| <i>Левинский Борис Нахимович</i> | – кандидат физико-математических наук, наставник, физик-исследователь, наставник в физической лаборатории, Холонский Технологический Институт, 58102, Холон, Израиль, borisl@hit.ac.il |
| <i>Нестеров Леонид Андреевич</i> | – ведущий инженер, Университет ИТМО, 197101, Санкт-Петербург, Россия; старший научный сотрудник, ОАО ГОИ им. С.И. Вавилова, 199034, Санкт-Петербург, Россия, nesterovla@mail.ru |

- Розанов Николай Николаевич** – доктор физико-математических наук, профессор, член-корр. РАН, начальник отдела, ОАО ГОИ им. С.И. Вавилова, 199034, Санкт-Петербург, Россия; заведующий кафедрой, Университет ИТМО, 197101, Санкт-Петербург, Россия, nrosanov@yahoo.com
- Файнберг Борис Д.** – Ph.D., профессор, полный профессор, Холонский Технологический Институт, 58102, Холон, Израиль; приглашенный профессор, Тель-Авивский университет, 69978, Тель-Авив, Израиль, fainberg@hit.ac.il
- Фёдоров Сергей Васильевич** – кандидат физико-математических наук, доцент, доцент, Университет ИТМО, 197101, Санкт-Петербург, Россия; ведущий научный сотрудник, ОАО ГОИ им. С.И. Вавилова, 199034, Санкт-Петербург, Россия, sfedorov2006@bk.ru
- Nikolai A. Veretenov** – junior scientific researcher, Vavilov State Optical Institute, 199034, Saint Petersburg, Russia; leading engineer, ITMO University, 197101, Saint Petersburg, Russia, torrek@gmail.com
- Boris N. Levinsky** – Ph.D., tutor, scientific researcher, Holon Institute of Technology, 58102, Holon, Israel, borisl@hit.ac.il
- Leonid A. Nesterov** – leading engineer, ITMO University, 197101, Saint Petersburg, Russia; senior scientific researcher, Vavilov State Optical Institute, 199034, Saint Petersburg, Russia, nesterovla@mail.ru
- Nikolai N. Rosanov** – D.Sc., Professor, RAS Corresponding member, Section head, Vavilov State Optical Institute, 199034, Saint Petersburg, Russia; Department head, ITMO University, 197101, Saint Petersburg, Russia, nrosanov@yahoo.com
- Boris D. Fainberg** – PhD, full professor, Holon Institute of Technology, 58102, Holon, Israel; visiting professor, Tel-Aviv University, 69978, Tel-Aviv, Israel, fainberg@hit.ac.il
- Sergei V. Fedorov** – PhD, Associate professor, Associate professor, ITMO University, 197101, Saint Petersburg, Russia; leading scientific researcher, Vavilov State Optical Institute, 199034, Saint Petersburg, Russia, sfedorov2006@bk.ru

*Принято к печати 30.06.14
Accepted 30.06.14*

UDC 543.424.2, 538.97

SERS OF BACTERIORHODOPSIN WITH OUT-DIFFUSED SILVER
NANOISLANDSFabian Heisler^{a,b}, Ekaterina S. Piliugina^c, Semen D. Chervinskii^{d,e}, Anton K. Samusev^b, Andrey A. Lipovskii^{c,d}^a Friedrich-Schiller-Universität Jena, 07743, Jena, Germany, fabian.heisler@uni-jena.de^b ITMO University, 197101, Saint-Petersburg, Russia^c Saint-Petersburg State Academic Univesity, 194021, Saint-Petersburg, Russia^d Saint-Petersburg State Polytechnic University, 195251, Saint-Petersburg, Russia^e University of Eastern Finland, 80101, Joensuu, Finland

Abstract. We present the studies on surface-enhanced Raman spectroscopy (SERS) of bacteriorhodopsin in purple membranes using self-assembled silver nanoisland films for Raman signal enhancement. These metal island films were fabricated on soda-lime glass slides subjected to silver-sodium ion exchange in molten $\text{Ag}_{0.05}\text{Na}_{0.95}\text{NO}_3$ at the temperature of 325°C for 20 minutes and subsequent treatment in hydrogen atmosphere at the temperature of 250°C for 10 minutes. The films typically consisted of 20–30 nm closely placed nanoislands. Being tested as SERS substrates for rhodamine 6G the nanoisland films gave the possibility to observe respective characteristic Raman lines from a dried drop of rhodamine 6G dissolved in water in the concentration of 10^{-6} M. Similarly fabricated substrates were used to obtain SERS spectra of bacteriorhodopsin in purple membranes dispersed in water, and Raman peaks at $1000\text{--}1020\text{ cm}^{-1}$, $1150\text{--}1220\text{ cm}^{-1}$ and $1530\text{--}1570\text{ cm}^{-1}$ were resolved. The substrates made it possible to register characteristic Raman peaks only for an order of magnitude lower concentration of bacteriorhodopsin in contrast to the virgin glass substrate, that is the enhancement of Raman signal was considerably less than for rhodomin 6G. This is supposed to be due to bacteriorhodopsin molecules packing in patches, and it prevents bacteriorhodopsin in purple membranes from penetration between the nanoislands where the local enhancement of the electric field of exciting light wave is maximal.

Keywords: SERS, silver nanoisland films, bacteriorhodopsin, rhodamine 6G

Acknowledgements. This work was financially supported by the Ministry of Education and Science of the Russian Federation (Projects 11.G34.31.0020 and 16.1233.2014/K) and the Government of the Russian Federation, Grant 074-U01. For providing the samples of purple membranes and assistance with SERS experiments the authors are grateful to Vitaly Polovinkin, Valentin Borshchevskiy and Valentin Gordeliy (Laboratory for Advanced Studies of Membrane Proteins, Moscow Institute of Physics and Technology, Dolgoprudny, Russia; Institut de Biologie Structurale J.-P. Ebel, Grenoble, France; Institute of Complex Systems (ICS), ICS-6: Structural Biochemistry, Research Centre Juelich, Juelich, Germany; Institute of Crystallography, University of Aachen (Rheinisch-Westfälische Technische Hochschule), Aachen, Germany).

УДК 543.424.2, 538.97

ГКР СПЕКТРОСКОПИЯ МОЛЕКУЛ БАКТЕРИОРОДОПСИНА,
АДСОРБИРОВАННЫХ НА СЕРЕБРЯНЫЕ НАНООСТРОВКОВЫЕ ПЛЕНКИ
Ф. Хайслер^{a,b}, Е.С. Пилигина^c, С.Д. Червинский^{d,e}, А.К. Самусев^b, А.А. Липовский^{c,d}^a Йенский Университет имени Фридриха Шиллера, 07743, Йена, Германия, fabian.heisler@uni-jena.de^b Университет ИТМО, 197101, Санкт-Петербург, Россия^c Санкт-Петербургский Академический университет – научно-образовательный центр нанотехнологий РАН, 194021, Санкт-Петербург, Россия^d Санкт-Петербургский государственный политехнический университет, 195251, Санкт-Петербург, Россия^e Университет Восточной Финляндии, 80101, Йоэнсуу, Финляндия

Аннотация. Представлены результаты экспериментов по поверхностно-усиленной рамановской (ГКР) спектроскопии молекул бактериородопсина в пурпурных мембранах, адсорбированных на самоорганизованных серебряных наноструктурных пленках. Островки серебра формировались на поверхности натрий-кальциевого силикатного стекла в результате ионного обмена в расплаве смеси нитратов серебра и натрия, $\text{Ag}_{0.05}\text{Na}_{0.95}\text{NO}_3$, при температуре 325°C в течение 20 минут и последующем отжиге в водородной атмосфере при температуре 250°C в течение 10 минут. Типичные пленки состояли из расположенных близко друг к другу островков размером 20–30 нм. В ходе тестирования в качестве подложек для ГКР спектроскопии островковые пленки позволили наблюдать соответствующие характерные линии комбинационного рассеяния от высушенной капли водного раствора родамина 6G с концентрацией 10^{-6} моль/л. Аналогичным образом изготовленные подложки были использованы для исследования ГКР спектров бактериородопсина в пурпурных мембранах, диспергированных в воде. Нами были зарегистрированы линии комбинационного рассеяния в областях $1000\text{--}1020\text{ см}^{-1}$, $1150\text{--}1220\text{ см}^{-1}$ и $1530\text{--}1570\text{ см}^{-1}$. Подложки со наноструктурными пленками серебра позволили зарегистрировать характерные линии комбинационного рассеяния бактериородопсина с концентрацией лишь на порядок меньшей, чем в случае стеклянной подложки. Таким образом, усиление сигнала комбинационного рассеяния оказалось существенно меньшим, чем в случае родамина 6G. Предположительно, это связано с тем, что молекулы бактериородопсина упакованы в белковый кристалл, что не позволяет им проникать в промежутки между островками серебра, где локальное усиление электрического поля падающей волны максимально.

Ключевые слова: ГКР, серебряные наноструктурные пленки, бактериородопсин, родамин 6G

Благодарности. Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (гранты №№ 11.G34.31.0020 и 16.1233.2014/K), а также государственной финансовой поддержке ведущих

университетов Российской Федерации (субсидия 074-U01). За предоставление образцов пурпурных мембран авторы благодарны Виталию Половинкину, Валентину Борщевскому и Валентину Горделию (Лаборатория «Перспективные исследования мембранных белков» Московского физико-технического института, г. Долгопрудный, Россия; Институт структурной биологии, Гренобль, Франция; Институт сложных систем ICS-6: структурная биология, Исследовательский центр Юлиха, Юлих, Германия; Институт кристаллографии, Университет Аахена (Rheinisch-Westfälische Technische Hochschule), Аахен, Германия).

Introduction

Bacteriorhodopsin (BR) is a light-driven proton pump from *Halobacterium salinarum*: it captures light energy and uses it to move protons through the membranes out of biological cells [1]. This gives the possibility for the conversion of light into chemical energy. In nature BR exists in so-called purple membranes (PM) which have patches consisting of BR incorporated into lipids with an approximate mass ratio BR: lipids equal to 3:1 [1]. BR is well studied by various biophysical methods [1–4] and belongs to the class of membrane proteins, which are of great interest to the pharmaceutical industry [5, 6]. This paper is devoted to the observation of surface-enhanced Raman scattering (SERS) [7] from BR in PM. The SERS is assisted by self-assembled silver nanoisland films formed on the surface of an ion-exchanged glass.

Silver nanoisland films: fabrication and characterization

The technique used to fabricate the silver nanoisland film on a glass substrate includes two steps (Fig. 1): silver ion exchange into glass followed by thermal treatment of the glass in a hydrogen atmosphere [8]. We used commercially available Menzel microscope slides [9] as glass substrates. These soda-lime glass substrates are placed for 20 minutes in $\text{Ag}_{0.05}\text{Na}_{0.95}\text{NO}_3$ ion exchange batch at the temperature of 325°C (Fig. 1, a). This temperature is sufficient to activate the diffusion of alkali ions out of the glass and their replacement with silver ions from the melted silver-sodium nitrate. This results in the enrichment of the subsurface layer of the glass with silver ions.

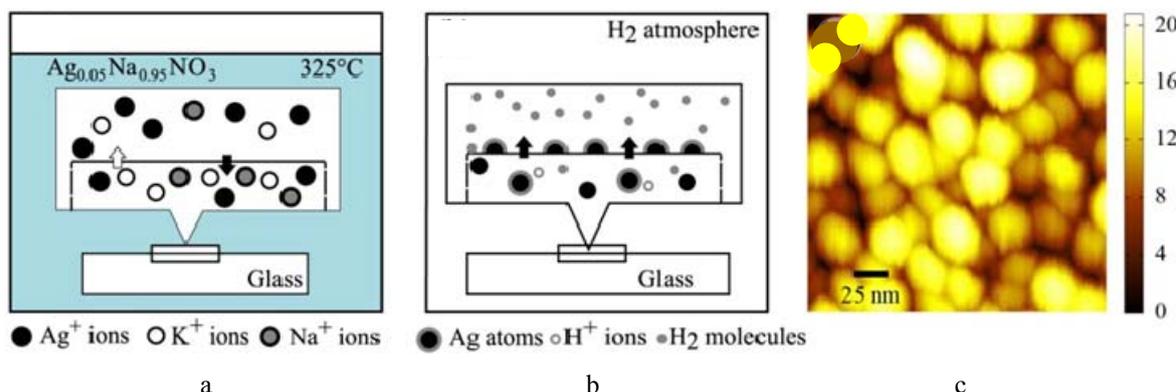
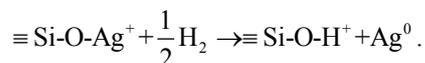


Figure 1: Silver nanoisland film fabrication: ion exchange between glass and molten salt (a); thermal treatment in hydrogen atmosphere (b); Typical AFM picture of the fabricated nanoisland film used for SERS measurements (c)

After the ion exchange the glass substrate is treated for 10 minutes in hydrogen atmosphere at the temperature of 250°C (Fig. 1, b). Hydrogen diffuses into the subsurface glass layer, reducing silver ions via the following reaction:



Due to the low solubility of neutral silver in the glass the silver atoms coagulate and form nanoparticles. In this soft mode of hydrogen treatment the nanoparticles are being formed on the glass surface rather than in the bulk of the glass and the self-arrangement of the formed nanoparticles results in the formation of the silver nanoisland film. A typical atomic force microscopy (AFM) image of the surface of the glass film with nanoislands is shown in Fig. 1, c. The size of the nanoislands obtained in the regime described above is about 20–30 nm.

Silver nanoisland films as SERS substrates. For the measurements of the Raman spectra we used the confocal Raman spectrometer Horiba LabRAM™ HR UV-VIS-NIR and a frequency-doubled continuous-wave Nd:YAG laser with a wavelength of 532 nm as a light source. The laser beam was focused at the sample surface with a Mitutoyo M Plan Apo 50x/0.55 objective lens. The beam waist was approximately 1.5 μm .

To perform the Raman experiments, we deposited drops of the analytes dissolved in water in different concentrations on the surface of the prepared glasses and dried them at room temperature; the diameter of the so formed spots of the analytes was about 7 mm. All the experiments were carried out with analytes deposited both on the nanoisland film and virgin glass surface.

To verify the ability of the prepared silver nanoisland films to enhance the Raman signal, we used rhodamine 6G (R6G) dissolved in water in the concentration of 10^{-6} M. The spectrum shown in Fig. 2 was

acquired from R6G on silver nanoisland film at 30 s exposition and averaged over 20 measurements while using 20 nW exciting laser radiation power at the sample. The observed characteristic peaks at 612 cm^{-1} , 772 cm^{-1} , 1362 cm^{-1} , 1510 cm^{-1} and 1647 cm^{-1} are in good agreement with literature data [10]. Similar BR SERS results were reported earlier in our paper related to silver nanoislands formed on the surface of a silver-containing phosphate glass using out-diffusion technique [11]. The reference spectra obtained from R6G deposited on a virgin SERS-inactive glass substrate demonstrated a strong luminescent background which gave no possibility to distinguish any Raman peaks and to estimate the degree of the enhancement provided by the nanoisland film.

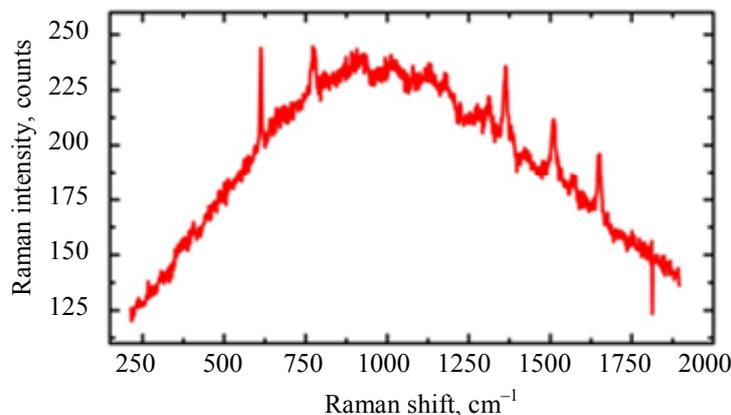


Figure 2: SERS spectrum from R6G on nanoisland film acquired at laser power of 20 nW, 30 s exposition.

Raman scattering of bacteriorhodopsin. The main goal of this study was to register SERS spectra of BR in PM assisted by fabricated silver nanoisland film. The experiments have been carried out using BR in concentrations of 0.48 and 0.048 mg/ml, see Fig. 3. We used $200\text{ }\mu\text{W}$ exciting laser power and acquired the spectrum within 50 s averaging it over 2 measurements. For the sample with a concentration of 0.48 mg/ml (Fig. 3, a) the observed Raman peaks at $1000\text{--}1020\text{ cm}^{-1}$, $1150\text{--}1220\text{ cm}^{-1}$ and $1530\text{--}1570\text{ cm}^{-1}$ are in good correspondence to the literature [12]. Raman peaks were registered using both virgin glass and the nanoisland film. Fig. 3, a, clearly shows that the nanoisland film makes it possible to resolve more Raman peaks than the virgin glass.

Raman measurements were performed using BR in concentration of 0.048 mg/ml. We have demonstrated that the spectrum obtained using the non-SERS-active glass substrate contains only background without any distinguishable peaks (Fig. 3, b). At the same time the Raman spectrum of BR acquired using the nanoisland film shows vivid peaks around 1020 cm^{-1} and 1530 cm^{-1} .

It is important to mention that BR molecules are packed in PM patches consisting of approximately 1000 molecules [1]. The thickness of each patch is about 5 nm and corresponds to the size of BR molecules, whereas the lateral size of each patch is of the order of several hundreds of nanometers. So BR molecules in PM are not located between the nanoislands where the local enhancement of the electric field in the light wave is maximal. These results in a weaker Raman signal were compared to the ones obtained for R6G molecules.

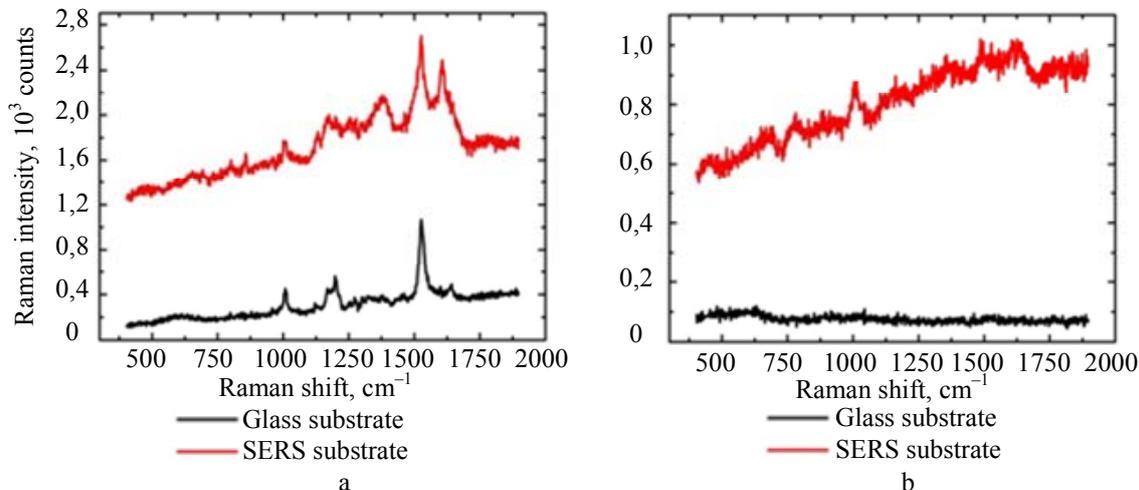


Figure 3. Raman spectra of BR in the concentrations of (a) 0.48 and (b) 0.048 mg/ml. The black and red curves correspond to reference Raman spectra of BR on virgin glass substrate and SERS spectra of BR deposited on the nanoisland film respectively. The spectra were acquired at laser power of $200\text{ }\mu\text{W}$, with an exposition of 50 s

Conclusion

Finally, we have fabricated silver nanoisland films by means of hydrogen/metal ions reactive diffusion in glasses followed by the transport of neutral metal to the glass surface. This gave the possibility to obtain uniform nanoislands with the size of the order of tens of nanometers.

We have performed SERS measurements of R6G in order to verify the capability of the fabricated silver nanoisland films to deliver enhancement for Raman spectroscopy. The BR in PM spot formed after drying a 5 μ l drop with BR in PM in concentration of 0.48 mg/ml demonstrated selective enhancement of Raman peaks already observed in experiments without SERS substrate. For an order of magnitude with lower concentration of BR in PM the usage of the fabricated silver nanoisland film gives the possibility to resolve characteristic Raman peaks in contrast to the virgin glass substrate.

References

1. Lanyi J.K. Bacteriorhodopsin // *Annual Review of Physiology*. 2004. V. 66. P. 665–688.
2. Maeda A. Application of FTIR spectroscopy to the structural study on the function of bacteriorhodopsin // *Israel Journal of Chemistry*. 1995. V. 35. N 3-4. P. 387–400.
3. Nabiev I.R., Efremov R.G., Chumanov G.D. The chromophore-binding site of bacteriorhodopsin. Resonance Raman and surface-enhanced resonance Raman spectroscopy and quantum chemical study // *Journal of Biosciences*. 1985. V. 8. N 1-2. P. 363–374.
4. Mathies R.A., Lin S.W., Ames J.B., Pollard W.T. From femtoseconds to biology: mechanism of bacteriorhodopsin's light-driven proton pump // *Annual Review of Biophysics and Biophysical Chemistry*. 1991. V. 20. P. 491–518.
5. Krogh A., Larsson B., Heijne v.G., Sonnhammer E.L.L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes // *Journal of Molecular Biology*. 2001. V. 305. N 3. P. 567–580.
6. Overington J.P., Al-Lazikani B., Hopkins A.L. How many drug targets are there? // *Nature Reviews Drug Discovery*. 2006. V. 5. N 12. P. 993–996.
7. Kneipp K., Moskowitz M., Kneipp H. *Surface Enhanced Raman Scattering. Physics and Applications*. NY: Springer, 2006. 460 p.
8. Chervinskii S., Sevriuk V., Reduto I., Lipovskii A. Formation and 2D-patterning of silver nanoisland film using thermal poling and out-diffusion from glass // *Journal of Applied Physics*. 2013. V. 114. N 22. Art. 224301.
9. Menzel-Glaser: Microscope slides [Электронный ресурс]. Режим доступа: <http://www.menzel.de/Microscope-Slides.687.0.html?&L=1>, свободный. Яз. англ. (дата обращения 16.06.2014)
10. Zhou Q., Li Z., Yang Y., Zhang Z. Arrays of aligned, single crystalline silver nanorods for trace amount detection // *Journal of Physics D: Applied Physics*. 2008. V. 41. N 15. Art. 152007.
11. Zhurikhina V.V., Brunkov P.N., Melehin V.G., Kaplas T., Svirko Yu., Rutckaia V.V., Lipovskii A.A. Self-assembled silver nanoislands formed on glass surface via out-diffusion for multiple usages in SERS applications // *Nanoscale Research Letters*. 2012. V. 7. N 1. P. 676.
12. Terner J., Champion A., El-Sayed M.A. Time-resolved resonance Raman spectroscopy of bacteriorhodopsin on the millisecond timescale // *Proc. National Academy of Sciences of the United States of America*. 1977. V. 74. N 12. P. 5212–5216.

- Fabian Heisler** – student, Friedrich-Schiller-Universität Jena, 07743, Jena, Germany; Trainee researcher, ITMO University, 197101, Saint Petersburg, Russia, fabian.heisler@uni-jena.de
- Ekaterina S. Piliugina** – student, Saint Petersburg State Academic University-Scientific Educational Center of Nanotechnologies RAS, 194021, Saint Petersburg, Russia, piliugina.ekaterina@gmail.com
- Semen D. Chervinskii** – engineer, Saint Petersburg State Polytechnic University, 195251, Saint Petersburg, Russia; early-stage researcher, University of Eastern Finland, 80101, Joensuu, Finland
- Anton K. Samusev** – PhD, scientific researcher, ITMO University, 197101, Saint Petersburg, Russia, a.samusev@phoi.ifmo.ru
- Andrey A. Lipovskii** – D.Sc., Professor, Professor, Saint Petersburg State Academic University, 194021, Saint Petersburg, Russia; Saint Petersburg State Polytechnic University, 195251, Saint Petersburg, Russia, lipovsky@spbau.ru
- Хайслер Фабиан** – студент, Йенский Университет имени Фридриха Шиллера, 07743, Йена, Германия; стажёр-исследователь, Университет ИТМО, 197101, Санкт-Петербург, Россия; fabian.heisler@uni-jena.de

- Пилюгина Екатерина Сергеевна* – студент, Санкт-Петербургский Академический университет – научно-образовательный центр нанотехнологий РАН, 194021, Санкт-Петербург, Россия, piliugina.ekaterina@gmail.com
- Червинский Семен Дмитриевич* – инженер, Санкт-Петербургский государственный политехнический университет, 195251, Санкт-Петербург, Россия; младший исследователь, Университет Восточной Финляндии, 80101, Йоэнсуу, Финляндия; semen.chervinsky@gmail.com
- Самусев Антон Кириллович* – кандидат физ.-мат. наук, научный сотрудник, Университет ИТМО, 197101, Санкт-Петербург, Россия, a.samusev@phoi.ifmo.ru
- Липовский Андрей Александрович* – доктор физико-математических наук, профессор, Санкт-Петербургский Академический университет – научно-образовательный центр нанотехнологий РАН, 194021, Санкт-Петербург, Россия; профессор, Санкт-Петербургский государственный политехнический университет, 195251, Санкт-Петербург, Россия; lipovsky@spbau.ru

Принято к печати 26.06.14

Accepted 26.06.14

УДК 538.913 538.971

АБЛЯЦИЯ ТВЕРДОГО ТЕЛА ПОД ДЕЙСТВИЕМ СВЕРХКОРОТКИХ
ЛАЗЕРНЫХ ИМПУЛЬСОВ: ИССЛЕДОВАНИЕ МЕТОДАМИ
МОЛЕКУЛЯРНОЙ ДИНАМИКИ

Д.С. Иванов^{а,с}, В.П. Вейко^б, Е.Б. Яковлев^б, Б. Ретфельд^а, М.Э. Гарсия^с

^а Технический университет Кайзерслаутерн и исследовательский центр ОПТИМАС, 67653, Кайзерслаутерн, Германия

^б Университет ИТМО, 197101, Санкт-Петербург, Россия, ebyakovlev@yandex.ru

^с Университет Кассель, 34109, Кассель, Германия

Аннотация. Методами молекулярной динамики исследован процесс лазерной абляции при воздействии на металлы сверхкоротких лазерных импульсов. Выясняется действенность и применимость гибридной атомистически-непрерывной модели для оценки оптимальных режимов обработки сверхкороткими лазерными импульсами. Предложено сочетание атомистической модели лазерно-индуцированного процесса неравновесного фазового перехода на атомном уровне с непрерывной двухтемпературной моделью для описания динамики фотовозбуждения свободных носителей. Показана применимость модели лазерной абляции на примере пленок алюминия и золота при воздействии импульсов с различной плотностью энергии и длительностью. Показано, что в зависимости от соотношения длительности лазерного импульса и характерного времени электрон-фононного взаимодействия материала реализуются фототермический и фотомеханический режимы разрушения, которые определяют качество и производительность лазерной обработки. Установлено, что при длительности лазерного импульса меньше времени электрон-фононного взаимодействия реализуется высокопроизводительный фотомеханический характер разрушения за счет возникших в области воздействия внутренних напряжений. Это подтверждается линейной зависимостью скорости абляции от поглощенной энергии. При длительности лазерного импульса больше времени электрон-фононного взаимодействия реализуется малопродуктивный фототермический режим разрушения. Результаты работы могут быть полезны специалистам, занимающимся разработкой лазерных технологий.

Ключевые слова: лазерная абляция, сверхкороткие лазерные импульсы, молекулярная динамика, фототермическое разрушение, фотомеханическое разрушение.

Благодарности. Работа выполнена при поддержке гранта Президента Российской Федерации НШ-1364.2014.2, грантов РФФИ 12-02-01194, 13-02-00033, 13-02-00971 и при государственной финансовой поддержке ведущих университетов Российской Федерации (субсидия 074-U01) и грантов DFG IV 122/1-1 и IV 122/1-2.

SOLID BODY ABLATION UNDER EXPOSURE TO ULTRA SHORT LASER
PULSES: STUDY BY MOLECULAR DYNAMICS METHODS

D.S. Ivanov^{а,с}, V.P. Veiko^б, E. B. Yakovlev^б, B. Rethfeld^а, M.E. Garcia^с

^а Technical University of Kaiserslautern and OPTIMAS research center, 67653, Kaiserslautern, Germany

^б ITMO University, 197101, Saint Petersburg, Russia, ebyakovlev@yandex.ru

^с University of Kassel, 34109, Kassel, Germany

Abstract. The process of laser ablation under the influence of ultra short laser pulses on metals is investigated by methods of molecular dynamics. The validity and applicability of the hybrid atomistic-continuous model for the estimation of optimum modes of ultra short laser pulses processing are explored. Combination of atomistic model of laser-induced non-equilibrium process of a phase transition at the atomic level with continuous two-temperature model for describing the dynamics of photo-excitation of free media is proposed. Applicability of laser ablation model on the example of aluminum films and gold under exposure to pulses with different energy density and duration is shown. It is indicated that, depending on the ratio of the laser pulse duration and the characteristic time of electron-phonon material interaction, photothermal and photomechanical modes of destruction are implemented that determine the quality and performance of the laser processing. It is established that at the duration of laser pulse less than the time of electron-phonon interaction high-performance photomechanical type of destruction is implemented by internal stresses arising in the area of exposure. This is confirmed by a linear dependence of the ablation rate from the absorbed energy. At the duration of laser pulse greater than the time of electron-phonon interaction inefficient photothermal mode of destruction is implemented. The results may be useful for specialists engaged in the development of laser technologies.

Keywords: laser ablation, ultra short laser pulses, molecular dynamics, photothermal destruction, photomechanical destruction.

Acknowledgements. The work is done under the grant support of the President of the Russian Federation NSH-1364.2014.2, RFBR grants 12-02-01194, 13-02-00033, 13-02-00971, and is partially financially supported by the Government of the Russian Federation (grant 074-U01), grants DFG IV 122/1-1 and IV 122/1-2.

Введение

Абляция при воздействии сверхкоротких лазерных импульсов широко используется в промышленности. Возможность локализации большого количества энергии в малом объеме хорошо зарекомендовала себя в качестве эффективного инструмента при лазерном сверлении, резке, сварке [1–3]. Оказалось, что

качество обработки сильно зависит от параметров лазерного воздействия. Например, было показано как экспериментально, так и теоретически, что время электрон-фононного взаимодействия, энергия и длительность импульсов могут играть важную роль в механизме лазерного разрушения металлов [4, 5]. В соответствии с этими работами, индуцированное изменение вещества, в том числе процессы плавления, откальвания, испарения и абляции, в целом определяются соотношением длительности импульса с характерным временем электрон-фононного взаимодействия. В других работах предполагается, что при оценке результата воздействия не менее важны свойства материала и другие параметры. Так, в ряде экспериментов [6, 7] четко определены два режима скорости абляции при низких и высоких плотностях энергии для таких металлов, как золото и серебро. Было сделано предположение, что низкая и высокая скорость абляции, в зависимости от поглощенной энергии, определяется фототермическим или фотомеханическим механизмом разрушения [8, 9].

Теоретические расчеты подтвердили, что качество и производительность обработки определяют преобладание фотомеханического или фототермического режима разрушения. Так, в частности, было определено, что условия возникновения локализации напряжений, реализующиеся при воздействии сверхкоротких лазерных импульсов лазерного излучения на вещество, являются основным фактором реализации разрушения отколом, которое носит скорее механический, нежели тепловой характер [10, 11]. При этом глубина удаления материала была больше, что является показателем эффективности обработки, с минимизацией теплового воздействия, а это является показателем качества. Условие для наступления соответствующего лазерного воздействия определяется соотношением [12, 13]

$$\max \{t_p, t_{e-ph}\} \approx t_{heating} \leq t_{mech} \approx \frac{L_{e-diff}}{C_s}, \quad (1)$$

где t_{mech} – характерное время механической разгрузки внутренних напряжений – оказывается больше, чем время нагрева $t_{heating}$, которое определяется максимальной из двух величин: либо длительностью импульса t_p , либо характерным временем электрон-фононной релаксации t_{e-ph} . Время механической разгрузки, с другой стороны, зависит от эффективной глубины поглощения лазерной энергии L_{e-diff} и скорости звука в материале C_s .

Хотя уравнение (1) и определяет условия для реализации механизма откола, к сожалению, его трудно применить для прогнозирования обработки материала напрямую, если говорить об эффективности и качестве. Известно, что существуют два основных механизма передачи поглощенной лазерной энергии в материал: быстрая теплопроводность горячих электронов, определяющая глубину распределения поглощенной энергии, и электрон-фононный теплообмен, определяющий скорость нагревания решетки [14]. Их зависимость от параметров облучения и свойств материала была недавно рассмотрена на основе упрощенного теоретического подхода с помощью двухтемпературной модели (ДТМ) в [12]. Важно, что свойства материала, такие как электронная теплопроводность K_e , электронная теплоемкость C_e , и параметр электрон-фононного взаимодействия G_{e-ph} , являются функциями температуры электронов [15–17] (см. рис. 1). Видно, что параметр электрон-фононного взаимодействия увеличивается с ростом температуры для обоих металлов. Для золота это увеличение сильнее выражено из-за наличия широкой d -зоны в его зонной структуре. С другой стороны, зависимости электронной теплопроводности от температуры для этих материалов значительно отличаются. Благодаря гораздо большей температуре Ферми (135 000 К) алюминия его теплопроводность линейно растет с электронной температурой, по крайней мере, до 30 000 К, в то время как для золота, обладающего сравнительно низкой температурой Ферми (56 000 К), на механизмы теплопроводности оказывает сильное влияние электрон-электронное взаимодействие уже при 20 000 К, что приводит к уменьшению ее значения. Динамика электронной температуры, в свою очередь, сильно зависит от длительности импульсов и поглощенной энергии. Исходя из этого, при таких сложных связях между свойствами материала и электронной температурой реализация одного или другого механизмов абляции в зависимости от параметров облучения неочевидна.

В работе с помощью гибридной атомистически-непрерывной модели, основанной на молекулярно-динамическом методе (МД), мы пытаемся выяснить действенность и применимость сделанных оценок, в частности, с помощью формулы (1), при определении параметров лазерного излучения для оптимизации режимов обработки сверхкороткими лазерными импульсами.

Модель

Идея построения гибридной модели основывается на способности МД метода, с одной стороны, отслеживать кинетику сверхбыстрых неравновесных фазовых переходов на атомном уровне и, с другой стороны, с помощью ДТМ в сплошной среде описывать динамику фотовозбужденных свободных носителей в твердом теле. В комбинированной МД-ДТМ-модели уравнение для температуры решетки ДТМ полностью заменяется уравнением движения атомов МД, а энергетический обмен между решеткой и горячими электронами учитывается в МД-части с помощью дополнительного коэффициента трения. Подробное описание МД-ДТМ-модели приведено в [5, 18]. Параметризация потенциала межатомного взаимодействия проведена согласно [19], что дает более точное воспроизведение свойств материала. Таким

образом, построенная модель, кроме присущих МД модели возможностей описания кинетики неравновесных процессов с атомным разрешением, способна описывать поглощение лазерного излучения, быструю электронную теплопроводность и электрон-фононную неравновесность. В связи с этим она может быть использована для исследования процессов, индуцированных сверхкоротким лазерным импульсом, включая плавление, испарение, отщепление (спалацию, откол) и абляцию тонких и толстых металлических мишеней [20].

Для выполнения задач нашего исследования мы выбрали два очень разных металла: алюминий и золото. Они относятся к так называемым переходной и благородной группам металлов соответственно. Алюминий имеет относительно малое (1 пс) характерное время электрон-фононной релаксации, а золото – большое (20 пс). В общем случае данные времена являются функцией плотности поглощенной энергии, однако в любом случае они различаются как минимум на порядок по величине. МД-ДТМ-вычислительные ячейки для обоих материалов состоят из $20 \times 20 \times 400$ узлов кристаллической решетки в направлениях X , Y и Z -осей, что соответствует расчету в общей сложности для 640 000 атомов. Следовательно, для параметров решетки 0,405643 нм для алюминия и 0,408472 нм для золота в МД-части построенная вычислительная сетка представляет пленку около 160 нм толщины. Однако, если мы применим динамические граничные условия для передачи давления на границе МД-части [21], то получим возможность, как было показано в недавней работе [22], моделирования взаимодействия лазерного излучения с толстыми металлическими мишенями. Так, непрерывная часть модели обеспечивает расчет теплового потока на расстоянии более 20 мкм в глубину. Для надежного определения параметров (температура, давление, плотность и т.д.) в МД-части модели принята сетка с размером ячейки 1 нм, что дает достаточно высокое пространственное разрешение для ДТМ-части модели и по-прежнему удовлетворяет размеру приемлемого статистического ансамбля: усреднение определяемых теплофизических величин ведется также по всем соседним ячейкам МД-части, насчитывающим в целом около 4000 частиц. Наконец, мы применяем свободные граничные условия на верхней границе образца для действия лазерного импульса и периодические граничные условия в направлениях X и Y , перпендикулярных направлению распространения лазерного импульса.

Результаты

Чтобы показать различие в развитии разрушения, вызванного воздействием лазерного излучения у двух выбранных материалов, мы воздействуем на алюминий импульсами длительностью t_p , равной 0,2 пс и 6,0 пс, а на золото – 6,0 пс и 60 пс. Таким образом, в обоих случаях мы имеем длительность лазерного воздействия значительно ниже и выше характерного времени электрон-фононного взаимодействия, даже с учетом его зависимости от плотности поглощенной лазерной энергии. Более того, судя по поведению функции электронной температуропроводности золота, существуют два крайних случая с точки зрения распределения поглощенной лазерной энергии по глубине (рис. 1): локализация энергии в непосредственной близости от поверхности материала, вследствие ограничения влияния теплопроводности, и увеличение глубины поглощения при высокой степени возбуждения носителей поглощенной энергией. Последовательные материалы будут подвергнуты воздействию лазерного излучения с различной плотностью поглощенной энергии Q , что и послужит выявлению режимов, определяющих механизм разрушения.

Мы начнем с простого случая лазерной абляции алюминия. Так как этот материал обладает характерным временем электрон-фононной релаксации 1 пс, то импульс длительностью 0,2 пс можно рассматривать как сверхкороткий, а импульс 6,0 пс – как длинный. Результаты взаимодействия этих импульсов с алюминиевым образцом представлены в виде полей температуры и давления в веществе на рис. 2, 3 для сверхкороткого (рис. 2, а, и рис. 3, а) и длинного (рис. 2, б, и рис. 3, б) импульса соответственно. Видно, что результаты воздействия этих двух импульсов сильно отличаются для одного и того же материала и одной и той же поглощенной энергии. Объяснить это можно на основе рис. 1, б, где приведены зависимости электронной теплопроводности и коэффициента электрон-фононного взаимодействия для алюминия от электронной температуры. В частности, при воздействии короткого импульса электронный газ нагревается до более высоких температур (около 30 000 К) (рис. 1, б, зеленый прямоугольник), что, в свою очередь, приводит к большей теплопроводности электронов. В связи с этим в случае коротких импульсов поглощенная энергия оказывается распределенной на большей глубине, на которой устанавливаются градиенты температур и давления. Кроме того, поскольку импульс короче, чем время электрон-фононного взаимодействия, то материал не имеет достаточно времени для теплового расширения в ответ на быстрый нагрев. Таким образом, в твердом теле происходит формирование области больших внутренних напряжений, что приводит к высокому давлению при относительно низкой температуре. Релаксация внутренних напряжений будет создавать при разгрузке волну давления, проходящую сквозь материал. Это может вызвать его расслоение на определенной глубине (даже в нескольких местах) там, где растягивающие напряжения при заданной температуре решетки достаточно велики. Такого рода фотомеханические повреждения приводят к большей абляции, т.е. более эффективны с точки зрения производительности и каче-

ства (чистоты удаления материала) вследствие меньшего термического повреждения окружающего материала (рис. 2, а, и рис. 3, а).

Противоположная ситуация возникает при воздействии лазерного импульса длительностью 6,0 пс с тем же материалом при той же поглощенной энергии. Поскольку импульс длинный, то в этом случае максимальная электронная температура из-за меньшей плотности мощности будет меньше (около 20 000 К), что, в свою очередь, снижает теплопроводность (рис. 1, а, красный прямоугольник). Как следствие, градиенты температуры и давления формируются на меньшей глубине по сравнению с предыдущим случаем, к тому же условия локализации формирования внутренних напряжений согласно формуле (1) в этом случае не выполняются. Высокая температура существует на меньшей глубине, что приводит к снижению давления и создает условия для преобладания термического разрушения. В этом режиме релаксация внутренних напряжений, сформировавшихся в приповерхностной области, вызывает активное разбрызгивание расплавленного материала, вклад механизма испарения в разрушение увеличивается, что приводит к уменьшению глубины абляции. Такой режим обработки с точки зрения эффективности и качества (рис. 2, б, и рис. 3, б) невыгоден.

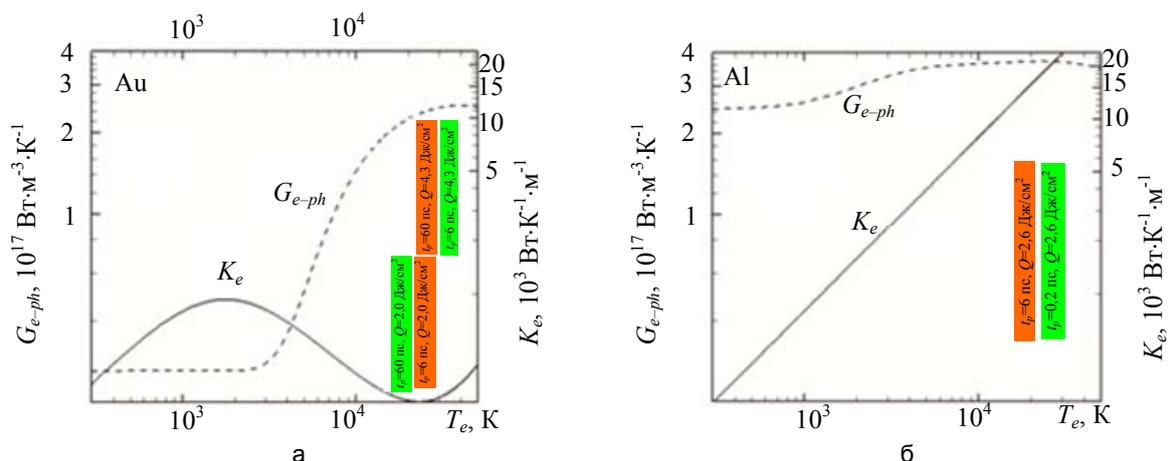


Рис. 1. Электронная теплопроводность K_e (сплошная линия) и коэффициент электрон-фононного взаимодействия G_{e-ph} (штриховая линия) для золота (а) и алюминия (б), как функция электронной температуры T_e [15–17] (температура атомной решетки $T = 300 \text{ К}$). Красные или зеленые прямоугольники указывают соответственно тепловой или механический режим разрушения, полученные в результате моделирования, для различных плотностей поглощенной энергии и времени воздействия лазерного импульса, используемых в работе. Положения прямоугольников по горизонтали указывают на характерные значения изменения электронной температуры в процессе моделирования

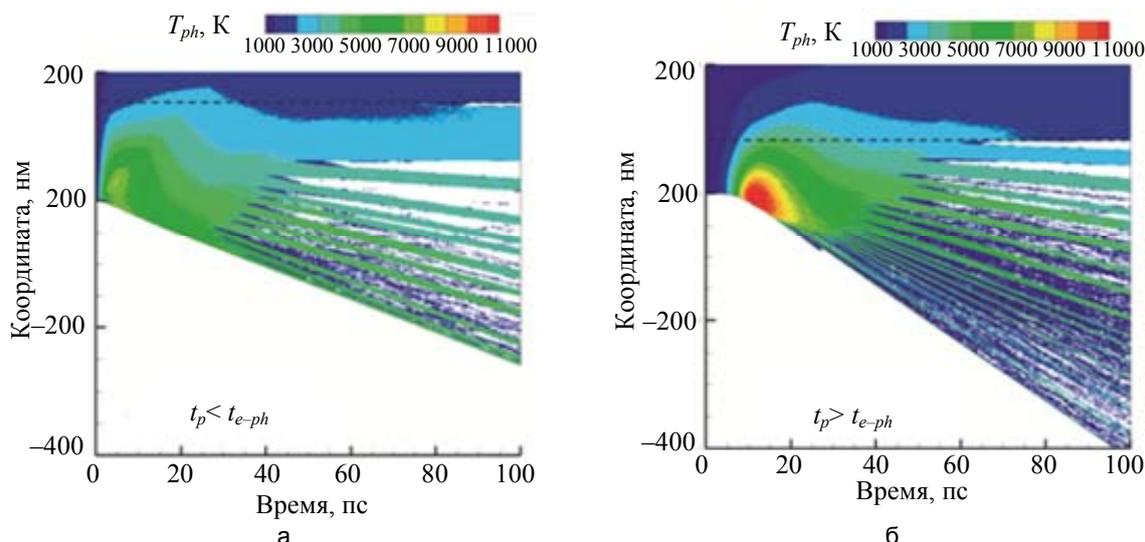


Рис. 2. Температурные поля T_{ph} внутри материала при воздействии импульсов длительностью 0,2 пс (а) и 6 пс (б) с толстой алюминиевой пленкой, при поглощенной плотности энергии 2,6 Дж/см². Пунктирная линия – глубина максимального разрушения

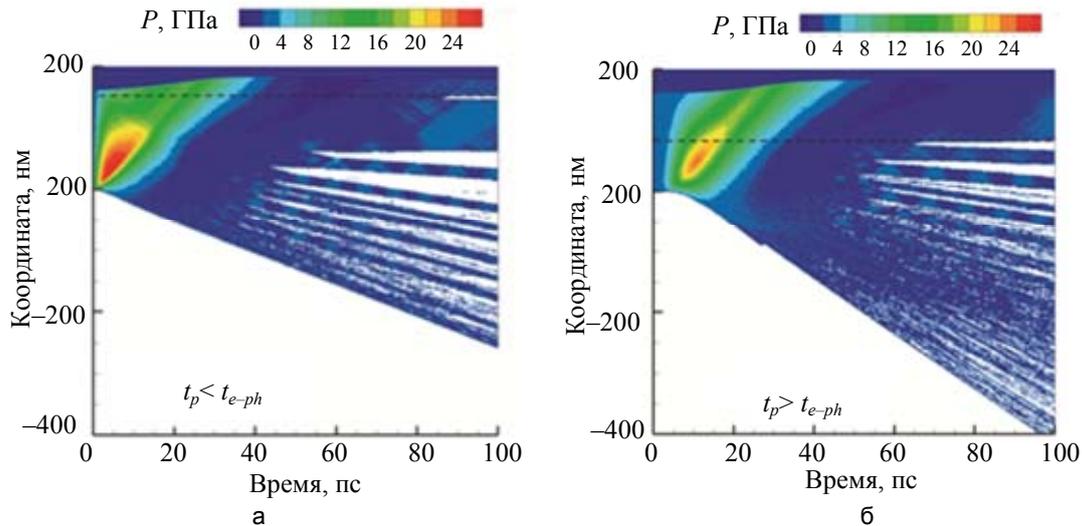


Рис. 3. Поле давлений P внутри материала при воздействии импульсов длительностью 0,2 пс (а) и 6 пс (б) с толстой алюминиевой пленкой, при поглощенной плотности энергии 2,6 Дж/см². Пунктирная линия – глубина максимального разрушения

Далее мы рассмотрим воздействие лазерных импульсов длительностью 6,0 и 60 пс на золото при поглощенной плотности энергии 2,0 Дж/см². Поскольку характерное время электрон-фононной релаксации золота 20 пс, то первый импульс можно считать коротким, а второй – длинным. Рассчитанные поля температур и давлений внутри материала для этого случая приведены на рис. 4, 5. На первый взгляд получен результат, противоположный предыдущему. Однако, если учесть изменения электронной теплопроводности и электрон-фононного взаимодействия для золота от электронной температуры (рис. 1, а), становится ясно, что чем короче импульс, тем выше электронная температура при одинаковой поглощенной энергии (рис. 4, а, и рис. 5, а). Достигнутая электронная температура (до 30 000 К) приводит к малой электронной теплопроводности при большом коэффициенте электрон-фононного взаимодействия из-за возбуждения электронов из d -зоны, что, в конечном счете, определяет малую глубину поглощения энергии (рис. 1, а, красный нижний прямоугольник). Это приводит к значительному тепловому разрушению образца при небольшой глубине удаления материала. При длительности же импульса 60 пс, хотя достигаемая электронная температура (20 000 К) ниже, электронная теплопроводность больше при меньшем коэффициенте электрон-фононного взаимодействия (рис. 1, а, зеленый прямоугольник). Таким образом, несмотря на то, что условия локализации формирования внутренних напряжений в данном случае не выполняются, характер разрушений больше обуславливается существующими напряжениями, приводящими к большему фотомеханическому повреждению при меньшем участии фототеплового (рис. 4, б, и рис. 5, б).

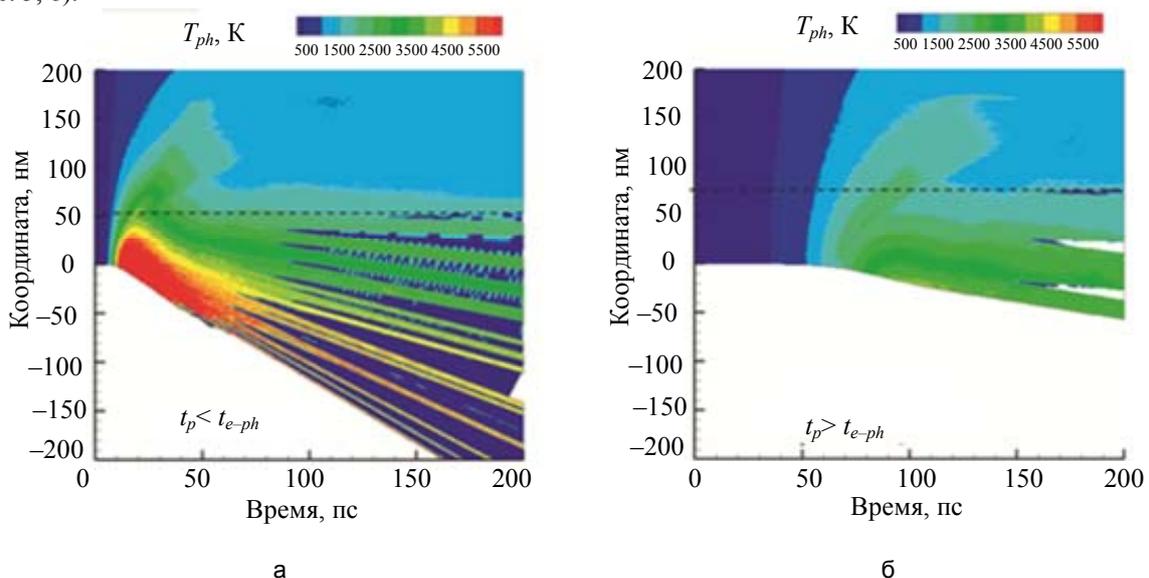


Рис. 4. Температурные поля внутри материала при воздействии импульсов длительностью 6,0 пс (а) и 60 пс (б) с толстой золотой пленкой, при поглощенной плотности энергии 2,0 Дж/см². Пунктирная линия – глубина максимального разрушения

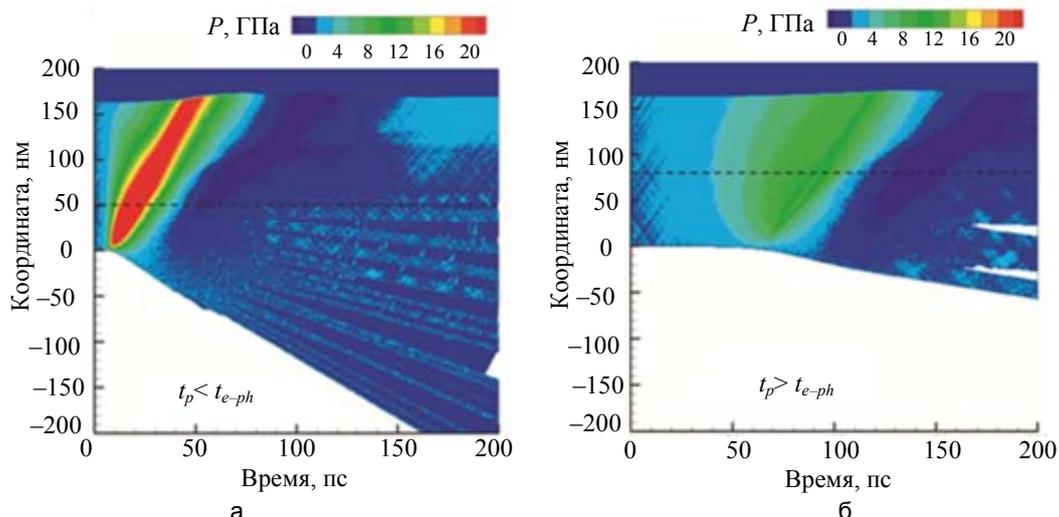


Рис. 5. Поле давлений внутри материала при воздействии импульсов длительностью 6,0 пс (а) и 60 пс (б) с толстой золотой пленкой, при поглощенной плотности энергии 2,0 Дж/см². Пунктирная линия – глубина максимального разрушения

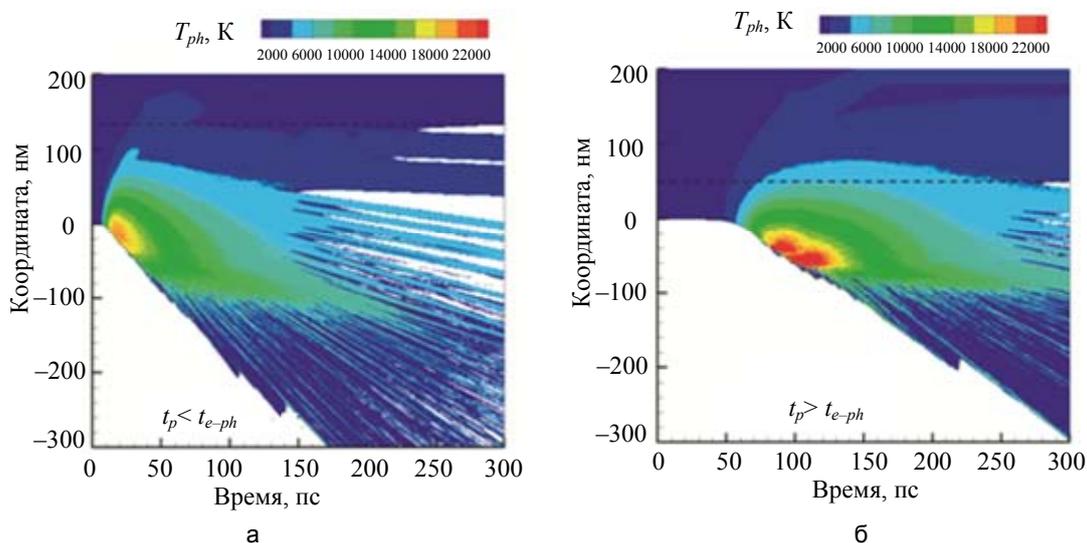


Рис. 6. Температурные поля внутри материала при воздействии импульсов длительностью 6,0 пс (а) и 60 пс (б) с толстой золотой пленкой, при поглощенной плотности энергии 4,3 Дж/см². Пунктирная линия – глубина максимального разрушения

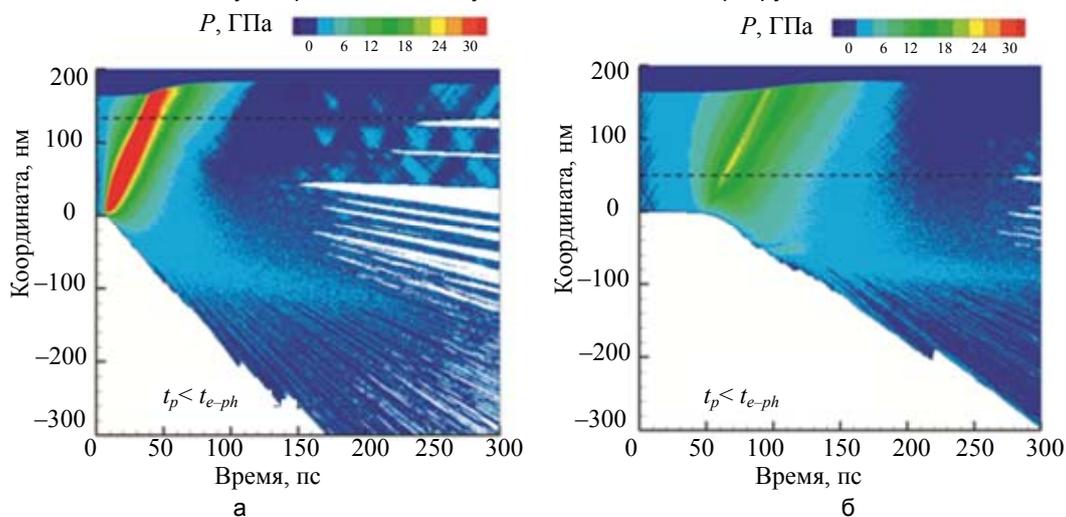


Рис. 7. Поле давлений внутри материала при воздействии импульсов длительностью 6,0 пс (а) и 60 пс (б) с толстой золотой пленкой, при поглощенной плотности энергии 4,3 Дж/см². Пунктирная линия – глубина максимального разрушения

Ситуация меняется при увеличении поглощенной плотности энергии до 4,3 Дж/см² (рис. 6, 7). При длительности импульса 6 пс электронная температура увеличивается, что вызывает резкий рост электронной теплопроводности золота (рис. 1, а, верхний зеленый прямоугольник), а коэффициент электрон-фононного взаимодействия при этом остается приблизительно таким же. Создаются условия для формирования локализации внутренних напряжений, в результате реализуется фотомеханическое разрушение с высокой производительностью и хорошим качеством (рис. 6, а, и рис. 7, а). Напротив, длинный импульс приводит к тепловому режиму, что определяет фототермическое разрушение облученных образцов с малой производительностью и неудовлетворительным качеством (рис. 6, б, и рис. 7, б).

Заключение

Обобщая полученные результаты, можно сформулировать обобщенный критерий преобладания фотомеханического режима разрушения, используя связь $G_{e-ph} \sim C_e t_{e-ph}$ [23] для оценки времени электрон-фононной релаксации, в выражении (1):

$$\max \{t_p, t_{e-ph}\} \approx t_{heating} \leq t_{mech} \approx \frac{L_{e-diff}}{C_s} \approx \frac{\sqrt{D_e t_{e-ph}}}{C_s} = \frac{\sqrt{t_{e-ph} K_e(Q)/C_e(Q)}}{C_s} \approx \frac{\sqrt{K_e(Q)/G_{e-ph}(Q)}}{C_s}, \quad (2)$$

где K_e , G_{e-ph} , и C_e – электронная теплопроводность, постоянная электрон-фононного взаимодействия и электронная теплопроводность соответственно. Для каждой длительности импульса t_p все эти параметры зависят от электронной температуры и плотности поглощенной энергии Q . Для гарантии высокой производительности и качества обработки с участием абляции необходимо минимизировать левую часть неравенства путем простого сокращения длительности импульса и максимизировать правую часть неравенства путем увеличения глубины поглощенной энергии в зависимости от свойств материала, например, путем увеличения плотности поглощенной энергии для увеличения ее эффективной глубины проникновения в материал.

Предложенная теория хорошо объясняет экспериментальные данные, полученные в работах [6–8] (рис. 8). В рассмотренных экспериментах по абляции длительность импульса может быть меньше, больше и порядка характерного времени электрон-фононного взаимодействия в зависимости от свойств материала. Оба импульса длительностью 0,2 пс и 10 пс – короткие для золота, а скорость абляции V_{abl} более или менее линейна в зависимости от поглощенной энергии, что подтверждает фотомеханический характер разрушения за счет локализации возникших в области воздействия внутренних напряжений. Для алюминия импульс длительностью 10 пс – длинный. В этом случае не будут реализованы условия возникновения локализации внутренних напряжений даже при высоких плотностях энергии. Исходя из этого, механизм разрушения является фототермическим, тогда как при длительности импульса 0,2 пс он будет фотомеханическим. Наконец, никель с характерным временем электрон-фононного взаимодействия 10 пс – граничный случай, поэтому задействованы оба механизма разрушения. В зависимости от поглощенной плотности мощности будет реализован либо фотомеханический, либо фототермический механизм, либо их совокупность.

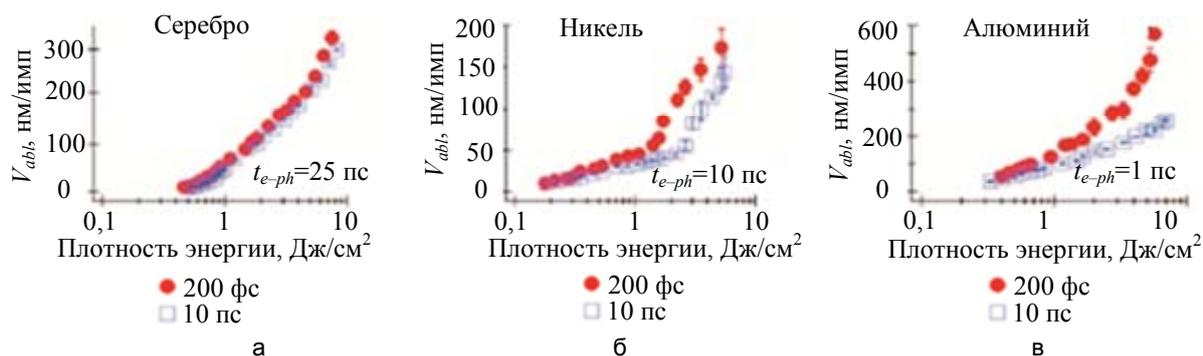


Рис. 8. Экспериментально измеренные скорости абляции серебра (а), никеля (б) и алюминия (в) в зависимости от поглощенной плотности энергии для длительности импульса 0,2 пс (круги) и 10 пс (квадраты) [7]

Полученная формула (2) также согласуется с теоретическими результатами [12], где моделирование диссипации поглощенной энергии лазерного импульса электрон-фононными столкновениями и теплопроводностью позволило рассмотреть оптимизацию длительности импульса для получения максимальной температуры на поверхности материала при лазерном облучении. Показано, что при малом коэффициенте электрон-фононного взаимодействия G_{e-ph} и большой электронной теплопроводности K_e глубина эффективного лазерного поглощения энергии больше, что, в свою очередь, может вызвать фотомеханическое разрушение.

Наконец, необходимо также отметить, что в зависимости от конкретных свойств материалов, полученные результаты при определении режимов абляции с помощью предложенной модели могут немного отличаться. Например, влияние шероховатости поверхности и степени кристалличности облучаемой мишени, рассмотренное в работе [13], может увеличить область режимов фотомеханического разрушения, в то время как оценки по более сложным моделям электронной теплопроводности [17] и электрон-фононного взаимодействия [16] могут дать ее уменьшение. В любом случае, основная идея настоящей работы заключается в том, что для длительности импульса, меньшей, чем характерное время установления электрон-фононного равновесия, как это описано с помощью уравнения (2), существует диапазон поглощенной плотности энергии, в котором при резком росте теплопроводности с температурой электронов можно запастись энергией лазерного излучения на глубине, достаточно большой для реализации условий возникновения локализации внутренних напряжений во время процесса абляции. Это способствует обработке с высокими скоростями удаления материала и одновременно с ее хорошим качеством в технологических применениях короткого лазерного импульса.

Литература

1. Mezzapesa F.P., Sibillano T., Columbo L.L., Di Niso F., Ancona A., Dabbicco M., De Lucia F., Lugarà P.M., Scamarcio G. Direct investigation of the ablation rate evolution during laser drilling of high aspect ratio micro-holes // *Proceedings of SPIE - The International Society for Optical Engineering*. 2012. V. 8243. Art. 82430S.
2. Lo Turco S., Nava G., Osellame R., Vishnubhatla K.C., Ramponi R. Femtosecond laser micromachining for optofluidic and energy applications // *Optical Materials*. 2013. V. 36. N 1. P. 102–105.
3. Rizvi N.H. Femtosecond laser micromachining: current status and applications // *Riken Review*. 2003. V. 50. P. 107–112.
4. Wellershoff S.-S., Hohlfeld J., Gudde J., Matthias E. The role of electron-phonon coupling in femtosecond laser damage of metals // *Applied Physics A: Materials Science and Processing*. 1999. V. 69. N 7. P. S99–S107.
5. Ivanov D.S., Zhigilei L.V. Combined atomistic-continuum modelling of short-pulse laser melting and disintegration of metal films // *Physical Review B – Condensed Matter and Materials Physics*. 2003. V. 68. N 6. Art. 064114. P. 064114-1–064114-22.
6. Nolte S., Momma C., Jacobs H., Tunnemann A., Chichkov B.N., Wellegehausen B., Welling H. Ablation of metals by ultrashort laser pulses // *Journal of Optical Society of America B: Optical Physics*. 1997. V. 14. N 10. P. 2716–2722.
7. Mannion P.T., Favre S., Ivanov D.S., O'Connor G.M., Glynn T.J. Experimental investigation of micromachining on metals with pulse durations in the range of the electron-phonon relaxation time (pico to sub-picosecond) // *Proc. 3rd International WLT-Conference on Lasers Manufacturing*. Munich, 2005. P. 521–526.
8. Momma C., Nolte S., Chichkov B.N., Alvensleben F.V., Tunnerbaum A. Precise laser ablation with ultrashort pulses // *Applied Surface Science*. 1997. V. 109–110. P. 15–19.
9. Toenshoff H.K., Ostendorf A., Nolte S., Korte F., Bauer T. Micromachining using femtosecond lasers // *Proceedings of SPIE – Laser Precision Manufacturing 2000 Conference*. Omiya, Japan, 2000. P. 136–139.
10. Leveugle E., Ivanov D.S., Zhigilei L.V. Photomechanical spallation of molecular and metal targets: molecular dynamics study // *Applied Physics A: Materials Science and Processing*. 2004. V. 79. N 7. P. 1643–1655.
11. Zhigilei L.V., Lin Z., Ivanov D.S. Atomistic modeling of short pulse laser ablation of metals: connections between melting, spallation, and phase explosion // *Journal of Physical Chemistry C*. 2009. V. 113. N 27. P. 11892–11906.
12. Ivanov D.S., Rethfeld B.C. The effect of pulse duration on the interplay of electron heat conduction and electron-photon interaction: photo-mechanical versus photo-thermal damage of metal targets // *Applied Surface Science*. 2009. V. 255. N 24. P. 9724–9728.
13. Ivanov D.S., Lipp V.P., Rethfeld B., Garcia M.E. Molecular-dynamics study of the mechanism of short-pulse laser ablation of singlecrystal and polycrystalline metallic targets // *Journal of Optical Technology*. 2014. V. 81. N 5. P. 250–253.
14. Zhigilei L.V., Ivanov D.S. Channels of energy redistribution in short-pulse laser interactions with metal targets // *Applied Surface Science*. 2005. V. 248. N 1–4. P. 433–439.
15. Anisimov S.I., Rethfeld B. On the theory of ultrashort laser pulse interaction with a metal // *Proc. SPIE*. 1997. V. 3093. P. 192–203.
16. Lin Z., Zhigilei L.V., Celli V. Electron-phonon coupling and electron heat capacity of metals under conditions of strong electron-phonon nonequilibrium // *Physical Review B - Condensed Matter and Materials Physics*. 2008. V. 77. N 7. Art. 075133.

17. Povarnitsyn M.E., Andreev N.E., Apfelbaum E.M., Itina T.E., Khishchenko K.V., Kostenko O.F., Levashov P.R., Veysman M.E. A wide-range model for simulation of pump-probe experiments with metals // *Applied Surface Science*. 2012. V. 258. N 23. P. 9480–9483.
18. Ivanov D.S., Kuznetsov A.I., Lipp V.P., Rethfeld B., Chichkov B.N., Garcia M.E., Schulz W. Short laser pulse nanostructuring of metals: direct comparison of molecular dynamics modeling and experiment // *Applied Physics A: Materials Science and Processing*. 2013. V. 111. P. 675–687.
19. Zhakhovskii V.V., Inogamov N.A., Petrov Yu.V., Ashitkov S.I., Nishihara K. Molecular dynamics simulation of femtosecond ablation and spallation with different interatomic potentials // *Applied Surface Science*. 2009. V. 255. N 24. P. 9592–9596.
20. Wu C., Zhigilei L.V. Microscopic mechanism of laser spallation and ablation of metal targets from large-scale molecular dynamics simulations // *Applied Physics A: Materials Science and Processing*. 2014. V. 114. N 1. P. 11–32.
21. Schäfer C., Urbassek H.M., Zhigilei L.V., Garrison B.J. Pressure-transmitting boundary conditions for molecular dynamics simulations // *Computational Materials Science*. 2002. V. 24. N 4. P. 421–429.
22. Zhigilei L.V., Ivanov D.S., Leveugle E., Sadigh B., Bringa E.M. Computer modeling of laser melting and spallation of metal targets // *Proceedings of SPIE – The International Society for Optical Engineering*. 2004. V. 5448. P. 505–519.
23. Anisimov S.I., Kapeliovich B.L., Perel'man T.L. Electron emission from metal surfaces exposed to ultrashort laser pulses // *ЖЭТФ*. 1974. V. 39. N 2. P. 375–377.

- Иванов Дмитрий Сергеевич** – PhD, старший научный сотрудник, Университет ИТМО, 197101, Санкт-Петербург, Россия; старший научный сотрудник, Технический Университет Кайзерслаутерн и исследовательский центр ОПТИМАС, 67653, Кайзерслаутерн, Германия; старший научный сотрудник, Университет Касселя, 34109, Кассель, Германия, Ivanov@uni-kassel.de
- Вейко Вадим Павлович** – доктор технических наук, профессор, заведующий кафедрой, 197101, Университет ИТМО, Санкт-Петербург, Россия, veiko@lastech.ifmo.ru
- Яковлев Евгений Борисович** – доктор технических наук, доцент, профессор, 197101, Университет ИТМО, Санкт-Петербург, Россия, yak@lastech.ifmo.ru (ebyakovlev@yandex.ru)
- Ретфельд Барбель** – PhD, профессор, профессор, Технический Университет Кайзерслаутерн и исследовательский центр ОПТИМАС, 67653, Кайзерслаутерн, Германия, rethfeld@physik.uni-kassel.de
- Гарсия Мартин Эзекель** – PhD, профессор, Университет Кассель, 34109, Кассель, Германия, Garcia@physik.uni-kassel.de
- Dmitry S. Ivanov** – PhD, senior scientific researcher, ITMO University. 197101, Saint Petersburg, Russia; Technical University of Kaiserslautern and OPTIMAS Research Center, Germany, 67653, Kaiserslautern, Germany; Kassel University, 34109, Kassel, Germany, Ivanov@uni-kassel.de
- Vadim P. Veiko** – D.Sc., Professor, ITMO University, 197101, Saint Petersburg, Russia, veiko@lastech.ifmo.ru
- Evgeny B. Yakovlev** – D.Sc., Associate professor, Professor, ITMO University, 197101, Saint Petersburg, Russia, yak@lastech.ifmo.ru, ebyakovlev@yandex.ru
- Baerbel Rethfeld** – PhD, full professor, Technical University of Kaiserslautern and OPTIMAS Research Center, 67653, Kaiserslautern, Germany, rethfeld@physik.uni-kassel.de
- Martin E. Carcia** – PhD, Professor, University Kassel Germany, 34109, Kassel, Germany, Garcia@physik.uni-kassel.de

Принято к печати 19.06.14
Accepted 19.06.14

УДК 62.50: 681.5.01

ФУНДАМЕНТАЛЬНАЯ МАТРИЦА ЛИНЕЙНОЙ НЕПРЕРЫВНОЙ СИСТЕМЫ
В ЗАДАЧЕ ОЦЕНКИ ЕЕ ТРАНСПОРТНОГО ЗАПАЗДЫВАНИЯ

Н.А. Дударенко^а, Н.А. Полинова^а, А.В. Ушаков^а

^аУниверситет ИТМО, 197101, Санкт-Петербург, Россия, ushakov-AVG@yandex.ru

Аннотация. Рассмотрена задача формирования количественной оценки транспортного запаздывания линейных непрерывных систем. Основным результатом получен с помощью фундаментальной матрицы решения системы линейных дифференциальных уравнений, заданной в нормальной форме Коши, для случая, как одномерных систем, так и для систем типа «многомерный вход-многомерный выход». Особенное свойство фундаментальной матрицы состоит в том, что весовая функция системы может быть сформирована как свободная составляющая движения этой системы, порождаемая вектором начального состояния, совпадающим с матрицей-столбцом входа исследуемой системы. Таким образом, использование свойств фундаментальной матрицы решения системы позволило решить задачу оценки транспортного запаздывания линейной непрерывной системы без использования помехонезащищенной в аппаратной среде операции дифференцирования и без формирования экзогенного воздействия. Полученные результаты апробированы на примере задачи оценки транспортного запаздывания системы, представляющей собой последовательное соединение аperiodических звеньев первого порядка с одинаковыми постоянными времени. Результаты моделирования подтвердили справедливость полученных математических выкладок. Знание транспортного запаздывания может быть использовано при настройке многоагрегатных технологических комплексов и при диагностике их возможного функционального вырождения при эксплуатации.

Ключевые слова: фундаментальная матрица, линейная непрерывная система, дельта-функция Дирака, весовая функция, транспортное запаздывание.

Благодарности. Работа поддержана Министерством образования и науки Российской Федерации (проект 14.Z50.31.0031).

FUNDAMENTAL MATRIX OF LINEAR CONTINUOUS SYSTEM
IN THE PROBLEM OF ESTIMATING ITS TRANSPORT DELAY

N.A. Dudarenko^a, N.A. Polinova^a, A.V. Ushakov^a

^аITMO University, 197101, Saint Petersburg, Russia, ushakov-AVG@yandex.ru

Abstract. The paper deals with the problem of quantitative estimation for transport delay of linear continuous systems. The main result is received by means of fundamental matrix of linear differential equations solutions specified in the normal Cauchy form for the cases of SISO and MIMO systems. Fundamental matrix has the dual property. It means that the weight function of the system can be formed as a free motion of systems. Last one is generated by the vector of initial system conditions, which coincides with the matrix input of the system being researched. Thus, using the properties of the system-solving for fundamental matrix has given the possibility to solve the problem of estimating transport linear continuous system delay without the use of derivation procedure in hardware environment and without formation of exogenous Dirac delta function. The paper is illustrated by examples. The obtained results make it possible to solve the problem of modeling the pure delay links using consecutive chain of aperiodic links of the first order with the equal time constants. Modeling results have proved the correctness of obtained computations. Knowledge of transport delay can be used when configuring multi-component technological complexes and in the diagnosis of their possible functional degeneration.

Keywords: fundamental matrix, linear continuous system, Dirac delta function, weight function, transport delay.

Acknowledgements. The work is supported by the Ministry of Education and Science of the Russian Federation (project 14.Z50.31.0031).

Введение

Проблема, вынесенная в заголовок работы, порождена необходимостью возрождения такой динамической характеристики линейной непрерывной системы, передаточная функция (ПФ) которой не имеет нулей, как транспортное запаздывание, которое широко использовалось для сравнения свойств динамических систем в период становления теории автоматического управления в 50–60-е годы прошлого столетия как самостоятельной науки [1]. Со временем необходимость в этой характеристике у практиков теории управления отпала, они в оценках временных характеристик систем управления стали обходиться только показателями кривой переходного процесса. Тем не менее, транспортное запаздывание является весьма содержательной характеристикой, потому что она позволяет провести простую аналогию между каналами связи и системами управления. В идеальном случае наиболее удобной для измерения транспортного запаздывания является пара сигналов в виде дельта-функции на входе и весовой функции на выходе. Однако такую пару сигналов на практике реализовать невозможно. Причиной тому являются два обстоятельства. Первое – невозможность реализации дельта-функции. Второе – невозможность инструментального формирования весовой функции путем дифференцирования переходной характеристики системы. Ситуация существенным образом изменилась, когда в теорию и практику систем автоматического управле-

ния, начиная с 70-х годов прошлого столетия стал внедряться метод пространства состояний. Обнаружилось, что фундаментальная матрица решения системы линейных дифференциальных уравнений [2, 3], заданных в нормальной форме Коши для описания линейной непрерывной системы управления, обладает свойством, состоящим в том, что весовая функция системы может быть сформирована как свободная составляющая движения этой системы, порождаемая вектором начального состояния, совпадающим с матрицей–столбцом входа исследуемой системы. Это свойство фундаментальной матрицы использовано авторами для оценки транспортного запаздывания линейной непрерывной системы. Знание транспортного запаздывания может быть использовано при настройке многоагрегатных технологических комплексов и при диагностике их возможного функционального вырождения при эксплуатации.

Постановка задачи

Рассмотрим непрерывную динамическую систему, описываемую системой линейных дифференциальных уравнений в канонической форме Коши [3–11]

$$\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{x}(t) + \mathbf{G}\mathbf{g}(t), \quad \mathbf{x}(0); \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t), \quad (1)$$

где $\mathbf{x}, \mathbf{g}, \mathbf{y}$ – соответственно векторные переменные состояния, экзогенного воздействия и выхода, обладающие размерностями $\dim(\mathbf{x}) = n, \dim(\mathbf{g}) = \dim(\mathbf{y}) = m$; матрицы $\mathbf{F}, \mathbf{G}, \mathbf{C}$ – соответственно состояния, входа и выхода размерностей $\dim(\mathbf{F}) = n \times n, \dim(\mathbf{G}) = n \times m, \dim(\mathbf{C}) = m \times n$. Поставим задачу найти явное решение системы дифференциальных уравнений (1) в форме

$$\mathbf{x}(t) = \mathbf{x}\{\mathbf{x}(0), \mathbf{g}(t), t\}; \quad \mathbf{y}(t) = \mathbf{C}\mathbf{x}(t). \quad (2)$$

Если воспользоваться *принципом суперпозиции*, который справедлив для линейных представлений, коим является (1), то можно записать $\mathbf{x}(t) = \mathbf{x}_c(t) + \mathbf{x}_b(t)$, где $\mathbf{x}_c(t)$ – свободная составляющая движения, порожденная $\mathbf{x}(0) \neq 0$, так что $\mathbf{x}_c(t) = \mathbf{x}\{\mathbf{x}(0), \mathbf{g}(t) \equiv 0, t\}$, $\mathbf{y}_c(t) = \mathbf{C}\mathbf{x}_c(t)$; $\mathbf{x}_b(t)$ – вынужденная составляющая движения, порожденная $\mathbf{g}(t) \neq 0$ так, что $\mathbf{x}_b(t) = \mathbf{x}\{\mathbf{g}(t), \mathbf{x}(0) \equiv 0, t\}$, $\mathbf{y}_b(t) = \mathbf{C}\mathbf{x}_b(t)$.

Для вычисления решения (2) системы уравнений (1) сформулируем утверждение.

Утверждение 1 (У.1). Общий вид явного решения (2) системы дифференциальных уравнений (1), записываемого в аддитивной форме, представим соотношениями

$$\left. \begin{aligned} \mathbf{x}(t) &= \Phi(t)\mathbf{x}(0) + \int_0^t \Phi(t, \tau)\mathbf{G}\mathbf{g}(\tau)d\tau = e^{\mathbf{F}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{F}(t-\tau)}\mathbf{G}\mathbf{g}(\tau)d\tau, \\ \mathbf{y}(t) &= \mathbf{C}\Phi(t)\mathbf{x}(0) + \int_0^t \mathbf{C}\Phi(t, \tau)\mathbf{G}\mathbf{g}(\tau)d\tau = \mathbf{C}e^{\mathbf{F}t}\mathbf{x}(0) + \int_0^t \mathbf{C}e^{\mathbf{F}(t-\tau)}\mathbf{G}\mathbf{g}(\tau)d\tau \end{aligned} \right\}, \quad (3)$$

где $\Phi(t) = e^{\mathbf{F}t}$, $\Phi(t, \tau) = \Phi(t)\Phi^{-1}(\tau) = e^{\mathbf{F}(t-\tau)}$. ■

Доказательство утверждения можно найти в [3–11].

Рассмотрим теперь вынужденное движение системы (1), порождаемое экзогенным воздействием вида дельта-функции Дирака [12]

$$\mathbf{g}(t) = \delta(t) = \begin{cases} \infty & \text{при } t = 0. \\ 0 & \text{при } t \neq 0. \end{cases}$$

Тогда в силу свойства $\delta(t)$ сворачивать интеграл [4, 13] в подынтегральное выражение соотношения (3) примут вид

$$\left. \begin{aligned} \mathbf{x}(t) &= e^{\mathbf{F}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{F}(t-\tau)}\mathbf{G}\delta(\tau)d\tau = e^{\mathbf{F}t}\mathbf{x}(0) + e^{\mathbf{F}t}\mathbf{G}\mathbf{1}_V \\ \mathbf{y}(t) &= \mathbf{C}e^{\mathbf{F}t}\mathbf{x}(0) + \int_0^t \mathbf{C}e^{\mathbf{F}(t-\tau)}\mathbf{G}\delta(\tau)d\tau = \mathbf{C}e^{\mathbf{F}t}\mathbf{x}(0) + \mathbf{C}e^{\mathbf{F}t}\mathbf{G}\mathbf{1}_V \end{aligned} \right\}. \quad (4)$$

В соотношениях (4) $\mathbf{1}_V = \text{col}\{\mathbf{1}_{V_i} = 1, i = \overline{1, n}\}$ – n -мерный вектор, составленный из единиц.

Аддитивные представления (3), (4) явного решения системы (1) позволяют ввести в рассмотрения три динамические матрицы системы (1):

1. $\Phi(t) = e^{\mathbf{F}t}$ – фундаментальная матрица системы;
2. $\Phi(t, \tau) = \Phi(t)\Phi^{-1}(\tau) = e^{\mathbf{F}(t-\tau)}$ – переходная матрица системы, $\Phi(t, 0) = \Phi(t)$;
3. $\mathbf{w}(t) = \mathbf{C}\Phi(t)\mathbf{G} = \mathbf{C}e^{\mathbf{F}t}\mathbf{G}$ – весовая матрица системы, составленная из $(m \times m)$ скалярных весовых (импульсных переходных) функций $w_{ji}(t) = \mathbf{C}^j\Phi(t)\mathbf{G}_i = \mathbf{C}^j e^{\mathbf{F}t}\mathbf{G}_i$ наблюдаемых на j -м выходе при подаче скалярной δ - функции на i -й вход.

Оценка транспортного запаздывания системы с помощью ее фундаментальной матрицы

Теперь поставим задачу оценки транспортного запаздывания системы τ_d с помощью ее фундаментальной матрицы $\Phi(t)$. Решение задачи начнем с определения транспортного запаздывания в системах автоматического управления. Транспортным запаздыванием в системах автоматического управления [14–16] в отличие от «чистого» запаздывания [17–20] называется явление, заключающееся в том, что с началом изменения сигнала на входе системы в силу конечной скорости его распространения сигнал на выходе системы начинает проявляться только спустя некоторое время. При этом наилучшей парой сигналов, по которым с наибольшей точностью может быть зафиксирована величина τ_d транспортного запаздывания, является сигнал вида δ -функции Дирака на входе и весовая функция (импульсная переходная) на выходе.

Очевидно, при оценке транспортного запаздывания τ_d рассматривается конкретный скалярный сепаратный канал системы (1), связывающий j -й выход $y_j(t)$ с i -м входом $g_i(t)$. При этом на i -й вход должен подаваться сигнал вида скалярной δ -функции Дирака так, что $g_i(t) = \delta(t)$, а на j -м выходе должен наблюдаться сигнал $y_j(t, g_i(t) = \delta(t), g_l(t) = 0|_{l \neq i}; \mathbf{x}(0) = 0)$, который является откликом на δ -функцию Дирака, т.е. ее выходным образом. Локализация этого отклика на временной оси определяет транспортное запаздывание, идентифицируемое по размещению на ней первого экстремального значения отклика.

Возникает первая проблема измерения транспортного запаздывания τ_d описанным способом, которая состоит в том, что δ -функция является физически нереализуемой, но отклик на нее может быть вычислен с использованием фундаментальной матрицы системы (1). Действительно, если воспользоваться представлением (4) применительно к j -му выходу при подаче на i -й вход скалярной δ -функции Дирака, то становится справедливой запись

$$y_j(t) = C^j e^{Ft} \mathbf{x}(0) + C^j e^{Ft} \mathbf{G}_i |_{\mathbf{x}(0)=0} = C^j e^{Ft} \mathbf{G}_i. \tag{5}$$

Представление (5) содержит доказательство следующего утверждения.

Утверждение 2 (У.2). Отклик $y_j(t, g_i(t) = \delta(t), g_l(t) = 0|_{l \neq i}; \mathbf{x}(0) = 0)$ на сигнал вида скалярная δ -функция Дирака, подаваемый на i -й вход системы (1) может быть сформирован как свободное движение системы (1), наблюдаемое на ее j -м выходе, при условии, что начальное состояние системы задано в форме $\mathbf{x}(0) = \mathbf{G}_i$, что приводит к системе соотношений

$$\dot{\mathbf{x}}(t) = \mathbf{F}\mathbf{x}(t), \mathbf{x}(0) = \mathbf{G}_i, y_j(t) = C^j \mathbf{x}(t) \Rightarrow y_j(t) = C^j \Phi(t) \mathbf{G}_i = C^j e^{Ft} \mathbf{G}_i. \quad \blacksquare$$

Применим полученный результат к решению задачи оценки транспортного запаздывания τ_d системы, представляющей собой последовательное соединение аperiodических звеньев первого порядка [21] с одинаковыми постоянными времени так, что для ее ПФ «вход–выход» может быть записана цепочка эквивалентных представлений

$$\Phi(s) = \frac{y(s)}{g(s)} = \left(\frac{1}{Ts + 1} \right) \Big|_{T=\alpha}^n = \left(\frac{\alpha}{s + \alpha} \right)^n = \alpha^n \left(\frac{1/s}{1 + \alpha/s} \right)^n. \tag{6}$$

Векторно-матричное представление (1) системы с ПФ (6) будет характеризоваться матричными компонентами

$$\left. \begin{aligned} \mathbf{F} = \mathbf{J}(\alpha) &= \begin{bmatrix} -\alpha & 1 & 0 & \dots & 0 & 0 \\ 0 & -\alpha & 1 & \dots & 0 & 0 \\ 0 & 0 & -\alpha & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\alpha & 1 \\ 0 & 0 & 0 & \dots & 0 & -\alpha \end{bmatrix} = -\alpha \mathbf{I} + \mathbf{J}(0); \\ \mathbf{G} &= \begin{bmatrix} \mathbf{0}_{(n-1) \times 1} \\ \alpha^n \end{bmatrix}; \mathbf{C} = \begin{bmatrix} 1 & \mathbf{0}_{1 \times (n-1)} \end{bmatrix} \end{aligned} \right\}. \tag{7}$$

Учитывая, что в аддитивном представлении (7) матрицы состояния системы аддитивные компоненты оказываются мультипликативно коммутативными [6–9, 22], то становится справедливым представление матричной экспоненты

$$e^{Ft} = e^{J(\alpha)t} = e^{-\alpha t} e^{J(0)t} = e^{-\alpha t} \begin{bmatrix} 1 & t & t^2/2 & \dots & t^{(n-1)}/(n-1)! \\ 0 & 1 & t & \dots & t^{(n-2)}/(n-2)! \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & t \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \quad (8)$$

Поставим задачу вычисления отклика в форме (5) системы типа «одномерный вход–одномерный выход» с ПФ (6) на входной сигнал типа δ -функции Дирака. Тогда с учетом (7) и (8) получим

$$y(t) = Ce^{Ft}G = \alpha^n \left(t^{(n-1)}/(n-1)! \right) e^{-\alpha t}. \quad (9)$$

Для оценки величины $\tau_d = \text{minarg} \left\{ \frac{dy(t)}{dt} = 0 \& t \neq 0 \right\}$ транспортного запаздывания продифференцируем по времени отклик (9) системы на сигнал типа δ -функции Дирака, в результате чего для τ_d можно записать

$$\tau_d = \text{minarg} \left\{ \frac{d \left(t^{(n-1)} e^{-\alpha t} \right)}{dt} = 0 \& t \neq 0 \right\} = \arg \{ (n-1) - \alpha t = 0 \} = \frac{n-1}{\alpha}. \quad (10)$$

Из выражения (10) следует, что в отдельно взятом апериодическом звене первого порядка транспортное запаздывание отсутствует, что можно установить, положив в (10) $n = 1$.

Теперь оценим величину отклика (9) в точке $t = t_d$ для $n > 1$ путем подстановки (10) в (9), в результате чего получим

$$y(t_d) = \alpha \left((n-1)^{(n-1)} / (n-1)! \right) e^{-(n-1)}.$$

В таблице приведены значения относительных величин отклика $y(t_d)/\alpha$ и коэффициентов нормализации K_n для значений $n = 2; 3; 5; 10; 20; 40$.

n	2	3	5	10	20	40
$y(t_d)/\alpha$	0,3679	0,2707	0,1954	0,1318	0,0911	0,0637
$K_n \alpha$	2,7181	3,6941	5,1177	7,5873	10,9769	15,6986

Таблица. Значения относительных величин отклика системы и коэффициентов нормализации для различных значений n

На рис. 1 приведены кривые откликов на экзогенное воздействие $g(t) = \delta(t)$ систем с ПФ (6) для $T = 1\text{с}$ ($\alpha = 1\text{с}^{-1}$) и значений $n = 2; 3; 5; 10; 20; 40$. Кривые четко подтверждают результат (10).

На рис. 2 приведены кривые модифицированных откликов на экзогенное воздействие $g(t) = \delta(t)$ систем с ПФ (6) для $T = 1\text{с}$ ($\alpha = 1\text{с}^{-1}$) и значений $n = 2; 3; 5; 10; 20; 40$, построенные на основе кривых рис. 1 и значений коэффициентов нормализации K_n , вычисляемые по данным таблицы.

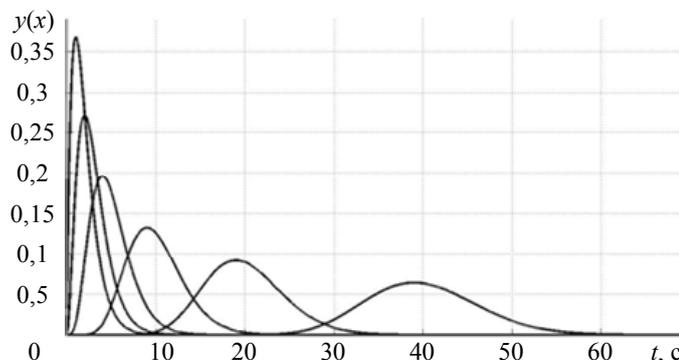


Рис. 1. Кривые откликов $y(t)$ на экзогенное воздействие $g(t) = \delta(t)$ систем с ПФ (6) для $T = 1\text{с}$ ($\alpha = 1\text{с}^{-1}$) и значений $n = 2; 3; 5; 10; 20; 40$

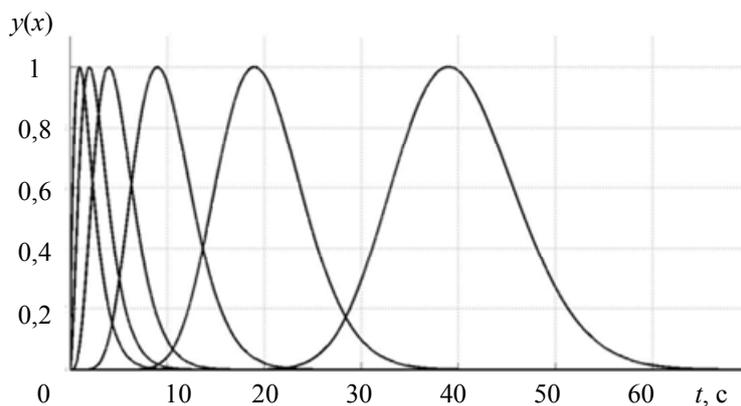


Рис. 2. Кривые модифицированных откликов $y(t)$ на $g(t) = \delta(t)$ систем с ПФ (6) для $T = 1\text{с}$ ($\alpha = 1\text{с}^{-1}$) и значений $n = 2; 3; 5; 10; 20; 40$

Для полноты картины на рис. 3 приведены кривые откликов на экзогенное единичное воздействие $g(t) = 1(t)$ тех же систем с ПФ (6) для $T = 1\text{с}$ ($\alpha = 1\text{с}^{-1}$) и значений $n = 2; 3; 5; 10; 20; 40$.

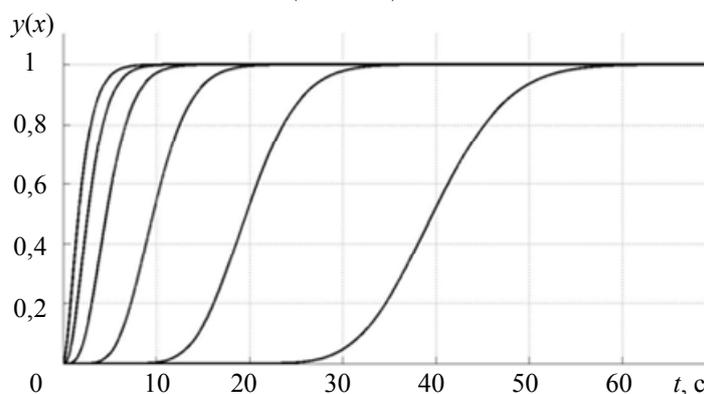


Рис. 3. Кривые откликов $y(t)$ на экзогенное единичное воздействие $g(t) = 1(t)$ тех же систем с ПФ (6) для $T = 1\text{с}$ ($\alpha = 1\text{с}^{-1}$) и значений $n = 2; 3; 5; 10; 20; 40$

Из сравнения кривых рис. 1 и кривых рис. 3 хорошо видно, что весовые функции систем, как отклики на экзогенное воздействие $g(t) = \delta(t)$, и переходные функции систем, как отклики на экзогенное единичное воздействие $g(t) = 1(t)$, связаны отношением дифференцирования по времени. Заметим, что использование свойств фундаментальной матрицы позволило решить поставленную задачу без использования помехозащищенной в аппаратной среде операции дифференцирования и без формирования экзогенного воздействия $g(t) = \delta(t)$.

Следует сказать, что совместное использование кривых рис. 1, 3, если их дополнить аналитическим выражением (10), позволяет достаточно элегантно решать задачу моделирования звеньев чистого запаздывания с помощью последовательной цепочки аperiodических звеньев первого порядка. При этом задача оказывается двухпараметрической с параметрами n – их числа и T – их постоянной времени.

Заключение

Фундаментальная матрица линейной непрерывной системы обнаружила свои дополнительные содержательные свойства, позволившие решить задачу формирования количественной оценки транспортного запаздывания системы «одномерный вход–одномерный выход». Перенос этих возможностей на системы «многомерный вход–многомерный выход» применительно к отдельным скалярным каналам позволит предупреждать развитие такого системного свойства как вырождения [23] систем «многомерный вход–многомерный выход», опираясь на достаточно простой системный показатель.

Литература

1. Дралюк Б.Н. Синайский Г.В. Системы автоматического регулирования объектов с транспортным запаздыванием. М.: Энергия, 1969. 72 с.
2. Красовский Н.Н. Теория управления движением. М.: Наука, 1968. 476 с.
3. Зубов В.И. Лекции по теории управления. 2-е изд. СПб.: Лань, 2009. 496 с.
4. Андреев Ю.Н. Управление конечномерными линейными объектами. М.: Наука, 1976. 424 с.
5. Арнольд В.И. Обыкновенные дифференциальные уравнения. 4-е изд. М.: МЦНМО, 2012. 380 с.
6. Bellman R. Introduction to Matrix Analysis. NY: McGraw-Hill, 1960. 328 p.
7. Гантмахер Ф.Р. Теория матриц. М.: Наука, 1973. 575 с.
8. Понтрягин Л.С. Обыкновенные дифференциальные уравнения. Ижевск: Регулярная и хаотическая динамика, 2001. 400 с.
9. Anderson B.D.O., Moore J.V. Linear Optimal Control. Prentice-Hall, 1971. 413 p.
10. Zadeh L.A., Desoer C.A. Linear System Theory: The State Space Approach. NY: McGraw-Hill, 1963. 628 p.
11. Дударенко Н., Слита О., Ушаков А. Современная теория многомерного управления: аппарат пространства состояний. Saarbrücken: LAP LAMBERT Academic Publishing, 2011. 418 с.
12. Дирак П.А. Принципы квантовой механики: Пер. с англ. М.: Наука, 1979. 480 с.
13. Математическая энциклопедия. М.: Советская энциклопедия, 1979. Т. 2. 1103 с.
14. Попов Е.П. Динамика систем автоматического регулирования. М.: Гостехиздат, 1954. 800 с.
15. Ротач В.Я. Расчет динамических промышленных автоматических систем регулирования. М.: Энергия, 1973. 440 с.
16. Шавров А.А. Компенсатор транспортного запаздывания в системах автоматического управления // Вестник РГАЗУ: Агроинженерия. 2004. С. 52.
17. Tang G.-Y., Fu P.-L. Suboptimal control approach of linear time-delay systems // Proc. 14th World Congress of IFAC. China, 1999. P. 99–103.
18. Cao Y.-Y., Lam J., Sun Y.-X. Robust control for uncertain systems with time-delay and jump parameters // Proc. 14th World Congress of IFAC. China, 1999. P. 191–196.
19. Jankovic M. Control of nonlinear systems with time delay // Proc. 42nd IEEE Conference on Decision and Control. Maui, USA, 2003. V. 5. P. 4545–4550.
20. Kharitonov V.L., Niculescu S.-I., Moreno J., Michiels W. Static output feedback stabilization: necessary conditions for multiple delay controllers // IEEE Transaction on Automatic Control. 2005. V. 50. N 1. P. 82–86.
21. Бесекецкий В.А., Попов Е.П. Теория систем автоматического регулирования. СПб: Профессия, 2003. 752 с.
22. Акунов Т.А., Дударенко Н.А., Полинова Н.А., Ушаков А.В. Исследование колебательности процессов в аperiodических непрерывных системах, порождаемой фактором кратности собственных чисел // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 3 (85). С. 55–61.
23. Дударенко Н., Ушаков А. Анализ многомерных динамических систем: технология контроля вырождения. Saarbrücken: LAP LAMBERT Academic Publishing, 2012. 232 с.

<i>Дударенко Наталия Александровна</i>	– кандидат технических наук, доцент, доцент, Университет ИТМО, 197101, Санкт-Петербург, Россия, dudarenko@yandex.ru
<i>Полинова Нина Александровна</i>	– магистрант, Университет ИТМО, 197101, Санкт-Петербург, Россия, polinova_nina@mail.ru
<i>Ушаков Анатолий Владимирович</i>	– доктор технических наук, профессор, профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия, ushakov-AVG@yandex.ru
<i>Natalia A. Dudarenko</i>	– PhD, Associate professor, Associate professor, ITMO University, 197101, Saint Petersburg, Russia, dudarenko@yandex.ru
<i>Nina A. Polinova</i>	– postgraduate, ITMO University, 197101, Saint Petersburg, Russia, polinova_nina@mail.ru
<i>Anatoly V. Ushakov</i>	– D.Sc., full professor, ITMO University, 197101, Saint Petersburg, Russia, ushakov-AVG@yandex.ru

Принято к печати 21.05.14
Accepted 21.05.14

УДК 621.362; 621.315.562

ТЕПЛОВЫЕ И ЭЛЕКТРИЧЕСКИЕ ПОЛЯ ПРИ ИСКРОВОМ ПЛАЗМЕННОМ
СПЕКАНИИ ТЕРМОЭЛЕКТРИЧЕСКИХ МАТЕРИАЛОВЛ.П. Булат^а, Д.А. Пшенай-Северин^{а, б}, И.А. Нефедова^а, А.В. Новотельнова^а, Ю.Г. Гуревич^с^а Университет ИТМО, 197101, Санкт-Петербург, Россия, novotelnova@yandex.ru^б Физико-технический институт им. А.Ф. Иоффе, 194021, Санкт-Петербург, Россия^с Исследовательский центр Национального политехнического института Мексики, 07738, Мехико, Мексика**Аннотация.**

Постановка проблемы. Повышение термоэлектрической добротности термоэлектриков связывают с использованием наноструктурированных термоэлектрических материалов, полученных из порошков методом искрового плазменного спекания. Сохранение наноструктуры порошка в процессе спекания возможно при оптимальных температурных режимах их получения. Выбор этих режимов осложнен свойствами термоэлектриков – анизотропных полупроводниковых материалов. Решение данной проблемы путем моделирования процесса спекания требует грамотного подхода к постановке задачи и корректного задания свойств термоэлектриков и материалов, входящих в рабочую установку, а также граничных условий. В представленной работе предложена работоспособная модель спекания термоэлектриков.

Методы. Рассматривается процесс спекания термоэлектрического материала – теллурида висмута с использованием установки SPS-511S. Учитываются температурные зависимости электро- и теплопроводности теллурида висмута, а также материалов, из которых изготовлены элементы установки. Показано, что распределение температуры в образце может быть определено в рамках стационарной задачи. Моделирование выполняется в программном продукте Comsol Multiphysics. Граничные условия включают конвективный теплообмен и излучение по закону Стефана–Больцмана.

Результаты. Проведено компьютерное моделирование электро- и теплофизических процессов, протекающих при искровом плазменном спекании. Получено распределение температуры и потенциала в образце при спекании. Показана определяющая роль в формировании температурного поля в образцах графитовой пресс-формы. Проанализировано влияние геометрических размеров графитовой пресс-формы на условия спекания наноструктурированных термоэлектриков.

Практическая значимость. Установлены оптимальные размеры цилиндрической пресс-формы для получения однородных по объему образцов на основе теллурида висмута. Показаны пути модификации процесса спекания, позволяющие создавать образцы термоэлектриков с прогнозируемыми свойствами.

Ключевые слова: искровое плазменное спекание, термоэлектрические материалы, наноструктуры, компьютерное моделирование, термоэлектрическая добротность, теплопроводность, тепловые и электрические поля.

Благодарности. Работа выполнена при поддержке Минобрнауки: субсидия № 14.579.0039 и задание № 3.912.2014/К.

THERMAL AND ELECTRIC FIELDS AT SPARK PLASMA SINTERING
OF THERMOELECTRIC MATERIALSL.P. Bulat^а, D.A. Pshenai-Severin^{а, б}, I.A. Nefedova^а, A.V. Novotelnova^а, Yu.G. Gurevich^с^а ITMO University, 197101, Saint Petersburg, Russia, novotelnova@yandex.ru^б Ioffe Physical Technical Institute, 194021, Saint Petersburg, Russia^с Center for Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV I.P.N.), 07738, Mexico-City, Mexico**Abstract.**

Problem statement. Improvement of thermoelectric figure of merit is connected with the usage of nanostructured thermoelectric materials fabricated from powders by the spark plasma sintering (SPS) method. Preservation of powder nanostructure during sintering is possible at optimum temperature modes of thermoelectrics fabrication. The choice of these modes becomes complicated because of anisotropic properties of semiconductor thermoelectric materials. The decision of the given problem by sintering process simulation demands the competent approach to the problem formulation, a correct specification of thermoelectric properties, the properties of materials forming working installation, and also corrects boundary conditions. The paper deals with the efficient model for sintering of thermoelectrics.

Methods. Sintering process of the bismuth telluride thermoelectric material by means of SPS-511S installation is considered. Temperature dependences of electric and thermal conductivities of bismuth telluride, and also temperature dependences of installation elements materials are taken into account. It is shown that temperature distribution in the sample can be defined within the limits of a stationary problem. The simulation is carried out in the software product Comsol Multiphysics. Boundary conditions include convective heat exchange and also radiation under Stefan-Boltzmann law.

Results. Computer simulation of electric and thermal processes at spark plasma sintering is carried out. Temperature and electric potential distributions in a sample are obtained at the sintering conditions. Determinative role of graphite compression mould in formation of the temperature field in samples is shown. The influence of geometrical sizes of a graphite compression mould on sintering conditions of nanostructured thermoelectrics is analyzed.

Practical importance. The optimum sizes of a cylindrical compression mould for fabrication of volume homogeneous samples based on bismuth telluride are determined. Ways of updating for the sintering process are shown giving the

possibility to fabricating thermoelectric samples with predicted properties.

Keywords: spark plasma sintering, thermoelectric materials, nanostructures, computer simulation, thermoelectric figure of merit, thermal conductivity, thermal and electric fields.

Acknowledgements. The work is supported by the Ministry of Education and Science of the Russian Federation (grant № 14.579.0039 and task № 3.912.2014/К).

Введение

Искровое плазменное спекание (spark plasma sintering, SPS-метод) все шире используется для создания самых различных функциональных материалов из микро- и нанопорошков. Спекание происходит в специальных установках путем воздействия на порошок импульсов постоянного электрического поля. Данный метод оказался очень удобным и эффективным для изготовления разного рода наноструктур и нанокомпозитов. Так, оказалось, что искровое плазменное спекание практически незаменимо для создания объемных наноструктурированных термоэлектрических материалов (см., например, [1–4]). Дело в том, что, в отличие от обычного горячего прессования, консолидация наночастиц в объемный образец с помощью искрового плазменного спекания препятствует процессам рекристаллизации [1–4]. А рекристаллизация приводит к увеличению размеров нанозерен в объемном образце от десятков до сотен нанометров, т.е. образец превращается из наноструктурного в образец с микронеоднородностями. В этом случае перестают реализовываться механизмы, которые приводят к увеличению термоэлектрической добротности материала в наноструктурах, в первую очередь речь идет об интенсивности дополнительного рассеяния фононов и носителей заряда на межзеренных границах [5, 6].

Однако искровое плазменное спекание определяется сложными быстропротекающими взаимозависимыми электрическими и тепловыми процессами. В связи с этим подбор режимов спекания на практике осуществляется эмпирическим путем, причем не только при изготовлении объемных образцов термоэлектрических наноматериалов, но и вообще при создании функциональных материалов различного назначения.

Итак, возникает вопрос о выборе таких режимов искрового плазменного спекания, которые могли бы обеспечить получение наноматериалов с оптимизированными термоэлектрическими свойствами. В настоящей работе сделан первый шаг для решения данной задачи – на примере термоэлектрических материалов на основе теллурида висмута Bi_2Te_3 выясняется, можно ли воспроизвести полученные экспериментально значения температуры образца при заданных значениях токов и разумных параметрах материалов и характеристик установки.

Особенности моделирования спекания термоэлектрических материалов на основе Bi_2Te_3 заключаются в том, что они представляют собой анизотропные полупроводниковые материалы. Их электропроводность и теплопроводность существенно зависят от температуры. Исходя из этого, решение данной задачи требует грамотного подхода к ее постановке и корректного задания свойств термоэлектриков и материалов, входящих в рабочую установку, а также граничных условий. В рассматриваемом случае спекание материала должно производиться при равномерном распределении температуры по объему образца. Температура во время спекания должна обеспечивать получение термоэлектрического материала с высокой эффективностью. Эффективность термоэлектрических материалов определяется безразмерной термоэлектрической добротностью $ZT = \frac{\alpha^2 \sigma}{\kappa} T$, где α – коэффициент термоЭДС, σ – коэффициент электропроводности, κ – коэффициент теплопроводности, T – абсолютная температура.

Для определенности расчеты выполняются для установки SPS-511S (Япония), с использованием которой производилось спекание термоэлектрических материалов на основе Bi_2Te_3 [1–4].

Экспериментально полученное и надежно подтвержденное наибольшее значение безразмерной термоэлектрической добротности в наноструктурированных материалах на основе Bi_2Te_3 представлено в работе [7]. Добротность достигала значений $ZT = 1,23$ при $T = 360$ К в образцах состава $\text{Bi}_{0,4}\text{Sb}_{1,6}\text{Te}_3$, полученных механоактивационной обработкой с последующим искровым плазменным спеканием под давлением 50 МПа при температуре спекания $T_s = 770$ К [7]. Эти данные принимались во внимание при оценке соответствия модельных и экспериментальных результатов.

Получаемая в итоге модель спекания термоэлектриков позволит направленно модифицировать процесс спекания с тем, чтобы создавать термоэлектрические материалы с прогнозируемыми свойствами.

Моделирование процесса искрового плазменного спекания

Установка SPS-511S состоит из вакуумной камеры с охлаждаемыми водой стальными стенками (рис. 1). Внутри установки между стальными электродами 1 и 6 размещается закрепленная графитовыми вставками 2 пресс-форма 5. Спекаемый образец 4 помещен между графитовыми пуансонами 3. Для обеспечения надежного контакта между стальными электродами и графитовыми вставками внутри пресс-формы устанавливаются прокладки 7 и 9 и трубку из графитовой фольги 8.

Отвод тепла от стальных электродов 1 и 6 происходит с помощью водяного охлаждения.

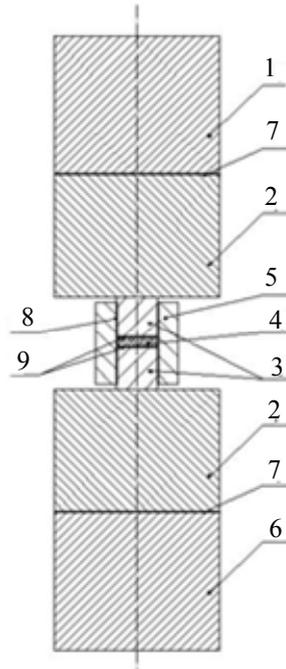


Рис. 1. Схематическая модель установки искрового плазменного спекания: 1 – верхний охлаждаемый электрод; 2 – графитовые вставки; 3 – верхний и нижний графитовые пуансоны; 4 – образец; 5 – пресс-форма; 6 – нижний охлаждаемый электрод; 7, 9 – прокладки из графитовой фольги; 8 – трубка из графитовой фольги

Использованная модель включает систему дифференциальных уравнений, описывающих распределение электрического потенциала и температуры. Плотности потоков заряда \mathbf{j} и тепла \mathbf{q} связаны с градиентами этих величин законами Ома и Фурье:

$$\mathbf{j} = \sigma(-\nabla V), \quad \mathbf{q} = \kappa(-\nabla T),$$

где σ и κ – соответственно коэффициенты электро- и теплопроводности. Законы сохранения заряда и энергии имеют следующий вид:

$$\operatorname{div} \mathbf{j} = 0,$$

$$c_p \rho \frac{\partial T}{\partial t} + \operatorname{div} \mathbf{q} = Q_j, \tag{1}$$

где c_p и ρ – удельная теплоемкость при постоянном давлении и плотность материала. В уравнении (1) учтен разогрев джоулевым теплом $Q_j = \mathbf{j} \cdot (-\nabla V)$.

Качественно данная модель учитывает протекание через прессформу и образец электрического тока, который приводит к генерации теплоты. Благодаря теплопроводности теплота распределяется по всему объему установки. Если бы пресс-форма была теплоизолирована, то температура образца в данной модели увеличивалась бы неограниченно. Практически конечная температура определяется:

- величиной тока и электропроводностями материалов, которые задают величину сгенерированного тепла;
- величинами теплопроводностей и тепловыми граничными условиями, определяющими скорость оттока теплоты.

Результаты сравнения температуры образца с полученной в эксперименте температурой должны быть довольно чувствительными к величинам кинетических коэффициентов и виду граничных условий.

В эксперименте использовался импульсный постоянный ток до 1000 А [1–4]. Длительность импульса составляла 3,3 мс, скважность – 2,2 мс. За пакетом из 12 импульсов следовали два периода отключения тока. Аналогичные параметры тока использовались в работе [8], где было показано, что при данных частотах скин-эффектом можно пренебречь. Исходя из этого, если мощность разогрева не превосходит возможностей теплоотвода, то в используемом подходе временные характеристики тока определяют только соответствующую зависимость температуры образца, но не ее конечное значение. В связи с этим максимальное значение температуры образца можно найти, решая стационарный вариант задачи (1).

Граничные условия для потенциала и тока соответствуют нулевому потенциалу на нижнем электроде и заданному значению тока через образец. В расчетах было использовано максимальное значение тока 832 А. При этом средняя температура образца достигала 770 К. Начальная температура установки

была равна $T_0 = 300$ К. Тепловые граничные условия соответствовали излучательному отводу тепла от боковых граней пресс-формы, графитовых вставок и электродов. Было использовано выражение для излучательного переноса тепла между двумя поверхностями, находящимися при температурах T и $T_0 < T$ [9] с коэффициентами излучения ε_1 и ε_2 . Величина оттока тепла от горячей поверхности равна

$$q_{rad} = \frac{\sigma_{SB} (T^4 - T_0^4)}{\varepsilon_1^{-1} + \varepsilon_2^{-1} - 1}.$$

Здесь $\sigma_{SB} = 5,670 \cdot 10^{-8}$ Вт·м⁻²·К⁻⁴ – постоянная Стефана–Больцмана, параметры ε для графита и стали были приняты равными 0,75 и 0,675 соответственно [10].

Для описания водяного охлаждения использовались граничные условия, соответствующие конвективному теплообмену,

$$q_{conv} = K_b (T - T_0),$$

где для контакта стали с водой был выбран коэффициент $K_b = 370$ Вт/(м²·К) [11].

Поскольку процесс искрового плазменного спекания протекает в широком диапазоне температур, при расчетах необходимо учитывать температурные зависимости кинетических коэффициентов. Эти зависимости для графита и стали, пересчитанные из [12–14], приведены на рис. 2, а, б, соответственно. Для графитовой фольги коэффициенты теплопроводности составляли 130 и 3 Вт/(м·К) вдоль и поперек плоскости соответственно, а электропроводность была равна 10^5 См/м [13]. Температурные зависимости для графитовой фольги были выбраны из [12] с учетом приведенных значений при комнатной температуре (рис. 2, а).

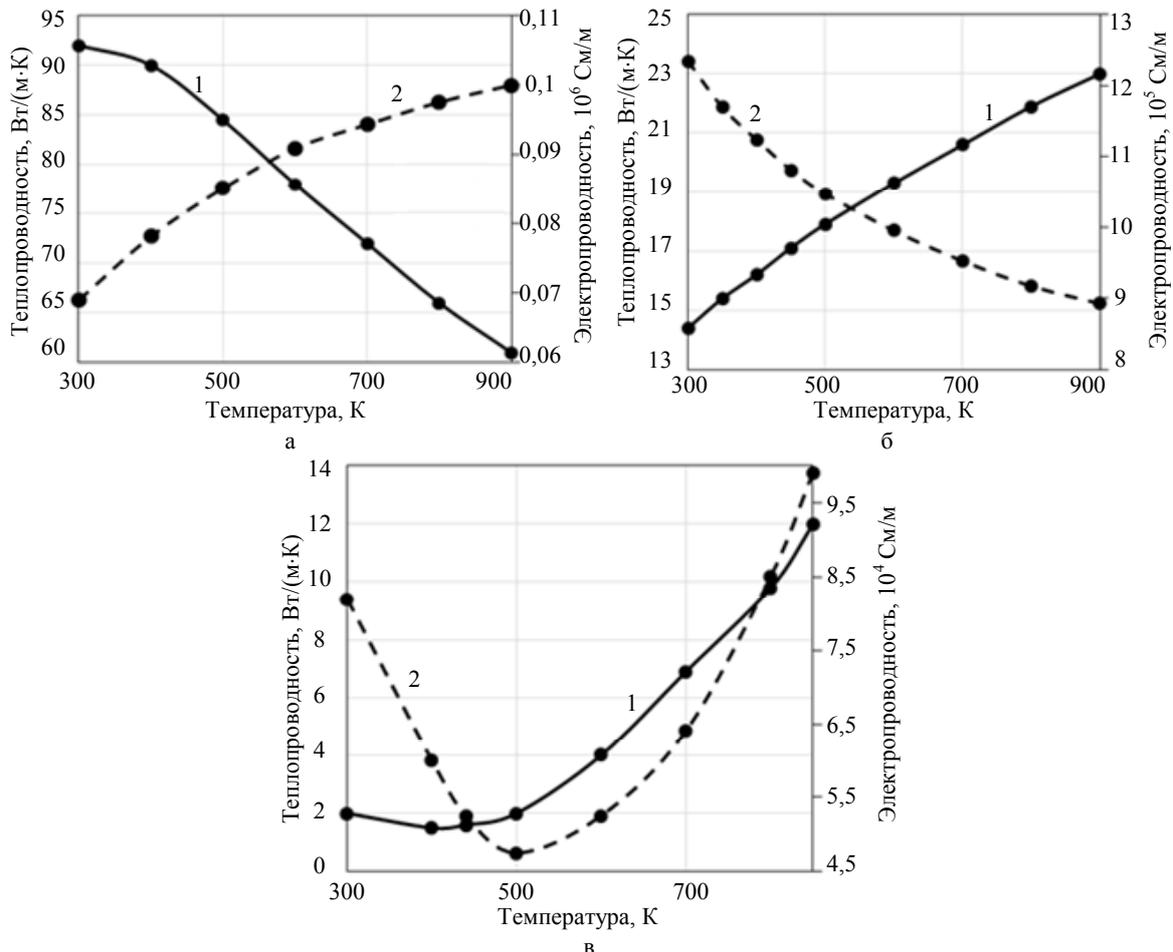


Рис. 2. Температурная зависимость теплопроводности (1) и удельного сопротивления (2): графита (а) [12]; стали (б) [14]; твердого раствора на основе Bi_2Te_3 (в) [15]

Кинетические коэффициенты для Bi_2Te_3 в плоскости скола в широком диапазоне температур измерены в [15]. Эти данные были пересчитаны с учетом анизотропии электропроводности $\sigma_{33} / \sigma_{11} = 0,36$ и теплопроводности $\kappa_{33} / \kappa_{11} = 0,47$ [16], которые для изотропного поликристалла дают значения

$\sigma_{poly} / \sigma_{11} = 0,74$ и $\kappa_{poly} / \kappa_{11} = 0,79$. Указанные зависимости приведены на рис. 2, в. Реальные значения сопротивления порошка на начальной стадии неизвестны, однако в использованном здесь стационарном подходе на конечном этапе прессования значения кинетических коэффициентов должны быть близки к приведенным.

С использованием приведенных данных и схемы установки (рис. 1) построена модель и рассчитаны распределения температуры и потенциала. Заданы следующие геометрические размеры элементов установки: графитовые вставки (диаметр $D = 80$ мм, толщина $h = 60$ мм); графитовые пуансоны ($D = 20$ мм, $h = 20$ мм); графитовая фольга ($D = 80$ мм, $h = 0,7$ мм); прокладки и трубка из графитовой фольги ($h = 0,2$ мм); пресс-форма (внешний диаметр – 40 мм, внутренний – 20 мм, $h = 40$ мм); спекаемый образец ($D = 20$ мм; $h = 5$ мм).

Моделирование тепловых и электрических полей выполнялось с помощью метода конечных элементов в программном продукте Comsol Multiphysics. Для расчетов использовалась осесимметричная модель, т.е. система уравнений (1) решалась в цилиндрических координатах.

Основные результаты моделирования тепловых и электрических полей

Распределения потенциала и температуры для случая граничных условий теплового излучения на боковых гранях и конвективного теплообмена с охлаждаемых электродов в условиях эксперимента приведены на рис. 3.

Температура образца, определенная путем усреднения по его объему, составила $T_s = 792$ К, что хорошо согласуется с результатами эксперимента [7].

Рис. 3 подтверждает тот факт, что областям наибольшего изменения электрического потенциала, т.е. зонам наибольшей напряженности электрического поля, соответствует наибольшая температура.

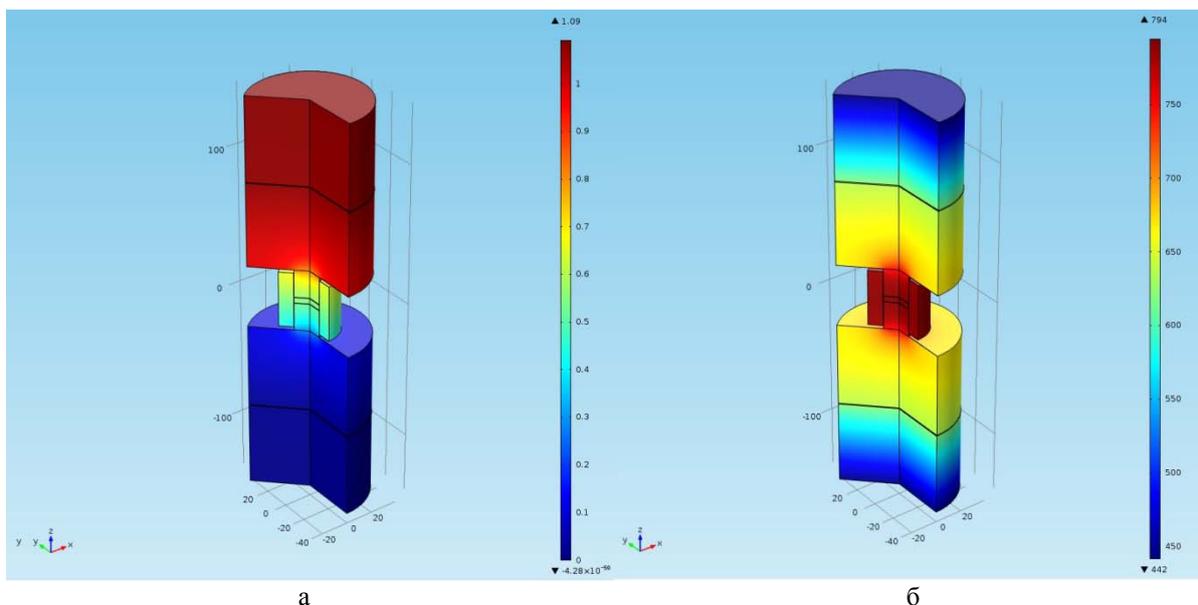


Рис. 3. Расчетное распределение потенциала (а) и температуры (б)

Экспериментальные результаты [7] показывают, что при данном химическом составе спекаемого термоэлектрического материала определяющее влияние на величину термоэлектрической добротности оказывает температура спекания T_s . С другой стороны, для практических целей важно иметь образцы с высокой степенью однородности термоэлектрических свойств, т.е. спекание всех точек образца должно проходить при одной температуре. Для проверки степени различия температуры спекания в зависимости от точки образца были проведены расчеты распределения температуры в образце в ходе искрового плазменного спекания в материалах на основе Bi_2Te_3 близких составов в условиях эксперимента. Расчетные зависимости распределения температуры в объеме спекаемого образца приведены на рис. 4. Легко убедиться, что, по крайней мере, для типичной SPS-установки с цилиндрическими пресс-формой и пуансонами (рис. 1) различие температуры в разных точках спекаемого образца невелико.

На температурное и электрические поля внутри спекаемого образца могут оказывать существенное влияние геометрические размеры пресс-формы. В рамках предложенной модели было проанализировано влияние внешнего диаметра графитовой пресс-формы на температуру в спекаемом образце (рис. 5) и распределение потенциала.

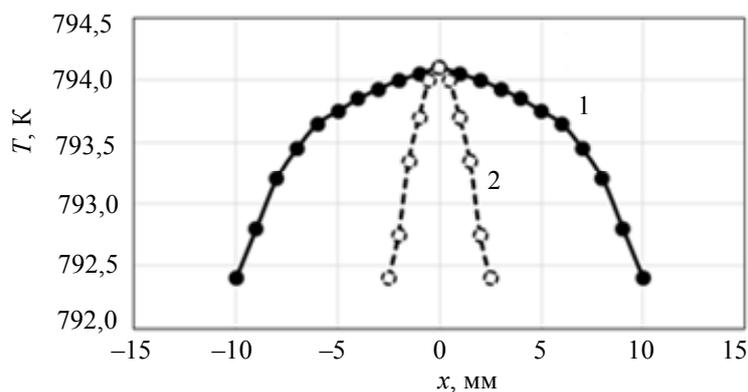


Рис. 4. Типичное распределение температуры в продольном (1) и поперечном (2) сечении образца в процессе спекания. Значению $x=0$ соответствует центр образца

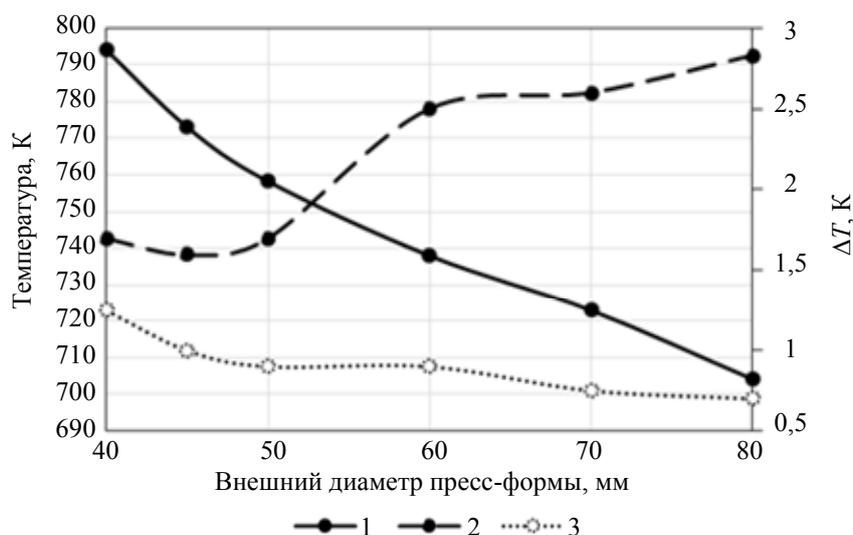


Рис. 5. Зависимость температуры в центре спекаемого образца (1) и разности температур между центральной областью образца и его внешней поверхностью в поперечном (2) и продольном (3) направлениях от внешнего диаметра пресс-формы

Из представленных на рис. 5 зависимостей видно, что увеличение диаметра пресс-формы приводит к снижению температуры в образце, при этом увеличивается перепад температуры между центром и боковыми поверхностями образца (увеличивается радиальный градиент температуры). Одновременно с этим осевой градиент температуры снижается. Разность потенциалов между верхним и нижним охлаждаемым электродом при изменении внешнего диаметра пресс-формы 40–80 мм увеличивается на 5%. Из представленных результатов видно, что существуют оптимальные геометрические размеры пресс-формы для получения однородного наноструктурированного материала. По результатам расчетов необходимых условий спекания Bi_2Te_3 в установке SPS-511S удастся достичь при использовании цилиндрической пресс-формы с внешним диаметром, лежащим в диапазоне 40–45 мм.

Заключение

С целью выбора режимов искрового плазменного спекания, которые могли бы обеспечить получение наноматериалов с оптимизированными термоэлектрическими свойствами, в настоящей работе решена следующая задача. На примере термоэлектрических материалов на основе Bi_2Te_3 выяснено, можно ли воспроизвести полученные экспериментально значения температуры образца при заданных значениях токов и разумных параметрах материалов и характеристик установки. Расчеты выполнены на примере установки SPS-511S, с помощью которой экспериментально производилось спекание термоэлектрических материалов на основе Bi_2Te_3 . Используемая модель включает систему дифференциальных уравнений, описывающих распределение электрического потенциала и температуры. Учтены температурные зависимости электро- и теплопроводности Bi_2Te_3 , а также графита и стали (материалов, из которых изготовлена установка). Максимальное значение температуры образца определялось в рамках стационарной задачи. Граничные условия включали конвективный теплообмен и излучение по закону Стефана–Больцмана.

Получены распределения электрического потенциала и температуры. Температура образца определенная путем усреднения по его объему, составила 792 К, что хорошо согласуется с результатами эксперимента. Для типичной SPS-установки различие температуры в разных точках спекаемого образца невелико.

Определяющую роль в формировании температурного поля в образцах играют геометрические размеры графитовой пресс-формы. Изменяя ее размеры и геометрию, можно воздействовать на условия процесса спекания, что позволит создавать образцы термоэлектриков с прогнозируемыми свойствами. Установлены оптимальные размеры графитовой пресс-формы для создания температурных условий, обеспечивающих в процессе спекания получение однородного наноструктурированного Bi_2Te_3 .

Литература

1. Bulat L.P., Drabkin I.A., Karatayev V.V., Osvenskii V.B., Parkhomenko Yu.N., Lavrentev M.G., Sorokin A.I., Pshenai-Severin D.A., Blank V.D., Pivovarov G.I., Bublik V.T., Tabachkova N.Yu. Structure and transport properties of bulk nanothermoelectrics based on $\text{Bi}_x\text{Sb}_{2-x}\text{Te}_3$ fabricated by SPS method // *Journal of Electronic Materials*. 2013. V. 42. N 7. P. 2110–2113.
2. Драбкин И.А., Освенский В.Б., Сорокин А.И., Булат Л.П., Пивоваров Г.И. Анизотропия термоэлектрических свойств объемного наноструктурированного материала на основе $(\text{Bi,Sb})_2\text{Te}_3$, полученного методом искрового плазменного спекания (SPS) / Труды межгосударственного семинара «Термоэлектрики и их применения». СПб, 2013. С. 29–34.
3. Бублик В.Т., Драбкин И.А., Каратаев В.В., Лаврентьев М.Г., Освенский В.Б., Булат Л.П., Пивоваров Г.И., Сорокин А.И., Табачкова Н.Ю. Объемный наноструктурированный термоэлектрический материал на основе $(\text{Bi,Sb})_2\text{Te}_3$, полученный методом искрового плазменного спекания (SPS) / Труды межгосударственного семинара «Термоэлектрики и их применения». СПб, 2013. С. 23–28.
4. Drabkin I.A., Osvenski V.B., Parkhomenko Yu.N., Sorokin A.I., Pivovarov G.I., Bulat L.P. Anisotropy of thermoelectric properties of p-type nanostructured material based on $(\text{Bi, Sb})_2\text{Te}_3$ // *Journal of Thermoelectricity*. 2013. N 3. P. 35–46.
5. Булат Л.П., Освенский В.Б., Пархоменко Ю.Н., Пшенай-Северин Д.А. Исследование возможностей увеличения термоэлектрической эффективности в наноструктурированных материалах на основе $\text{Bi}_2\text{Te}_3\text{-Sb}_2\text{Te}_3$ // *Физика твердого тела*. 2012. Т. 54. № 11. С. 2036–2042.
6. Булат Л.П., Освенский В.Б., Пшенай-Северин Д.А. Влияние распределения зерен по размерам на решеточную теплопроводность наноструктурированных материалов на основе $\text{Bi}_2\text{Te}_3\text{-Sb}_2\text{Te}_3$ // *Физика твердого тела*. 2013. Т. 55. № 12. С. 2323–2330.
7. Bulat L.P., Drabkin I.A., Karatayev V.V., Osvenskii V.B., Parkhomenko Yu.N., Pshenay-Severin D.A., Sorokin A.I. The influence of anisotropy and nanoparticle size distribution on lattice thermal conductivity and the thermoelectric figure of merit in nanostructured $(\text{Bi,Sb})_2\text{Te}_3$ // *Journal of Electronic Materials*. 2014. V. 43. N 6. P. 2121–2126.
8. Anselmi-Tamburini U., Gennari S., Garay J.E., Munir Z.A. Fundamental investigations on the spark plasma sintering/synthesis process: II. Modeling of current and temperature distributions // *Materials Science and Engineering*. 2005. V. 394. N 1–2. P. 139–148.
9. Cengel Y.A., Ghajar A.J. *Heat and Mass Transfer*. 4th ed. McGraw-Hill, 2011. 960 p.
10. Total Emissivity of Various Surfaces [Электронный ресурс]. 2003. Режим доступа: <http://www.contika.dk/Download/litteratur/emission.pdf>, свободный. Яз. англ. (дата обращения 11.05.2014).
11. Overall Heat Transfer Coefficients for Some Common Fluids and Heat Exchanger Surfaces [Электронный ресурс]. Режим доступа: http://www.engineeringtoolbox.com/overall-heat-transfer-coefficients-d_284.html, свободный. Яз. англ. (дата обращения 11.05.2014).
12. Hust J.G. Standard Reference Materials: a Fine-Grained, Isotropic Graphite for Use as NBS Thermophysical Property RM's from 5 to 2500 K [Электронный ресурс]. NBS Special Publication 260-89. 1984. Режим доступа: <http://www.nist.gov/srm/upload/SP260-89.PDF>, свободный. Яз. англ. (дата обращения 11.05.2014).
13. Графитовая фольга графлекс [Электронный ресурс]. 2010. Режим доступа: http://traverss.ru/grafitovaya_folga_graf, свободный. Яз. рус. (дата обращения 11.05.2014).
14. Hust J.G., Giarratano P.J. Standard Reference Materials: Thermal Conductivity and Electrical Resistivity Standard Reference Materials: Austenitic Stainless Steel, SRM's 735 and 798, from 4 to 1200 K [Электронный ресурс]. NBS Special Publication 260-46. 1975. Режим доступа: <http://www.nist.gov/srm/upload/SP260-46.PDF>, свободный. Яз. англ. (дата обращения 11.05.2014).
15. Магомедов Я.Б., Гаджиев Г.Г., Омаров З.М. Температурная зависимость теплопроводности и электропроводности Bi_2Te_3 и его расплава // Фазовые переходы, упорядоченные состояния и новые материалы. 2013. № 9. С. 1–5.

16. Stecker K., Süßmann H., Eichler W., Heiliger W., Stordeur M. Martin-Luther Univ. Halle-Wittenberg, Math-Naturwiss. 1978. V. 27. N 5. P. 5.

- Булат Лев Петрович** – доктор физико-математических наук, профессор, заведующий кафедрой, Университет ИТМО, 197101, Санкт-Петербург, Россия, bulat@irbt-itmo.ru
- Пшенай-Северин Дмитрий Александрович** – кандидат физико-математических наук, старший научный сотрудник, ведущий инженер, Университет ИТМО, 197101, Санкт-Петербург, Россия; старший научный сотрудник, Физико-технический институт им. А.Ф. Иоффе, 194021, Санкт-Петербург, Россия, d.pshenay@gmail.com
- Нефедова Ирина Александровна** – аспирант, Университет ИТМО, 197101, Санкт-Петербург, Россия, nefedosita@rambler.ru
- Новотельнова Анна Владимировна** – кандидат технических наук, доцент, доцент, Университет ИТМО, 197101, Санкт-Петербург, Россия, novotelnova@yandex.ru
- Гуревич Юрий Генрихович** – доктор физико-математических наук, профессор, профессор-исследователь, Исследовательский центр Национального политехнического института Мексики, 07738, Мехико, Мексика, gurevich@fis.cinvestav.mx
- Lev P. Bulat** – D.Sc., Professor, Department head, ITMO University, 197101, Saint Petersburg, Russia, bulat@irbt-itmo.ru
- Dmitry A. Pshenai-Severin** – PhD, leading engineer, ITMO University, 197101, Saint Petersburg, Russia; senior scientific researcher, Ioffe Physical Technical Institute, 194021, Saint Petersburg, Russia, d.pshenay@gmail.com
- Irina A. Nefedova** – postgraduate, ITMO University, 197101, Saint Petersburg, Russia, nefedosita@rambler.ru
- Anna V. Novotelnova** – PhD, Associate professor, ITMO University, 197101, Saint Petersburg, Russia, novotelnova@yandex.ru
- Yuri G. Gurevich** – D.Sc., Professor-Researcher, CINVESTAV I.P.N. (Center for Research and Advanced Studies of the National Polytechnic Institute), 07738, Mexico-City, Mexico, gurevich@fis.cinvestav.mx

Принято к печати 12.05.14
Accepted 12.05.14

УДК 666.1:666.3

ВЛИЯНИЕ ОДНОСТАДИЙНОЙ ОБРАБОТКИ КВАРЦЕВОЙ КЕРАМИКИ ГЕЛЕОБРАЗУЮЩИМИ ВОДНЫМИ РАСТВОРАМИ НА ПРОЧНОСТНЫЕ ХАРАКТЕРИСТИКИ

С.К. Евстропьев^a, В.М. Волынкин^b, А.В. Шашкин^b, К.В. Дукельский^c, А.Г. Коробейников^{d,e},
Ю.А. Гатчин^d, В.И. Поляков^d

^a ОАО «ГОИ им. С.И. Вавилова», 192171, Санкт-Петербург, Россия

^b ОАО «НИТИОМ ВНЦ «ГОИ им. С.И. Вавилова», 192171

^c Санкт-Петербургский государственный университет аэрокосмического приборостроения, 190000, Санкт-Петербург, Россия

^d Университет ИТМО, 197101, Санкт-Петербург, Россия, Korobeynikov_A_G@mail.ru

^e Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им.Н.В. Пушкова Российской академии наук (СПбФ ИЗМИРАН), 199034, Санкт-Петербург, Россия

Аннотация. Приведены основные результаты исследования влияния обработки кварцевой керамики кремнийсодержащими и алюминийсодержащими гелеобразующими водными растворами на ее прочность и пористость. Для пропитки образцов пористой кварцевой керамики предложены и изготовлены водно-спиртовые растворы на основе тетраэтоксисилана (TEOS, $\text{Si}(\text{C}_2\text{H}_5\text{O})_4$) с добавками нитрата алюминия. Образцы керамики пропитывались растворами при комнатной температуре в течение 12 мин. После пропитки образцы подвергались сушке при комнатной температуре в течение 24 ч и термообработке в электрической муфельной печи. Проведенные эксперименты показали, что пропитка гелеобразующими растворами пористых образцов кварцевой керамики даже без дополнительной термообработки приводит к увеличению прочности необожженных образцов в 6–7 раз. Высокотемпературная термообработка предварительно пропитанных керамических образцов приводит к разложению нитрата алюминия и удалению органических остатков, а также упрочнению образовавшихся дополнительных связей между частицами материала. Это также значительно улучшает прочностные характеристики кварцевой керамики. Таким образом, экспериментально показана возможность значительного упрочнения пористой кварцевой керамики и увеличения стабильности ее прочностных свойств путем предварительной пропитки кремнийсодержащими и алюминийсодержащими гелеобразующими растворами даже без дополнительной термообработки. Выявлено, что пропитка образцов пористой кварцевой керамики этими растворами приводит лишь к незначительному уменьшению пористости образцов. Последующая термообработка пропитанных пористых керамических образцов при температуре 900–1200 °С приводит к дополнительному значительному увеличению их механической прочности.

Ключевые слова: кварцевая керамика, пропитывающий раствор, прочность, температура.

INFLUENCE OF QUARTZ CERAMICS SINGLE-STAGE PROCESSING BY GEL-FORMING WATER SOLUTIONS ON ITS STRENGTH CHARACTERISTICS

S. K. Evstropiev^a, V. M. Volynkin^a, A. V. Shashkin^a, K. B. Dukel'sky^b,
A. G. Korobeynikov^{c,d}, Yu. A. Gatchin^c, V. I. Polyakov^c

^a Vavilov State Optical Institute, 192171, Saint Petersburg, Russia, evstropiev@bk.ru

^b Saint Petersburg State University of Aerospace Instrumentation, 190000, Saint Petersburg, Russia, KDukel@GOI.ru

^c ITMO University, 197101, Saint Petersburg, Russia, Korobeynikov_A_G@mail.ru

^d Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences St. Petersburg Branch (St.P. FIZMIRAN), 199034, Saint Petersburg, Russia, Korobeynikov_A_G@mail.ru

Abstract. The main research results of the influence of quartz ceramics processing by silicon- and aluminum-containing gel-forming water solutions on its durability and porosity are given. Aqueous-alcoholic solutions based on tetraethoxysilane (TEOS, $\text{Si}(\text{C}_2\text{H}_5\text{O})_4$) with additives of aluminum nitrate are proposed for impregnation of porous quartz ceramics samples. Ceramic samples are being impregnated with solutions at the room temperature for 12 minutes. After impregnation they are being exposed to drying at the room temperature for 24 hours and heat treatment in the electric muffle furnace. The made experiments show that impregnation of quartz ceramics porous samples by gel-forming solutions leads to durability growth of not burned samples by 6-7 times even without additional heat treatment. High-temperature heat treatment of previously impregnated ceramic samples leads to decomposition of aluminum nitrate and removal of fossils, and also to hardening of the formed additional bonds between material particles. It considerably improves strength characteristics of quartz ceramics as well. Thus, the possibility of considerable hardening of porous quartz ceramics and stability growth of its strength properties by preliminary impregnation of silicon- and aluminum-bearing gel-forming solutions even without additional heat treatment is experimentally shown. It is revealed that impregnation of porous quartz ceramic samples by these solutions leads only to insignificant reduction of porosity of samples. Subsequent heat treatment of the impregnated porous ceramic samples at the temperatures, equal to 900-1200°C, results in additional significant increase in their mechanical durability.

Keywords: quartz ceramics, impregnating solution, durability, temperature.

Введение

Кварц широко применяется при производстве кварцевого стекла [1] и кремнеземистых огнеупоров. Например, кварцевая керамика является хорошо известным материалом, применяемым в различных областях промышленности. Традиционно в качестве исходного материала для получения кварцевой керамики используются шликеры на основе дисперсных порошков кварцевого стекла [2]. Для создания стек-

лоприпаса (тигли, мешалки и др.) в качестве метода формования изделий используется шликерное литье в гипсовые формы с последующей сушкой и термообработкой.

Решению проблемы увеличения прочности и уменьшения пористости различных изделий из кварцевой керамики посвящено большое количество научных статей и патентов (например, [3–7]). Одним из наиболее эффективных способов упрочнения кварцевой керамики является ее пропитка растворами, содержащими модифицирующие компоненты [4, 8–10]. Для пропитки используются водные растворы солей хрома [9, 10], алюминия [4], а также растворы, содержащие соединения щелочных или щелочноземельных металлов, фосфора или бора [8].

Физико-химические процессы, протекающие при сушке и термообработке пропитанной кварцевой керамики и определяющие ее упрочнение, включают осаждение в порах материала модифицирующих соединений и их дальнейшую эволюцию в процессе термообработки. В [4] показано, что при сушке и последующей термообработке кварцевой керамики, обработанной раствором соли алюминия, происходит формирование наночастиц $\alpha\text{-Al}_2\text{O}_3$ и их внедрение в зоны стыка зерен кварцевого стекла за счет массопереноса.

Хорошо известным связующим материалом являются золи кремнезема [11]. На практике в качестве исходного материала для коллоидного кремнезема используются растворы жидкого стекла (см., например, [12]). Для получения стеклоприпаса, свободного от примесей щелочных металлов, в качестве исходного материала для получения золь кремнезема можно использовать кремнийорганические соединения. Использование кремнийорганических связующих для производства технической оксидной керамики описано в литературе. Так, в [13] приведены результаты разработок по использованию тетраэтоксисилана (TEOS) и других органосиликатных соединений в качестве связующего для порошкообразных материалов при изготовлении композиционных материалов.

В водных растворах TEOS гидролизуетеся с образованием коллоидных частиц кремнезема, размер и степень агрегации которых зависят от концентрации и pH растворов. Влияние pH раствора на протекание процессов гидролиза TEOS и конденсации частиц кремнезема подробно описано в литературе (см., например, [10]).

В [14] описано применение в качестве пропитывающей жидкости TEOS для упрочнения безобжиговой кварцевой керамики. При этом технологический процесс осуществляется в два этапа. На первом этапе пористая заготовка из кварцевой керамики пропитывается чистым (негидролизованным) TEOS. Гидролиз TEOS и формирование коллоидных частиц кремнезема осуществляется в щелочной среде на втором этапе путем дополнительной обработки уже пропитанной заготовки водным раствором аммиака в течение 24 ч. Необходимость использования двухстадийного и длительного процесса пропитки керамического материала является существенным недостатком способа упрочнения, описанного в [14].

В настоящей работе представлены результаты исследования влияния одностадийной обработки кварцевой керамики гелеобразующими водными растворами, содержащими TEOS и нитрат алюминия, на прочность и пористость материала.

Методика эксперимента

В качестве исходного материала были использованы образцы кварцевой керамики, полученные методом шликерного литья в гипсовые формы и подвергнутые сушке при комнатной температуре. Образцы имели форму штабиков размерами $65 \times 8 \times 8$ мм и характеризовались пористостью около 25%.

Для пропитки образцов пористой кварцевой керамики были изготовлены водно-спиртовые растворы на основе тетраэтоксисилана (TEOS, $\text{Si}(\text{C}_2\text{H}_5\text{O})_4$) с добавками нитрата алюминия. Химический состав растворов приведен в табл. 1. Образцы керамики пропитывались растворами при комнатной температуре в течение 12 мин. После пропитки образцы подвергались сушке при комнатной температуре в течение 24 ч и термообработке в электрической муфельной печи. Для испытаний в каждом технологическом режиме обработки (температура и длительность термообработки; состав пропитывающего раствора) использовались серии из 5 образцов.

Номер пропитывающего раствора	Химический состав растворов			
	Вода	Этанол	TEOS	$\text{Al}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$
1	10	25	29	4
2	9	25	30	5
3	9	25	30	10

Таблица 1. Химический состав пропитывающих растворов, г

Определение кажущейся плотности и открытой пористости проведено методом гидростатического взвешивания в соответствии с ГОСТ 2409-80. Определение предела прочности при изгибе проведено ме-

тодом трехточечного изгиба. Морфология материалов исследовалась методом сканирующей электронной микроскопии на приборе SUPRA 55VP-25-78.

Экспериментальные результаты и обсуждение

Проведенные эксперименты показали, что пропитка гелеобразующими растворами пористых образцов кварцевой керамики даже без дополнительной термообработки приводит к значительному увеличению их прочности. В табл. 2 приведены экспериментально определенные значения прочности исходных и пропитанных различными растворами образцов кварцевой керамики.

Номер пропитывающего раствора	Прочность, МПа	Среднеквадратичное отклонение σ , МПа	Относительное стандартное отклонение σ/P_{cp}
Без пропитки	2,06	0,2	0,09
1	13,44	0,23	0,17
2	15,00	0,18	0,12
3	13,04	0,7	0,05

Таблица 2. Прочность керамических образцов до термообработки

В [12, 15] существенное упрочнение кварцевой керамики без стадии высокотемпературного обжига достигалось при ее автоклавной обработке. Полученные нами данные (табл. 2) свидетельствуют о том, что пропитка гелеобразующими растворами увеличивает прочность необожженных образцов в 6–7 раз. Достигнутое значительное упрочнение материала позволяет рассматривать пропитку гелеобразующими растворами в качестве эффективного метода обработки необожженных изделий из кварцевой керамики, снижающего риск их повреждения при транспортировке или последующей обработке.

Высокотемпературная термообработка предварительно пропитанных керамических образцов приводит к разложению нитрата алюминия и удалению органических остатков, а также упрочнению образовавшихся дополнительных связей между частицами материала. На рис. 1 представлены зависимости прочности исходного (кривая 1) и пропитанного раствором № 1 (см. табл. 1) (кривая 2) керамических образцов от температуры термообработки. Видно, что упрочнение исходных непропитанных образцов наблюдается лишь при температурах термообработки более 1000–1100 °С. Термообработка непропитанных керамических образцов при температуре 900 °С не приводит к изменению их прочности. Аналогичная термообработка пропитанных образцов приводит к значительному их упрочнению.

Проведенные нами эксперименты показали также, что предварительная пропитка пористых образцов уменьшает величины относительного отклонения значений их прочности. Для пропитанных образцов эти значения значительно меньше, чем для необработанной керамики, во всем использованном диапазоне температур термообработки (900–1200 °С). Это свидетельствует о том, что обработка кварцевой керамики предварительной пропиткой гелеобразующими растворами обеспечивает не только значительное увеличение прочности, но и существенное уменьшение разброса ее значений, т.е. возрастает стабильность прочностных характеристик материала.

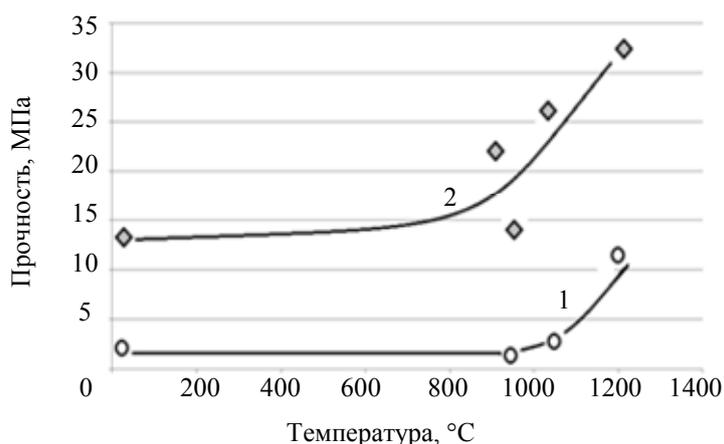


Рис. 1. Зависимость прочности на изгиб от температуры термообработки традиционной кварцевой керамики (кривая 1) и кварцевой керамики, предварительно пропитанной гелеобразующим раствором (кривая 2)

Измерения плотности и пористости образцов показали, что однократная пропитка лишь незначительно уменьшает пористость керамики. Так, после термообработки при 1050 °С в течение 2 ч пористость непропитанной керамики составляет 20%, а предварительная пропитка раствором № 1 (см. табл. 1) с последующей аналогичной термообработкой уменьшает пористость образцов до 18%. Аналогично,

лишь небольшое уменьшение пористости керамики при предварительной их однократной пропитке наблюдается и при других температурах термообработки.

При сушке и последующей высокотемпературной термообработке из пропитывающего раствора в порах керамического материала формируются оксидные соединения кремния и алюминия. При этом объем образующейся твердой фазы мал по сравнению с объемом пропитывающего раствора и с объемом заполненных этим раствором пор соответственно. В связи с этим пористость керамического материала при однократной его пропитке и последующей термообработке уменьшается незначительно. Наблюдаемое при предварительной пропитке значительное упрочнение керамики при лишь небольшом уменьшении ее пористости может объясняться тем, что формирующийся из раствора оксидный материал заполняет мелкие поры и дополнительно связывает частицы керамики.

На рис. 2 приведены электронномикроскопические снимки поверхности образцов исходной кварцевой керамики (рис. 2, а) и образца керамики, предварительно пропитанного раствором № 1 (рис. 2, б). Оба образца керамики были подвергнуты термообработке при 1050 °С. Сравнение приведенных снимков показывает, что принципиальных изменений в структуре материала не наблюдается. Однако обращает на себя внимание визуально проявляемая несколько большая связанность структуры материала, подвергнутого предварительной пропитке.

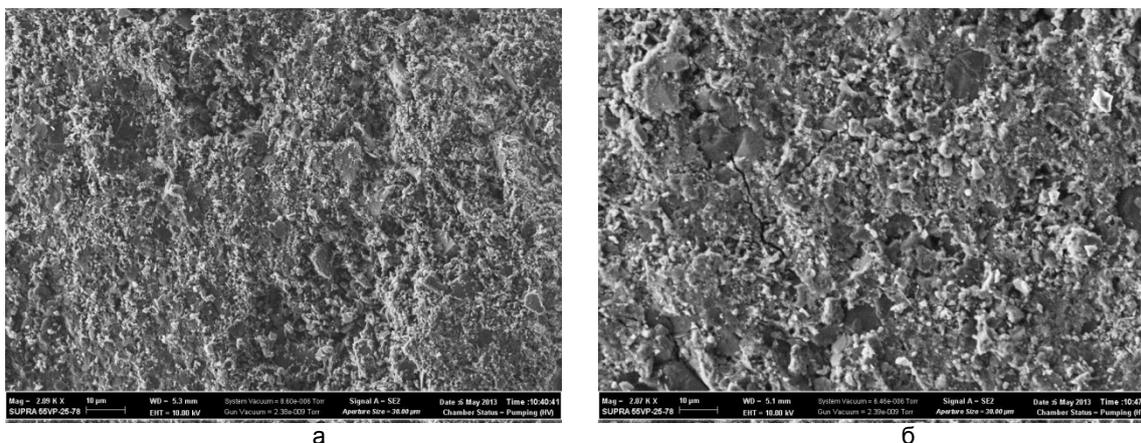


Рис. 2. Электронно-микроскопический снимок поверхности кварцевой керамики после термообработки при 1050°С: не прошедшей обработку пропитывающим раствором (а); после обработки пропитывающим раствором (б)

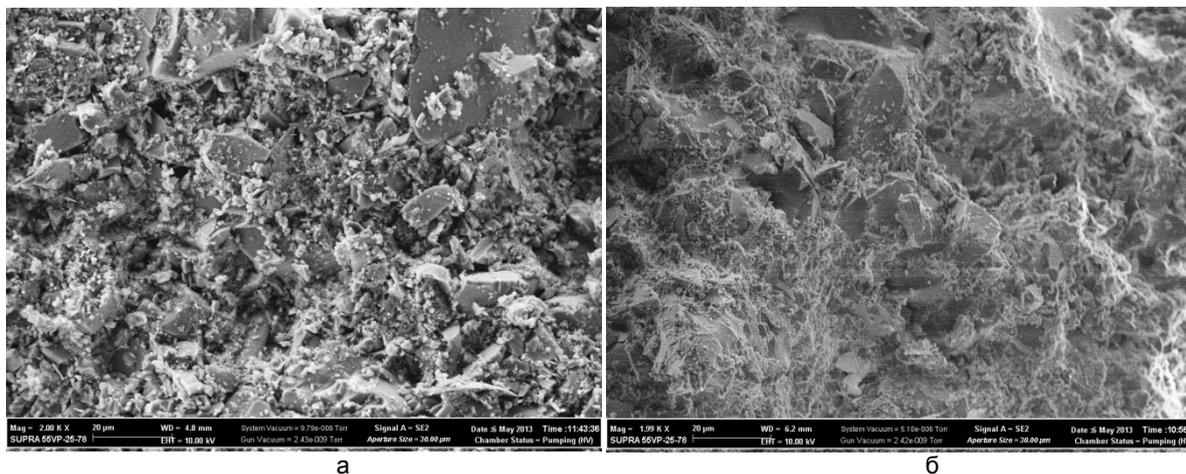


Рис. 3. Электронно-микроскопический снимок поверхности кварцевой керамики после термообработки при 1050°С: не прошедшей обработку пропитывающим раствором (а); после обработки пропитывающим раствором (б)

Существенно другая картина наблюдается для электронно-микроскопических снимков областей излома образцов (рис. 3). Если для образца керамики, не подвергнутого упрочняющей обработке, наблюдается полное разрушение структуры материала (рис. 3, а), то образец, подвергнутый упрочняющей обработке (рис. 3, б), сохраняет связанные воедино кремнеземистые частицы.

Заключение

Экспериментально показана возможность значительного упрочнения кварцевой керамики и увеличения стабильности ее прочностных свойств путем предварительной пропитки кремнийсодержащими и

алюминийсодержащими гелеобразующими растворами до стадии высокотемпературного спекания материала. Пропитка этими растворами приводит к дополнительному связыванию частиц керамического материала и лишь незначительно уменьшает его пористость. Последующая термообработка пропитанных керамических образцов при температурах 900–1200 °С приводит к дополнительному значительному увеличению их механической прочности.

Литература

1. Коробейников А.Г., Гатчин Ю.А., Дукельский К.В., Тер-Нерсисянц Е.В. Технологические методы снижения уровня оптических потерь в микроструктурированных волоконных световодах // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 1 (89). С. 31–38.
2. Пивинский Ю.Е., Ромашин А.Г. Кварцевая керамика. М.: Металлургия, 1974. 264 с.
3. Kanzaki S., Tabata H., Abe O., Shigetoshi O., Chitake Y., Senshu M. Process for producing alumina silica sintered ceramics having improved high-temperature strength. Patent US N4895814, filling date 24.07.1987.
4. Бородай Ф.Я., Викулин В.В., Иткин С.М., Ляшенко Л.П., Шкарупа И.Л., Самсонов В.И. Наномодифицированная кварцевая керамика с повышенной высокотемпературной прочностью. Патент РФ №2458022. Бюл. 2012. №22.
5. Байкова Л.Г., Мамалимов Р.И., Песина Т.И., Чмель А.Е., Щербаков А.И. Структурные превращения в процессе термической обработки кварцевой керамики // Стекло и керамика. 2013. № 8. С. 32–34.
6. Викулин В.В., Бородай Ф.Я., Иткин С.М., Русин М.Ю., Шкарупа И.Л. Способ получения кварцевой керамики и изделий из нее. Патент РФ №2380341. Бюл. 2010. №3.
7. Иткин С.М., Викулин В.В., Русин М.Ю., Шкарупа И.Л. Способ обработки изделий из кварцевой керамики. Патент РФ №2267470. Бюл. 2006. № 1.
8. Sato N., Emori I., Takada K., Inomoto S., Kawazu M. Method for producing porous silica ceramic material. Patent US N 20090206525, filling date 23.06.2006.
9. Викулин В.В., Бородай Ф.Я., Бородай С.П., Шкарупа И.Л. Способ получения оболочек антенных обтекателей из кварцевой керамики. Патент РФ №2436206. Бюл. 2011. № 34.
10. Church P.K., Knutson O.J. Method of impregnating porous refractory bodies with inorganic chromium compound. Patent US N 3789096, filling date 28.12.1967.
11. Айлер Р. Химия кремнезёма: Пер. с англ. / Под ред. В.П. Прянишникова. М.: Мир, 1982. Ч. 2. 712 с.
12. Бевз В.А., Пивинский Ю.Е. Способ изготовления кварцевой керамики // А.с. СССР № 804607. Бюл. 1981. № 6.
13. Яо И.М. Композиционные керамические материалы на основе кремнийорганического связующего и тугоплавких бескислородных наполнителей. Автореф. дис. ... канд. техн. наук. Казань, Казанский государственный технологический университет, 2000. 18 с.
14. Матусевич И.С., Ахьян А.М. Способ упрочнения необожженной кварцевой керамики // А.с. СССР № 1137091. Бюл. 1985. № 4.

- Евстропьев Сергей Константинович** – доктор химических наук, ведущий научный сотрудник, ОАО «ГОИ им. С.И. Вавилова», 192171, Санкт-Петербург, Россия, evstropiev@bk.ru
- Волынкин Валерий Михайлович** – кандидат химических наук, ведущий научный сотрудник, ОАО «НИТИОМ ВНЦ «ГОИ им. С.И. Вавилова», 192171, Санкт-Петербург, Россия, volynkin@yandex.ru
- Шашкин Александр Викторович** – кандидат химических наук, начальник сектора, ОАО «НИТИОМ ВНЦ «ГОИ им. С.И. Вавилова», 192171, Санкт-Петербург, Россия, Shashkin2000@mail.ru
- Дукельский Константин Владимирович** – кандидат технических наук, доцент, доцент, Санкт-Петербургский государственный университет аэрокосмического приборостроения, 190000, Санкт-Петербург, Россия, KDukel@GOI.ru
- Коробейников Анатолий Григорьевич** – доктор технических наук, профессор, профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия; зам. директора по науке, Санкт-Петербургский филиал Федерального Государственного бюджетного учреждения науки Института Земного магнетизма, ионосферы и распространения радиоволн им.Н.В.Пушкова Российской академии наук (СПбФ ИЗМИРАН), 199034, Санкт-Петербург, Россия, Korobeynikov_A_G@mail.ru
- Sergei K. Evstropiev** – D.Sc., leading scientific researcher, Vavilov State Optical Institute, 192171, Saint Petersburg, Russia, evstropiev@bk.ru
- Valery M. Volynkin** – PhD, leading scientific researcher, Vavilov State Optical Institute, 192171, Saint Petersburg, Russia, volynkin@yandex.ru
- Alexander V. Shashkin** – PhD, Section head, Vavilov State Optical Institute, 192171, Saint Petersburg, Russia, Shashkin2000@mail.ru

- Konstantin V. Dukel'sky*** – PhD, Associate professor, Associate professor, Saint Petersburg State University of Aerospace Instrumentation, 190000, Saint Petersburg, Russia, KDukel@GOI.ru
- Anatoly G. Korobeynikov*** – D.Sc., Professor, ITMO University, 197101, Saint Petersburg, Russia; Deputy director for science, Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the Russian Academy of Sciences St. Petersburg Branch (St.P. FIZMIRAN), 199034, Saint Petersburg, Russia, Korobeynikov_A_G@mail.ru
- Yury A. Gatchin*** – D.Sc., Professor, Department head, ITMO University, 197101, Saint Petersburg, Russia, Gatchin@mail.ifmo.ru
- Vladimir I. Polyakov*** – PhD, Associate professor, Associate professor, ITMO University, 197101, Saint Petersburg, Russia, v_i_polyakov@mail.ru

Принято к печати 02.07.14
Accepted 02.07.14

УДК 678. 58

ИССЛЕДОВАНИЕ СОРБЦИОННЫХ ХАРАКТЕРИСТИК ПОЛИМЕРНЫХ МИНЕРАЛ-НАПОЛНЕННЫХ КОМПОЗИТОВ ДЛЯ МЕДИЦИНЫ

Ю.А. Игнатьева^а, М.В. Успенская^а, О.В. Борисов^б, Р.О. Олехнович^а, Р.А. Евсеев^с,
К.Н. Касанов^с^а Университет ИТМО, 197101, Россия, Санкт-Петербург, Alissia87@mail.ru^б Университет г. Пау, 64053, г. Пау, Франция^с Военно-медицинская академия имени С.М. Кирова, 194044, Санкт-Петербург, Россия

Аннотация. Методом радикальной полимеризации в водной среде получены и изучены полимерные композиции на основе акриловых производных и модифицированных ионами серебра частиц бентонита с различной долей и дисперсностью. В качестве исходных веществ выбраны производные частично нейтрализованной акриловой кислоты, акриламида и метиленбисакриламида, а также наполнителя – бентонита с массовой долей 0–0,05%. Продемонстрировано влияние концентрации наполнителя на абсорбционные характеристики полимерных материалов в дистиллированной воде. Показано, что увеличение доли минерал-содержащего наполнителя до 5 мас.% приводит к увеличению значений равновесной степени набухания в 1,5–2 раза по сравнению с ненаполненной полимерной матрицей. Наибольшими кинетическими характеристиками обладают акриловые нанокomпозиции с массовой долей модифицированного бентонита 0,01. Исследованы кинетические зависимости набухания новых композиционных материалов в физиологическом растворе от дисперсности наполнителя. Показано, что в области высокой дисперсности (менее 0,25 мкм) доля минерал-содержащего наполнителя в количестве 1 мас.% в составе композиции приводит к значительному увеличению значений равновесной степени набухания по сравнению с ненаполненным образцом (в 1,5 раза). Изучен эффект полиэлектролитного подавления набухания полимерных композиций в физиологическом растворе, что приводит к уменьшению в среднем на порядок значений равновесной степени набухания по сравнению с этими значениями в дистиллированной воде. Показана перспективность использования полученных композиций в качестве полимерных матриц при создании раневых повязок для лечения ран различной этиологии. Результаты работы рекомендуются к использованию в медицинской практике для оптимизации протекания раневого процесса.

Ключевые слова: сшитые сополимеры, акриловые полимеры, гидрогели, набухание, сорбция, бентонит, серебро.

Благодарности. Работа выполнена при государственной финансовой поддержке ведущих университетов Российской Федерации (субсидия 074-U01).

INVESTIGATION OF SORPTION CHARACTERISTICS OF POLYMERIC MINERAL-FILLED COMPOSITES FOR MEDICINE

Yu. A. Ignatieva^a, M.V. Uspenskaya^a, O.V. Borisov^b, R.O. Olekhovich^a, R.A. Evseev^c, K.N. Kasanov^c^a ITMO University, 197101, Saint Petersburg, Russia, Alissia87@mail.ru^b Université de Pau et des Pays de l'Adour, 64053, Pau, France^c Military Medical Academy named after S.M. Kirov, 194044, Saint Petersburg, Russia

Abstract. The polymer compositions on the base of acrylic derivatives and bentonite particles modified by silver ions with various share and dispersion are received and studied by radical polymerization in the water. Partially neutralized acrylic acid, acrylamide and methylene-bis-acrylamide and particles of bentonite with fraction 0 - 0,05 mass.% are chosen as initial substances. The influence of bentonite concentration on absorbing characteristics of polymer materials in the distilled water is shown. It is demonstrated that the increase of bentonite fraction up to 5 mass.% leads to the rise of degree of equilibrium swelling by 1,5 – 2 times in comparison with an unfilled polymer matrix. The acrylic nanocompositions with a mass fraction of bentonite equal to 0,01 mass.% possess the greatest kinetic characteristics. Kinetic dependences of new composite materials swelling in physiological solution from a filler dispersion part are investigated. It is shown that in high dispersion (with particle size less than 0,25 μm) a part of mineral-containing filler equal to 1 mass.% leads to significant increase in values of equilibrium swelling degree in comparison with an unfilled sample (by 1,5 times). The effect of polyelectrolyte suppression of polymer composition swelling in physiological solution is studied. It results in values reduction of equilibrium swelling degree in comparison with these values in the distilled water. Application prospects for the received compositions are shown at bandages creation for wounds treatment of various etiologies. Research results are recommended for usage in medical practice for optimization of wound process march.

Keywords: cross-linked copolymers, acrylic polymers, hydro-gels, swelling, sorption, bentonite, silver.

Acknowledgements. The work was partially financially supported by the Government of the Russian Federation (grant 074-U01).

Введение

Сшитые сополимеры, называемые также гидрогелями, все глубже внедряются в различные аспекты нашей жизни, находя применение в промышленности, сельском хозяйстве, медицине [1–5]. В частности, в медицине гидрогели используются в качестве носителей лекарственных препаратов, биочипов, ионитов для сепарации белков, при создании предметов санитарно-гигиенического назначения и т.п. [6–10]. Особым классом в последнее время являются материалы для создания покрытий при лечении ран различной природы, полученных при природных и техногенных катастрофах, конфликтах и на производстве. Основными требованиями к таким материалам – способность защиты ран от внешней среды, повышенная сорбция раневого экссудата, а также лечебное действие, обеспечивающее как борьбу с инфекциями, так и их профилактику [11–14].

Большое количество статей посвящено созданию полимерных сорбционных материалов медицинского назначения при получении раневых покрытий для лечения травм различной природы [10, 15–17]. Однако большинство из создаваемых и применяемых на практике раневых покрытий отвечает только одному, реже – двум требованиям. Чаще всего основным требованием к раневому материалу считается его сорбционная способность – вальнеросорбция. С другой стороны, высокая технологичность производства, в частности, использование биологических субстратов и клеточных препаратов, повышает стоимость товара, что также ограничивает внедрение в повседневную жизнь новых видов материалов.

Для оптимизации протекания раневого процесса и получения недорогостоящих покрытий необходим комплексный подход. Одним из путей решения является создание так называемых «универсальных» материалов на основе гидрогелевого композита, модифицированных наночастицами. Широко известно, что введение серебра в состав композиций придает материалам антимикробные свойства [15, 16].

Целью представленной работы является исследование абсорбционных характеристик полимерных композитов на основе акрилового сополимера и модифицированного ионами серебра бентонита.

Методики синтеза и исследования полимерных композитов

В ходе работы были исследованы полимерные композиты, полученные на основе акриловых сополимеров, модифицированных Ag-содержащими бентонитовыми глинами с различной дисперсностью. Полимерные акриловые сополимеры были получены на основе частично нейтрализованной акриловой кислоты (степень нейтрализации 0,8) и акриламида в соотношении 3:1 путем радикальной полимеризации в водной среде с конверсией мономеров 99,8%. В качестве иницирующей системы была выбрана окислительно-восстановительная система персульфат аммония – тетраметилэтилендиамин. Метиленибисакриламид с 0,3 мас.% использовался в качестве сшивающего агента. Температура синтеза составляла 45°C, продолжительность полимеризации – 3 ч. Методика синтеза и методы исследования полимерных композиционных материалов представлены в работах [10, 13].

В работе были получены образцы полимерных композитов, содержащих в качестве наполнителя модифицированный ионами серебра бентонит двух фракций – с дисперсностью частиц бентонита менее 0,25 мкм (I фракция) и 0,25–0,5 мкм (II фракция) – и с содержанием серебра 13,5% и 20,72%. Доля модифицированного наполнителя варьировалась в пределах 0–5 мас.%. Наполнитель был предоставлен кафедрой общей химии Белгородского государственного университета.

В ходе работы были исследованы сорбционные характеристики акриловых композиций с процентным содержанием бентонита 1, 2, 3 и 5 мас.% с различной дисперсностью при температуре 20°C в дистиллированной и физиологическом растворах. Расчет степени набухания производили по формуле

$$Q = \frac{m_t - m_0 \times \gamma}{m_0 \times \gamma},$$

где m_t и m_0 – масса стандартного образца гидрогеля после и до набухания материала; γ – влагосодержание исследуемого образца. Влагосодержание определяли отношением

$$\gamma = \frac{m_0 - m'}{m'},$$

где m' – масса высушенного образца полимера при 120°C до постоянного веса, г; m_0 – масса исходного образца материала, г. Константа скорости набухания на начальном этапе процесса (K), характеризующая способность материала к набуханию, определялась путем интегрирования выражения для нахождения скорости набухания полимера и вычислялась по формуле

$$K = \frac{1}{e} \times \ln \left[\frac{Q_m}{Q_m - Q} \right],$$

где Q_m – предельное набухание; Q – количество растворителя, поглощенного одним граммом полимера за время t .

Изучение набухания композитов в дистиллированной воде

В ходе работы нами были исследованы сорбционные характеристики полимерных композиционных материалов медицинского назначения. На рис. 1 представлены кинетические зависимости набухания полимерных акриловых композиций с размером частиц бентонита менее 0,25 мкм, модифицированных серебром, с массовой долей 0,135 в дистиллированной воде при температуре окружающей среды 20°C. Как видно из рис. 1, введение бентонит-содержащего наполнителя приводит к повышению значений равновесной степени набухания в дистиллированной воде в 1,5–2 раза при доле бентонита в полимерной смеси, равной 2 мас.%. Это объясняется тем, что сам бентонит обладает сорбционной способностью в воде и водных растворах, что при небольшом наполнении полимерной матрицы приводит к повышению общей поглощающей способности материала.

Модификация частиц бентонита ионами серебра, как видно из рис. 1, значительно увеличивает водопоглощающую способность полимерной композиции. При этом, чем больше дисперсность, тем выше значение равновесной степени набухания материала медицинского назначения.

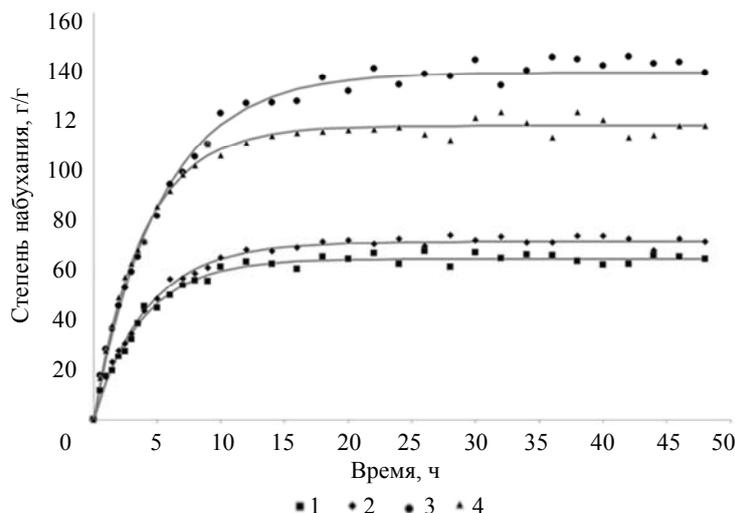


Рис. 1. Кинетические кривые набухания акрилового композита, модифицированного частицами бентонита с массовой долей 2 мас.% при температуре 20°C в дистиллированной воде: 1 – акриловый гидрогель без наполнителя; 2 – гидрогель, модифицированный натриевым бентонитом; 3 – гидрогель, модифицированный Ag-содержащим бентонитом с дисперсностью частиц $d < 0,25$ мм; 4 – гидрогель, модифицированный Ag-содержащим бентонитом с дисперсностью частиц $d = 0,25 - 0,5$ мм

При переходе к изучению процессов набухания в физиологическом растворе видно (рис. 2), что значения равновесных степеней набухания при наличии ионов металлов в водных растворах солей резко уменьшаются (до значений 10–20 г/г), что связано с эффектом полиэлектролитного подавления [2, 4]. Как в физиологическом растворе, так и в дистиллированной воде при 20°C наибольшей сорбционной способностью обладает полимерный нанокompозитный гидрогель, модифицированный Ag-содержащим бентонитом с дисперсностью $d < 0,25$ мм.

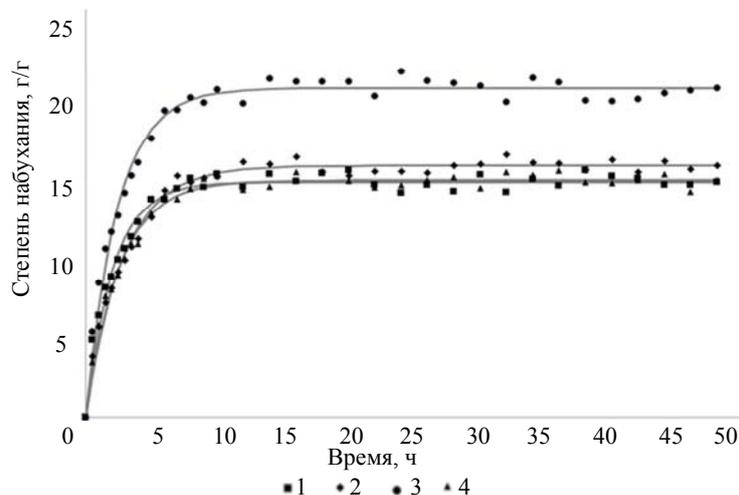


Рис. 2. Кинетические кривые набухания акрилового композита, модифицированного частицами бентонита с массовой долей 2 мас.% при температуре 20°C в физиологическом растворе: 1 – акриловый гидрогель без наполнителя; 2 – гидрогель, модифицированный натриевым бентонитом; 3 – гидрогель, модифицированный Ag-содержащим бентонитом с дисперсностью частиц $d < 0,25$ мм; 4 – гидрогель, модифицированный Ag-содержащим бентонитом с дисперсностью частиц $d = 0,25 - 0,5$ мм

На рис. 3 представлены графики зависимости сорбционных параметров акриловых гидрогелевых композитов с дисперсностью частиц $d < 0,25$ мм и различным процентным содержанием, модифицированных частицами серебра бентонита в полимерной матрице, от времени водопоглощения при температуре 20°C. Как видно из рис. 3, увеличение доли наполнителя до 5 мас.% приводит к повышению значений равновесной степени набухания, что объясняется гидрофильностью и сорбционной способностью самого бентонита. Однако увеличение доли бентонита свыше 5 мас.% приводит к снижению значений набухания полимерного материала, что объясняется ростом среднего эффективного числа физических

узлов сетки при взаимодействии цепей полимера с увеличивающейся суммарно поверхностью частиц модификатора – бентонита, и, как следствие, к ограничению подвижности цепей в ходе формирования поверхностного слоя. Величины значений констант скорости набухания в дистиллированной воде при температуре 20°C для полимерных модифицированных образцов приведены в таблице.

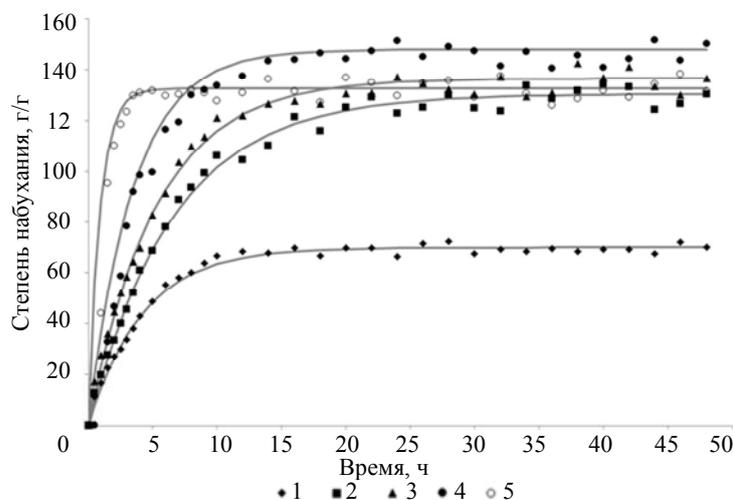


Рис. 3. Кинетические кривые набухания акрилового композита, модифицированного частицами бентонита с дисперсностью $d < 0,25$ мм при температуре 20°C в дистиллированной воде с различной массовой долей (мас.%): 1 – 0; 2 – 1; 3 – 2; 4 – 3; 5 – 5

Дисперсность бентонита, мм	Содержание бентонита, мас.%	Константа скорости, 1/ч
$d < 0,25$	1	0,0047
	2	0,0031
	3	0,0018
	5	0,0033
$d = 0,25-0,5$	1	0,0043
	2	0,0036
	3	0,0034
	5	0,0035

Таблица. Константы скорости набухания исследуемых образцов акрилового композита

Как видно из таблицы, наибольшим значением констант скорости набухания и, как следствие, максимальной скоростью набухания на начальном этапе обладают полимерные композиционные материалы с долей бентонита – 1 мас.%.

Заключение

В ходе работы изучены кинетические характеристики набухания полимерных акриловых композитов, модифицированных Ag-содержащими бентонитами с различной дисперсностью, в физиологическом растворе и дистиллированной воде. Показано, что набухание в физиологическом растворе значительно ниже, чем в дистиллированной воде, и достигает в среднем значения равновесной степени набухания, равной 20 г/г. Показана перспективность использования полимерного композита в качестве матрицы для перспективных медицинских приложений.

Литература

1. Лукьянов Г.Н., Успенская М.В. Количественное описание нелинейной динамики пористой акриловой тонкой пленки // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 2 (78). С. 84–87.
2. Будтова Т.В., Сулейменов И.Е., Френкель С.Я. Сильнонабухающие полимерные гидрогели – некоторые современные проблемы и перспективы // Журнал прикладной химии. 1997. Т. 70. № 4. С. 529–539.
3. Polotsky A.A., Plamper F.A., Borisov O.V. Collapse-to-swelling transitions in pH- and thermoresponsive microgels in aqueous dispersions: the thermodynamic theory // Macromolecules. 2013. V. 46. N 21. P. 8702–8709.
4. Итин А.Л., Лукин С.Б., Успенская М.В., Соловьев В.С. Исследование оптических свойств акрилового гидрогеля для систем индикации загрязнений // Изв. вузов. Приборостроение. 2012. Т. 55. № 7. С. 85–90.
5. Sandu T., Sârбу A., Constantin F., Vulpe S., Iovu H. Acrylic hydrogels-based biocomposites: synthesis and characterization // Journal of Applied Polymer Science. 2013. V. 127. N 5. P. 4061–4071.

6. Yang S., Park K., Rocca J.G. Semi-interpenetrating polymer network superporous hydrogels based on poly(3-sulfopropyl acrylate, potassium salt) and poly(vinyl alcohol): synthesis and characterization // Journal of Bioactive and Compatible Polymers. 2004. V. 19. N 2. P. 81–100.
7. Павлюченко В.Н., Иванчев С.С. Композиционные полимерные гидрогели // Высокомолекулярные соединения. 2009. Т. 51. № 7. С. 1075–1095.
8. Ahmad M.B., Shameli K., Darroudi M., Yunus W.M.Z., Abraham N.A., Hamid A.A., Zargar M. Synthesis and antibacterial activity of silver/montmorillonite nanocomposites // Research Journal of Biological Sciences. 2009. V. 4. N 9. P. 1032–1036.
9. Borisova O.V., Zaremski M.Y., Borisov O.V., Billon L. The well-defined bootstrap effect in the macroinitiator-mediated pseudoliving radical copolymerization of styrene and acrylic acid // Polymer Science – Series B. 2013. V. 55. N 11–12. P. 573–576.
10. Касанов К.Н., Попов В.А., Евсеев Р.А., Андреев В.А., Везенцев А.И., Пономарева Н.Ф., Игнатъева Ю.А., Успенская М.В., Хрипунов А.К. Модифицированный серебром монтмориллонит: получение, антимикробная активность и медицинское применение в биоактивных раневых покрытиях // Научные ведомости Белгородского государственного университета. Серия: Медицина. Фармация. 2013. Т. 23. № 18. С. 188–197.
11. Попов В.А., Игнатъева Ю.А., Успенская М.В., Касанов К.Н. Синтез сорбирующих полимеров медицинского назначения // Известия СПбГТИ(ТУ). 2014. № 23 (49). С. 23–25.
12. Fong J., Wood F. Nanocrystalline silver dressings in wound management: a review // International Journal of Nanomedicine. 2006. V. 1. N 4. P. 441–449.
13. Касанов К.Н., Попов В.А., Евсеев Р.А., Игнатъева Ю.А., Успенская М.В. Биоактивное гидрогелевое раневое покрытие. Заявка № 2013149052, опубл. 06.11.2013.
14. Pillai J.J., Thulasidasan A.K.T., Anto R.J., Chithralekha D.N., Narayanan A., Kumar G.S.V. Folic acid conjugated cross-linked acrylic polymer (FA-CLAP) hydrogel for site specific delivery of hydrophobic drugs to cancer cells // Journal of Nanobiotechnology. 2014. V. 12. N 1. Art. 25.
15. Wright J.B., Lam K., Hansen D., Burrell R.E. Efficacy of topical silver against fungal burn wound pathogens // American Journal of Infection Control. 1999. V. 27. N 4. P. 344–350.
16. Baker C., Pradhan A., Pakstis L., Pochan D.J., Shah S.I. Synthesis and antibacterial properties of silver nanoparticles // Journal of Nanoscience and Nanotechnology. 2005. V. 5. N 2. P. 244–249.
17. Banerjee I., Mishra D., Das T., Maiti T.K. Wound pH-responsive sustained release of therapeutics from a poly(NIPAAm-co-AAc) hydrogel // Journal of Biomaterials Science, Polymer Edition. 2012. V. 23. N 1–4. P. 111–132.

- Игнатъева Юлия Андреевна** – аспирант, Университет ИТМО, 197101, Россия, Санкт-Петербург, Alissia87@mail.ru
- Успенская Майя Валерьевна** – доктор технических наук, профессор, заведующий кафедрой, Университет ИТМО, 197101, Санкт-Петербург, Россия, Mv_uspenskaya@mail.ru
- Борисов Олег Владимирович** – директор исследовательского центра, Университет г. Пау, 64053, г. Пау, Франция, oleg.borisov@univ-pau.fr
- Олежников Роман Олегович** – кандидат технических наук, доцент, Университет ИТМО, 197101, Санкт-Петербург, Россия, r.o.olekhnovich@mail.ru
- Евсеев Роман Анатольевич** – курсант, Военно-медицинская академия имени С.М. Кирова, 194044, Санкт-Петербург, Россия, rdelend@gmail.com
- Касанов Кирилл Николаевич** – младший научный сотрудник, Военно-медицинская академия имени С.М. Кирова, 194044, Санкт-Петербург, Россия, Kasanov_kiria@mail.ru
- Yuliya A. Ignatieva** – postgraduate, ITMO University, 197101, Saint Petersburg, Russia, Alissia87@mail.ru
- Maya V. Uspenskaya** – D.Sc., Professor, Department head, ITMO University, 197101, Saint Petersburg, Russia, Mv_uspenskaya@mail.ru
- Oleg V. Borisov** – Directeur de Recherche CNRS, Université de Pau et des Pays de l'Adour, 64053, Pau, France, oleg.borisov@univ-pau.fr
- Roman O. Olekhnovich** – PhD, Associate professor, ITMO University, 197101, Saint Petersburg, Russia r.o.olekhnovich@mail.ru
- Roman A. Evseev** – student, Military Medical Academy named after S.M. Kirov, 194044, Saint Petersburg, Russia, rdelend@gmail.com
- Kirill N. Kasanov** – PhD, junior research scientist, Military Medical Academy named after S.M. Kirov, 194044, Saint Petersburg, Russia, Kasanov_kiria@mail.ru

Принято к печати 21.07.14
Accepted 21.07.14

УДК 541.183.2.678

КИНЕТИКА КРИСТАЛЛИЗАЦИИ ПОЛИМЕРНЫХ НАНОКОМПОЗИТОВ НА ОСНОВЕ ПОЛИАМИДА 12, МОДИФИЦИРОВАННОГО НАНОЧАСТИЦАМИ Cr₂O₃

Е.С. Шаповал^а, В.В. Зуев^а^а Университет ИТМО, 197101, Санкт-Петербург, Россия, katenka-shapoval@yandex.ru

Аннотация. Методом полимеризации *in situ* получены полимерные композиты на основе матрицы полиамида 12, наполненные 0,1–10 вес.% наноразмерными частицами Cr₂O₃. В результате выполненных исследований был разработан метод синтеза полимерных нанокompозитов на основе матрицы полиамида 12, наполненной наноразмерными магнитными частицами Cr₂O₃, обеспечивающий равномерное встраивание наполнителя в полимерную матрицу без образования агломератов наночастиц. Проведены механические испытания на сжатие образцов и показано, что механические свойства полимерных композитов (модуль Юнга, предельная прочность) понижаются на 20–30% по сравнению с немодифицированным полиамидом 12, синтезированным по выбранной методике. С помощью электронной микроскопии и дифференциальной сканирующей калориметрии определено влияние наполнителя на морфологию и кинетику кристаллизации полимерных нанокompозитов. Получены значения степени кристаллизации, константы скорости кристаллизации для различных интервалов переохлаждения и параметры уравнений Аврами. Показано, что первичная нуклеация происходит по атермическому механизму и наноразмерные частицы не являются зародышами кристаллообразования. Установлено, что наночастицы встраиваются в полимерную матрицу и равномерно распределяются в кристаллитах. Результаты работы могут быть полезны при создании сенсоров электрических и магнитных полей, создании микроразмерных механических устройств, а также использованы при разработке новых материалов для 3D-принтеров.

Ключевые слова: наночастицы Cr₂O₃ оксида хрома (III), полимерные нанокompозиты, механические характеристики нанокompозитов, кинетика кристаллизации, уравнение Аврами.

CRYSTALLIZATION KINETICS OF POLYMERIC NANOCOMPOSITES BASED ON POLYAMIDE 12 MODIFIED BY Cr₂O₃ NANOPARTICLES

E.S. Shapoval^a, V.V. Zuev^a^a ITMO University, 197101, Saint Petersburg, Russia, katenka-shapoval@yandex.ru

Abstract. *In situ* polymerization method is used for obtaining polymeric composites based on polyamide12 matrix (PA 12), filled with Cr₂O₃ nanoparticles. The carried out researches result in synthesis method development for polymeric nanocomposites based on PA 12 matrix filled with nano-sized Cr₂O₃ magnetic particles providing uniform embedding of the filler into polymeric matrix without formation of nanoparticles agglomerates. Mechanical tests on samples compression are carried out. It is shown that mechanical properties of polymeric composites (Young's modulus, durability limit) are decreased for 20-30 % as compared with not modified PA 12 synthesized by means of the chosen method. The influence of the filler on crystallization morphology and kinetics of polymeric nanocomposites is determined by electron microscopy and differential scanning calorimetry. The values of crystallization degree, crystallization rate constant for different supercooling intervals and parameters of Avrami equation are obtained. The initial nucleation is shown to be going on according to non-thermal mechanism, and nanoparticles are not the germs of crystallization. It is stated that nanoparticles are embedded into polymeric matrix and uniformly allocated in crystallites. Research results can find their application at creation of electric and magnetic fields, micro-sized mechanical devices, and at development of new materials for 3D printers.

Keywords: nanoparticles of chromium (III) oxide Cr₂O₃, polymeric nanocomposites, mechanical properties of nanocomposites, crystallization kinetics, Avrami equation.

Введение

Создание инженерных конструкционных материалов, которые могут быть использованы как датчики в различных сенсорных устройствах, является одним из основных направлений современной науки о материалах [1, 2]. Перспективным в этом направлении является получение полимерных композитных материалов, которые сочетают в себе гибкость, легкую перерабатываемость и прозрачность, восприимчивы к электромагнитным воздействиям и обладают электропроводностью. Восприимчивость к действию магнитных полей может быть придана полимерам путем введения магнитных частиц [3]. При правильном методе синтеза композитов наноразмерность обеспечивает однородное распределение наполнителя в объеме полимера и достижение желаемых свойств при относительно малом (доли весовых процентов) количестве модификатора [4].

В качестве объекта исследований для получения нанокompозитов авторами был выбран наноразмерный оксид хрома (III). Этот материал обладает набором привлекательных характеристик – легко диспергируется в полимерах, являясь зеленым пигментом, придает материалу устойчивость к истиранию, является связующим для композитов с повышенной термостойкостью, обладает свойствами полупроводника и является антиферромагнетиком с температурой Нееля немногим выше 27°C [5]. Полиамид 12 (ПА-12) является конструкционным материалом с высокими механическими характеристиками, который легко окрашивается различными пигментами, поэтому он и был выбран в качестве полимерной матрицы [6, 7]. В качестве зеленого красителя для ПА-12 используется и оксид хрома (III) [8]. При этом окрашивание полимеров производится смешением готового ПА-12 с пигментом в расплаве. В то же время синтез композитов с целью создания новых материалов на основе ПА-12 до сих пор не проводился, хотя полимер-

ные композиты, наполненные наноразмерным оксидом хрома (III), известны (на основе полиамида 6, полистирола и полиметилметакрилата [9]). Известны лишь работы, посвященные исследованию кинетики полимеризации додекалактама (мономер ПА-12) в присутствии оксида хрома (III) [10].

Целью настоящей работы является синтез и изучение механических свойств полимерных нанокompозитов на основе ПА-12, модифицированного наноразмерными частицами Cr_2O_3 . Так как степень кристалличности полимера влияет на его оптические, механические и диэлектрические свойства [11], то в работе изучаются степень кристалличности композитов при различном уровне наполнения и кинетика кристаллизации композитов в различных условиях.

Экспериментальная часть

Наноразмерный Cr_2O_3 с размером частиц 1–5 нм был предоставлен ООО «ИнформТехСистемы» (ООО ИТС) (Санкт-Петербург, Россия). Наночастицы получались растворным методом с последующим термическим отжигом. Полимерные нанокompозиты получали методом полимеризации *in situ* [12] после смешивания наполнителя и мономера с использованием в качестве катализатора фосфорной кислоты и нагреванием в течение 12 ч при 300°C. Модуль Юнга и предел прочности определяли на разрывной машине UTS 10 (UTStestsysteme, Германия) при сжатии для образцов цилиндрической формы диаметром 10 мм и высотой 20 мм в интервале нагрузок 0,1 Н до 20 кН при скорости сжатия 1 мм/мин. Все измерения проводились для серий не менее чем 5 образцов, полученных при разных синтезах нанокompозитов.

Калориметрические исследования проводились на дифференциальном сканирующем калориметре Netzsch DSC 204 F1 Phoenix. Температуры плавления определялись при скорости нагрева 10 °C/мин в токе инертного газа. Изотермическая кинетика кристаллизации изучалась по следующей методике: образцы нагревали до 180 °C (со скоростью 10 °C/мин) а затем быстро охлаждали (со скоростью 40 °C/мин) до температуры кристаллизации от 145 до 165 °C.

Электронные фотографии сколов композитов после напыления платиной получены на электронном сканирующем микроскопе Supra 55VP с увеличением 2000–150000 \times .

Инфракрасные спектры записаны на приборе Bruker Tensor 27 для порошков методом нарушенного полного внутреннего отражения.

Результаты и обсуждение

Важнейшей задачей при синтезе полимерных нанокompозитов является достижение равномерного распределения наноразмерного наполнителя в полимерной матрице, так как в противном случае все преимущества использования наноматериалов пропадают [13]. Для контроля распределения наполнителя нами была использована электронная спектроскопия (рис. 1).

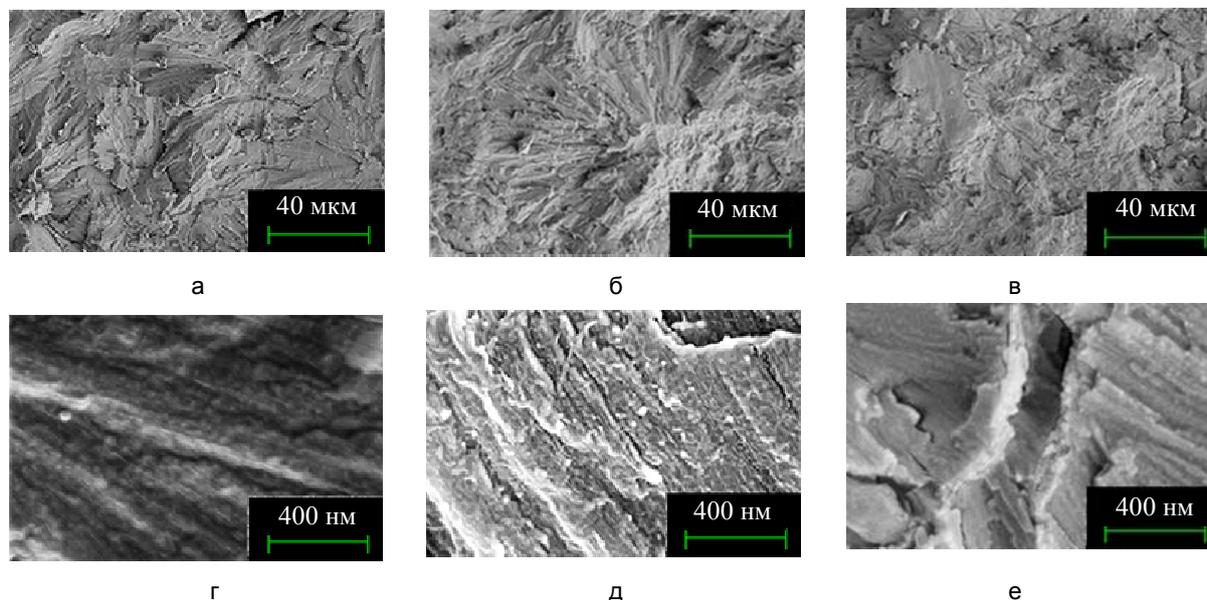


Рис. 1. Электронные микрофотографии сколов нанокompозитов: ненаполненный ПА-12 (а); ПА-12, содержащий: 1% Cr_2O_3 (б); 10% Cr_2O_3 (в); 1% Cr_2O_3 (г); 5% Cr_2O_3 (д); 10% Cr_2O_3 (е); а–в – увеличение 2000 \times ; г–е – увеличение 15000 \times

Как видно из рис. 1, г–е, даже при концентрации наноразмерного Cr_2O_3 10 мас.% не наблюдается какой-либо агрегации наполнителя, при этом сами частицы практически не удается обнаружить. Причину этого следует искать как в преимуществах разработанного нами метода синтеза полимерных нанокompозитов [12], так и в присущих Cr_2O_3 амфотерных свойствах и склонности образовывать устойчивые ком-

плексы с различными азотсодержащими лигандами [11], что способствует его совместимости с матрицей ПА-12. Однако может возникнуть опасение, что в процессе синтеза наночастицы растворяются в полимерной матрице либо, активно взаимодействуя с ней, вызывают ее деградацию. Чтобы исключить эту возможность, нами были записаны инфракрасные (ИК) спектры чистого ПА-12 и полимерных нанокомпозитов различного состава (рис. 2).

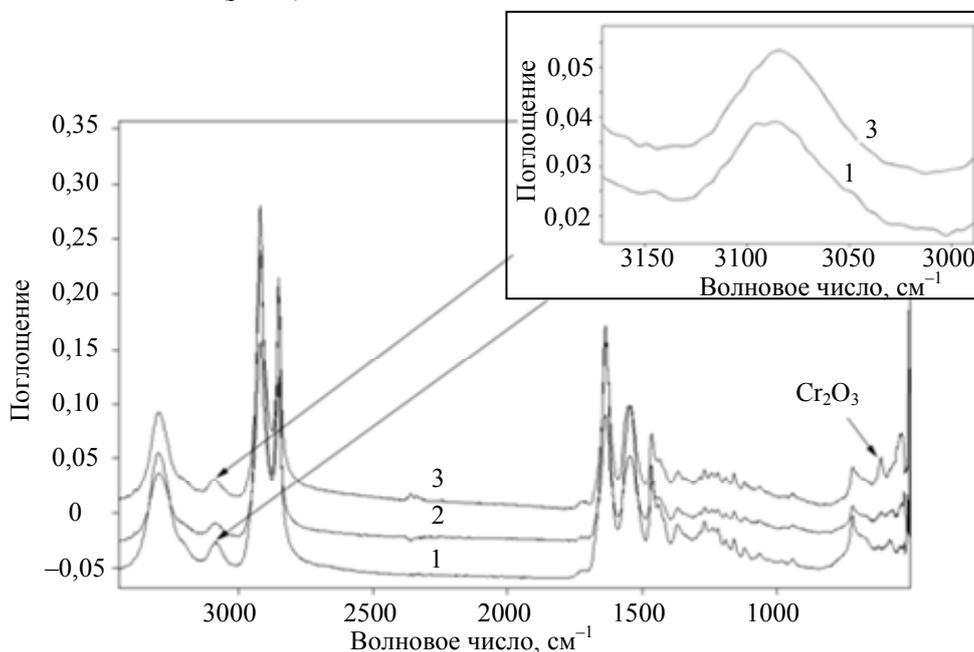


Рис. 2. Инфракрасные спектры ненаполненного ПА-12 (1); ПА-12, содержащий: 1% Cr_2O_3 (2); 10% Cr_2O_3 (3)

Сравнение ИК спектров показывает, что не наблюдается никаких признаков деградации полимерной основы, а в спектре композита с 10 мас.% Cr_2O_3 отчетливо проявляется полоса при 613 см^{-1} , которая отвечает отдельной фазе наноразмерного оксида хрома (III) [14]. Таким образом, разработанный нами метод синтеза полимерных нанокомпозитов позволяет получить системы с распределением наполнителя на уровне отдельных частиц. Это позволяет перейти к анализу влияния наполнителя на механические свойства полимерных композитов. На рис. 3 и в таблице представлены результаты механических испытаний.

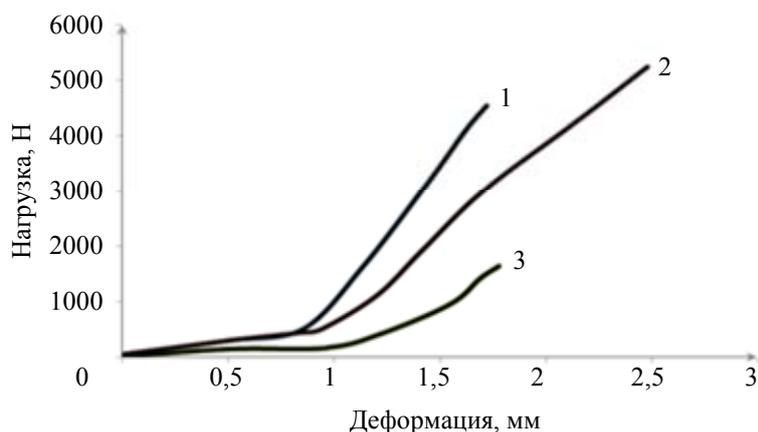


Рис. 3. Кривые сжатия для ПА-12, наполненного Cr_2O_3 , с концентрацией: 0,1% (1), 1% (2) и 10% (3)

Как видно из рис. 3, нанокомпозиты обладают худшими механическими свойствами по сравнению с ПА-12, полученным в аналогичных условиях. Чтобы выяснить причину падения механических характеристик, нами были изучены особенности формирования кристаллической структуры нанокомпозитов, которая для частично-кристаллических полимеров оказывает существенное влияние на механические характеристики. Анализ электронных микрофотографий (рис. 1, а–в) показывает, что как ПА-12, так и нанокомпозиты являются высококристаллическими полимерами сходной морфологии. В то же время для нанокомпозита с 10 мас.% Cr_2O_3 отчетливо проявляется новая фаза (рис. 1, в), присутствие которой не наблюдалось в ПА-12. Известно, что ПА-12 может существовать в виде двух кристаллических модификаций – α и γ [15]. α -модификация ПА-12 является более высокоплавкой и отличается более высокими

механическими свойствами, тогда как γ -модификация придает ПА-12 большую пластичность. Признаком α -модификации ПА-12 является пик плавления, снятый при помощи дифференциального сканирующего калориметра (ДСК), на термограмме при 174°C (рис. 4) и наличие полосы амидной связи в ИК спектре при 3060 см^{-1} (рис. 2) [15].

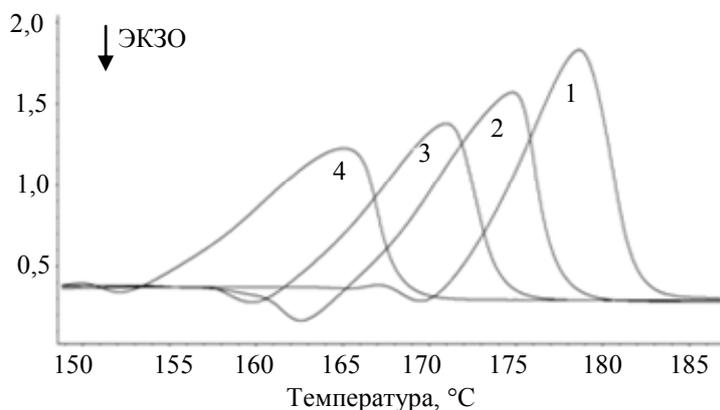


Рис. 4. ДСК термограмма плавления для ненаполненного ПА-12 (1); ПА-12, содержащего: 1% Cr_2O_3 (2); 5% Cr_2O_3 (3); 10% Cr_2O_3 (4). Стрелкой показано направление экзотермического процесса

Введение наноразмерного Cr_2O_3 в матрицу ПА-12 приводит к снижению температуры плавления композитов до 164°C (температура плавления γ -модификации) (рис. 4) и смещению составной полосы амидной связи в ИК спектре к 3080 см^{-1} (рис. 2). Таким образом, введение наноразмерного Cr_2O_3 в ПА-12 вызывает его кристаллизацию в γ -форме, что и приводит к падению механических свойств нанокompозитов. Подобный эффект у нанонаполнителей нередок. Например, наноразмерный оксид цинка, металлическое серебро, наноглины вызывают селективную кристаллизацию полиамида 6 в γ -форме, тогда как фуллерен C_{60} , углеродные нанотрубки, графит – в α -модификации [15].

Концентрация наполнителя	Модуль Юнга, ГПа	Прочность, МПа	Степень кристалличности, %	Показатель n в уравнении Аврами при температуре 158°C	Константа скорости кристаллизации при температуре 158°C, мин^{-1}
Ненаполненный ПА-12	0,97	55,6	63,0	2,8	1,87
0,1% Cr_2O_3	0,65	34,8	63,0	2,8	1,77
1% Cr_2O_3	0,48	45,4	53,0	1,9	2,89
5% Cr_2O_3	0,47	49,4	51,2	1,9	3,30
10% Cr_2O_3	0,36	13,0	54,3	2,9	2,38

Таблица. Свойства нанокompозитов

Нами была определена степень кристалличности синтезированных нанокompозитов:

$$\chi = \frac{\Delta H_{\text{пл}}}{\Delta H_{100}(1 - W_{\text{Cr}_2\text{O}_3})} \times 100\%,$$

где χ – степень кристалличности; $W_{\text{Cr}_2\text{O}_3}$ – массовая доля наполнителя; $\Delta H_{\text{пл}}$ – теплота плавления композита; ΔH_{100} – теплота плавления ПА-12 при 100% кристалличности, равная 95 Дж/г [16]. Как видно из таблицы, степень кристалличности нанокompозитов ниже, чем у чистого ПА-12, синтезированного в тех же условиях. Чтобы понять причину этого, нами была изучена кинетика кристаллизации нанокompозитов в изотермических условиях. Как скорость кристаллизации рассматривалась величина, обратная времени достижения 10% степени кристалличности ($v = 1/\tau_{0,1}$). При этом скорость кристаллизации подчиняется кинетике первого порядка, а значит, константа скорости может быть определена как $k = \ln 2/\tau_{0,5}$ ($\tau_{0,5}$ – время достижения 50% степени кристалличности) [17]. Как видно из рис. 5, скорость кристаллизации практически не зависит ни от интервала переохлаждения, ни от концентрации наполнителя. Это косвенно говорит об атермической природе зародышеобразного при кристаллизации.

Для подтверждения этого нами был проведен анализ кристаллизации в соответствии с уравнением Аврами (зависимость степени кристалличности от времени) [17, 18]:

$$\chi = 1 - \exp[-Kt^n],$$

где K – константа скорости кристаллизации, t – время при заданной температуре, показатель n связан с механизмом зародышеобразования и роста кристаллов.

На рис. 6 показаны зависимости степени кристалличности от времени ($\lg(-\ln(1-\chi))$ от $\lg t$), которые позволяют определить значения n для исследованных нами систем. Эти зависимости линейны, следовательно, уравнение Аврами адекватно описывает кристаллизацию исследуемых систем при разных температурах и составах композитов. Полученные значения показателя n примерно одинаковы для ПА-12 и нанокompозитов и равны 2–3. Это значение соответствует механизму атермического зародышеобразования при кристаллизации в форме усеченной сферы [18], что подтверждается и формой кристаллитов (рис. 1).

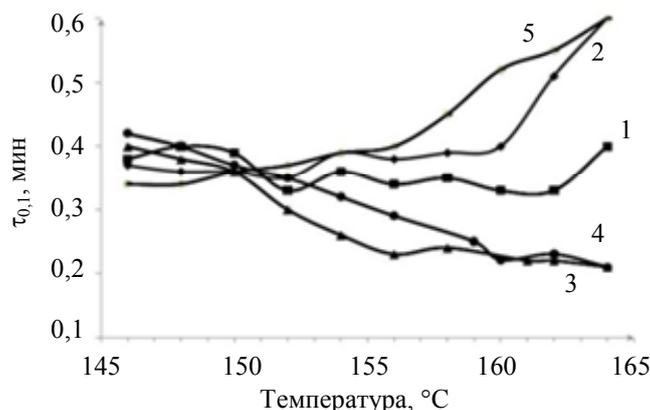


Рис. 5. Время 10% кристаллизации для ненаполненного ПА-12 (1); ПА-12, содержащего: 0,1 мас.% Cr₂O₃ (2); 1 мас.% Cr₂O₃ (3); 5 мас.% Cr₂O₃ (4); 10 мас.% Cr₂O₃ (5)

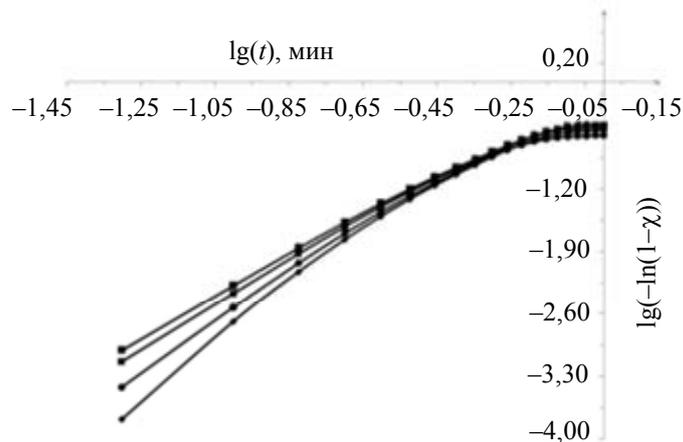


Рис. 6. Зависимость Аврами для степени кристалличности для нанокompозита 5 мас.% Cr₂O₃ при температурах: 159°C (1); 160°C (2); 162°C (3); 164°C (4)

Обсуждение результатов

Полученные результаты показывают активную роль наноразмерного оксида хрома (III) в формировании кристаллической структуры ПА-12, что полностью отсутствует в случае макроразмерного аналога [9]. Таким образом, наблюдается выраженный наноэффект наполнителя. В то же время наночастицы встраиваются в полимерную матрицу и никак не влияют на кинетику кристаллизации и механизм зародышеобразования, т.е. наночастицы не являются центрами роста кристаллов, а равномерно распределяются в кристаллитах. Это особенно важно при высоких степенях наполнения нанокompозитов, так как позволяет, сохраняя свойства полимеров (пластичность, стойкость к растрескиванию и ударопрочность), придавать композитам новые качества – восприимчивость к действию магнитных и электрических полей – и делает подобные нанокompозиты перспективными материалами для использования в различных сенсорных устройствах. Сохранение основных параметров кинетики кристаллизации позволяет использовать полученные композиты для замены ПА-12 в традиционных областях его использования без изменения существующих технологий, что особенно важно при их применении в 3D-принтерах, где как материал для изготовления полимерных деталей сложной геометрии широко используется ПА-12.

Заключение

В результате выполненных исследований был разработан метод синтеза полимерных нанокомпози- тов на основе матрицы полиамида 12, наполненной наноразмерными магнитными частицами Cr_2O_3 , обеспечивающий равномерное встраивание наполнителя в полимерную матрицу без образования агломе- ратов наночастиц. Изучены механические и термические свойства полученных нанокомпози- тов, а также кинетика кристаллизации в изотермическом режиме. Механические свойства нанокомпози- тов снижаются, что связано с изменением типа кристаллической формы полимерной матрицы. Введение наноразмер- ного Cr_2O_3 в полиамид 12 вызывает его кристаллизацию в γ -форме, что и приводит к падению механиче- ских свойств нанокомпози- тов. Этот эффект отличен от наблюдаемого при введении макроразмерных на- полнителей, где при сильном связывании наполнителя и матрицы происходит перенос прочностных свойств наполнителя на матрицу и наблюдается эффект усиления. Получены кинетические параметры кристаллизации полимерных нанокомпози- тов – скорость кристаллизации и константа скорости кристал- лизации. Показано, что введение наночастиц размерностью 1–5 нм практически не влияет на величину константы скорости. Как и для чистого полиамида 12, значения константы скорости кристаллизации со- ставляют 2–3 мин^{-1} . Полученный результат показывает, что при изготовлении литых изделий из данных нанокомпози- тов нет необходимости менять технологию, используемую для полиамида 12. Кинетика кри- сталлизации была успешно описана в рамках формализма Аврами. Показано, что первичная нуклиация происходит по атермическому механизму и наноразмерные частицы не являются зародышами кристалло- образования. В результате исследования выяснено, что наночастицы встраиваются в полимерную матри- цу и равномерно распределяются в кристаллитах. Результаты работы могут быть полезны при создании сенсоров электрических и магнитных полей, создании микроразмерных механических устройств, и ис- пользованы при разработке новых материалов для 3D-принтеров.

Литература

1. Treacy M.M.J., Ebbesen T.W., Gibson J.M. Exceptionally high Young's modulus observed for individual carbon nanotubes // *Nature*. 1996. V. 381. N 6584. P. 678–680.
2. Spitalsky Z., Tasis D., Papagelis K., Galiotis C. Carbon nanotube–polymer composites: chemistry, processing, mechanical and electrical properties // *Progress in Polymer Science*. 2010. V. 35. N 3. P. 357–401.
3. Yoo H., Moon S.-K., Hwang T., Kim Y.S., Kim J.-H., Choi S.-W., Kim J.H. Multifunctional magnetic nano- particles modified with polyethylenimine and folic acid for biomedical theranostics // *Langmuir*. 2013. V. 29. N 20. P. 5962–5967.
4. Zuev V.V. The mechanisms and mechanics of the toughening of epoxy polymers modified with fullerene C_{60} // *Polymer Engineering and Science*. 2012. V. 52. N 12. P. 2518–2522.
5. Makhlof S.A., Bakr Z.H., Al-Attar H., Moustafa M.S. Structural, morphological and electrical properties of Cr_2O_3 nanoparticles // *Materials Science and Engineering B: Solid-State Materials for Advanced Technology*. 2013. V. 178. N 6. P. 337–343.
6. Athreya S.R., Kalaitzidou K., Das S. Mechanical and microstructural properties of Nylon-12/carbon black composites: selective laser sintering versus melt compounding and injection molding // *Composites Science and Technology*. 2011. V. 71. N 4. P. 506–510.
7. Razzaq M.Y., Anhalt M., Frommann L., Weidenfeller B. Thermal, electrical and magnetic studies of magnet- ite filled polyurethane shape memory polymers // *Materials Science and Engineering A*. 2007. V. 444. N 1. P. 227–235.
8. Vuxbaum G., Pfaff G. *Industrial Inorganic Pigments*. 3rd ed. Wiley-VCH, 2005. 315 p.
9. Park D.-H., Hwang S.-J., Oh J.-M., Yang J.-H., Choy J.-H. Polymer-inorganic supramolecular nanohybrids for red, white, green, and blue applications // *Progress in Polymer Science*. 2013. V. 38. N 10–11. P. 1442–1486.
10. Горбунова Е.В., Деев Е.С., Куличихин С.Г., Рябов Е.А. Кинетические особенности полимеризации капро- и додекалактама в присутствии окислов титана, хрома, цинка и циркония // *Пластические мас- сы*. 1981. № 10. С. 12–19.
11. Зуев В.В., Иванова Ю.Г. Полимерные нанокомпози- ты на основе полиамида 6, модифицированного фуллероидными наполнителями // *Высокомолекулярные соединения*. 2011. Т. 53. № 5. С. 733–738.
12. Шаповал Е.С., Зуев В.В. Полимерные нанокомпози- ты на основе полиамида 12, модифицированного наночастицами // *Научно-технический вестник информационных технологий, механики и оптики*. 2013. № 4 (86). С. 92–95.
13. Zhang J., Xu S., Kumacheva E. Polymer microgels: reactors for semiconductor, metal, and magnetic nano- particles // *Journal of American Chemical Society*. 2004. V. 126. N 25. P. 7908–7914.
14. Помогайло А.Д., Розенберг А.С., Уфлянд И.Е. *Наночастицы металлов в полимерах*. М.: Химия, 2000. 672 с.

15. Naffakh M., Marco C., Gomez M.A., Jimenez I. Novel melt-processable nylon 6/ inorganic fullerene-like WS₂ nanocomposites: complex isothermal crystallization kinetics and melting behaviour // *Materials Chemistry and Physics*. 2011. V. 128. N 1–2. P. 265–273.
16. Plummer C.J.G., Zanetto J.-E., Bourban P.-E., Manson J.-A.E. The crystallization kinetics of polyamide-12 // *Colloid and Polymer Science*. 2001. V. 279. N 4. P. 312–322.
17. Avrami M. Kinetics of phase change. I General theory // *The Journal of Chemical Physics*. 2004. V. 7. N 12. P. 1103–1112.
18. Вундерлих Б. Физика макромолекул. Зарождение, рост и отжиг кристаллов. Т. 2. М.: Мир, 1979. 575 с.

Шаповал Екатерина Сергеевна – студент, Университет ИТМО, 197101, Санкт-Петербург, Россия, katenka-shapoval@yandex.ru

Зуев Вячеслав Викторович – доктор химических наук, профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия, zuev@hq.macro.ru

Ekaterina S. Shapoval – student, ITMO University, 197101, Saint Petersburg, Russia, katenka-shapoval@yandex.ru

Vjacheslav V. Zuev – D.Sc., Professor, ITMO University, 197101, Saint Petersburg, Russia, zuev@hq.macro.ru

Принято к печати 15.05.14

Accepted 15.05.14

УДК 535.6

НЕОРГАНИЧЕСКИЕ ЛЮМИНОФОРЫ В СТЕКЛЕ НА ОСНОВЕ СВИНЦОВО-СИЛИКАТНЫХ СТЕКОЛ

В.А. Асеев^а, А.Ю. Биби^а, Е.В. Колобкова^а, Я.А. Некрасова^а, Н.В. Никоноров^а, А.Е. Романов^б,
Ю.В. Тузова^а, М.А. Швалева^а

^а Университет ИТМО, 197101, Санкт-Петербург, Россия, Anastasiya.bibik@list.ru

^б Тартуский университет, 50090, Тарту, Эстония

Аннотация. Разработан и синтезирован люминесцирующий композит типа «люминофор в стекле» на основе высокопреломляющей свинцово-силикатной матрицы и мелкодисперсного порошка кристалла алюмо-иттриевого граната, активированного церием. В качестве стеклообразной матрицы была выбрана свинцово-силикатная система $(40\text{SiO}_2-20\text{PbO}-(40-x)\text{PbF}_2-x\text{AlF}_3, x=25,15,10,5,0)$. Для получения композита «люминофор в стекле» исходное стекло измельчалось до порошкообразного состояния (фритта) с размером частицы порядка 50 мкм и затем смешивалось с порошком коммерческого люминофора АИГ:Ce³⁺ в соотношении 30:70. Получившаяся смесь запрессовывалась в форму под давлением и спекалась на подложке из кварцевого стекла в течение 30 мин при температуре 823 К. Таким образом, была получена плоскопараллельная пластинка композита «люминофор в стекле» диаметром 10 мм. Произведена оптимизация состава стекла с целью снижения потерь света при рассеянии на границе раздела стекло–люминофор путем подбора показателя преломления. Проведены рентгенофазовые и спектрально-люминесцентные исследования полученного композита. На основании результатов данных исследований показано, что при спекании не происходит деградации порошка YAG:Ce. Исследована зависимость интенсивности люминесценции от температуры в диапазоне от комнатной температуры до 473 К. Показано, что использование люминофора в стекле снижает температурное тушение люминесценции по сравнению с силиконом. С использованием композита «люминофор в стекле» на основе свинцово-силикатных стекол с низкой температурой стеклования создан макет белого светодиода. Получившийся светодиод излучает белый свет с цветовой температурой 4370 К, световая отдача составляет 58 лм/Вт. Разработанный композит может быть использован для производства мощных светодиодов белого света.

Ключевые слова: люминофор в стекле, свинцово-силикатное стекло, алюмо-иттриевый гранат с церием, спектры люминесценции, световая отдача, белые светодиоды.

Благодарности. Работа выполнена при государственной финансовой поддержке, выделяемой на реализацию программы развития международных научных лабораторий Университета ИТМО в соответствии с пунктом «Правил распределения и предоставления субсидий на государственную поддержку ведущих университетов Российской Федерации в целях повышения их конкурентоспособности среди ведущих мировых научно-образовательных центров».

INORGANIC PHOSPHORS IN GLASS BASED ON LEAD SILICATE GLASSES

V.A. Aseev^a, A.Yu. Bibik^a, E.V. Kolobkova^a, Ya. A. Nekrasova^a, N.V. Nikonorov^a, A.E. Romanov^b,
Ju.V. Tuzova^a, M.A. Shvaleva^a

^a ITMO University, 197101, Saint Petersburg, Russia, Anastasiya.bibik@list.ru

^b University of Tartu, 50090, Tartu, Estonia

Abstract. We created and synthesized luminescent composite of the "phosphor in glass" type, based on the lead-silicate matrix and fine-dispersed powder of cerium-activated yttrium-aluminum garnet crystal. Lead-silicate system $(40\text{SiO}_2-20\text{PbO}-(40-x)\text{PbF}_2-x\text{AlF}_3, x = 0-25)$ was chosen as the glassy matrix. Initial glass was reduced to powder (frit) for "phosphor in glass" composite with a particle size about 50 μm . Glass frit and powder of commercial YAG:Ce³⁺ phosphor were mixed in a ratio of 30 to 70 (wt %). Then this composite was pressed in a tablet and sintered on a quartz substrate at 823 K for 30 minutes. Thus, the plane parallel sheet for composite of the "phosphor in glass" was obtained with a diameter equal to 10 mm. For the purpose to reduce the loss of light in the presence of dispersion at a glass-phosphor boundary, optimization of glass mixture was done by adjusting the refractive index. X-ray phase and spectral-luminescent analysis of the derived composite were done. The results of these studies showed that there was no degradation of YAG:Ce powder during sintering. Dependence of luminescence intensity from temperature in the range from room temperature to 473 K was studied. It was shown, that with the phosphor in glass usage thermal quenching of luminescence was reduced in comparison with the silicone. The model of white LED was created with the "phosphor in glass" composite based on lead-silicate glasses with low temperature of vitrifying. The derived LED emits white light with a color temperature of 4370 K, and the luminous efficiency is equal to 58 lm/W. The developed luminescent composite based on the lead-silicate matrix can be used for the production of high-power white light LED.

Keywords: phosphor in glass, lead silicate glass, yttrium-aluminum garnet doped with cerium, luminescence spectra, luminous efficiency, white LEDs.

Acknowledgements. The work was partially financially supported by the state program for development of international scientific laboratories at ITMO University according to "Rules of grants distribution for the Russian Federation Government financial support of the leading universities aimed at their competitiveness growth among the chief world scientific educational centers".

Введение

К настоящему времени одним из традиционных методов получения светодиодов белого свечения является использование комбинации синего светодиода на основе кристалла InGaN с зелено-желтым люминофором. В зависимости от соотношения зеленых и желтых полос люминесценции, возбуждаемых синим излучением кристалла, возможно получение белого света различной цветовой температуры. Сего-

дня коммерческие люминофоры представляют собой мелкодисперсные порошки кристалла алюмоиттриевого граната, активированного трехвалентным церием (АИГ:Ce³⁺). Этот порошок получают методом твердотельной реакции между оксидами иттрия, алюминия и церия. Для фиксации люминесцентного порошка на поверхности кристалла светодиода его смешивают с жидким полимером – оптическим силиконом, который играет роль буферного связующего материала, и такой композит «люминофор в полимере» наносят на поверхность светоизлучающего кристалла нитрида индия. После затвердевания композита на него наносится линза, формирующая диаграмму направленности белого излучения. Несмотря на простоту технологии получения композита «люминофор в полимере» и нанесения его на поверхность кристалла, а также сравнительно невысокую стоимость, такой подход имеет серьезные недостатки, которые, прежде всего, проявляются в процессе эксплуатации белого светодиода. Так, например, в процессе эксплуатации композит «люминофор в полимере» подвергается длительному воздействию синего излучения кристалла нитрида индия, а также нагревается до значительных температур (более 423 К). В результате этого силикон изменяет свои свойства: появляются центры окраски, дефекты и деформация, которые снижают коэффициент полезного действия светодиода. Еще одним из недостатков является высокий уровень рассеяния композита, которое возникает на границе раздела фаз полимер–кристалл, поскольку показатель преломления кристалла АИГ:Ce³⁺ составляет 1,83, а полимера 1,54 [1]. В научно-информационных источниках были опубликованы данные, по которым одним из самых лучших показателей преломления люминофоров для использования в светодиодной промышленности является значение показателя, большее 1,4 [2–6]. Значительная разница между показателями преломления приводит к высокому рассеянию, что ухудшает излучательные характеристики светодиода.

Для решения указанной проблемы в работах [7–9] были изучены различные связующие неорганические материалы, которые имеют высокую термическую устойчивость по сравнению с полимерными материалами. Для дальнейшего повышения эффективности белых светодиодов необходим неорганический материал, например, стекло, у которого термическая и оптическая прочности могут в несколько раз превышать аналогичные характеристики для органических материалов. Более того, изменение состава стекла позволяет изменять показатель преломления в широких пределах, приближая его к показателю преломления кристалла АИГ:Ce³⁺, и, таким образом, снижать светорассеяние в композите. В научно-информационных источниках в качестве матриц для люминофоров упоминаются следующие виды стекол: P₂O₅-MgO-CaO-SrO-BaO-Dy-Tm [10], SiO₂-Na₂O₃-Al₂O₃-CaO [11], SiO₂-B₂O₃-RO (R = Ba, Zn) [10], SnO-ZnO-P₂O₅ [12], SiO₂-Al₂O₃-B₂O₃-ZnO-Li₂O-BaO [13]. Еще один пример использования стекла продемонстрирован в работе [14], в которой было показано, что в качестве связующего буферного слоя с порошком АИГ:Ce³⁺ может служить силикатное стекло. В этом случае композит «люминофор в стекле» (или, в англоязычной версии, Phosphor-in-Glass, PiG) получается при смешивании двух мелкодисперсных порошков стекла и люминофора АИГ:Ce³⁺ с последующим их спеканием. В основном в качестве исходных стекол для фритты используют свинцово-силикатные стекла. Так, в работе [15] показана реализация PiG на стеклах и проведены испытания светодиодов на их основе. В работе [16] показано, что такая замена приводит к увеличению выхода люминесценции по сравнению с традиционным силиконом на 7%. К сожалению, на сегодняшний день публикации в этом направлении практически нет.

Целью настоящей работы была разработка композита «люминофор в стекле» на основе высокопреломляющих свинцово-силикатных стекол (40SiO₂-20PbO-(40-x)PbF₂-xAlF₃) с низкой температурой стеклования для мощных белых светодиодных матриц.

Экспериментальная часть

В качестве стеклообразной матрицы была выбрана свинцово-силикатная система (40SiO₂-20PbO-(40-x)PbF₂-xAlF₃, x = 25, 15, 10, 5, 0). Варьирование концентраций фторидов свинца и алюминия позволяет изменять показатель преломления исходного стекла в широких пределах, в результате можно добиться полного совпадения показателей преломления кристаллического люминофора АИГ:Ce³⁺ и матрицы стекла для снижения потерь света на границе раздела люминофор–стекло. Также введение фторидов свинца снижает температуру стеклования, что позволяет снизить температуру спекания образцов для предотвращения взаимодействия компонентов и сохранения кристаллической фазы люминофора. Методом дифференциального термического анализа был определен диапазон температур стеклования для стекол такой системы – от 593 К до 678 К. Температура синтеза стекол, активированных фторидами редкоземельных элементов, составляла 1173 К. Время синтеза – 30 мин. Варка стекол в открытых корундовых тиглях в атмосфере воздуха обеспечивала хорошие показатели качества стекла (в первую очередь прозрачность) и соответствовала окислительным условиям. Очевидно, что химический состав по синтезу не соответствовал конечному составу стекла, так как в процессе варки осуществлялось значительное улетучивание фтора преимущественно в форме тетрафторида кремния SiF₄. Для свинцовофторосиликатных стекол было проанализировано влияние содержания фторидов на изменение некоторых физических свойств. Увеличение общего содержания фторидов в стекле производилось путем замещения оксида свинца PbO на фторид свинца PbF₂. Непосредственно после выработки стекла подвергались отжигу в муфельной печи.

Для создания экспериментальных образцов было выбрано стекло синтезированной серии, обладающее наиболее близким составом к требуемому. Стекло содержит 15% AlF_3 , 45% $\text{PbF}_2 + \text{PbO}$ и обладает температурой стеклования 663 К.

Для получения композита «люминофор в стекле» исходное стекло измельчалось до порошкообразного состояния (фритта) с размером частицы порядка 50 мкм и затем смешивалось с порошком коммерческого люминофора АИГ: Ce^{3+} в соотношении 30:70. Такое соотношение было выбрано исходя из состава стандартных смесей люминофора с оптическим силиконом для белых светодиодов. Получившаяся смесь запрессовывалась в форму под давлением и спекалась на подложке из кварцевого стекла в течение 30 мин при температуре 823 К. Таким образом, была получена плоскопараллельная пластинка композита «люминофор в стекле» диаметром 10 мм.

Рентгенофазовый анализ проводился на рентгеновском дифрактометре «Rigaku UltimaIV» – дифрактометр общего назначения типа ДРОН с использованием CuK_α -излучения. Напряжение на трубке составляло 40 кВ, ток трубки 40 мА, общая выходная мощность 1,6 кВт. Применялась фокусировка лучей по методу Брэгга–Брентано. Измерения производились со скоростью 1 °/мин. Спектры возбуждения и люминесценции записывались на спектрофлуориметре Perkin-ElmerLS-50b в диапазоне 400–550 нм и 550–700 нм с шагом 0,5 нм и скоростью сканирования 60 нм/мин.

Для исследования температурной зависимости люминесценции использовалась печь замкнутого объема, в которую помещался образец. Температура внутри печи задавалась цифровым контроллером и отслеживалась с помощью термодатчика. Градиент температуры по объему печи не превышал 1 К. Люминесценция возбуждалась лазером ($\lambda = 457$ нм) и регистрировалась при помощи монохроматора (Acton Research 300) и фотоэлектронного умножителя (Hamamatsu R928).

Запись интегральных спектров люминесценции люминофора и излучения диода проводилась при возбуждении синим диодом (модель X10 фирмы Оптоган) с максимумом длины волны 456 нм. Для этого на синий светодиод помещались пластинки люминофора, и в интегрирующей сфере диаметром 50 мм производилась регистрация суммарного спектра диода и желтого люминофора при помощи программно-аппаратного комплекса, состоящего из спектрометра и программного обеспечения (ПО) SpectraWiz фирмы StellarNet.

Световая отдача светодиодного модуля, состоящего из синтезируемого люминофора и синего светодиода X10, измерялась на оптической сфере Gamma Scientific с использованием программы SpectralSuite 3.0. Для этого полученный светодиодный модуль помещали в сферу, и при регистрации спектров люминесценции считалась световая отдача.

Результаты и обсуждение

Замещение фторида свинца на фторид алюминия позволяет плавно варьировать показатель преломления (рис. 1). Это связано с более высокой удельной рефракцией ионов свинца по сравнению с ионами алюминия.

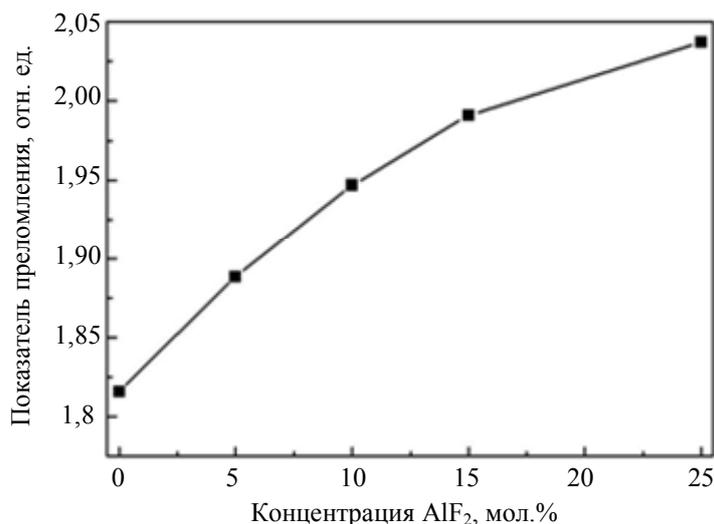


Рис. 1. Зависимость показателя преломления стекла от содержания AlF_2

Из рис. 1 видно, что увеличение количества фторида алюминия приводит к росту показателя преломления от 1,82 до 2,08. Изменяя соотношение алюминий–свинец, можно добиться полного совпадения показателей преломления кристаллического люминофора АИГ: Ce^{3+} и матрицы стекла. В нашем случае показатель преломления стекла 1,83, совпадающий с показателем преломления АИГ: Ce^{3+} , достигается при содержании фторида алюминия порядка 5 мол. %.

Одним из требований, предъявляемых к стеклянной фритте, является отсутствие взаимодействия с порошком кристаллического люминофора АИГ:Ce³⁺ при спекании. Это связано с тем, что широкая полоса люминесценции церия в желто-зеленой области принадлежит переходу $5d-4f$, и ее интенсивность сильно зависит от окружающей структуры. Церий, находящийся в структуре граната (Y₃Al₅O₁₂), обладает одной из наиболее эффективных полос люминесценции, и модификация структуры граната приводит к уменьшению интенсивности люминесценции. Чтобы выяснить, не нарушается ли при взаимодействии кристаллическая структура граната, образец композита «люминофор в стекле» был исследован методом рентгенофазового анализа. Для сравнения был добавлен спектр исходного порошка АИГ:Ce³⁺ (рис. 2).

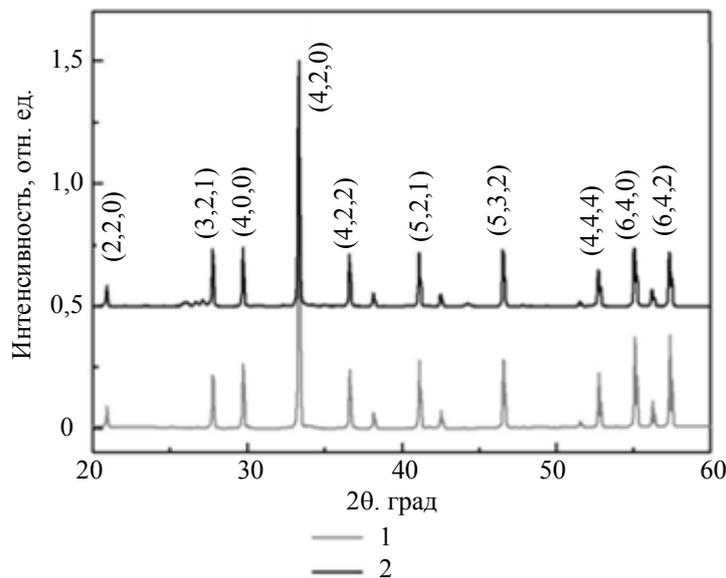


Рис. 2. Рентгенограмма порошка кристалла Y₃Al₅O₁₂:Ce (1) и композита «люминофор в стекле» (2). В скобках обозначены индексы Миллера

На рентгенограмме видно, что положение пиков в любой матрице совпадает, что свидетельствует о присутствии одинаковой кристаллической фазы Y₃Al₅O₁₂. Разница же в интенсивности пиков свидетельствует о различном объеме кристаллической фазы в образцах.

Сравним спектры возбуждения (рис. 3, а) и люминесценции (рис. 3, б) исходного порошка кристалла АИГ:Ce³⁺ со спектрами АИГ:Ce³⁺ в различных матрицах – коммерческом силиконе и разработанном свинцово-силикатном стекле.

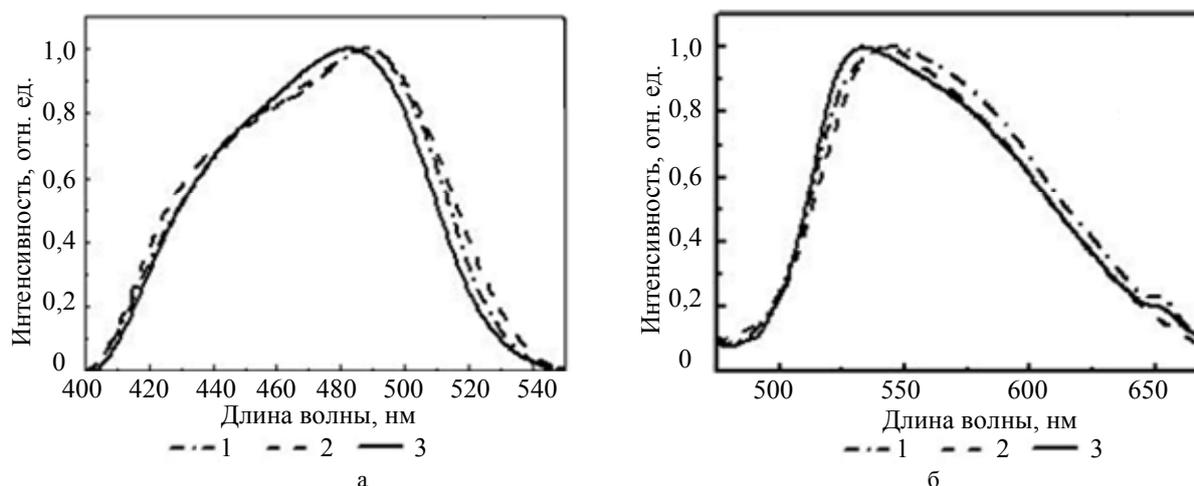


Рис. 3. Спектр возбуждения (а) и люминесценции (б): исходного порошка АИГ:Ce³⁺ (1); композита «люминофор в стекле» (2); «люминофор в полимере» (3)

На рис. 3 видно, что форма спектров практически не зависит от состава буферного связующего материала. Эти результаты также свидетельствуют о том, что значительного взаимодействия стекла и кристаллической фазы при спекании не происходит.

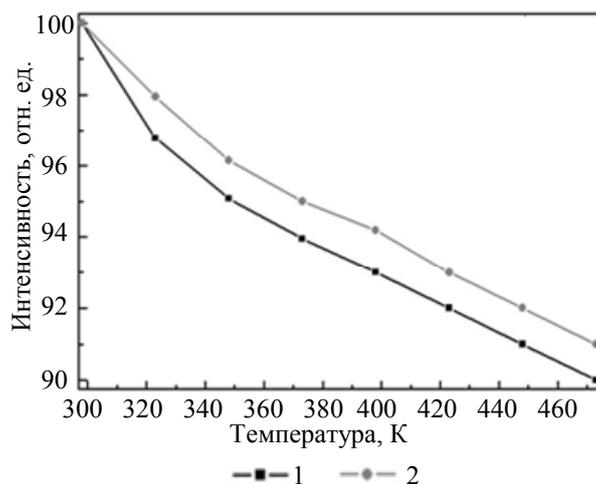


Рис. 4. Зависимость интенсивности люминесценции от температуры для «люминофора в силиконе» (1) и композита «люминофор в стекле» (2)

Рис. 4 отображает зависимость интенсивности люминесценции от температуры, которая менялась в диапазоне от комнатной температуры до 200°C. Интенсивность люминесценции Pr^{3+} для 473 К уменьшилась на 9%, а для $\text{YAG}:\text{Ce}^{3+}$ в полимере – на 10% относительно комнатной температуры. Ослабление температурного тушения для «люминофора в стекле» происходит за счет того, что матрица стекла обладает более высокой теплопроводностью (порядка 0,8 Вт/(м·К)), чем силикон (0,2 Вт/(м·К)). Это приводит к более равномерному распределению тепла по всей поверхности образца, что снижает локальный перегрев люминофора.

Следующим этапом работы явилось тестирование светодиода с композитом «люминофор в стекле». На рис. 5 показан интегральный спектр излучения, состоящий из синей полосы излучения кристалла InGaN и широкой полосы АИГ: Ce^{3+} в зелено-желтой области спектра. На основании интегрального спектра при помощи ПО SpectraWiz была определена цветовая температура, эффективность преобразования излучения из синего в желто-зеленое и световая отдача.

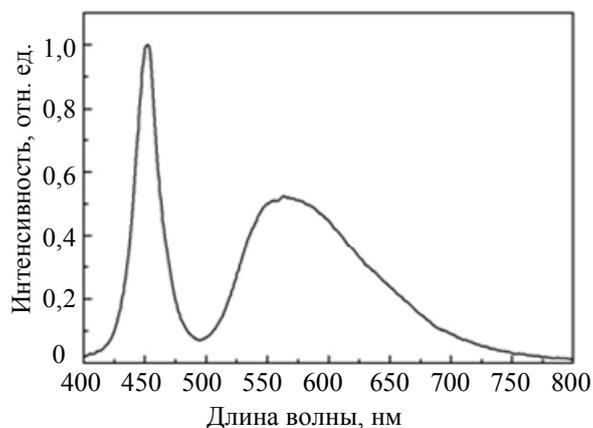


Рис. 5. Интегральный спектр излучения кристалла InGaN и композита «люминофор в стекле»

Получившийся светодиод излучает белый свет с цветовой температурой 4370 К, световая отдача составила 58 лм/Вт. Значения для традиционных светодиодов с силиконом – порядка 85 лм/Вт при соизмеримой цветовой температуре [7].

Заключение

Разработан композит «люминофор в стекле» на основе высокопреломляющей свинцово-силикатной матрицы для белых светодиодов. В исходном стекле оптимизировано соотношение оксидов и фторидов свинца и их суммарная концентрация (45 мол.%), с целью выравнивания показателя преломления стеклообразной матрицы и кристаллического порошка АИГ: Ce^{3+} для уменьшения светорассеяния на границе двух фаз «стекло–кристалл». На основе разработанного композита «люминофор в стекле» создан макет белого светодиода, обладающего значениями световой отдачи 58 лм/Вт и цветовой температуры 4370 К. Разработанный материал может быть использован в качестве люминофора для мощных светодиодов белого излучения.

Литература

1. Shin-Etsu Silicone. Product Information [Электронный ресурс]. Режим доступа: <http://www.silicone.jp/e/index.shtml>, свободный. Яз. англ. (дата обращения 21.03.2014).
2. Zhao H., Liu G., Zhang J., Poplawsky J.D., Dierolf V., Tansu N. Approaches for high internal quantum efficiency green InGaN light emitting diodes with large overlap quantum wells // *Optics Express*. 2011. V. 19. N 14. P. A991–A1007.
3. Farrell R.M., Young E.C., Wu F., Denbaars S.P., Speck J.S. Materials and growth issues for high-performance nonpolar and semipolar light-emitting devices // *Semiconductor Science and Technology*. 2012. V. 27. N 2. Art. 024001.
4. Zhang J., Tansu N. Improvement in spontaneous emission rates for InGaN quantum wells on ternary InGaN substrate for light-emitting diodes // *Journal of Applied Physics*. 2011. V. 110. N 11. Art. 113110.
5. Ee Y.-K., Biser J.M., Cao W., Chan H.M., Vinci R.P., Tansu N. Metalorganic vapor phase epitaxy of III-nitride light-emitting diodes on nanopatterned AGOG sapphire substrate by abbreviated growth mode // *IEEE Journal of Selected Topics in Quantum Electron*. 2009. V. 15. N 4. P. 1066–1072.
6. Li X.H., Song R., Ee Y.-K., Kumnorkaew P., Gilchrist J.F., Tansu N. Light extraction efficiency and radiation patterns of III-nitride light-emitting diodes with colloidal microlens arrays with various aspect ratios // *IEEE Photonics Journal*. 2011. V. 3. N 3. P. 489–499.
7. Tanabe S., Fujita S., Sakamoto A., Yamamoto S. Glass ceramics for solid state lighting // *Ceramic Transactions*. 2006. V. 173. P. 19–25.
8. Fujita S., Sakamoto A., Tanabe S. Luminescence characteristics of YAG glass-ceramic phosphor for white LED // *IEEE Journal on Selected Topics in Quantum Electronics*. 2008. V. 14. P. 1387–1391.
9. Nishiura S., Tanabe S. Preparation and optical properties of Eu^{2+} and Sm^{3+} co-doped glass ceramic phosphors emitting white color by violet laser excitation // *Journal of the Ceramic Society of Japan*. 2008. V. 116. N 1358. P. 1096–1099.
10. Tsai C.-C., Cheng W.-C., Chang J.-K., Chen Y.-L., Chen J.-H., Hsu Y.-C., Cheng W.-H. Ultra-high thermal-stable glass phosphor layer for phosphor-converted white light-emitting diodes // *IEEE/OSA Journal of Display Technology*. 2013. V. 9. N 6. P. 427–432.
11. Tsvetkova M.N., Korsakov V.G., Sychev M.M., Chernovets B.V., Itkinson G.V. Study of photophosphors for white LEDs // *Journal of Optical Technology*. 2011. V. 78. N 6. P. 403–407.
12. Tsai C.-C., Chung C.-H., Wang J., Cheng W.-C., Chen M.-H., Liou J.-S., Chang J.-K., Hsu Y.-C., Huang S.-C., Lee C.-W., Hu H.-L., Huang S.B., Kuang J.-H., Cheng W.-H. High thermal stability of high-power phosphor based white-light-emitting diodes employing Ce:YAG-doped glass // *Proc. Electronic Components and Technology Conference*. 2010. P. 700–703.
13. Tsai C.-C., Liou J.-S., Cheng W.-C., Chung C.-H., Chen M.-H., Wang J., Cheng W.-H. High humidity resistance of high-power white-light-emitting diode modules employing Ce:YAG doped glass // *Proc. Electronic Components and Technology Conference*. 2011. P. 1626–1630.
14. Lee Y.K., Lee J.S., Heo J., Im W.B., Chung W.J. Phosphor in glasses with Pb-free silicate glass powders as robust color-converting materials for white LED applications // *Optics Letters*. 2012. V. 37. N 15. P. 3276–3278.
15. Allen S.C., Steckl A.J. A nearly ideal phosphor-converted white light-emitting diode // *Applied Physics Letters*. 2008. V. 92. N 14. Art. 143309.
16. Setlur A.A. Phosphors for LED-based solid-state lighting // *Electrochemical Society Interface*. 2009. V. 18. N 4. P. 32–36.

<i>Асеев Владимир Анатольевич</i>	– кандидат физико-математических наук, ассистент, Университет ИТМО, 197101, Санкт-Петербург, Россия, Aseev@oi.ifmo.ru
<i>Бибики Анастасия Юрьевна</i>	– студент, инженер, Университет ИТМО, 197101, Санкт-Петербург, Россия, Anastasiya.bibik@list.ru
<i>Колобкова Елена Вячеславовна</i>	– доктор химических наук, доцент, профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия; Kolobok106@rambler.ru
<i>Некрасова Яна Андреевна</i>	– аспирант, инженер, Университет ИТМО, 197101, Санкт-Петербург, Россия, Nekrasova@oi.ifmo.ru
<i>Никоноров Николай Валентинович</i>	– доктор физико-математических наук, профессор, зав. кафедрой, Университет ИТМО, 197101, Россия, Санкт-Петербург, Nikonogov@oi.ifmo.ru
<i>Романов Алексей Евгеньевич</i>	– доктор физико-математических наук, приглашенный профессор, Тартуский Университет, 50090, Тарту, Эстония, Alexey.romanov@niuitmo.ru
<i>Тузова Юлия Вячеславовна</i>	– студент, Университет ИТМО, 197101, Санкт-Петербург, Россия, july2006624@gmail.com
<i>Швалева Мария Александровна</i>	– ассистент кафедры, Университет ИТМО, 197101, Санкт-Петербург, Россия, Shvalevama@niuitmo.ru

- Vladimir A. Aseev* – PhD, assistant, ITMO University, 197101, Saint Petersburg, Russia, Aseev@oi.ifmo.ru
- Anastasiya Yu. Bibik* – student, engineer, ITMO University, 197101, Saint Petersburg, Russia, Anastasiya.bibik@list.ru
- Elena V. Kolobkova* – D.Sc., Associate professor, Professor, ITMO University, 197101, Saint Petersburg, Russia, Kolobok106@rambler.ru
- Yana A. Nekrasova* – postgraduate, engineer, ITMO University, 197101, Saint Petersburg, Russia, Nekrasova@oi.ifmo.ru
- Nikolai V. Nikonorov* – D.Sc., Professor, Department head, ITMO University, 197101, Saint Petersburg, Russia, Nikonorov@oi.ifmo.ru
- Alexey A. Romanov* – D.Sc., Visiting professor, University of Tartu, Tartu, 50090, Estonia, Alexey.romanov@niuitmo.ru
- Julia V. Tuzova* – student, ITMO University, 197101, Saint Petersburg, Russia, july2006624@gmail.com
- Mariya A. Shvaleva* – assistant, ITMO University, 197101, Saint Petersburg, Russia, Shvalevama@niuitmo.ru

Принято к печати 31.05.14
Accepted 31.05.14

УДК 004.932.2

УЛУЧШЕНИЕ КАЧЕСТВА РАСПОЗНАВАНИЯ В СЕТЯХ ГЛУБОКОГО ОБУЧЕНИЯ С ПОМОЩЬЮ МЕТОДА ИМИТАЦИИ ОТЖИГА

А.С. Потапов^а, В.В. Батищева^б, Ш. Пан^с

^а Университет ИТМО, 197101, Санкт-Петербург, Россия, pas.aicv@gmail.com

^б СПбГУ, 199034, Санкт-Петербург, Россия

^с Цзилиньский университет, 130012, Цзилинь, КНР

Аннотация. Предметом исследования в работе стали методы глубокого обучения, в которых происходит автоматическое построение признаков преобразований при решении задач распознавания образов. В качестве конкретного типа сетей глубокого обучения были взяты многослойные автоэнкодеры, выполняющие нелинейное преобразование признаков, с логистической регрессией в качестве верхнего слоя, выполняющего классификацию. В целях проверки гипотезы о возможности повышения вероятности распознавания образов в сетях глубокого обучения, традиционно обучаемых послойно методом градиентного спуска, путем глобальной оптимизации параметров сети разработан и реализован оригинальный вариант метода имитации отжига применительно к настройке весов связей автоэнкодеров при дообучении слоя логистической регрессии с помощью стохастического градиентного спуска. Тестирование, проведенное на стандартной базе рукописных символов MNIST, показало уменьшение ошибок распознавания в 1,1–1,5 раза на тестовой выборке в случае модифицированного метода по сравнению с исходным методом, основанным на локальной оптимизации. Таким образом, не возникает эффект чрезмерно близкой подгонки, и подтверждается возможность улучшения качества обучения (в терминах повышения вероятности распознавания) сетей глубокого обучения с помощью методов глобальной оптимизации. Результаты работы могут быть использованы для повышения вероятности распознавания образов в областях, требующих автоматического построения нелинейных признаков преобразований, в том числе при распознавании изображений.

Ключевые слова: распознавание образов, глубокое обучение, автоэнкодер, логистическая регрессия, имитация отжига.

Благодарности. Работа выполнена при поддержке Министерства образования и науки Российской Федерации и Совета по грантам Президента Российской Федерации (грант МД-1072.2013.9) и частично при государственной финансовой поддержке ведущих университетов Российской Федерации (субсидия 074-U01).

IMPROVEMENT OF RECOGNITION QUALITY IN DEEP LEARNING NETWORKS BY SIMULATED ANNEALING METHOD

A.S. Potapov^a, V.V. Batishcheva^b, Sh. Pang^c

^a ITMO University, 197101, Saint Petersburg, Russia, pas.aicv@gmail.com

^b Saint Petersburg State University, 199034, Saint Petersburg, Russia

^c Jilin University, 130012, Changchun, Jilin Province, P.R. China

Abstract. The subject of this research is deep learning methods, in which automatic construction of feature transforms is taken place in tasks of pattern recognition. Multilayer autoencoders have been taken as the considered type of deep learning networks. Autoencoders perform nonlinear feature transform with logistic regression as an upper classification layer. In order to verify the hypothesis of possibility to improve recognition rate by global optimization of parameters for deep learning networks, which are traditionally trained layer-by-layer by gradient descent, a new method has been designed and implemented. The method applies simulated annealing for tuning connection weights of autoencoders while regression layer is simultaneously trained by stochastic gradient descent. Experiments held by means of standard MNIST handwritten digit database have shown the decrease of recognition error rate from 1.1 to 1.5 times in case of the modified method comparing to the traditional method, which is based on local optimization. Thus, overfitting effect doesn't appear and the possibility to improve learning rate is confirmed in deep learning networks by global optimization methods (in terms of increasing recognition probability). Research results can be applied for improving the probability of pattern recognition in the fields, which require automatic construction of nonlinear feature transforms, in particular, in the image recognition.

Keywords: pattern recognition, deep learning, autoencoder, logistic regression, simulated annealing.

Acknowledgements. The work is supported by the Ministry of Education and Science of the Russian Federation and the Russian Federation President's Council for Grants (grant MD-1072.2013.9), and partially financially supported by the Government of the Russian Federation (grant 074-U01).

Введение

Необходимость в распознавании образов возникает в разных областях – от геологической разведки до медицинской диагностики. К какой бы сфере ни относилась задача, при распознавании образов перед построением классификатора зачастую необходимо строить существенные (инвариантные) признаки на основе некоторых исходных признаков, что нередко делается вручную. Задача автоматического выделения признаков является особенно актуальной для систем компьютерного зрения [1].

Последнее связано с тем, что классы объектов, распознаваемых по их изображениям, крайне редко бывают линейно разделимы в пространстве первичных признаков – яркостей пикселей. Построение нелинейных разделяющих поверхностей возможно с помощью таких традиционных методов распознавания образов, как, например, метод обобщенных решающих функций или машин опорных векторов. Однако применение этих методов равносильно построению линейной разделяющей поверхности в расширенном пространстве признаков, где дополнительные нелинейные признаки заданы априорно (явно или неявно – через ядра) [2]. Типичные нелинейные признаковые преобразования (такие как полиномиальное) не обеспечивают последующей линейной разделимости классов образов, в связи с чем и возникает потребность в новых признаках, построенных с учетом особенностей распределения образов в исходном пространстве признаков.

Одним из современных подходов к проблеме автоматического выбора нелинейных признаков является подход на основе сетей глубокого обучения, ставших весьма популярными [3–6] благодаря своей способности решать задачи распознавания образов, в том числе в области компьютерного зрения, без использования вручную построенных признаков. К примеру, сети глубокого обучения были использованы для создания системы распознавания дорожных знаков, впервые в истории искусственного интеллекта продемонстрировавшей качество решения задачи распознавания выше, чем у человека [3]. Другим впечатляющим примером является использование сетей глубокого обучения совместно с методами обучения с подкреплением при обучении компьютера игре в видеоигры с использованием только необработанных видеоданных и внутриигрового подкрепления от совершаемых действий (при этом в ряде случаев достигаемая эффективность игры оказывалась выше, чем у человека) [4]. Стоит также отметить, что в регулярно проводимом соревновании по распознаванию объектов на изображениях «Large Scale Visual Recognition Challenge» последние годы программы-победители зачастую основаны на принципах глубокого обучения. Таким образом, методы на основе сетей глубокого обучения являются наилучшими известными методами распознавания образов, по крайней мере, в ряде предметных областей.

Хотя сети глубокого обучения уже привели к значительному успеху благодаря своей способности строить выразительные признаковые представления путем использования большого числа слоев, реализующих композицию нелинейных преобразований [5, 6], их исследование и совершенствование продолжается. Помимо принципиальных проблем [7, 8], требующих, видимо, существенного выхода за рамки имеющейся парадигмы глубокого обучения, один из вопросов, который можно сформулировать по отношению к сетям данного типа, заключается в том, строится ли в результате их обучения оптимальное признаковое преобразование. Данный вопрос связан с тем, что обучение в этих сетях обычно происходит послойно, а в каждом слое – при помощи метода градиентного спуска [9–11], что является локальной («жадной») оптимизационной процедурой и не гарантирует нахождения глобального оптимума. Возможным оказывается обучить глубокую сеть и традиционным методом обратного распространения ошибки [12], который, однако, также основан на градиентном спуске. В то же время существуют методы оптимизации, например метаэвристические, реализующие поиск глобально оптимальных решений и широко использующиеся для обучения искусственных нейронных сетей [13, 14]. В связи с этим целью работы стало исследование гипотезы о возможности повышения качества (вероятности распознавания) сетей глубокого обучения с помощью одного из таких методов – имитации отжига.

Описание многослойных автоэнкодеров

Одним из вариантов сетей глубокого обучения являются многослойные автоэнкодеры с добавлением слоя логистической регрессии при решении задач распознавания. Именно этот тип сетей был использован в настоящей работе.

Одиночный автоэнкодер состоит из трех слоев – входного, скрытого и выходного (слоя реконструкции). Активности нейронов входного слоя устанавливаются в значения признаков исходного образа – вектора $\mathbf{x} \in [0, 1]^N$ размерности N , который преобразуется в вектор активностей нейронов скрытого слоя $\mathbf{y} = s(\mathbf{W}\mathbf{x} + \mathbf{b})$, $\mathbf{y} \in [0, 1]^d$, соответствующих значениям новых признаков, где \mathbf{W} – матрица весов связей размером $d \times N$; \mathbf{b} – вектор порогов нейронов; s – активационная сигмоидальная функция. Активности нейронов слоя реконструкции получаются аналогично как $\mathbf{z} = s(\mathbf{W}'\mathbf{y} + \mathbf{b}')$, $\mathbf{z} \in [0, 1]^N$.

Отличие автоэнкодеров от прочих сетей прямого распространения сигнала заключается в методе обучения, определяемом их предназначением, заключающимся в построении таких скрытых признаков (число которых фиксировано и, как правило, меньше, чем число исходных признаков), что отклонение результата реконструкции \mathbf{z} от исходного образа \mathbf{x} (на обучающей выборке) минимально. В этой связи матрица \mathbf{W}' – это матрица обратного преобразования. Обычно она принимается равной \mathbf{W}^T . Обучение автоэнкодера традиционно проводится методом стохастического градиентного спуска, в котором для каждого образа обучающей выборки \mathbf{x}_i по очереди оценивается отклонение $L_i(\mathbf{W}, \mathbf{b}) = (\mathbf{x}_i - \mathbf{z}_i(\mathbf{x}_i | \mathbf{W}, \mathbf{b}))^2$, и далее параметры сети \mathbf{W} , \mathbf{b} смещаются в направлении, обратном направлению градиента $\nabla L_i(\mathbf{W}, \mathbf{b})$.

Многослойные автоэнкодеры состоят из нескольких автоэнкодеров, где каждый последующий автоэнкодер принимает на вход в качестве образа вектор признаков, формируемый скрытым слоем (а не слоем реконструкции) предыдущего автоэнкодера, и выполняет новое нелинейное преобразование вектора признаков предыдущего уровня. Обучение такой сети происходит послойно: каждый последующий слой обучается после того, как закончено обучение предыдущего слоя. В случае применения многослойных автоэнкодеров в распознавании образов при обучении с учителем результаты реконструкции используются только в процессе обучения для настройки признаковых преобразований, а основным продуктом являются активности нейронов скрытого слоя автоэнкодера верхнего уровня, которые подаются в качестве входных векторов признаков на некоторый классификатор. В качестве такового часто применяется сеть прямого распространения, которая обучается после автоэнкодеров.

Здесь обычно используется логистическая регрессия (представимая сетью прямого распространения без скрытого слоя, т.е. проводящая линейную разделяющую поверхность во входном пространстве признаков). В логистической регрессии вычисляются вероятности принадлежности образа к различным классам на основе следующего уравнения:

$$p(y = c | \mathbf{x}, \mathbf{W}) = \frac{\exp(\mathbf{w}_c^T \mathbf{x} + b_c)}{\sum_{c'=1}^C \exp(\mathbf{w}_{c'}^T \mathbf{x} + b_{c'})},$$

где \mathbf{x} – входной вектор признаков; c – индекс класса; C – общее число классов; \mathbf{W} – весовая матрица, составленная из C векторов \mathbf{w}_c ; b_c – пороги нейронов. Параметры этого классификатора также обычно настраиваются стохастическим градиентным спуском с целью минимизации отрицательного логарифма правдоподобия.

Одной из распространенных модификаций автоэнкодеров [15], которая была применена и нами, являются автоэнкодеры с подавлением шума, в которых результат реконструкции в процессе обучения вычисляется на дополнительно зашумленных образах, а ошибки реконструкции вычисляются относительно исходных образов без добавленного шума. Таким образом, автоэнкодеры обучаются восстанавливать незашумленный образ по зашумленному.

Описание разработанного метода

Для исследования возможности повышения вероятности распознавания сетями описанного типа путем отказа от сугубо локальной оптимизации весов связей сети был использован метод имитации отжига. Суть данного метода заключается в том, что при оптимизации некоторой функции $f(\mathbf{x})$, трактуемой как энергия системы \mathbf{x} , текущее решение \mathbf{x}_i на i -й итерации заменяется новым решением \mathbf{x}_{i+1} в соответствии с распределением вероятностей, например, задаваемым в форме

$$p(\mathbf{x}_{i+1} | \mathbf{x}_i) = \frac{1}{(2\pi T_i)^{D/2}} \exp\left(-\frac{|\mathbf{x}_{i+1} - \mathbf{x}_i|^2}{2T_i}\right) \frac{1}{1 + \exp(\Delta E / T_i)},$$

где $\Delta E = f(\mathbf{x}_{i+1}) - f(\mathbf{x}_i)$ – изменение энергии системы; T_i – «температура», управляющий параметр, задающий величину характерного изменения состояния системы за одну итерацию и уменьшающийся с номером итерации.

Оптимизация параметров сети методом имитации отжига может вестись по-разному, в зависимости от того, что именно принимать за функцию энергии f и состояние системы \mathbf{x} . В настоящей работе предложено следующее оригинальное применение метода имитации отжига. В качестве состояния системы \mathbf{x} использованы матрицы весов и пороги нейронов всех слоев автоэнкодеров, но без слоя логистической регрессии. Начальное состояние \mathbf{x}_0 определялось с помощью стандартной, «жадной» оптимизации сети (послойное обучение стохастическим градиентным спуском). В качестве функции энергии f выступала вероятность правильного распознавания образов обучающей выборки. При ее вычислении для каждого рассматриваемого состояния системы \mathbf{x} (значений параметров автоэнкодеров) проводилось дообучение слоя логистической регрессии. Иными словами, метод имитации отжига использовался для уточнения признаковых преобразований, задаваемых автоэнкодерами, с точки зрения их полезности для распознавания заданных образов.

Экспериментальная проверка

Экспериментальное исследование с целью проверки возможности глобальной оптимизации признаковых преобразований проводилось с использованием традиционной для тестирования методов машинного обучения базы изображений рукописных символов MNIST [16] (рисунок). Обучение реализованной нейронной сети (с разными размерами скрытых слоев в автоэнкодерах) сначала производилось на обучающей выборке изображений обычным способом, затем сеть обучалась с помощью модификации метода имитации отжига на той же выборке, и в итоге проводилось сравнение работы обученных сетей на одинаковой тестовой выборке.

Были использованы следующие параметры метода имитации отжига: начальная температура – 1,0 (в связи с тем, что активности нейронов находятся в диапазоне от 0 до 1), закон понижения температуры с номером итерации – линейный (поскольку использование логарифмического закона, характерного для больцмановского отжига, ведет к слишком медленной сходимости), число итераций – 200 (что обеспечивает достаточную точность настройки весов для признаков преобразований). Автоматическая настройка параметров метода имитации отжига не требовалась.

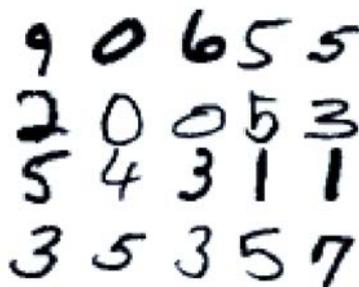


Рисунок. Примеры изображений рукописных символов из базы MNIST [16]

Экспериментальные данные позволяют провести сравнение работы автоэнкодеров с использованием имитации отжига и без него. В приведенной ниже таблице указана частота ошибок (отношение числа неверно классифицированных образов к общему числу образов в выборке), полученных при следующих параметрах: размер обучающей выборки – 40000 образов, размер тестовой выборки – 10000 образов. Число входных нейронов для первого автоэнкодера было равно числу пикселей на изображении и составляло $28 \times 28 = 784$ признака. Выходное число нейронов в слое логистической регрессии всегда было равно 10, т.е. числу классов (цифры 0–9). Число скрытых слоев и число нейронов в них варьировалось (что отражено в числах в левом столбце таблицы) для проверки повышения вероятности распознавания при разных конфигурациях сети.

Число нейронов в скрытых слоях	Исходный метод		Модифицированный метод	
	Частота ошибок (обучающая выборка)	Частота ошибок (тестовая выборка)	Частота ошибок (обучающая выборка)	Частота ошибок (тестовая выборка)
80, 70, 60, 50, 40	0,340	0,333	0,241	0,238
180, 170, 160, 150, 140	0,181	0,172	0,117	0,117
180, 140	0,102	0,092	0,078	0,082
140	0,082	0,080	0,070	0,076
400, 350, 300	0,079	0,075	0,045	0,057
600, 400	0,061	0,061	0,037	0,055
400, 300	0,067	0,066	0,042	0,052

Таблица. Частота ошибок, полученных до и после применения имитации отжига

Как видно из таблицы, частота ошибок на обучающей выборке с использованием модифицированного метода уменьшается, что свидетельствует о наличии нескольких экстремумов целевой функции, в связи с чем оправдано применение методов глобальной оптимизации.

Самым важным результатом является то, что после применения имитации отжига сокращается частота ошибок и на тестовой выборке, т.е. не возникает эффекта переобучения (чрезмерно близкой подгонки, в результате которой улучшение качества распознавания на обучающей выборке приводит к ухудшению качества распознавания на тестовой выборке). В связи с тем, что во всех семи случаях при произвольном выборе параметров сети получено уменьшение частоты ошибки, гипотеза о возможности уменьшения вероятности ошибки при использовании глобальной оптимизации многослойной сети может быть принята на уровне значимости $\alpha=0,01$ по критерию Уилкоксона. Достигнутое снижение частоты ошибок важно для потенциального применения на практике и свидетельствует о возможности дальнейшего повышения вероятностей распознавания в сетях глубокого обучения за счет глобальной оптимизации параметров сети. При этом время обучения в модифицированном алгоритме при использовании имитации отжига возрастает в 3–3,5 раза (абсолютное значение времени обучения в зависимости от параметров сети варьировалось от нескольких секунд до нескольких суток на процессоре Core i5 3,33 ГГц, что, однако, не имеет существенного значения, поскольку при прикладном использовании обучение подобных сетей обычно проводится с использованием графических карт). Такое замедление обучения не является критичным, хотя в ряде случаев может быть нежелательным.

Заключение

В работе было проведено исследование возможности использования метаэвристического метода имитации отжига для обучения промежуточных слоев автоэнкодеров в сетях глубокого обучения в дополнение к традиционно используемой локальной послойной оптимизации градиентным спуском. Результаты показывают, что разработанный метод позволяет снизить частоту ошибок распознавания, оценивавшуюся по тестовой выборке на базе MNIST, в 1,1–1,5 раза. Таким образом, возможна глобальная оптимизация весов связей в сетях глубокого обучения без возникновения эффекта чрезмерно близкой подгонки (переобучения) при использовании вероятности распознавания образов обучающей выборки в качестве целевой функции. Данный результат указывает на направление возможного дальнейшего развития методов глубокого обучения, однако необходима разработка вычислительно эффективных методов обучения с использованием глобальной оптимизации. Результаты работы могут быть использованы для повышения вероятности распознавания образов в областях, требующих автоматического построения нелинейных признаков преобразований, в том числе при распознавании изображений.

Литература

1. He Y., Kavukcuoglu K., Wang Y., Szlam A., Qi Y. Unsupervised Feature Learning by Deep Sparse Coding [Электронный ресурс]. 2013. Режим доступа: <http://arxiv.org/pdf/1312.5783v1>, свободный. Яз. англ. (дата обращения 03.07.2014).
2. Arnold L., Rebecchi S., Chevallier S., Paugam-Moisy H. An introduction to deep learning // Proc. 19th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2011). Bruges, Belgium, 2011. P. 477–488.
3. Ciresan D.C., Meier U., Masci J., Schmidhuber J. Multi-column deep neural network for traffic sign classification // Neural Networks. 2012. V. 32. P. 333–338.
4. Mnih V., Kavukcuoglu K., Silver D., Graves A., Antonoglou I., Wierstra D., Riedmiller M. Playing Atari with Deep Reinforcement Learning [Электронный ресурс]. 2013. Режим доступа: <http://arxiv.org/pdf/1312.5602v1.pdf>, свободный. Яз. англ. (дата обращения 03.07.2014).
5. Le Roux N., Bengio Y. Representational power of restricted boltzmann machines and deep belief networks // Neural Computation. 2008. V. 20. N 6. P. 1631–1649.
6. Gregor K., Mnih A., Wierstra D., Blundell C., Wierstra D. Deep Autoregressive Networks [Электронный ресурс]. 2013. Режим доступа: <http://arxiv.org/pdf/1310.8499v2>, свободный. Яз. англ. (дата обращения 03.07.2014).
7. Tenenbaum J.B., Kemp C., Griffiths T.L., Goodman N.D. How to grow a mind: statistics, structure, and abstraction // Science. 2011. V. 331. N 6022. P. 1279–1285.
8. Szegedy Ch., Zaremba W., Sutskever I., Bruna J., Erhan D., Goodfellow I., Fergus R. Intriguing properties of neural networks [Электронный ресурс]. 2014. Режим доступа: <http://arxiv.org/pdf/1312.6199v4>, свободный. Яз. англ. (дата обращения 03.07.2014).
9. Bengio Y., Lamblin P., Popovici D., Larochelle H. Greedy layer-wise training of deep networks // Advances in Neural Information Processing Systems. 2007. V. 19. P. 153–160.
10. Hinton G.E., Osindero S., Teh Y.-W. A fast learning algorithm for deep belief nets // Neural Computation. 2006. V. 18. N 7. P. 1527–1554.
11. Ranzato M.A., Poultney Ch., Chopra S., LeCun Y. Efficient learning of sparse representations with an energy-based model // Advances in Neural Information Processing Systems. 2007. V. 19. P. 1137–1144.
12. Ciresan D.C., Meier U., Gambardella L.M., Schmidhuber J. Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition [Электронный ресурс]. 2010. Режим доступа: <http://arxiv.org/pdf/1003.0358>, свободный. Яз. англ. (дата обращения 03.07.2014).
13. Царев Ф.Н. Совместное применение генетического программирования, конечных автоматов и искусственных нейронных сетей для построения системы управления беспилотным летательным аппаратом // Научно-технический вестник СПбГУ ИТМО. 2008. № 8 (53). С. 42–60.
14. Бондаренко И.Б., Гатчин Ю.А., Гераничев В.Н. Синтез оптимальных искусственных нейронных сетей с помощью модифицированного генетического алгоритма // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 2 (78). С. 51–55.
15. Vincent P., Larochelle H., Bengio Y., Manzagol P.-A. Extracting and composing robust features with denoising autoencoders // Proc. 25th International Conference on Machine Learning. Helsinki, Finland, 2008. P. 1096–1103.
16. LeCun Y., Cortes C., Burges C.J.C. The MNIST Database of handwritten digits [Электронный ресурс]. Режим доступа: <http://yann.lecun.com/exdb/mnist/>, свободный. Яз. англ. (дата обращения 03.07.2014).

- Потапов Алексей Сергеевич*** – доктор технических наук, доцент, профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия, pas.aicv@gmail.com
- Батищева Вита Вячеславовна*** – студент, СПбГУ, 199034, Санкт-Петербург, Россия, elokkuu@mail.ru
- Пан Шуцао*** – студент, Цзилиньский университет, 130012, Цзилинь, КНР, pangshuchao1212@sina.com
- Alexey S. Potapov*** – D.Sc., Associate professor, Professor, ITMO University, 197101, Saint Petersburg, Russia, pas.aicv@gmail.com
- Vita V. Batishcheva*** – student, Saint Petersburg State University, 199034, Saint Petersburg, Russia, elokkuu@mail.ru
- Shu-Chao Pang*** – student, Jilin University, 130012, Changchun, Jilin Province, P.R. China, pangshuchao1212@sina.com

Принято к печати 07.07.14
Accepted 07.07.14

УДК 057.087.1:621.391.26

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ПРОГРАММНО-АППАРАТНЫХ СРЕДСТВ ВВОДА И ВЫДЕЛЕНИЯ РЕЧЕВОГО СИГНАЛА ПОЛЬЗОВАТЕЛЯ СИСТЕМ ГОЛОСОВОЙ АУТЕНТИФИКАЦИИ

О.Н. Файзулаева^а, И.Ш. Невлюдов^а

^аХарьковский национальный университет радиоэлектроники, 61166, Харьков, Украина, Olga_Kharkov_2014@mail.ru

Аннотация. Рассматривается научная задача повышения отношения сигнал/шум речевого сигнала пользователя компьютерных систем и сетей в процессе его голосовой аутентификации. Объект исследования – процесс ввода и выделения речевого сигнала пользователя системы аутентификации в компьютерных системах и сетях. Исследуются методы и средства ввода и выделения речевого сигнала на фоне внешних помеховых сигналов. Экспериментально исследуются пути повышения качества речевого сигнала пользователя в системах голосовой аутентификации. Рассматриваются программно-аппаратные средства экспериментальной установки для ввода и выделения речевого сигнала на фоне внешних помеховых воздействий. Поскольку современные компьютерные средства, в том числе и мобильные, оснащены двухканальной звуковой картой, при вводе речевого сигнала использовались два микрофона. Расстояние между приемниками акустических волн составляло 20 мм и обеспечивало формирование одного лепестка диаграммы направленности микрофонной решетки в требуемой области регистрации речевого сигнала (от 100 Гц до 8 кГц). Как свидетельствуют результаты экспериментальных исследований, использование направленных свойств предложенной микрофонной решетки и применение пространственно-временной обработки регистрируемых сигналов с адаптивными весовыми коэффициентами позволило существенно снизить влияние помеховых сигналов. Представлены результаты экспериментальных исследований предложенных программно-аппаратных средств ввода и выделения речевых сигналов на фоне помеховых воздействий. Предложенные решения позволяют повысить значение отношения сигнал/шум регистрируемых полезных сигналов до 20 дБ при воздействии внешних помеховых сигналов в диапазоне частот от 4 до 8 кГц. Полученные результаты могут быть полезными специалистам, выполняющим исследование в области распознавания речи, а также различения диктора.

Ключевые слова: аутентификация, антенная решетка, диаграмма направленности, речевой сигнал, квадратурная обработка, микрофон.

EXPERIMENTAL STUDY OF FIRMWARE FOR INPUT AND EXTRACTION OF USER'S VOICE SIGNAL IN VOICE AUTHENTICATION SYSTEMS

O.N. Faizulaieva^a, I. Sh. Nevlyudov^a

^aKharkiv National University of Radioelectronics, 61166, Kharkiv, Ukraine, Olga_Kharkov_2014@mail.ru

Abstract. Scientific task for improving the signal-to-noise ratio for user's voice signal in computer systems and networks during the process of user's voice authentication is considered. The object of study is the process of input and extraction of the voice signal of authentication system user in computer systems and networks. Methods and means for input and extraction of the voice signal on the background of external interference signals are investigated. Ways for quality improving of the user's voice signal in systems of voice authentication are investigated experimentally. Firmware means for experimental unit of input and extraction of the user's voice signal against external interference influence are considered. As modern computer means, including mobile, have two-channel audio card, two microphones are used in the voice signal input. The distance between sonic-wave sensors is 20 mm and it provides forming one direction pattern lobe of microphone array in a desired area of voice signal registration (from 100 Hz to 8 kHz). According to the results of experimental studies, the usage of directional properties of the proposed microphone array and space-time processing of the recorded signals with implementation of constant and adaptive weighting factors has made it possible to reduce considerably the influence of interference signals. The results of firmware experimental studies for input and extraction of the user's voice signal against external interference influence are shown. The proposed solutions will give the possibility to improve the value of the signal/noise ratio of the useful signals recorded up to 20 dB under the influence of external interference signals in the frequency range from 4 to 8 kHz. The results may be useful to specialists working in the field of voice recognition and speaker discrimination.

Keywords: authentication, array, direction pattern, voice signal, quadrature processing, microphone.

Введение

В последнее время исследователи большое внимание уделяют совершенствованию систем аутентификации вычислительных систем и сетей на основе использования биометрических признаков пользователя. В ряде случаев отдается предпочтение системам голосовой аутентификации. Они имеют ряд преимуществ – простота устройств ввода, удобство ввода парольного сообщения, которое при необходимости может наращиваться, устойчивая работа в каналах связи и др. Вместе с тем современные системы голосовой аутентификации имеют показатель равной ошибки (EER – Equal Error Rates) в диапазоне 3–5%, что подчеркивает актуальность исследований в области их совершенствования [1].

В [2] предложено для совершенствования программно-аппаратных средств ввода и выделения речевого сигнала в системах голосовой аутентификации использовать два канала регистрации акустических волн, микрофоны в которых расположены на расстоянии $d = 20$ мм, с последующей квадратурной обработкой материалов регистрации.

Целью настоящей работы являются экспериментальные исследования предложенных программно-аппаратных решений и оценка их адекватности, достоверности и эффективности. Здесь же отметим, что в большинстве работ для голосовой аутентификации используются преимущественно спектральные характеристики речевого сигнала пользователя. Например, в [3] использовалась огибающая спектра голосового источника, в [4] предложен метод кепстрального преобразования спектра речевых сигналов, в [5] применялась модель, в которой спектрально-временные характеристики речевого сигнала анализируются гребенкой фильтров. Исходя из этого, ниже будут подвергаться сравнительному анализу спектральные характеристики речевого сигнала, зарегистрированного с помощью одного канала (микрофона) и предложенной системой ввода и выделения речевого сигнала на базе двух каналов (использование для регистрации двух микрофонов) [2]. При этом основное внимание будем уделять анализу диапазона спектра до 8 кГц, что обусловлено наличием отличительных признаков речевого сигнала пользователя в диапазоне от 100 Гц до 8 кГц [6, 7]. Для этого рассчитанный спектр, диапазон изменения которого определяется половиной частоты временной дискретизации, будем ограничивать частотой 8 кГц («короткий» спектр).

Программно-аппаратная схема экспериментальных исследований

Структурная схема программно-аппаратных средств экспериментальной установки представлена на рис. 1. Основу экспериментальной установки составляет ноутбук, к которому подключается выносное устройство с двумя микрофонами и схемой усиления регистрируемых колебаний. Питание выносного устройства осуществляется от USB-порта ноутбука, а регистрируемые стереосигналы поступают на вход его звуковой карты. В ноутбуке установлены и используются два класса программных средств – COOL EDIT и MATLAB. Рассмотрим более подробно назначение и функции экспериментальной установки.

В качестве приемника и усилителя акустических сигналов для каждого канала регистрации в выносном устройстве использовалась одна из практических схем микрофонного усилителя.

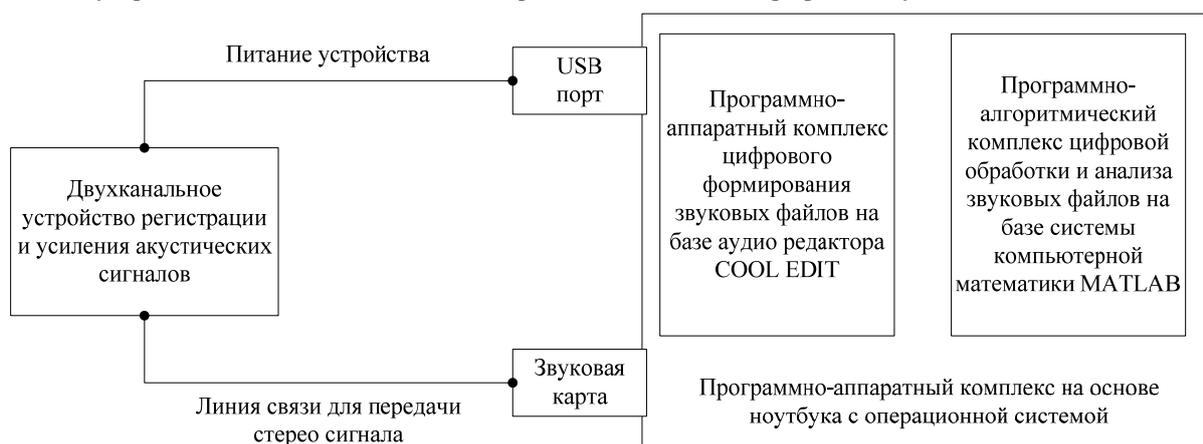


Рис. 1. Структурная схема экспериментальной установки

В ноутбуке в качестве программно-аппаратного обеспечения используется аудиоредактор COOL EDIT. Редактор является свободно распространяемым (ShareWare), его можно найти в Интернете на сайте www.syntrillium.com [8]. Он предельно прост в использовании, сохраняя при этом наиболее верные и качественные алгоритмы обработки звука. Редактор может работать с любым типом звуковых карт, при этом расширяя свои возможности в зависимости от класса качества звуковой карты. Аудиоредактор позволяет задать для звуковой карты ноутбука вид регистрируемого сигнала (моно или стерео), частоту временной дискретизации и количество бит на отсчет при регистрации, а также сформировать звуковой файл.

Сформированный файл далее поступает на цифровую обработку, процедуры которой реализуются в системе компьютерной математики MATLAB [9]. Преимущества системы MATLAB обусловлены тем, что можно использовать алгоритмический язык для реализации уникальных процедур цифровой обработки данных, а также применять готовые к использованию типовые функции для решения задач анализа цифровых данных, например, реализацию преобразования Гильберта, оценку спектра обрабатываемого сигнала и т.д.

Методика и результаты обработки материалов регистрации

Структурная схема системы выделения речевого сигнала, в которой реализована квадратурная обработка материалов регистрации, представлена на рис. 2 [2].

Ввод речевого сигнала пользователя осуществляется с расстояния 0,7–1 м по нормали к оси двух-элементной антенной решетки, т.е. в области максимума ее диаграммы направленности. Помеховый сигнал воздействует под углом к нормали оси антенной решетки.

Для формирования квадратурной составляющей используется преобразование Гильберта–Хуанга, которое не только обеспечивает качественную обработку сигналов с нестационарными и нелинейными искажениями, что присуще акустическим сигналам, но и дает возможность повысить отношение сигнал/шум (ОСШ) за счет исключения при необходимости зашумленных первых модовых функций [10–12].

Для формирования весовых коэффициентов ω_i используются адаптивные процедуры их расчета, которые имеют название алгоритма минимума среднеквадратического отклонения Уидроу–Хоффа [2]. При этом расчет адаптивных весовых коэффициентов осуществляется по шумовому сигналу, например, на этапе идентификации пользователя, а затем на этапе верификации последнему предлагается произнести заданную последовательность цифр.

Оценка качества ввода речевого сигнала производится на фоне помехового воздействия интенсивных шумов улицы, а также в замкнутом помещении. В последнем случае в качестве помеховых источников выступает шум работы коммутационной аппаратуры и серверов. В качестве полезного сигнала используется речевой сигнал цифр (от 0 до 9).

Далее представлены результаты экспериментальных исследований, полученные при частоте временной дискретизации 44 и 64 кГц.

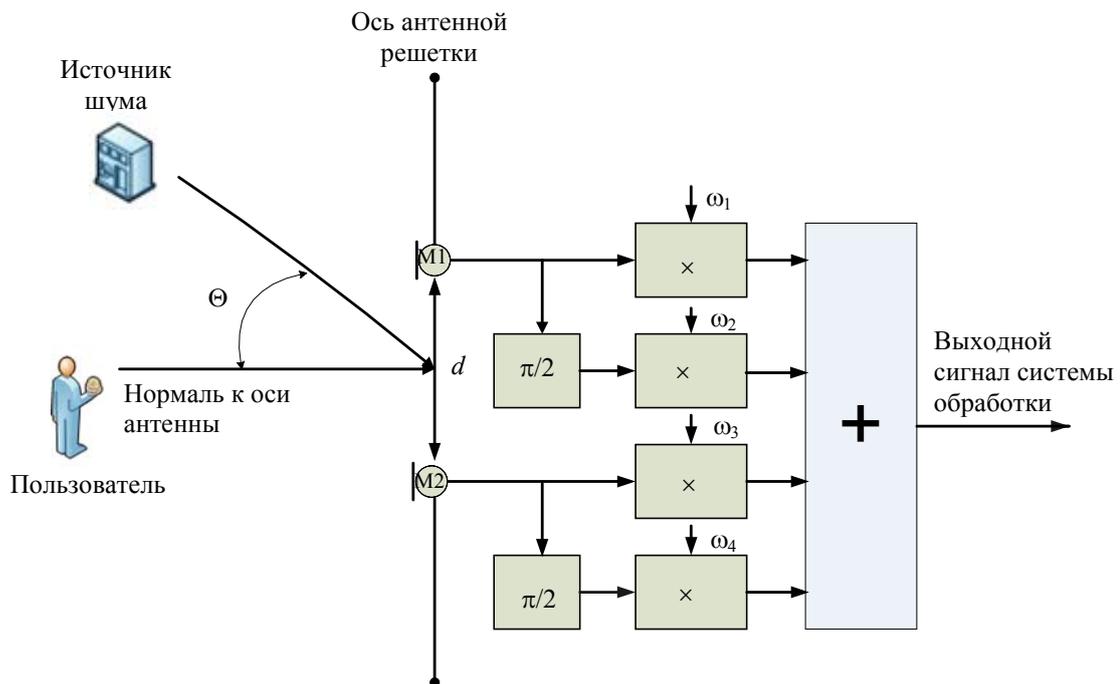


Рис. 2. Структурная схема системы ввода и выделения речевого сигнала: M1 и M2 – микрофоны; $(\pi/2)$ – программные формирователи квадратурной составляющей; ω_i – весовые коэффициенты; $i = 1, \dots, 4$; (\times) – умножители, реализующие весовую обработку; $(+)$ – сумматор

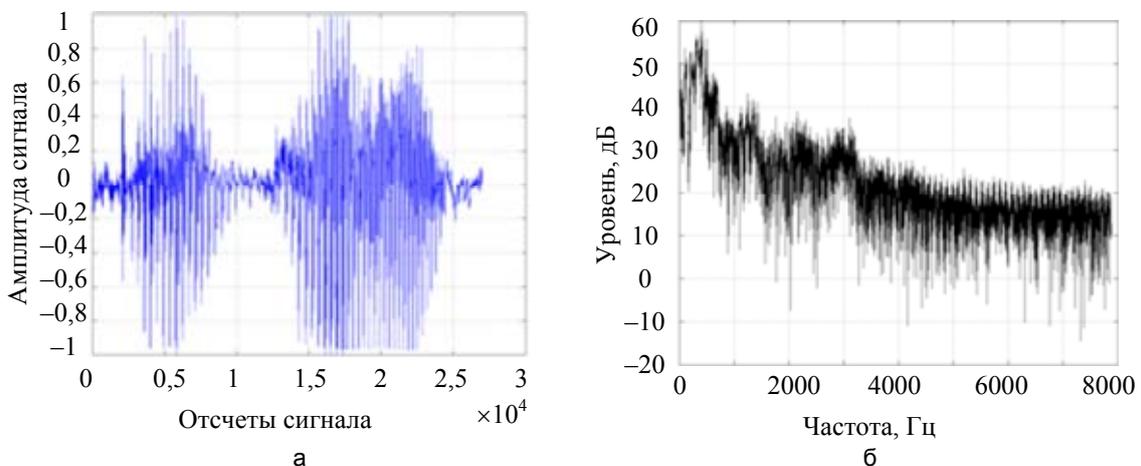


Рис. 3. Речевой сигнал цифры «1» пользователя системы аутентификации и его «короткий» спектр: речевой сигнал (а); «короткий» спектр (б). Частота дискретизации 44 кГц

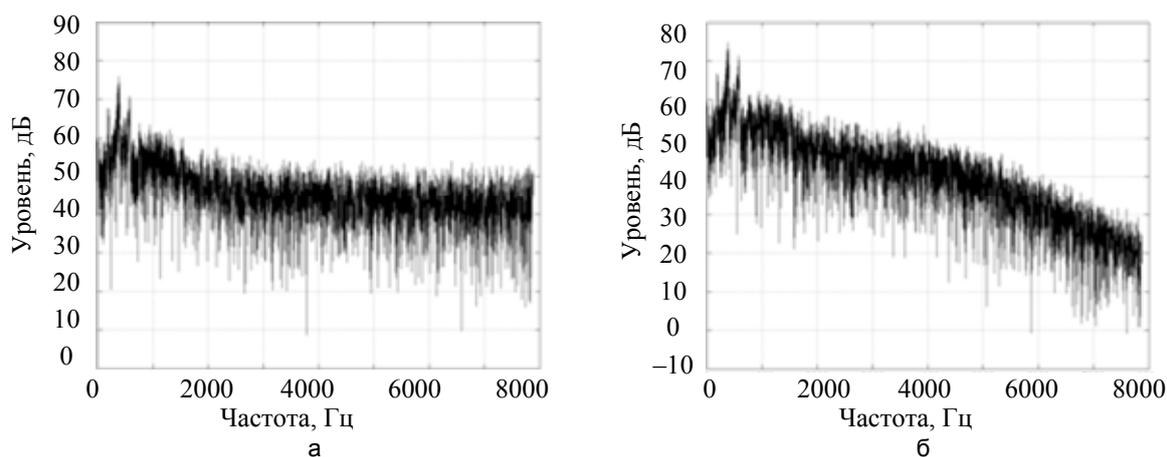


Рис. 4. Спектр смеси, зарегистрированной одним каналом (микрофоном), и спектр выходного сигнала предлагаемой системы цифровой обработки двух каналов: спектр смеси, зарегистрированной одним каналом (а); спектр выходного сигнала системы обработки (б). Частота дискретизации 44 кГц

Для примера на рис. 3, а, представлен речевой сигнал цифры «1» с частотой временной дискретизации 44 кГц, а на рис. 3, б, его «короткий» спектр.

На рис. 4, а, представлен «короткий» спектр смеси указанного речевого сигнала и шума улицы, который зарегистрирован одним из двух каналов (один микрофон). ОСШ рассматриваемой смеси составляет единицы децибел. На рис. 4, б, представлен «короткий» спектр результатов цифровой обработки двух каналов с учетом квадратурных составляющих и последующего их весового суммирования, т.е. выходной сигнал предлагаемой системы обработки (рис. 2).

Качественный анализ спектров рис. 3, б, и рис. 4, а, свидетельствует, что в области высоких частот (более 4 кГц) спектральная плотность мощности для одного канала регистрации увеличивается примерно на 20 дБ. В то же время результаты предлагаемой системы выделения речевого сигнала свидетельствуют о значительном подавлении высокочастотных составляющих, обусловленных влиянием шума.

Для количественного анализа результатов регистрации и обработки были использованы следующие показатели оценки качества ввода и выделения речевого сигнала на фоне шумовых последовательностей: коэффициент взаимной корреляции и средний «квадрат невязки».

Коэффициент взаимной корреляции оценивался между последовательностью «короткого» спектра речевого сигнала цифр (от 0 до 9) и последовательностью смеси указанного сигнала и помехового воздействия на выходе одного канала регистрации (или результатов предлагаемой схемы обработки цифровых данных двух каналов). Для расчета коэффициента взаимной корреляции использовалось известное соотношение для двух дискретных последовательностей

$$k = \frac{\sum_{i=1}^N (K_i - m_e) \cdot (\hat{K}_i - m_r)}{\sqrt{(K_i - m_e)^2 \cdot (\hat{K}_i - m_r)^2}},$$

где K_i и \hat{K}_i – анализируемые цифровые последовательности; $i = 1, \dots, N$ – номер отсчета анализируемой последовательности; N – количество анализируемых отсчетов; m_e , m_r – оценки математического ожидания анализируемых последовательностей.

Зависимость коэффициента взаимной корреляции от ОСШ представлены для различных частот временной дискретизации на рис. 5.

Из анализа представленных зависимостей следует, что коэффициент взаимной корреляции слабо изменяется при низком ОСШ. По этой причине был введен еще один показатель – средний «квадрат невязки» (η), для расчета которого использовалось соотношение

$$\eta = \frac{\sum_{i=1}^N (K_i - \hat{K}_i)^2}{N}.$$

При построении зависимостей среднего «квадрата невязок» от ОСШ выполнялась известная нормировка:

$$\hat{\eta} = \frac{\eta - \eta_{\min}}{\eta_{\max} - \eta_{\min}},$$

где η_{\min} , η_{\max} – минимальное и максимальное значение анализируемых показателей. Указанная нормировка позволит привести средний «квадрат невязок» к диапазону от 0 до 1 (нормированная «невязка»), которая, в свою очередь, упростит процесс сравнительного анализа полученных результатов.

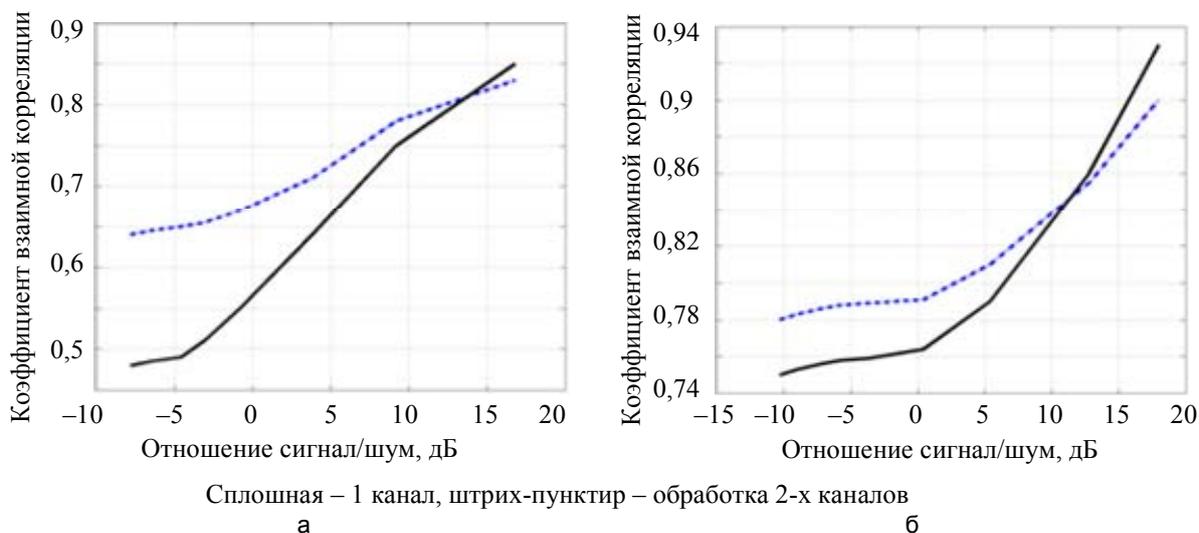


Рис. 5. Экспериментальные зависимости коэффициента взаимной корреляции от отношения сигнал/шум: частота временной дискретизации 44 кГц (а); частота временной дискретизации 64 кГц (б)

Зависимости нормированной «невязки» для частот временной дискретизации 44 и 64 кГц от ОСШ представлены на рис. 6, а, б, соответственно. Анализ представленных зависимостей позволяет сделать вывод, что предложенная система ввода и выделения речевого сигнала пользователя является более эффективной по отношению регистрации на один микрофон при обработке последовательностей с ОСШ менее 10 дБ. При ОСШ более 10 дБ возникают проблемы с формированием адаптивных весовых коэффициентов по слабому шумовому сигналу, а также сказывается влияние неидентичности характеристик каналов регистрации. В этом случае при низкой частоте временной дискретизации (менее 44 кГц) можно ограничиться применением одного канала регистрации.

Представленные результаты относятся к случаю выделения речевого сигнала цифры «1» на фоне интенсивных шумов улицы. Аналогичные зависимости получаются при обработке речевых сигналов других цифр на фоне интенсивных шумов коммутационной аппаратуры и сервера.

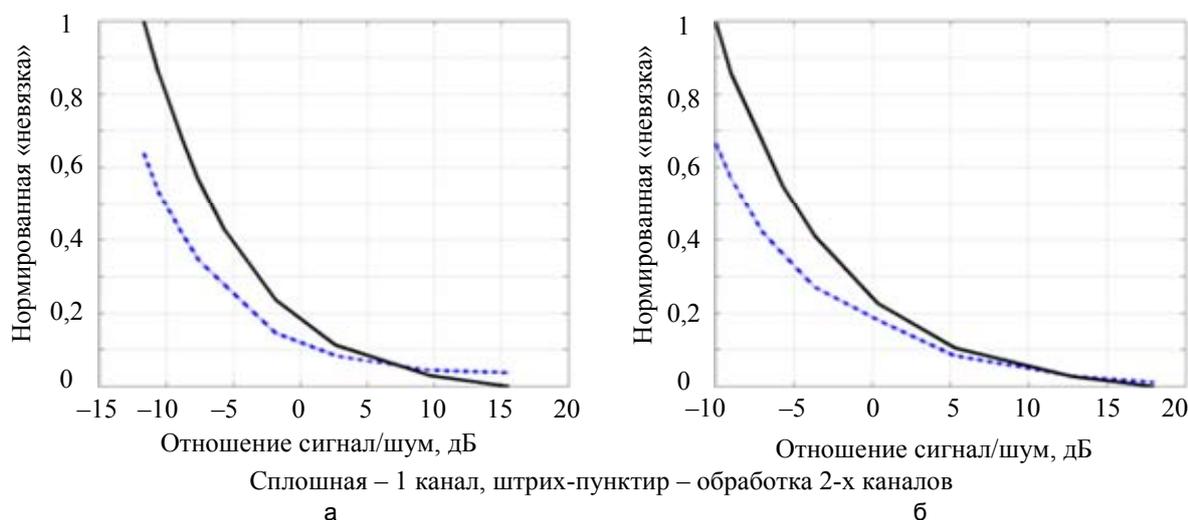


Рис. 6. Экспериментальные зависимости нормированной «невязки» от отношения сигнал/шум: частота временной дискретизации 44 кГц (а); частота временной дискретизации 64 кГц (б)

Заключение

Рассмотрены результаты экспериментальных исследований программно-аппаратных средств ввода и выделения речевого сигнала пользователя в системах голосовой аутентификации. Аппаратные средства ввода речевого сигнала включали два микрофона, расположенные на расстоянии 20 мм, и схемы предварительного усиления сигналов звуковой частоты каждого канала, которые подключались к звуко-

вой карте ноутбука. Программные средства содержали аудиоредактор COOL EDIT и систему компьютерной математики MATLAB. Аудиоредактор COOL EDIT позволял сформировать звуковой файл с заданными характеристиками, а в системе MATLAB была реализована цифровая обработка, включающая учет квадратурных составляющих регистрируемых последовательностей с использованием преобразования Гильберта–Хуанга и расчет адаптивных весовых коэффициентов на основе алгоритма минимума среднеквадратического отклонения Уидроу–Хоффа.

Предложенные технические решения позволяют повысить значение отношения сигнал/шум регистрируемых речевых сигналов до 20 дБ в диапазоне частот от 4 до 8 кГц. Особо эффективными предложенные технические решения являются при обработке цифровых последовательностей регистрируемых сигналов при отношении сигнал/шум менее 10 дБ.

Полученные результаты могут оказаться полезными и при решении иных задач, связанных с обработкой речевых сигналов в компьютерах, например, при различении диктора, построении систем физического доступа.

Дальнейшие исследования будут связаны с разработкой процедур компенсации помеховых сигналов в области низких частот.

Литература

1. Сорокин В.Н., Вьюгин В.В., Тананыкин А.А. Распознавание личности по голосу: аналитический обзор // Информационные процессы. 2012. Т. 12. № 1. С. 1–30.
2. Файзулаева О.Н., Невлюдов И.Ш. Пути улучшения качества речевого сигнала пользователя систем голосовой аутентификации // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 2 (90). С. 118–123.
3. Sorokin V.N., Tsyplikhin A.I. Speaker verification using the spectral and time parameters of voice signal // Journal of Communications Technology and Electronics. 2010. V. 55. N 12. P. 1561–1574.
4. Davis S.B., Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences // IEEE Transactions on Acoustics, Speech, Signal Process. 1980. V. ASSP-28. N 4. P. 357–366.
5. Patterson R.D., Holdsworth J. A functional model of neural activity patterns and auditory images: In Advances in Speech, Hearing and Language Processing. 1996. V. 3. P. 547–563.
6. Besacier L., Bonastre J.-F. Subband architecture for automatic speaker recognition // Signal Processing. 2000. V. 80. N 7. P. 1245–1259.
7. Lu X., Dang J. An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification // Speech Communication. 2007. V. 50. N 4. P. 312–322.
8. Петелин Р.Ю., Петелин Ю.В. Cool Edit Pro 2. Секреты мастерства. СПб. БХВ-Петербург, Арлит, 2002. 432 с.
9. Дьяконов В.П. MATLAB 6/6.1/6.5 + Simulink 4/5 Основы применения. М.: СОЛОН-Пресс, 2004. 768 с.
10. Белоусова Е.Э., Пастушенко Н.С., Пастушенко О.Н. Анализ влияния частоты дискретизации на качество формирования квадратурной составляющей аналитического сигнала // Восточно-Европейский журнал передовых технологий. 2013. Т 1. № 9 (61). С. 8–13.
11. Белоусова Е.Э., Пастушенко О.Н. Анализ влияния частоты дискретизации на качество формирования квадратурной составляющей для некоторых сигналов // Радиотехника. 2013. № 172. С. 141–146.
12. Файзулаева О.Н. Автоматизация процедур принятия решения об исключении из обработки первой модовой функции при использовании преобразования Гильберта-Хуанга // Радиотехника. 2013. № 175. С. 147–153.

Файзулаева Ольга Николаевна	–	аспирант, Харьковский национальный университет радиоэлектроники, 61166, Харьков, Украина, Olga_Kharkov_2014@mail.ru
Невлюдов Игорь Шакирович	–	доктор технических наук, профессор, заведующий кафедрой, Харьковский национальный университет радиоэлектроники, 61166, Харьков, Украина, tapr@kture.kharkov.ua
Olha N. Faizulaieva	–	postgraduate, Kharkiv National University of Radioelectronics, 61166, Kharkiv, Ukraine, Olga_Kharkov_2014@mail.ru
Ihor Sh. Nevlyudov	–	D.Sc., Professor, Department head, Kharkiv National University of Radioelectronics, 61166, Kharkiv, Ukraine, tapr@kture.kharkov.ua

Принято к печати 18.06.14
Accepted 18.06.14

УДК 002.53: 004.89

СОЗДАНИЕ СИСТЕМ ПОДДЕРЖКИ СОВМЕСТНЫХ НАУЧНЫХ ИССЛЕДОВАНИЙ В НАУКАХ О ЗЕМЛЕ В СРЕДЕ COLLA НА ОСНОВЕ МЕТОДА КУСТА СОБЫТИЙ

П. Дивьякко^a, К.А. Пшеничный^b, Б. Бенке^c, О.М. Канжелева^d

^a Национальный институт океанографии и экспериментальной геофизики, 34010, Триест, Италия

^b Университет ИТМО, 197101, Санкт-Петербург, Россия, cpshenichny@yandex.ru

^c Национальный институт геофизики и вулканологии Италии, Катанийский филиал, 95125, Катания, Италия

^d независимый исследователь, 98000-98099, Редмонд, США

Аннотация.

Постановка проблемы. Использование традиционных средств коммуникации в глобальных сетях для поддержки научно-исследовательских проектов не может решить специфических проблем, возникающих при совмещении знаний в одной и той же области, но выработанных разными научными школами, в разных парадигмах и на различной концептуальной основе. Существующие синхронные (например, Webex или ShowDocument) или асинхронные (Zimbra, Google Docs и другие) программные продукты эффективны лишь тогда, когда пользователи однозначно трактуют контекст исследования и не привносят в него собственных неформализуемых смыслов. В то же время для эффективной работы над научными проектами необходим инструмент, позволяющий в явном виде привносить подобные смыслы и связывать их с общим контекстом.

Методы. Представлено решение проблемы с помощью специализированных средств поддержки совместных научных исследований в среде COLLA, использующих методы представления и инженерии знаний. Показано, что эффективное моделирование многих областей знания требует применения специфических средств моделирования динамического знания, из которых наиболее разработанным является метод куста событий.

Результаты. Рассмотрены вопросы совмещения метода куста событий и системы поддержки совместных научных исследований COLLA на примере научно-исследовательских проектов в сейсмологии и вулканологии. Подход апробирован на примере двух актуальных задач наук о Земле – оценки сценариев развития геологических катастроф, связанных с извержением одного из самых опасных европейских вулканов (Этна), и одного из самых опасных эффектов, возникающих при землетрясениях (эффекта местоположения).

Практическая значимость. Метод куста событий может использоваться для количественной оценки сходства-различия моделируемых явлений на основе экспертного знания о них, что может иметь практическое значение далеко за пределами наук о Земле.

Ключевые слова: системы поддержки совместных научных исследований, представление знаний, инженерия знаний, динамическое знание, куст событий, науки о Земле, вулканология, Этна.

SYSTEMS FOR SUPPORT OF COLLABORATIVE STUDIES IN THE COLLA ENVIRONMENT BASED ON THE EVENT BUSH METHOD

P. Diviaco^a, C. A. Pshenichny^b, B. Behncke^c, O.M. Kanzheleva^d

^a Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, 34010, Trieste, Italy

^b ITMO University, 197101, Saint Petersburg, Russia, cpshenichny@yandex.ru

^c Sezione di Catania, Istituto Nazionale di Geofisica e Vulcanologia, 95125, Catania, Italy

^d Independent researcher, 98000-98099, Redmond, USA

Abstract.

Problem statement. Conventional tools of online communication in global networks for support of scientific and research projects fail to handle specific issues that arise at scientific collaboration of different scientific schools. They cannot be used to bring together the knowledge relating to the same field but generated in various scientific schools, under different paradigms and on different conceptual grounds. Existing software solutions, both synchronous (e.g., Webex or ShowDocument) and asynchronous (Zimbra, Google Docs and others) are efficient only when the users share similar understanding of the context and do not bring their own non-formalized meanings in it. Meanwhile, a tool is needed for scientific cooperation that makes it possible to bring one's own meanings and relate them to the common context.

Methods. Special tools for support of collaborative research in the COLLA environment using the methods of knowledge engineering and knowledge representation are given. Effective modeling of many knowledge domains is shown to require special methods of dynamic knowledge modeling; one of the well-developed is the event bush method.

Results. The paper deals with issues of combination of the event bush method and COLLA collaborative research support technique on the example of S&R projects in seismology and volcanology. Approach approbation is done on the example of two urgent tasks for geosciences: estimation of progress scenarios for geological disasters caused by Etna eruption, one of the most dangerous European volcanoes, and one of the most perilous earthquake effects (location effect).

Practical application. Event bush method can be used for quantitative assessment of similarity-difference for modeled effects based on the expert knowledge about them. This opportunity can find its substance in many fields of science far beyond geosciences.

Keywords: collaborative research support systems, knowledge representation, knowledge engineering, dynamic knowledge, event bush, geosciences, volcanology, Etna.

Введение

Современные научные исследования, как правило, представляют собой крупные совместные проекты, зачастую включающие десятки коллективов. Эти коллективы происходят из различных научных

школ, пользуются разной терминологией и, что наиболее важно, разной концептуальной базой. Для эффективной координации их работы необходимы технологии, использующие глобальные компьютерные сети. При этом обычные средства коммуникации (электронная почта, социальные сети, форумы и др.) могут решить эту задачу только на внешнем, самом простом уровне – на уровне технических вопросов (организации встреч, распределения средств и тому подобного). Использование их для ведения собственно научных дискуссий, как показывает опыт многочисленных научных форумов и блогов, приводит к смешению тематик, уходу от предмета обсуждения, подмене понятий и тезиса и, в конечном счете, не ведет к установлению взаимопонимания между учеными. Причина неэффективности обычных средств сетевого общения кроется в том, что они не позволяют вскрыть смысловые особенности коммуницируемой информации.

В последние годы получили распространение специализированные средства поддержки совместных исследований, как синхронные, требующие одновременной работы пользователей над одним проектом – например, Webex (<http://www.webex.com>) или ShowDocument (www.showdocument.com), так и асинхронные, предполагающие работу «по очереди» – Zimbra (www.zimbra.com), Google Docs (docs.google.com) и др. Как показывает опыт, асинхронные инструменты оказываются в большинстве случаев удобнее. Однако даже упомянутые программные продукты удовлетворяют, скорее, потребности промышленных и торговых компаний, нежели ученых, поскольку не могут достаточно «выпукло» показать взаимоотношения смыслов и контекстов. В настоящее время активно разрабатывается специализированный асинхронный инструмент для поддержки научных исследований COLLA [1], представляющий собой среду, в которой будут работать различные средства представления знаний.

Как отмечают многие специалисты в области методологии науки, найти общий язык для участников совместных исследований возможно только при условии, что каждый из них будет продолжать существовать в своем интеллектуальном пространстве [2, 3]. Выполнить данное условие крайне трудно, и для этого необходимы, как отмечают многие ученые, средства представления и инженерии знаний, сочетающие структурирование знаний и их визуализацию [4]. В настоящее время одним из основных методов, активно используемых для поддержки совместных научных исследований, являются онтологии [5]. Широко применяются также другие средства, основанные на классической логике – концептуальные графы [6], OWL, KIF, RDF, ER, N3 и др. [7].

Эти методы позволяют эффективно решать широкий круг задач, в которых сущности, их свойства и отношения рассматриваются как неизменные (построение моделей данных, алгоритмов поиска и другие). В то же время есть не меньший класс задач, в которых данное условие не только может не выполняться, но и должно нарушаться – и именно эти задачи представляют, как правило, особый интерес в научных исследованиях [8]. Для их решения требуются *динамические* методы представления и инженерии знаний, из которых семантически и синтаксически наиболее разработанным в настоящее время представляется метод куста событий [9].

Целью настоящей работы является разработка правил преобразования информации для создания систем поддержки совместных научных исследований в науках о Земле в среде COLLA на основе метода куста событий. Данная цель предполагает решение следующих задач:

- адаптировать метод куста событий для задач поддержки совместных научных исследований;
- внедрить существующую компьютерную реализацию метода куста событий в среду поддержки совместных научных исследований COLLA;
- валидировать предлагаемый подход на примере конкретных систем поддержки совместных научных исследований в науках о Земле.

Метод куста событий

Метод куста событий – это метод инженерии знаний о динамических сущностях (событиях, процессах, сценариях), основанный на представлении предметной области в виде «многопоточковой структуры» [9]. С помощью этого метода можно графически описать рассматриваемую предметную область на основе событий и связей между ними (или, иными словами, процессов).

Событие E определяется как высказывание вида $S - P_1, P_2, \dots, P_n$, где S – субъект, а P_1, P_2, \dots, P_n – предикаты, возможно, с отрицанием. На это определение вводятся следующие ограничения:

1. в каждом событии присутствует только один субъект;
2. субъект не может отрицаться;
3. все предикаты в событии также не могут отрицаться, т.е. в списке P_1, P_2, \dots, P_n должен быть, по крайней мере, один предикат без отрицания. Например, событием является высказывание «Сильное государство (S) не имеет больших запасов сырья ($-P_1$), не разрабатывает запасы сырья ($-P_2$), допускает возможность вмешаться в дела слабого государства (P_3), допускает возможность поглотить слабое государство (P_4)» [10].

События делятся на четыре типа (рис. 1):

1. первичные внутренние (тип ia),

2. первичные внешние (тип *ib*),
3. вторичные (тип *ii*),
4. третичные (тип *iii*).

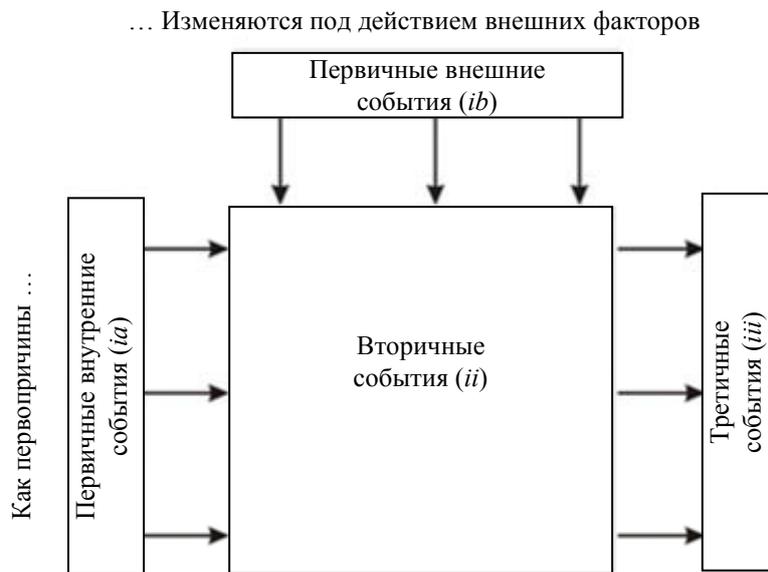


Рис. 1. Базовый синтаксис куста событий [8]

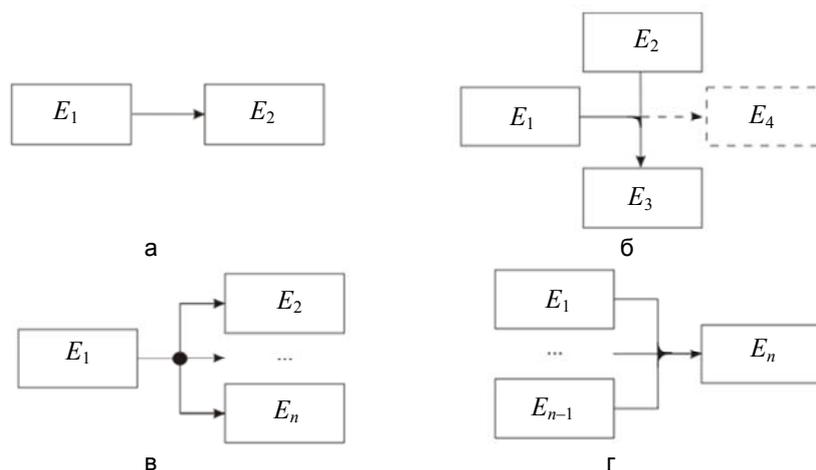


Рис. 2. Союзы куста событий: течение (а); влияние (б); разлив (в); слияние (г) [9]; *E* – событие

События связываются друг с другом союзами четырех видов – течения, влияния, слияния и разлива. Наличие первых двух является обязательным в кусте событий. Графические нотации союзов представлены на рис. 2 [9]. Детальное изложение метода куста событий приводится в работах [9] и [11].

С каждым объектом куста событий – субъектом или предикатом события, событием в целом, союзом, последовательностью союзов (т.е. сценарием) и, наконец, со всем кустом – можно связать данные любого формата. Связывая с одним и тем же объектом различные данные, пользователи структурируют и систематизируют общее знание по проекту на всех уровнях, устанавливая точки противоречия и возможности компромисса. Кроме того, если разница в подходах у ученых слишком велика, инженер по знаниям, сопровождающий проект, может предложить вместо одного куста семейство кустов событий с одинаковой семантикой (т.е. набором событий), но различным синтаксисом (т.е. организацией этих событий в сценарии посредством союзов). Это сохранит возможности связывания альтернативных данных с объектами кустов событий, но откроет возможности для формализации дискуссии и поиска общего языка на самом базовом, фундаментальном уровне понимания предметной области.

Компьютерная реализация метода куста событий в среде COLLA

К настоящему времени куст событий реализуется в виде векторной графики в соответствующих графических редакторах (например, Corel Draw, Microsoft Visio или Dia). Векторный файл конвертируется в формат SVG, что позволяет осуществить в среде COLLA интерактивную визуализацию куста собы-

тий (как и любого другого графического построения) и связывать с любым из его объектов внешние данные, а также уменьшать или увеличивать масштаб и просматривать куст по частям на экране (что представляется крайне важным ввиду сложности структуры и большого размера многих «рабочих» кустов). Файл SVG импортируется в формат COLLA, где становится основой нового проекта COLLA (рис. 3).

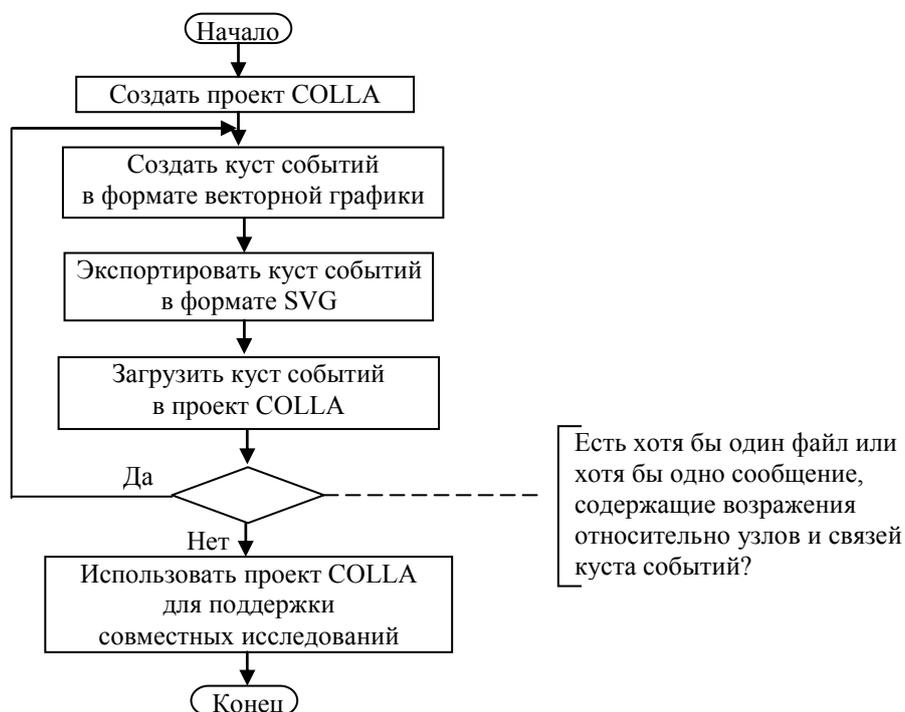


Рис. 3. Принципиальная схема проекта на основе куста событий в среде COLLA

COLLA предлагает два вида данных, которые можно связывать с графическими объектами – файлы и сообщения (рис. 3). И те, и другие могут содержать Интернет-ссылки. Файлы могут иметь любой формат. Сообщения создаются пользователями в проекте COLLA сразу в привязке к конкретному (выделенному пользователем) графическому или текстовому объекту. Сообщение появляется в среде COLLA, хранится в базе данных проекта, отображается на web-странице, соответствующей тому объекту, к которому оно относится, и дублируется по электронной почте всем участникам проекта. Адрес электронной почты необходим при регистрации в проекте COLLA. В теле письма, кроме текста сообщения в COLLA, содержится «волшебная ссылка», позволяющая автоматически регистрироваться в системе и немедленно приступить к работе по обсуждаемому проекту.

При импорте куста событий все партнеры видят его на экране и могут работать с ним, связывая с субъектами, предикатами, союзами, сценариями и кустом в целом файлы и сообщения. Щелкнув по тому или иному объекту, пользователь видит, какие файлы и сообщения с ним связаны, и может добавить свои. При этом все участники проекта, включая его самого, получают по электронной почте уведомление о подгруженном файле или комментарии к соответствующему объекту куста событий.

Системы поддержки совместных исследований в науках о Земле

Предлагаемый подход был применен в рамках нескольких международных исследовательских инициатив в науках о Земле – вулканологии (рис. 4) и сейсмологии [12].

Система поддержки вулканологических исследований преследует целью объединение и согласование представлений о характере извержений вулкана Этна (Италия). Для этого поведение вулкана моделируется средствами куста событий [13]. Как показано на рис. 4, куст событий (а) внедряется в среду COLLA в виде самостоятельного проекта в этой среде (б). В этом проекте к каждому объекту куста событий могут подгружаться сообщения (в) и файлы (г). Подгружая свои файлы, пользователь тем самым передает свой уникальный контекст, в котором существует его видение предмета обсуждения. Одновременно этот контекст – точнее, та его часть, которую сам пользователь посчитал значимой для совместного исследования – помещается в более общий контекст данного исследования, очерчиваемый кустом событий.

Описанная система действует в Национальном институте океанографии и экспериментальной геофизики в Триесте (Италия) и используется международным научным сообществом для прогнозной оценки активности кратеров Этны.

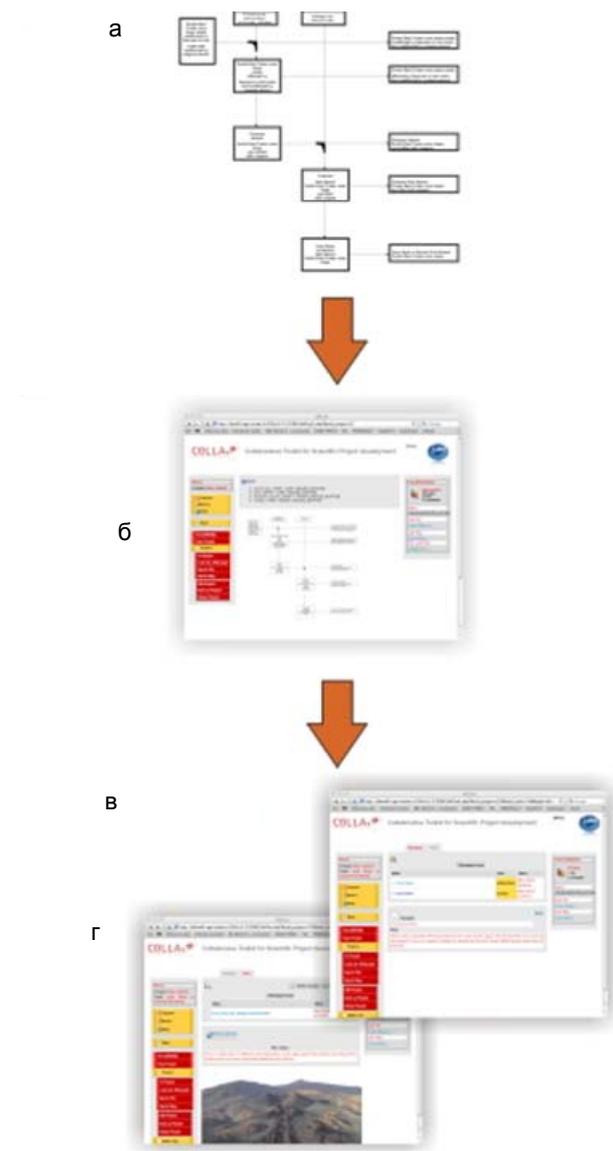


Рис. 4. Общая схема системы поддержки вулканологических исследований на вулкане Этна (Италия): куст событий (а); куст событий, внедренный в среду COLLA в том виде, в каком его видит пользователь (б); текстовые сообщения, связанные в среде COLLA с элементами куста событий (в); файлы, связанные в среде COLLA с элементами куста событий (г)

Еще одним примером использования куста событий в среде COLLA для решения задач в науках о Земле является создание системы оптимизации моделирования эффектов местоположения (site effects), заключающихся в многократном усилении сейсмической волны на ограниченных территориях вследствие особенностей геологического строения. Работа над данной системой была начата в рамках проекта Европейской комиссии CRODINAS (2009–2011 гг.), а в настоящее время она также развернута в итальянском Национальном институте океанографии и экспериментальной геофизики.

Моделирование эруптивной активности вулкана Этна

Пример построения куста событий по наблюдаемым на поверхности Этны породам и формам рельефа (рис. 5) на основе знания о породивших их процессах приведен на рис. 6. Подобным образом был построен общий куст событий, описывающий полную группу сценариев извержений Юго-Восточного кратера Этны [13]. Затем, путем удаления ненаблюдавшихся сценариев, из него были получены его частные случаи – кусты событий по трем последовательным извержениям, 16, 19 и 24 ноября 2006 г.

Одной из важнейших задач, решаемых для составления прогнозной оценки хода извержения, является определение сходства начинающегося или предполагаемого извержения с уже произошедшими. До настоящего времени данная задача решалась в вулканологии интуитивно, на основе зрительного опыта ученых, а в случае его отсутствия – на основе текстовых описаний. Авторами был произведен опрос спе-

циалистов, знакомых с тремя моделируемыми извержениями. Подавляющее большинство высказало мнение, что извержения 16 и 19 ноября гораздо более схожи друг с другом, нежели 16 и 24 ноября и 19 и 24 ноября 2006 г. Метод куста событий открывает теоретическую возможность проводить подобное сравнение на гораздо более строгом и беспристрастном уровне, оценивая меру сходства/различия и качественно, и количественно.

Для оценки различия кустов событий определим бинарное отношение \leq на множестве [14], в данном случае – на множестве кустов событий. Для заданных кустов A и B будем говорить: если $A \leq B$, то куст A не превосходит куст B . Это значит, что, во-первых, множество узлов куста A V_A является подмножеством множества узлов куста B V_B : $V_A \subseteq V_B$; во-вторых, множество ребер куста A E_A является подмножеством множества ребер куста B E_B : $E_A \subseteq E_B$.

Отметим, что множество кустов событий является частично упорядоченным множеством [15] относительно заданного выше бинарного отношения \leq . Действительно, на множестве кустов событий для данного отношения выполнены рефлексивность, транзитивность и антисимметричность. Таким образом, при сравнении двух кустов событий, описывающих извержения или эруптивные центры, нетрудно сформулировать соотношения между ними [16].



Рис. 5. Рельеф и породы, созданные эруптивными процессами на южном склоне вулкана Этна (Сицилия) вблизи Рефуджио Сапиенца. Фото К.А. Пшеничного

1. Отношение тождества: $A = B$. Кусты совпадают, т.е. совпадают и набор узлов, и все ребра между ними.
2. Отношение подчинения: $A > B$. Один куст является частью другого. Все его узлы и ребра присутствуют в другом кусте, но не все узлы и (или) ребра другого куста присутствуют в нем. Именно в таком отношении находятся кусты отдельных извержений к кусту по Юго-Восточному кратеру.
3. Отношение пересечения, или общего положения. A и B находятся в общем положении, т.е. существуют, по крайней мере, один узел (или ребро), принадлежащий исключительно кусту A , узел (или ребро), принадлежащий исключительно кусту B , а также узел, принадлежащий обоим кустам.
4. Отношение несравнимости. Кусты A и B не пересекаются, т.е. у кустов нет ни одного одинакового узла. Заметим, что отсутствие одинаковых узлов гарантирует отсутствие одинаковых ребер.

Первый случай тривиален, в остальных случаях можно оценить количество и характер шагов (введение/удаление узлов и связей), необходимых, чтобы из одного куста получить другой – или, иными словами, количественно определить, насколько один куст похож на другой. Соответственно, если считать куст событий моделью наших представлений о явлении, есть основания надеяться, что данный подход позволит определить тот «поправочный коэффициент», который надо ввести, чтобы применить вероятностное высказывание об одном вулкане или извержении к другому вулкану или извержению.

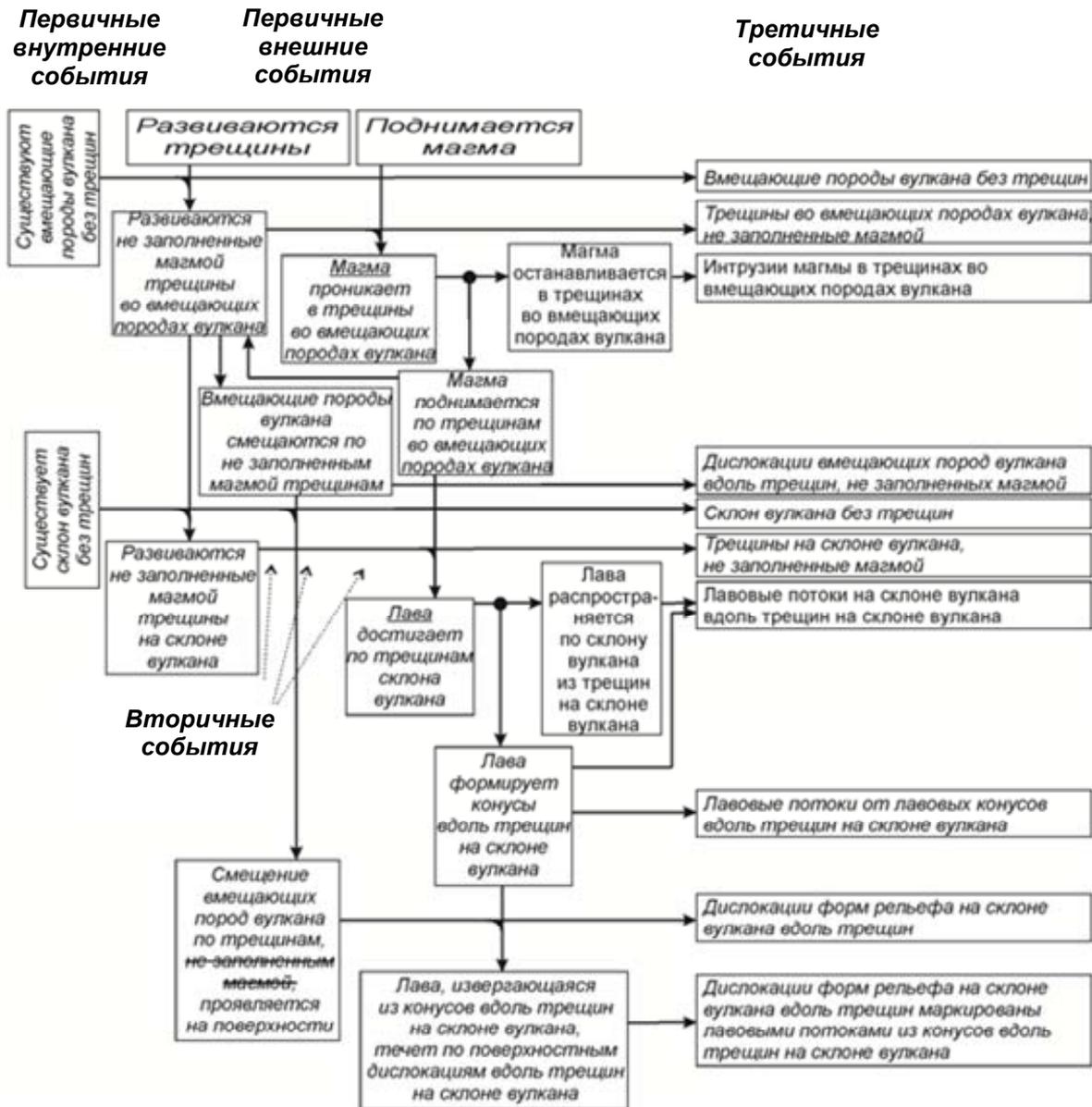


Рис. 6. Куст событий, описывающий формирование рельефа и пород, показанных на рис. 5. Зачеркнуты признаки, оказывающиеся не релевантными для следующего шага

Поправочный коэффициент является функцией, аргументами которой являются количество различных узлов $((V_A \setminus V_B) \cup (V_B \setminus V_A))$ и количество различных ребер $((E_A \setminus E_B) \cup (E_B \setminus E_A))$.

В простейшем случае данная функция может выглядеть следующим образом:

$$k(A, B) = 1 - \frac{|A \setminus B| + |B \setminus A|}{|A| + |B|},$$

где $|A \setminus B|$ – это количество узлов и ребер, которые есть в кусте событий A , но не присутствуют в кусте событий B ; $|B \setminus A|$ – это количество узлов и ребер, которые есть в кусте событий B , но не присутствуют в кусте событий A ; $|A|$, $|B|$ – это общее количество узлов и ребер кустов событий A , B соответственно. Данный поправочный коэффициент отражает схожесть кустов событий: если кусты равны, коэффициент равен 1, если у кустов нет ни одного одинакового узла, коэффициент равен 0.

Коэффициент был посчитан для кустов событий, имевших место 16, 19 и 24 ноября 2006 г.

Наименьшее значение имеет коэффициент для кустов событий 16 и 19 ноября, его значение $\approx 0,5$. Сравнение кустов событий для 16 и 24 ноября дает коэффициент $\approx 0,68$. Наибольшим сходством обладают кусты событий для 19 и 24 ноября, коэффициент $\approx 0,82$. С вулканологической точки зрения эти значения представляются вполне обоснованными, хотя интуитивно далеко не очевидными. Действительно, подавляющее большинство событий 19 ноября (связанных с истечением лавы или стромболианской активностью) в том виде, в каком они представлены в рассматриваемом кусте событий, повторились и

24 ноября. В то же время другие события третьего извержения (а именно, практически все необычные процессы перемещения вулканического материала) имеют гораздо больше общего с извержением 16 ноября. Тем не менее, интуитивное решение большинства экспертов-вулканологов относительно попарного сходства рассматриваемых извержений было иным.

Подобное сравнение можно производить не только для кустов событий, описывающих различные извержения одного вулкана, но и для двух любых кустов событий, например, для двух кустов событий, описывающих поведение различных вулканов. В будущем мы предполагаем разработать более тонкие меры сходства/различия, лучше учитывающие семантические различия между узлами. Такой сравнительный анализ был бы чрезвычайно полезен в решении ряда прикладных задач – таких, как оценка вулканической опасности или геотермального потенциала питающих магматических систем. Подобное сравнение можно осуществлять и в других областях наук о Земле, что значительно повысит надежность экспертных оценок при переносе знания по одному объекту на другой.

Кусты событий по трем последовательным извержениям – 16, 19 и 24 ноября 2006 г. – представляют собой частные случаи общего куста событий, построенного для описания полной группы эруптивных сценариев юго-восточного кратера Этны. За период с 2006 по 2014 гг. было описано несколько сотен больших и малых извержений юго-восточного, а затем и нового юго-восточного кратеров Этны. В ходе многих из них наблюдались события, отсутствовавшие при трех упомянутых извержениях – превращение потока лавы в поток обломков в результате взаимодействия лавы и снега, фонтанирование лавы из вновь раскрывшихся трещин и другие – но не было зафиксировано ни одного сценария, не содержащегося в общем кусте событий. Это означает, что за восемь лет не возникло ситуации, когда данные, подгружаемые участниками проекта в среду COLLA в виде файлов или сообщений, не могли бы быть адекватно соотнесены с принятой моделью вулканической активности.

Заключение

Как показывают проведенные исследования, метод куста событий, являющийся на данный момент основным методом инженерии динамических знаний, хорошо адаптируется для задач поддержки совместных научных исследований. Получаемый в результате применения данного метода графический конструктор (куст событий) легко разбивается на объекты различной сложности, каждый из которых соответствует тому или иному блоку знаний. Синтаксис и семантика куста событий облегчают задачу связывания с этим блоком знаний специалистов в предметной области с сохранением того интеллектуального пространства, в котором они были созданы.

Несмотря на то, что еще не создано специализированное программное обеспечение, позволяющее строить кусты событий, существующая компьютерная реализация данного метода в виде обыкновенной векторной графики легко внедряется через формат SVG в среду поддержки совместных научных исследований COLLA.

Данный подход был опробован на примере двух актуальных задач наук о Земле – оценки сценариев развития геологических катастроф, связанных с извержением одного из самых опасных европейских вулканов (Этны) и одного из самых опасных эффектов, возникающих при землетрясениях (эффекта местоположения). Созданные проекты позволили значительно повысить качество научной дискуссии и структурировать соответствующие области знания под решение конкретных задач, интересующих исследователей. Формализованное сравнение и количественная оценка сходства–различия обстановок и объектов на основе метода куста событий позволяют существенно повысить надежность экспертных оценок в науках о Земле.

Литература

1. Diviacco P. Addressing conflicting cognitive models in collaborative e-research: a case study in exploration geophysics / In: Collaborative and Distributed E-Research: Innovations in Technologies, Strategies and Applications. IGI Global, 2012. P. 247–275.
2. Kuhn T.S. The Structure of Scientific Revolutions. University of Chicago Press, 1962. 226 p.
3. Lakatos I., Musgrave A. Falsification and the methodology of scientific research programmes // Criticism and the Growth of Knowledge. Cambridge University Press, 1970. P. 91–195.
4. Star S.L., Griesemer J.R. Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's museum of vertebrate zoology // Social Studies of Science. 1989. V. 19. N 3. P. 387–420.
5. Gavrilova T., Bolotnikova E., Leshcheva I., Blagov E., Yanson A. Measuring psychological impact on group ontology design and development: an empirical approach // Communications in Computer and Information Science. 2013. V. 394. P. 29–43.
6. Sowa J.F. Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks/Cole Publ., 2000. 594 p.
7. Martin P. Knowledge representation in CGLF, CGIF, KIF, Frame-CG and Formalized-English // Lecture Notes in Computer Science. 2002. V. 2393. P. 77–91.

8. Pshenichny C.A., Mouromtsev D.I. Representation of the event bush approach in terms of directed hypergraphs // Lecture Notes in Computer Science. 2013. V. 7735. P. 289–300.
9. Pshenichny C.A., Kanzheleva O.M. Theoretical foundations of the event bush method // Special Paper of the Geological Society of America. 2011. V. 482. P. 139–164.
10. Соломин К.И., Пшеничный К.А. Опыт применения метода куста событий для анализа исторического контекста // Сборник трудов молодых ученых, аспирантов и студентов научно-педагогической школы кафедры ПБКС «Информационная безопасность, проектирование и технология элементов и узлов компьютерных систем» / Под ред. Ю.А. Гатчина. СПб: НИУ ИТМО, 2013. Вып. 1. С. 91–100.
11. Pshenichny C.A., Nikolenko S.I., Carniel R., Vaganov P.A., Khrabrykh Z.V., Moukhachov V.P., Akimova-Shterkhun V.L., Rezyapkin A.A. The event bush as a semantic-based numerical approach to natural hazard assessment (exemplified by volcanology) // Computers and Geosciences. 2009. V. 35. N 5. P. 1017–1034.
12. Diviaco P., Pshenichny C.A., Carniel R., Behncke B. A case study of the use of event bushes as a formal representation for computer-supported collaborative work in the geosciences // European Geoscience Union General Assembly. 2011. V. 13. EGU2011–1893.
13. Behncke B., Pshenichny C. Modeling unusual eruptive behavior of Mt. Etna, Italy, by means of event bush // Journal of Volcanology and Geothermal Research. 2009. V. 185. N 3. P. 157–171.
14. Кантор Г. Труды по теории множеств. М.: Наука, 1985. 431 с.
15. Александров П.С. Введение в теорию множеств и общую топологию. М.: Наука, 1977. 368 с.
16. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. 7-е изд. М.: ФИЗМАТЛИТ, 2004. 572 с.

- | | |
|-------------------------------------|---|
| <i>Дивьякко Паоло</i> | – PhD, научный сотрудник, Национальный институт океанографии и экспериментальной геофизики, 34010, Триест, Италия, pdiviaco@ogs.trieste.it |
| <i>Пшеничный Кирилл Анатольевич</i> | – кандидат геолого-минералогических наук, доцент, Университет ИТМО, 197101, Санкт-Петербург, Россия, cpshenichny@yandex.ru |
| <i>Бенке Борис</i> | – PhD, научный сотрудник, Национальный институт геофизики и вулканологии Италии, Катанийский филиал, 95125, Катания, Италия, Boris.behncke@nosпамct.ingv.it |
| <i>Канжелева Оксана Михайловна</i> | – PhD, независимый исследователь, Редмонд, США, Oksana.kanzheleva@gmail.com |
| <i>Paolo Diviaco</i> | – Scientific researcher, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, 34010, Trieste, Italy, pdiviaco@ogs.trieste.it |
| <i>Cyril A. Pshenichny</i> | – PhD, Associate professor, ITMO University, 197101, Saint Petersburg, Russia, cpshenichny@yandex.ru |
| <i>Boris Behncke</i> | – PhD, scientific researcher, Sezione di Catania, Istituto Nazionale di Geofisica e Vulcanologia, 95125, Catania, Italy, Boris.behncke@nosпамct.ingv.it |
| <i>Oksana M. Kanzheleva</i> | – PhD, Independent researcher, 98000-98099, Redmond, USA, Oksana.kanzheleva@gmail.com |

Принято к печати 23.04.14
Accepted 23.04.14

УДК 20.19.29

ДВУЯЗЫЧНАЯ МНОГОМОДАЛЬНАЯ СИСТЕМА ДЛЯ АУДИОВИЗУАЛЬНОГО СИНТЕЗА РЕЧИ И ЖЕСТОВОГО ЯЗЫКА ПО ТЕКСТУ

А.А. Карпов^{a, b}, М. Железны^c

^a Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), 199178, Санкт-Петербург, Россия

^b Университет ИТМО, 197101, Санкт-Петербург, Россия, karpov@iias.spb.su

^c Западночешский Университет, 30614, Пльзень, Чехия, zelezny@kky.zcu.cz

Аннотация. Представлена концептуальная модель, архитектура и программная реализация многомодальной системы аудиовизуального синтеза речи и жестового языка по входному тексту. Основными компонентами разработанной многомодальной системы синтеза (жестовый аватар) являются: текстовый процессор анализа входного текста; имитационная трехмерная модель головы человека; компьютерный синтезатор звучащей речи; система синтеза аудиовизуальной речи; имитационная модель верхней части тела и рук человека; многомодальный пользовательский интерфейс, интегрирующий компоненты генерации звучащей, визуальной и жестовой речи по тексту. Предложенная система выполняет автоматическое преобразование входной текстовой информации в речевую (аудиоинформацию) и жестовую (видеоинформацию), объединение и вывод ее в виде мультимедийной информации. На вход системы подается произвольный грамматически корректный текст на русском или чешском языке, который анализируется текстовым процессором для выделения предложений, слов и букв. Далее полученная текстовая информация преобразуется в символы жестовой нотации (используется международная «Гамбургская система нотации» – HamNoSys, которая описывает основные дифференциальные признаки каждого жеста рук: форму кисти, ориентацию руки, место и характер движения), на основе которых трехмерный жестовый аватар воспроизводит элементы жестового языка. Виртуальная трехмерная модель головы и верхней части тела человека реализована на языке моделирования виртуальной реальности VRML и управляется программно средствами графической библиотеки OpenGL. Предложенная многомодальная система синтеза является универсальной, она предназначена как для обычных пользователей, так и для людей с ограниченными возможностями здоровья (в частности, глухих и незрячих людей) и служит для целей мультимедийного аудиовизуального вывода вводимой текстовой информации.

Ключевые слова: многомодальные интерфейсы пользователя, человеко-машинное взаимодействие, жестовый язык, синтез речи, трехмерные модели, ассистивные технологии, жестовый аватар.

Благодарности. Исследование выполнено при частичной финансовой поддержке Правительства Российской Федерации (грант № 074-U01), фонда РФФИ (проект № 12-08-01265_a) и Европейского фонда регионального развития (ЕФРР), проект «Новые технологии для информационного общества» (NTIS), Европейский центр передового опыта, ED1.1.00/02.0090.

BILINGUAL MULTIMODAL SYSTEM FOR TEXT-TO-AUDIOVISUAL SPEECH AND SIGN LANGUAGE SYNTHESIS

A.A. Karpov^{a, b}, M. Zelezny^c

^a Saint Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 199178, Saint Petersburg, Russia;

^b ITMO University, 197101, Saint Petersburg, Russia, karpov@iias.spb.su

^c University of West Bohemia, 30614 Pilsen, Czech Republic, zelezny@kky.zcu.cz

Abstract. We present a conceptual model, architecture and software of a multimodal system for audio-visual speech and sign language synthesis by the input text. The main components of the developed multimodal synthesis system (signing avatar) are: automatic text processor for input text analysis; simulation 3D model of human's head; computer text-to-speech synthesizer; a system for audio-visual speech synthesis; simulation 3D model of human's hands and upper body; multimodal user interface integrating all the components for generation of audio, visual and signed speech. The proposed system performs automatic translation of input textual information into speech (audio information) and gestures (video information), information fusion and its output in the form of multimedia information. A user can input any grammatically correct text in Russian or Czech languages to the system; it is analyzed by the text processor to detect sentences, words and characters. Then this textual information is converted into symbols of the sign language notation. We apply international «Hamburg Notation System» - HamNoSys, which describes the main differential features of each manual sign: hand shape, hand orientation, place and type of movement. On their basis the 3D signing avatar displays the elements of the sign language. The virtual 3D model of human's head and upper body has been created using VRML virtual reality modeling language, and it is controlled by the software based on OpenGL graphical library. The developed multimodal synthesis system is a universal one since it is oriented for both regular users and disabled people (in particular, for the hard-of-hearing and visually impaired), and it serves for multimedia output (by audio and visual modalities) of input textual information.

Keywords: multimodal user interfaces, human-computer interaction, sign language, speech synthesis, 3D models, assistive technologies, signing avatar.

Acknowledgements. The research is partially financially supported by the Government of the Russian Federation (grant № 074-U01), by the Russian Foundation for Basic Research (project № 12-08-01265_a) and European Foundation of Regional Progress, project “New Technologies for Information Society” (NTIS), European Centre of Advanced Experience, ED1.1.00/02.0090.

Введение

Основным способом межчеловеческой коммуникации в обществе глухих и слабослышащих людей является жестовый язык (ЖЯ), в котором каждому смысловому понятию (или группе синонимичных понятий) соответствует определенный уникальный жестовый эквивалент. В нем для передачи информации используются различные визуально-кинестические средства естественного межчеловеческого общения: жесты рук, мимика и эмоции лица, артикуляция губ. ЖЯ не является универсальным во всех странах мира, так как он возникает и развивается естественным путем в разных локальных сообществах и изменяется со временем с появлением новой лексики. По данным международного интернет-каталога Ethnologue (www.ethnologue.com/subgroups/deaf-sign-language), в разных странах мира насчитывается порядка 140 различных жестовых языков, не считая их региолекты.

По статистике Министерства здравоохранения Российской Федерации (РФ), в России насчитывается около 200 тысяч глухих и слабослышащих граждан, имеющих инвалидность по слуху, и около 95% инвалидов по слуху общаются посредством языка жестов. Без его использования не обойтись также на приеме у врача, у нотариуса, на судебном процессе, на официальных лекциях, при бытовых разговорах и т.д. Также, согласно последней переписи населения 2010 г., в России насчитывается более 120 тысяч человек, владеющих русским ЖЯ. Для сравнения, в Чехии по статистике насчитывается около 7500 носителей чешского ЖЯ (это около 0,07% населения) и около 500 тысяч слабослышащих людей (4,7% всего населения страны) [1].

При этом нужно отметить, что ЖЯ является вторым государственным языком устного общения граждан в США, Финляндии, Испании, Чехии и т.д., что закреплено Конституцией этих стран. До недавнего времени русский ЖЯ не имел в России никакого официального статуса, однако в самом конце 2012 г. Президент РФ подписал закон, определяющий официальный статус русского ЖЯ в России как языка общения при наличии у людей нарушений слуха или речи, в том числе в сферах устного использования государственного языка РФ. В результате вступления в силу данного закона глухие люди получают возможность, например, обращения в государственные учреждения на ЖЯ. В дальнейшем предполагается создание автоматизированных систем субтитрования и сурдоперевода телевизионных программ и кинофильмов.

Основной задачей автоматизированных систем, основанных на компьютерной обработке ЖЯ и речи, является обеспечение равноправной коммуникации слышащих и глухих людей с нарушениями слуха, которых во всем мире насчитываются десятки миллионов человек. Одним из самых эффективных средств обучения и взаимодействия являются мультимедийные компьютерные системы, поэтому создание информационных приложений, способных работать с ЖЯ, является одной из приоритетных задач при работе с глухими и слабослышащими людьми. Особый интерес в этой области представляют системы компьютерного синтеза жестового языка и речи по входному тексту.

Одним из наиболее эффективных вариантов для реализации компьютерных систем синтеза ЖЯ является использование трехмерных анимированных моделей человека (так называемых жестовых аватаров (signing avatar)), которые могут управляться посредством символов жестовой нотации, описывая требуемые конфигурации рук и различные типы движений. Жесты из лексикона в такой системе синтеза представляют собой цепочку символов в выбранной нотации, поэтому словарь может легко модифицироваться и расширяться без использования специального оборудования. Специфика автоматических систем человеко-машинного взаимодействия и коммуникации состоит в том, что ЖЯ и жестовый словарь должны быть определенным образом формализованы, чтобы компьютер мог обрабатывать и синтезировать жесты. Для описания жеста по его визуальным признакам существуют несколько различных систем нотации (например, HamNoSys или Sign Writing), позволяющих зафиксировать описание жеста. Довольно широкое распространение в последнее время в мире (особенно в странах Европы) получила «Гамбургская система нотации» (HamNoSys) [2, 3]. Эта система отличается наибольшей проработанностью инвентаря знаков и пригодна для использования в компьютерных приложениях за счет того, что ее знаки переведены в компьютерную систему кодировки Юникод с соответствующими компьютерными шрифтами. Инвентарь HamNoSys позволяет записать практически любой жест, выполняемый одной или двумя руками, что делает эту систему универсальной и пригодной для записи практически любого ЖЯ мира. Таким образом, наиболее перспективным вариантом использования трехмерных виртуальных аватаров для синтеза ЖЯ является их управление посредством символов нотации жестов, которые описывают требуемые конфигурации рук и различные типы движений. Словарь жестов в такой системе представляет собой цепочку символов в одной системе нотации, поэтому может легко модифицироваться и расширяться без использования специального оборудования.

Используя такой подход, за последние годы на волне создания ассистивных технологий был разработан ряд моделей компьютерного синтеза жестовой речи для нескольких ЖЯ, включая американский, британский, французский, чешский ЖЯ и др. Среди известных зарубежных компьютерных систем синтеза жестовой речи, использующих различные анимированные аватары, следует отметить системы, разработанные в рамках различных проектов в Евросоюзе и США: DePaul ASL Synthesizer

(<http://asl.cs.depaul.edu>), аватары европейских проектов Dicta-Sign (www.dictasign.eu) [4], SIGNSPEAK (www.signspeak.eu/en) [5], SignCom [6], Italian SL [7], ViSiCAST (аватар Visia, www.visicast.co.uk), eSign (аватар vGuido, www.sign-lang.uni-hamburg.de/esign), аватары Sign Smith и Sign4Me компании Vcom3D (www.vcom3d.com), SiSi от IBM (www-03.ibm.com/press/us/en/pressrelease/22316.wss), американскую систему iCommunicator (www.icommunicator.com) и ряд других.

Модель и архитектура многомодальной системы синтеза речи и жестового языка

В ходе исследований авторами была разработана модель универсального многомодального человеко-машинного интерфейса (для вывода мультимедийной информации) и архитектура многомодальной системы синтеза аудиовизуальной речи и ЖЯ по тексту. Концептуальная модель универсального многомодального интерфейса пользователя представлена на рис. 1. Интерфейс выполняет автоматическое преобразование входной текстовой информации T в жестовую G , а также речевую аудиоинформацию A и видеоинформацию V , объединение и вывод ее в виде мультимедийной информации M (при этом преобразование g является функцией автоматической обработки текста, а f выполняет функцию объединения разнотипной информации в мультимедийное представление):

$$T \xrightarrow{g} \langle G, A, V \rangle \xrightarrow{f} M .$$



Рис. 1. Концептуальная модель универсального многомодального человеко-машинного интерфейса на основе синтеза аудиовизуальной речи и жестового языка

Такой пользовательский интерфейс вывода информации является универсальным, так как он предназначен для вывода входных текстовых данных посредством синтеза звучащей речи, артикуляции губ аватара и жестового языка как для обычных пользователей, так и для людей с ограниченными возможностями (глухих и незрячих людей).

На рис. 2 представлена разработанная архитектура многомодальной системы синтеза аудиовизуальной речи и ЖЯ по тексту [8]; ее основными компонентами являются:

- текстовый процессор анализа входного текста для последующего аудиосинтеза звучащей речи (по словам) и видеосинтеза жестовой и дактильной речи (показ фраз по словам или по буквам);
- имитационная трехмерная модель головы человека [9];
- аудиосинтезатор звучащей речи, осуществляющий преобразование текст–речь по входному тексту [10, 11];
- компьютерная система синтеза аудиовизуальной речи (говорящая голова (talking head)) на основе виртуальной объемной модели головы человека и машинного синтеза речи [12, 13];
- имитационная трехмерная модель верхней части тела и рук человека, в которой настраиваются параметры движений рук для синтеза элементов жестового языка на основе управляющих символов жестовой нотации HamNoSys [2, 14];
- многомодальный пользовательский интерфейс, интегрирующий компоненты генерации звучащей, визуальной и жестовой речи по входному тексту [1].

В многомодальной системе артикуляция губ, которые являются видимой частью органов речеобразования, сопровождается также синтезированной речью, которая может и не быть доступна полностью глухим людям, однако для слышащих людей синтез аудиовизуальной речи доступен и даже необходим для повышения разборчивости и естественности синтезируемой компьютером речи.

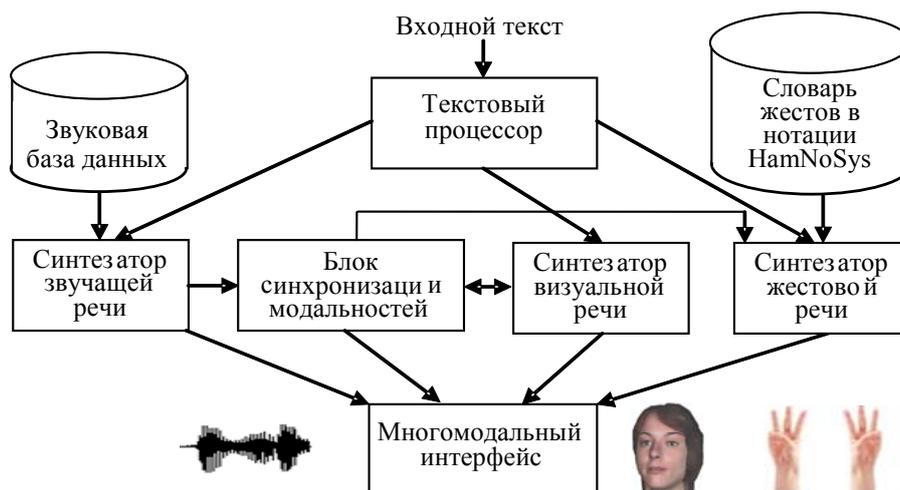


Рис. 2. Архитектура многомодальной системы синтеза аудиовизуальной речи и жестового языка по тексту

Программная реализация двуязычной многомодальной системы синтеза

В качестве практической реализации универсального многомодального интерфейса была программно разработана двуязычная многомодальная система синтеза аудиовизуальной речи и элементов ЖЯ. За основу была взята компьютерная система видеосинтеза чешского ЖЯ [15], созданная ранее Западночешским университетом в рамках исследовательского проекта Musslap (<http://musslap.zcu.cz>), которая была адаптирована к обработке русского ЖЯ. В итоге была разработана не просто система видеосинтеза жестов рук трехмерного аватара, а многомодальная система синтеза, в которой, помимо видеобработки, выполняется также аудиовизуальный синтез звучащей речи (хотя для глухих людей она и не доступна, но необходима для правильного синтеза артикуляции губ и мимики лица).

На вход системы подается произвольный грамматически корректный текст (на русском или чешском языке), который анализируется текстовым процессором, в нем выделяются предложения, слова (для аудиосинтеза речи и видеосинтеза артикуляции губ аватара) и буквы (для машинного синтеза дактильной речи [16]), которые автоматически преобразуются в символы жестовой нотации, на основе которой аватар воспроизводит мануальные жесты, декодируя символы нотации. Элементы жестовой речи в системе описываются при помощи системы записи жестов HamNoSys, отражающей основные дифференциальные признаки каждого жеста: форму кисти, ориентацию руки, место и характер движения. Виртуальная модель головы и верхней части туловища человека реализована на языке моделирования виртуальной реальности (Virtual Reality Modelling Language, VRML) и управляется программно средствами графической библиотеки OpenGL под управлением операционной системы семейства Microsoft Windows. Создан трехмерный жестовый аватар (signing avatar), который может иметь светлую либо черную одежду (в зависимости от предпочтений пользователя). Данный аватар, демонстрирующий финальную статью динамического жеста для числительного «16», показан в различных проекциях на рис. 3.



Рис. 3. Трехмерный жестовый аватар в различных проекциях

Подсистема аудиовизуального синтеза речи (так называемая говорящая голова [9, 17]) производит компиляционный синтез звучащей разговорной речи по тексту, совмещенный с движениями губ и лицевых органов трехмерной виртуальной головы человека. Для синтеза ЖЯ визуальная модель головы человека объединена с моделью туловища и рук (включая пальцы) человека. Как уже говорилось, ЖЯ складывается из комбинации динамических жестов, воспроизводимых обеими руками (либо одной рукой), и артикуляции губ, проговаривающих показываемое слово. При этом многие глухие люди, которые в детстве обладали слухом, способны «читать речь по губам» собеседника даже без использования жестов руками, поэтому такая речевая модальность должна являться неотъемлемым элементом компьютерной системы синтеза жестовой речи.

Также для системы многомодального синтеза разработан специальный метод синхронизации выходных аудио- и видеомодальностей. Синхронизация звучащей речи и жестов в системе осуществляется на основе временных меток начала и конца слов звучащей речи, синтезированной системой по тексту. Так как звучащая речь в среднем имеет более высокий темп воспроизведения, чем жестовая речь, то виртуальный аватар последовательно проговаривает и артикулирует с невысокой скоростью изолированные слова звучащей речи, дожидаясь окончания жестикуляции соответствующего слова (может включать в себя несколько последовательных букв дактильной азбуки), плавно переходя к следующему жесту слитной жестовой речи.

Разработанный жестовый аватар максимально имитирует стиль жестикуляции живых людей. Так, все жесты следуют в речи без пауз с соблюдением «плавности» и «текучести» жестов, что позволяет оформлять целые фразы и лексемы, а не набор изолированных друг от друга жестов. На данный момент словарь жестов системы составляет несколько сотен жестов для наиболее распространенных слов, цифр, букв и т.д.

Следует отметить, что компьютерный синтез ЖЯ с использованием трехмерного аватара обладает рядом достоинств при организации вывода информации пользователям с нарушениями слуха, в частности:

1. позволяет просматривать видеосинтез жестовой речи с разных сторон и углов обзора, что дает возможность лучше воспринимать пространственную информацию, например, степень удаленности рук от туловища и друг относительно друга (в отличие от двухмерных моделей);
2. дает возможность относительно легко пополнять и корректировать словарь жестов, так как вместо видеозаписей реального человека в словаре присутствуют компьютерные анимированные аватары, соответственно, для расширения словаря не обязательно записывать того же самого человека-демонстратора жестов в той же самой одежде и с той же прической, а также уровнем освещенности;
3. позволяет выполнять слитный синтез ЖЯ, в котором отдельные слова во фразах стыкуются бесшовно, т.е. не видны явные границы между соседними словами;
4. дает возможность заменять виртуальный аватар, используя новые высокореалистичные модели людей (мужчин или женщин, а также любых персонажей);
5. позволяет воспроизводить жестовую речь на экране с любой необходимой скоростью, как замедляя, так и ускоряя видеоряд.

Демонстрация и тестирование системы в Санкт-Петербурге были организованы при помощи сотрудников «Всероссийского общества глухих». Отзывы и качественная оценка системы потенциальными пользователями позволяют говорить об обеспечении естественности и разборчивости синтезированных элементов русского ЖЯ и дактильной речи (дактилологии), а также артикуляции и мимики губ виртуального аватара при речеобразовании.

Разработанная многомодальная система аудиовизуального синтеза речи и жестового языка (жестовый аватар) предназначена для организации универсальных человеко-машинных интерфейсов с целью коммуникации с людьми, имеющими тяжелые нарушения слуха и полностью глухими, посредством элементов разговорного ЖЯ (калькирующей жестовой речи и дактильной речи, воспроизводимых движениями/жестами рук виртуального помощника-аватара) и визуальной речи (артикуляции губ, обязательно сопутствующей жестовой модальности), а также речевой коммуникации со слепыми и слабовидящими людьми и полноценного мультимедийного общения со зрячими и слышащими пользователями. Данная система может использоваться для задач организации коммуникации человек-человек и человек-машина [1], в системах электронного обучения [18, 19], машинного перевода [20], виртуальной и дополненной реальности и т.д.

Представленная многомодальная система аудиовизуального синтеза речи и ЖЯ входит в состав большего комплекса – универсальной ассистивной информационной технологии (assistive technology) для людей с ограниченными возможностями здоровья [21], в состав которой также входят многомодальная система для бесконтактной работы с компьютером [22, 23], система аудиовизуального распознавания речи [24], система автоматического распознавания элементов жестового языка [25] и модель ассистивного интеллектуального пространства [26].

Заключение

В работе был представлен универсальный многомодальный интерфейс и разработана программная компьютерная система для аудиовизуального синтеза элементов жестового языка и звучащей речи по тексту, объединяющая бимодальную систему синтеза речи (виртуальная говорящая голова), обеспечивающую аудиовизуальный синтез речи, и модель тела и рук человека (жестовый аватар), выполняющую видеосинтез динамических жестов. Система предназначена для вывода входных текстовых данных посредством синтеза звучащей речи, артикуляции губ аватара и жестового языка как для обычных пользователей, так и для людей с ограниченными возможностями (глухих и незрячих).

Следующим этапом научно-исследовательских работ будет являться разработка системы синтеза разговорного жестового языка и речи по тексту. Ее создание осложняется необходимостью средств машинного перевода текста на жестовый язык, обладающий собственной структурой и грамматикой (отличной от письменного или устного языка), которые пока слабо исследованы лингвистами и недостаточно формализованы, но исследования в данном направлении активно ведутся, что позволяет говорить о скором решении данной проблемы. Дальнейшие исследования, разработки и внедрение в жизнь глухих людей автоматизированных компьютерных систем должны дополнительно привлечь к этой проблеме внимание общественности, а также обратить усилия ученых, разработчиков и компьютерных лингвистов к междисциплинарным исследованиям и работам в этой сфере обработки естественного языка.

Литература

1. Karpov A., Krnoul Z., Zelezny M., Ronzhin A. Multimodal synthesizer for Russian and Czech sign languages and audio-visual speech // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2013. V. 8009 LNCS. Part 1. P. 520–529.
2. Hanke T. HamNoSys – representing sign language data in language resources and language processing contexts // *Proc. International Conference on Language Resources and Evaluation, LREC 2004*. Lisbon, Portugal, 2004. P. 1–6.
3. Карпов А.А., Кагиров И.А. Формализация лексикона системы компьютерного синтеза языка жестов // *Труды СПИИРАН*. 2011. № 1 (16). С. 123–140.
4. Efthimiou E. et al. Sign language technologies and resources of the dicta-sign project // *Proc. 5th Workshop on the Representation and Processing of Sign Languages*. Istanbul, Turkey, 2012. P. 37–44.
5. Caminero J., Rodríguez-Gancedo M., Hernández-Trapote A., López-Mencía B. SIGNSPEAK project tools: a way to improve the communication bridge between signer and hearing communities // *Proc. 5th Workshop on the Representation and Processing of Sign Languages*. Istanbul, Turkey, 2012. P. 1–6.
6. Gibet S., Courty N., Duarte K., Naour T. The SignCom system for data-driven animation of interactive virtual signers: methodology and evaluation // *ACM Transactions on Interactive Intelligent Systems*. 2011. V. 1. N 1. Art. 6.
7. Borgotallo R., Marino C., Piccolo E. et al. A multi-language database for supporting sign language translation and synthesis // *Proc. 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*. Malta, 2010. P. 23–26.
8. Карпов А.А. Компьютерный анализ и синтез русского жестового языка // *Вопросы языкознания*. 2011. № 6. С. 41–53.
9. Železný M., Krňoul Z., Císař P., Matoušek J. Design, implementation and evaluation of the Czech realistic audio-visual speech synthesis // *Signal Processing*. 2006. V. 86. N 12. P. 3657–3673.
10. Tihelka D., Kala J., Matoušek J. Enhancements of viterbi search for fast unit selection synthesis // *Proc. 11th Annual Conference of the International Speech Communication Association, INTERSPEECH-2010*. Makuhari, Japan, 2010. P. 174–177.
11. Hoffmann R., Jokisch O., Lobanov B., Tsurulnik L., Shpilevsky E., Piurkowska B., Ronzhin A., Karpov A. Slavonic TTS and SST conversion for let's fly dialogue system // *Proc. 12th International Conference on Speech and Computer SPECOM-2007*. Moscow, Russia, 2007. P. 729–733.
12. Krňoul Z., Železný M., Müller L. Training of coarticulation models using dominance functions and visual unit selection methods for audio-visual speech synthesis // *Proc. Annual Conference of the International Speech Communication Association INTERSPEECH*. Pittsburgh, USA, 2006. V. 2. P. 585–588.
13. Karpov A., Tsurulnik L., Krňoul Z., Ronzhin A., Lobanov B., Železný M. Audio-visual speech asynchrony modeling in a talking head // *Proc. Annual Conference of the International Speech Communication Association INTERSPEECH*. Brighton, UK, 2009. P. 2911–2914.
14. Krňoul Z., Železný M. Translation and conversion for Czech sign speech synthesis // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2007. P. 524–531.
15. Krňoul Z., Kanis J., Železný M., Müller L. Czech text-to-sign speech synthesizer // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2008. V. 4892 LNCS. P. 180–191.

16. Карпов А.А. Машинный синтез русской дактильной речи по тексту // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2013. № 1. С. 20–26.
17. Карпов А.А., Цирульник Л.И., Железны М. Разработка компьютерной системы «говорящая голова» для аудиовизуального синтеза русской речи по тексту // Информационные технологии. 2010. № 8. С. 13–18.
18. Borgia F., Bianchini C.S., De Marsico M. Towards improving the e-learning experience for deaf students: e-LUX // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2014. V. 8514 LNCS. Part 2. P. 221–232.
19. Тампель И.Б., Краснова Е.В., Панова Е.А., Левин К.Е., Петрова О.С. Использование информационно-коммуникационных технологий в электронном обучении иностранным языкам // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 2 (84). С. 154–160.
20. Hruz M., Campr P., Dikici E. et. al. Automatic fingersign to speech translation system // Journal on Multimodal User Interfaces. 2011. V. 4. N 2. P. 61–79.
21. Karpov A., Ronzhin A. A universal assistive technology with multimodal input and multimedia output interfaces // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2014. V. 8513 LNCS. Part 1. P. 369–378.
22. Карпов А.А. ICanDo: Интеллектуальный помощник для пользователей с ограниченными физическими возможностями // Вестник компьютерных и информационных технологий. 2007. № 7. С. 32–41.
23. Karpov A., Ronzhin A., Kipyatkova I. An assistive bi-modal user interface integrating multi-channel speech recognition and computer vision // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2011. V. 6762. Part 2. P. 454–463.
24. Karpov A., Ronzhin A., Markov K., Zelezny M. Viseme-dependent weight optimization for CHMM-based audio-visual speech recognition // Proc. 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010. Makuhari, Japan, 2010. P. 2678–2681.
25. Kindiroglu A., Yalcin H., Aran O., Hruz M., Campr P., Akarun L., Karpov A. Automatic recognition of fingerspelling gestures in multiple languages for a communication interface for the disabled // Pattern Recognition and Image Analysis. 2012. V. 22. N 4. P. 527–536.
26. Карпов А.А., Акарун Л., Ронжин Ал.Л. Многомодальные ассистивные системы для интеллектуально-го жилого пространства // Труды СПИИРАН. 2011. № 4 (19). С. 48–64.

Карпов Алексей Анатольевич	– доктор технических наук, доцент, ведущий научный сотрудник, Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), 199178, Санкт-Петербург, Россия; профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия
Железны Милош	– PhD, доцент, доцент, Западночешский Университет, 306 14, Пльзень, Чехия, zelezny@kky.zcu.cz
Alexey A. Karpov	D.Sc., Associate professor, leading scientific researcher, Saint Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 199178, Saint Petersburg, Russia; Professor, ITMO University, 197101, Saint Petersburg, Russia, karpov@iias.spb.su
Milos Zelezny	PhD, Associate professor, Associate professor, University of West Bohemia, 306 14 Pilsen, Czech Republic, zelezny@kky.zcu.cz

*Принято к печати 11.07.14
Accepted 11.07.14*

УДК 004.056

СОВЕРШЕНСТВОВАНИЕ POLICE OFFICE MODEL ДЛЯ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ РОЕВЫХ РОБОТОТЕХНИЧЕСКИХ СИСТЕМ

И.А. Зикратов^а, А.В. Гуртов^{б, в, с}, Т.В. Зикратова^д, Е.В. Козлова^а

^а Университет ИТМО, 197101, Санкт-Петербург, Россия, zikratov@cit.itmo.ru

^б Хельсинский институт информационных технологий, 00014, University of Helsinki, Finland

^с Аалто Университет, FI-00076, Aalto, Finland

^д Военный институт (Военно-морской политехнический) ВУНЦ ВМФ «Военно-морская академия», 196602, г. Пушкин, Россия

Аннотация. Рассматриваются аспекты информационной безопасности групповых мобильных робототехнических комплексов с роевым интеллектом. Обсуждаются способы осуществления скрытых атак противоборствующей стороны на роевой алгоритм. Выполнено численное моделирование возможных деструктивных информационных воздействий на муравьиный алгоритм поиска кратчайшего пути. Приведена демонстрация последствий атак на муравьиный алгоритм при различной концентрации в рое роботов-диверсантов. Предложены подходы к обеспечению информационной безопасности в роевых робототехнических системах, основанные на реализации принципов централизованного управления безопасностью мобильных агентов. Разработан метод формирования самоорганизующейся системы управления информационной безопасностью роботов-агентов в роевых коллективах, реализующий Police Office Model – модель обеспечения безопасности на основе полицейских участков, которая применяется для обеспечения информационной безопасности мультиагентных систем. В основу метода положено использование в узлах графа сети полицейских участков, на которые возложены функции идентификации и аутентификации агентов, а также выявления диверсантов как по их формальным признакам, так и по их поведению в рое. Предложен перечень программно-аппаратных компонентов полицейских участков, состоящий из каналов связи между роботами-полицейскими, реестра узлов, базы данных агентов и механизмов шифрования и дешифрования. Предложены варианты логики функционирования механизма информационной безопасности роевых систем, отличающиеся временными диаграммами обмена данными между полицейскими участками. Представлен сравнительный анализ реализаций защищенных роевых систем в зависимости от логики функционирования полицейских участков, интегрированный в роевую систему. Показано, что модель безопасности сохраняет способность функционирования в условиях помех при длительности помехи, сопоставимой со временем преодоления агентом пути между полицейскими участками.

Ключевые слова: информационная безопасность, робототехнический комплекс, рой роботов, муравьиный алгоритм, групповая робототехника, модель Police Office Model, защищенный рой, уязвимость, атака.

POLICE OFFICE MODEL IMPROVEMENT FOR SECURITY OF SWARM ROBOTIC SYSTEMS

I.A. Zikratov^а, A.V. Gurtov^{б, в, с}, T.V. Zikratova^д, E.V. Kozlova^а

^а ITMO University, 197101, Saint Petersburg, Russia, zikratov@cit.itmo.ru

^б Helsinki Institute for Information Technology (HIIT), 00014 Helsinki, Finland

^с Aalto University, FI-00076 Aalto, Finland

^д Military Institute (Naval Polytechnic) Military Educational and Scientific Center of the Navy "Naval Academy", 196602, Pushkin, Saint Petersburg, Russia

Abstract. This paper focuses on aspects of information security for group of mobile robotic systems with swarm intellect. The ways for hidden attacks realization by the opposing party on swarm algorithm are discussed. We have fulfilled numerical modeling of potentially destructive information influence on the ant shortest path algorithm. We have demonstrated the consequences of attacks on the ant algorithm with different concentration in a swarm of subversive robots. Approaches are suggested for information security mechanisms in swarm robotic systems, based on the principles of centralized security management for mobile agents. We have developed the method of forming a self-organizing information security management system for robotic agents in swarm groups implementing POM (Police Office Model – a security model based on police offices), to provide information security in multi-agent systems. The method is based on the usage of police station network in the graph nodes, which have functions of identification and authentication of agents, identifying subversive robots by both their formal characteristics and their behavior in the swarm. We have suggested a list of software and hardware components for police stations, consisting of: communication channels between the robots in police office, nodes register, a database of robotic agents, a database of encryption and decryption module. We have suggested the variants of logic for the mechanism of information security in swarm systems with different temporary diagrams of data communication between police stations. We present comparative analysis of implementation of protected swarm systems depending on the functioning logic of police offices, integrated in swarm system. It is shown that the security model saves the ability to operate in noisy environments, when the duration of the interference is comparable to the time necessary for the agent to overcome the path between police stations.

Keywords: IT security, robotic complex, swarm of robots, ant algorithm, group robotics, Police Office Model, protected swarm, vulnerability, attack.

Введение

В работе исследуются вопросы применимости механизма безопасности мультиагентных информационных систем (МАС), основанного на модели полицейских участков – Police Office Model (POM) – для решения задач информационной безопасности (ИБ) в мультиагентных робототехнических системах

(МРТС) с роевым интеллектом. В отличие от механизмов жесткой безопасности, таких как шифрование канала связи, схемы криптографической аутентификации и авторизации, политики для предоставления полномочий, использование цифровых подписей и сертификатов и т.д., метод РОМ относят к механизмам мягкой безопасности [1], которые противостоят вредоносным информационным воздействиям, осуществляемым со стороны роботов-диверсантов. Под вредоносным информационным воздействием (атакой) будем понимать деятельность робота-диверсанта, направленную на реализацию угрозы ИБ в отношении роботов-агентов и осуществляемую с использованием информационных средств и технологий, в результате которой выбранное агентами новое действие не будет способствовать приращению целевого функционала МРТС в имеющихся условиях [2]. Иначе говоря, механизмы мягкой безопасности позволяют дать ответ на вопрос: а действительно ли робот-агент, получивший доступ в систему, выполняет действия, направленные на решение стоящих перед МАС (МРТС) задач?

Наиболее распространенными механизмами мягкой безопасности в МАС являются метод защищенных состояний агентов [3], метод Ксюдонга [4], «товарищеская» модель безопасности (Buddy Security Model, BSM) [5, 6], которые хорошо согласуются с принципами построения децентрализованных систем. Условно эти механизмы можно разделить на две группы – основанные на принципах децентрализованного и централизованного управления ИБ соответственно.

В методах первой группы функции ИБ возлагаются на агентов МАС. При этом выявление действий злоумышленника осуществляется в процессе интенсивных межагентных коммуникаций, что предъявляет относительно высокие требования к аппаратной составляющей МАС и приводит соответственно к удорожанию всей системы.

При централизованном управлении решение задач ИБ возлагается на специально созданные структуры, которые в некоторых моделях называются полицейскими участками, Police Office (PO). Суть централизованных механизмов ИБ в МАС состоит в том, что мультиагентная система разбивается на несколько областей (участков), в каждом из которых имеется модуль, осуществляющий функции идентификации и аутентификации агентов, а также анализирующий их деятельность. Очевидно, что правила выделения областей и МАС, а также порядок взаимодействия членов коллектива с РО будут иметь особенности для различных роевых алгоритмов.

Цель настоящей работы – разработка механизма ИБ роевых робототехнических систем, предназначенных для выявления и нейтрализации угроз, связанных с осуществлением скрытых атак на рой. В данной работе предложена модель, основанная на модели Ксюдонга, предназначенная для выявления роботов-диверсантов в роевых робототехнических системах. Показано использование модели РОМ в МРТС на основе численного моделирования широко известного классического муравьиного алгоритма поиска кратчайшего пути [7, 8]. Понимание сути модели РОМ позволит легко реализовать этот механизм в других популярных роевых алгоритмах – пчелином, рое частиц и т.д.

Атаки на муравьиный алгоритм

Муравьиный алгоритм – метаэвристический алгоритм, в котором колония искусственных муравьев в кооперации находит хорошее решение сложных дискретных задач оптимизации. Используя этот алгоритм при выборе направления следования, агенты не только выбирают кратчайший путь, но и ориентируются на опыт предшественников, оставляющих за собой на пути особый фермент – феромон. В работе [9] приведены описание и результаты моделирования атак на муравьиный алгоритм в задаче выбора кратчайшего пути, а также выполнен анализ их уязвимостей (рис. 1).

Исследуемым параметром являлась вероятность выбора P агентом кратчайшего маршрута для следующих исходных данных. N роботов, образующих рой, последовательно перемещается из узла i в узел o в условиях агрессивной среды (при атаках противоборствующей стороны). Количество агентов в рое $N=2000$, одновременно на участке находятся 50 агентов. Цель роя – выбрать кратчайший путь следования из узла i в узел o (маршрут А, рис. 1). Целью противоборствующей стороны является создание условий для увеличения времени на поиск оптимального маршрута или для выбора неоптимального маршрута (маршрут В, С или D, рис. 1), но который рой будет принимать за оптимальный. Выполнение замысла злоумышленника приведет к увеличению расхода энергии агентами, что, в свою очередь, ограничит их радиус действия и (или) будет способствовать невыполнению группировкой конечной задачи.

Несмотря на очевидную простоту примера, ему присущи типовые элементы более сложных графов, таких как наличие нескольких вершин и ребер, однозначно определяемый кратчайший (маршрут А) и альтернативные маршруты (маршруты В, С и D), типовые действия роя при выборе кратчайшего пути, что позволяет моделировать атаки на уязвимости классического муравьиного алгоритма Ant System.

Для моделирования использовался полигон моделей Kilobot в среде V-REP [10, 11]. Моделировались действия роя в штатном режиме и при атаке путем «внедрения» одного или нескольких роботов.

В штатном режиме изменения вероятностей выбора маршрутов А, В, С и D при последовательном прохождении N роботами – членами роя представлены на рис. 2. На рисунке видно, что процесс выбора пути условно можно разбить на две части – переходный процесс, когда количество отложенного феромо-

на на ребрах в значительной степени носит случайный характер (в данной реализации он закончился примерно при прохождении узла двухсотым роботом), и установившийся режим, когда длинные пути «исчезают» вследствие все большей привлекательности кратчайшего пути (более высокой концентрации феромона на маршруте А), а вероятность выбора кратчайшего пути асимптотически приближается к 1. Так, на рис. 2 видно, что при прохождении участка последним агентом, вероятность выбора им маршрута А составляет 0,93.

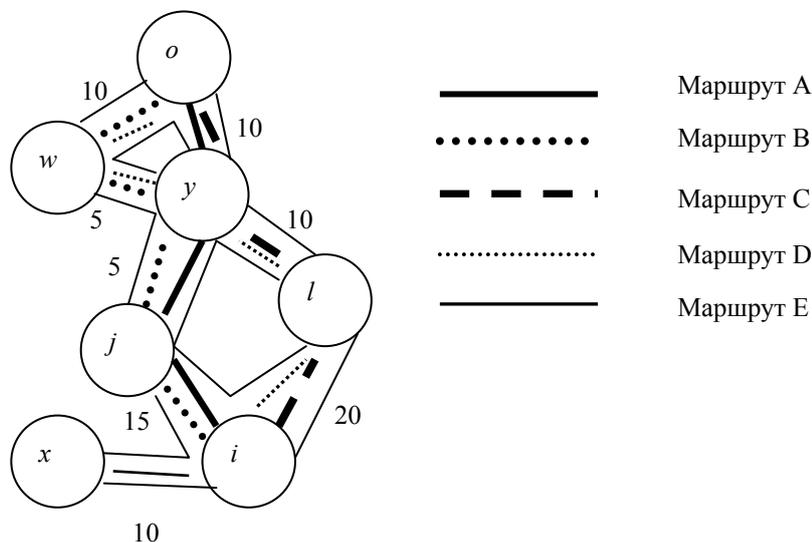


Рис. 1. Муравьиный алгоритм в задаче выбора кратчайшего пути (маршрут А) при наличии иных путей (маршруты В,С,Д) и ложного маршрута (маршрут Е)

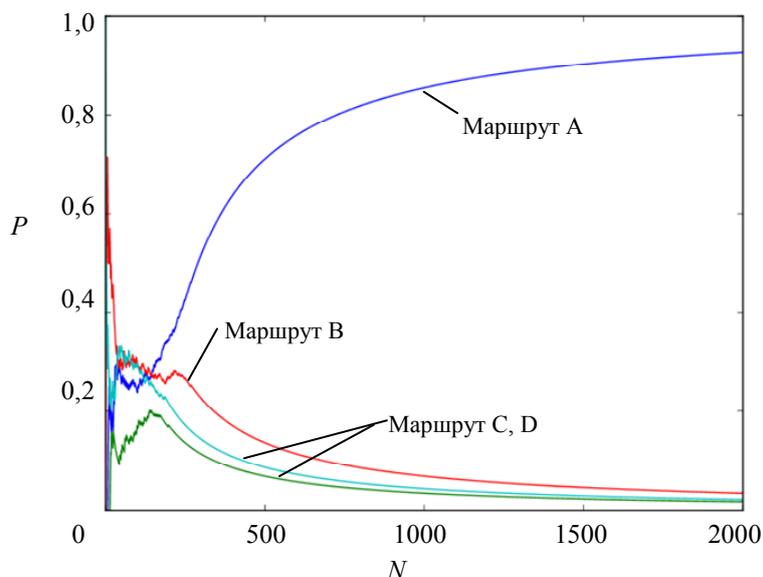


Рис. 2. Зависимость вероятности P выбора роем рационального пути от количества N роботов, прошедших по ребру без воздействия злоумышленников

При реализации атаки злоумышленником внедряются роботы, являющиеся для данной системы «инородными». Эти роботы начинают свой путь по нерациональному для роя маршруту – ребру (i, l) . В зависимости от технических возможностей злоумышленника и от количества времени, в течение которого требуется направить систему по ложному пути, количество роботов-диверсантов может быть различным. Целью фланирования роботов-диверсантов по маршруту является создание более высокой концентрации феромона на ребре (i, l) .

На рис. 3 представлены графики, иллюстрирующие влияние роботов-диверсантов, действующих на ребрах (i, l) , (l, y) , (y, w) и (w, o) (рис. 1), при различных значениях их концентрации по отношению к числу агентов, находящихся на рассматриваемом участке. Так, при наличии на «длинных» ребрах пяти диверсантов (концентрация 1:10, рис. 3, а) видно, что переходный процесс становится более продолжительным, а вероятность выбора маршрута А последним агентом снизилась до 0,59. Увеличение концентрации диверсантов на маршруте приводит практически к равновероятному выбору кратчайшего и иных

маршрутов (рис. 3, б), а уже при концентрации 1:2 возможно снижение вероятности выбора кратчайшего маршрута до пренебрежимо малого значения (рис. 3, в).

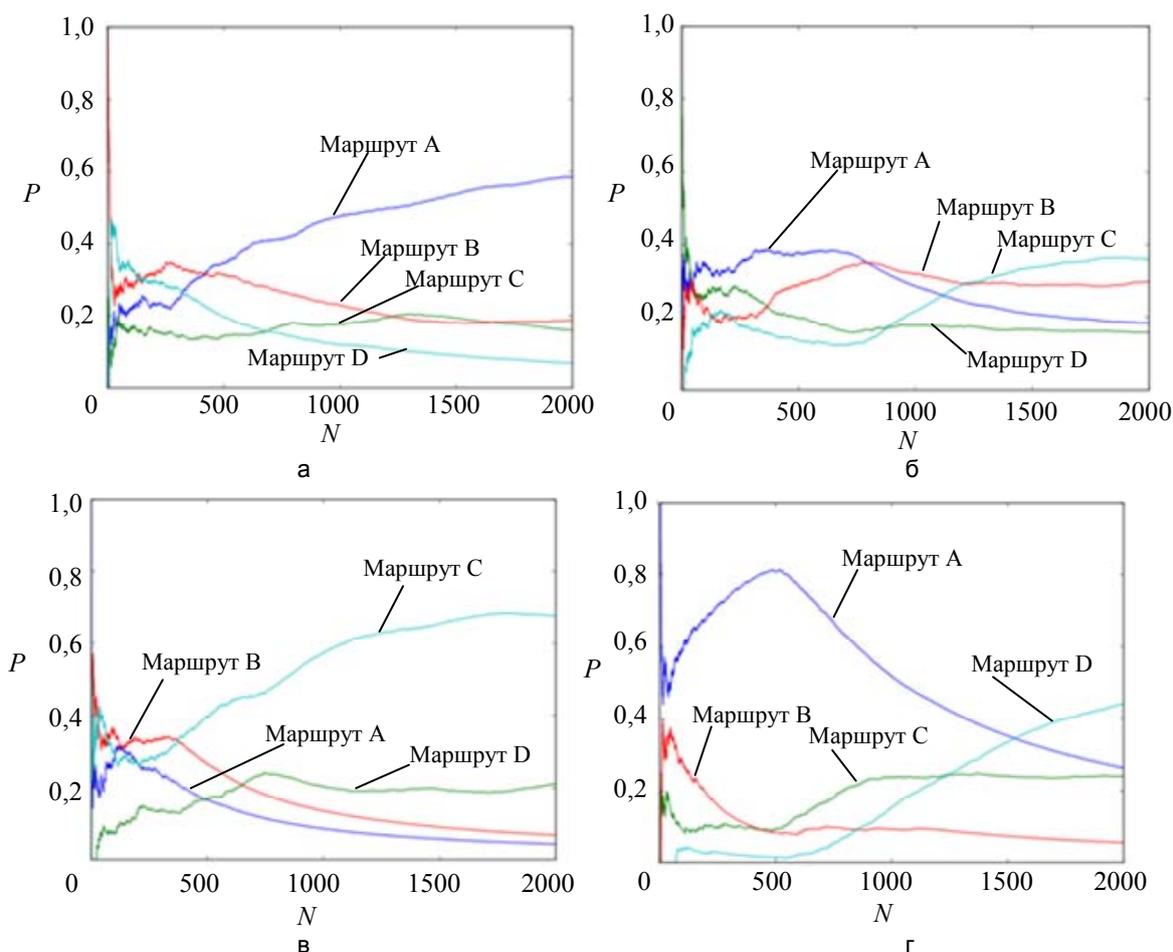


Рис. 3. Значения вероятности P выбора роём из N агентов рационального пути при действиях диверсантов на этапе поиска кратчайшего пути (примеры реализации) при разной концентрации диверсантов в составе роёв: 1:10 (а); 3:10 (б); 1:2 (в); 1:2. Внедрение диверсантов в установившемся режиме (г)

Следует обратить внимание, что внедрение диверсантов в установившемся режиме также приводит к резкому снижению эффективности алгоритма. Так, на реализации, представленной на рис. 3, г, видно, что инъекция в роё 25 диверсантов в установившемся режиме (после прохождения участка пятьюстами агентами) привело к тому, что вероятность выбора кратчайшего маршрута стала меньше вероятности выбора агентами нерационального маршрута С.

Из представленных результатов видно, что уязвимости, присущие муравьиному алгоритму, позволили осуществить деструктивное воздействие на роё путем информационной воздействия через сенсоры агентов без атаки на физическую и (или) программную компоненту робота. Физическое внедрение в роё роботов-диверсантов способно оказать дестабилизирующее влияние на процесс движения группировки как на этапе переходного процесса (в процессе поиска кратчайшего пути), так и в установившемся режиме, когда кратчайший путь движения уже найден. Следовательно, механизмы обеспечения ИБ должны осуществляться непрерывно на всех этапах выполнения задачи, быть масштабируемыми по пространству и обеспечивать устойчивость функционирования во времени.

Метод формирования системы информационной безопасности роботов-агентов в роевых коллективах, реализующий РОМ

Очевидно, что для предотвращения указанных выше атак роевые робототехнические комплексы должны оснащаться механизмами ИБ, отвечающими некоторым специфическим требованиям по сравнению с аналогичными механизмами защиты МАС [12–17]. Для обеспечения ИБ МАС систем различают два подхода к реализации таких механизмов – централизованный и децентрализованный. В частности, децентрализованный подход основан на принципе обеспечения взаимной безопасности и представляет собой такую систему безопасности, в которой агенты отвечают за безопасность друг друга, отслеживая происходящие в системе события и взаимодействуя между собой и внешней средой. В ходе межагентных

коммуникаций агенты системы обмениваются по каналам связи информацией о своем состоянии и выполняемых действиях, а также специальными сообщениями, которые несут в себе секретную информацию о состояниях известных им агентов и о возможных угрозах с их стороны либо со стороны узлов сети. Таким образом, все агенты системы получают информацию о потенциальных угрозах их безопасности. Информирова своих соседей о возможной опасности (например, при появлении «чужого» агента в системе), каждый из агентов несет ответственность за безопасность своего окружения и всей системы в целом. Привлекательность этой модели состоит в отсутствии какого-либо единого центра безопасности, что приводит к невозможности разрушения модели. Однако применение подобных систем для роевых коллективов, не оснащенных каналами связи, может быть сопряжено с трудностями технического характера и удорожанием системы.

В настоящей работе рассмотрены варианты построения самоорганизующегося механизма ИБ роевых систем, основанного на модели РО и предназначенного для предотвращения скрытых атак на роевые алгоритмы со стороны роботов-диверсантов. Самоорганизация в данном случае заключается в автоматическом формировании полицейских участков в узлах графа маршрутов (рис. 1) в результате работы муравьиного алгоритма. Централизация предусматривает освобождение роботов-агентов от функций ИБ и возложение этих функций на интегрированные в роевую модель специальные аппаратно-программные комплексы, реализующие функции РО и обеспечивающие поддержку межагентного взаимодействия и самоорганизации роя.

В состав каждого РО входят:

- агенты-полицейские по числу выходящих из узла (вершины графа) маршрутов (дуг);
- реестр узлов;
- база данных роботов-агентов;
- модуль шифрования данных (при необходимости).

Соседние РО связаны между собой каналами связи. Логика функционирования такой системы может быть основана на известной логике функционирования системы безопасности мобильных агентов информационных систем, предложенной в работе [18]. Применительно к рассматриваемой модели защищенного роя логика взаимодействия роботов-агентов с полицейскими участками РО может быть следующей.

Способ 1. Пусть некоторый робот-агент R_k , находящийся в узле i , выбрал согласно описанному выше алгоритму путь движения в направлении узла j . Агент R_k , отправляет запрос $Z1$ агенту-полицейскому AP_i участка PO_i на разрешение осуществления миграции на соответствующий узел. Агент-полицейский AP_i выполняет запрос, проверяет в своем реестре узлов существование узла j и в случае положительного ответа формирует и выдает роботу-агенту R_k уникальный сертификат, который содержит идентификатор агента, остаток энергоресурса, информацию о точке отправления и выбранном маршруте и время выдачи сертификата. При необходимости осуществляется шифрование данных уходящего робота-агента. После этого агент-полицейский дает разрешение на миграцию R_k . Прибыв в узел j , робот-агент R_k предъявляет агенту-полицейскому AP_j этого узла свой сертификат. Агент-полицейский AP_j участка PO_j на основе информации, содержащейся в предъявленном сертификате, осуществляет проверку в своем реестре узлов существования в системе узла i и робота-агента R_k в удостоверяющем центре полицейских участков. Если информация подтверждается, то AP_j обращается к агенту-полицейскому AP_i участка PO_i , с которого мигрировал агент R_k , который и выполняет функции удостоверяющего центра роботов-агентов, с запросом на подтверждение существования агента R_k и того факта, что ему было разрешено мигрировать на узел j . Если агент AP_i подтверждает факт существования и миграции агента R_k на узел j , агент-полицейский AP_j осуществляет расчеты временных и энергетических параметров движения робота-агента из узла i в узел j с целью проверки соответствия:

- фактического времени прихода агента и остатка энергоресурса расчетному времени прибытия и остатку энергоресурса, которые могут быть вычислены исходя из времени выхода из узла убытия и расстояния до него, а также величины остатка энергии в момент убытия робота из узла i ;
- соответствия фактического пути следования и пути, заявленного в сертификате.

При отсутствии противоречивых данных агент-полицейский AP_j заносит в свою базу данных агентов робота-агента R_k и предоставляет ему доступ к ресурсам узла, необходимым для решения стоящей перед роем задачи. Например, это может быть информация о длине путей, выходящих из узла, которая потребуется для реализации алгоритма поиска кратчайшего пути. При необходимости на участке PO_i производится дешифрование данных прибывшего агента.

Если AP_j не получил соответствующего подтверждения о существовании агента R_k в системе или обнаружено несоответствие по одному из пунктов проверки (фактические изменения остатка энергии, скорость и направление следования не соответствуют заявленным в сертификате значениям), то агент R_k блокируется, а доступ к ресурсам узла для него запрещается. Агент-полицейский AP_j в этом случае заносит агента R_k в «черный список» и информирует о присутствии «инородного» агента в системе всех ему известных полицейских участков. Аналогичные действия агент-полицейский AP_j осуществляет, если робот-агент R_k в качестве дальнейшего пути следования выбирает тот маршрут, по которому он прибыл в узел j , так как это служит признаком осуществления атаки.

Для предотвращения атак типа «ложный путь» [9] на агентов-полицейских возлагается функция блокирования выбора роботами-агентами тех маршрутов, на которых прибывающие на участок роботы-агенты имеют сертификат, выданный этим же полицейским участком.

Оценка производительности защищенного роевого алгоритма

Оценку производительности защищенного роевого алгоритма произведем путем сравнения времени выполнения задачи защищенным роем и незащищенным роем [18]. Для этого рассмотрим особенности работы систем на этапе выбора дуги маршрута, на этапе перемещения по дуге и на этапе прибытия в узел маршрута.

Способ 1. Этап выбора пути следования. Для незащищенной системы работное время T_1^H робота-агента будет состоять из времени T_{RV} – времени, потребного для вычисления кратчайшего маршрута согласно муравьиному алгоритму. Для защищенной системы добавляется время переговоров с агентом-полицейским – T_{RAP} , время работы агента-полицейского с реестром узлов – T_{APU} , время формирования сертификата для робота-агента – T_{RU} , и, возможно, шифрования данных агента – T_{KR} . Тогда для защищенной системы работное время агента на первом этапе будет равно

$$T_1^3 = T_{RV} + T_{RAP} + T_{APU} + T_{RU} + T_{KR}. \quad (1)$$

Этап движения по пути следования. На данном этапе работное время незащищенной и защищенной роевых систем совпадает и равно времени перемещения робота агента из узла i в узел j :

$$T_2^H = T_2^3 = T_{ij}. \quad (2)$$

Этап прибытия на узел назначения. По прибытии в вершину графа робот-агент приступает к выбору дальнейшего пути следования. Следовательно, работное время робота-агента T_3^H для незащищенного алгоритма в узле будет состоять из времени T_{RV} .

Для защищенной системы временные затраты будут определяться следующим соотношением:

$$T_3^3 = T_{RV} + T_{RAP} + T_{DKR} + T_{APU} + T_{UU} + T_{CalcR} + T_{BD} + T_{AR}, \quad (3)$$

где T_{RAP} – время переговоров с агентом-полицейским; T_{DKR} – время дешифрования данных (при необходимости); T_{APU} – время обращения агента-полицейского к реестру узлов; T_{UU} – время обращения к удостоверяющему центру (полицейскому участку узла убывтия робота-агента); T_{CalcR} – время, затрачиваемое на расчет параметров движения робота-агента по данным сертификата; T_{BD} – время обращения к базе данных агентов; T_{AR} – время принятия решения в отношении робота-агента по результатам аутентификации.

Очевидно, что общее время работы зависит как от сложности маршрута и потребного количества узлов размещения РО соответственно, так и от количества агентов в рое. Однако из формул (1)–(3) следует, что наибольшие затраты времени в защищенном роевом алгоритме падают на проведение процедур, связанных с вопросами обеспечения ИБ в узле прибытия робота-агента: T_{UU} , T_{CalcR} , T_{BD} и T_{AR} . Вместе с тем следует учесть, что время, затрачиваемое роботом-агентом на преодоление расстояния между узлами T_{ij} , всегда превышает величину $T_{UU} + T_{CalcR} + T_{BD} + T_{AR}$.

Исходя из этого, можно предложить следующую логику функционирования РО в составе защищенного роя.

Способ 2. Первый этап – выбор маршрута. Пусть некоторый робот-агент R_k , находящийся в узле i , выбрал путь движения в направлении узла j . Агент R_k отправляет запрос агенту-полицейскому AP_i участка PO_i на разрешение миграции на соответствующий узел. Агент-полицейский AP_i проверяет в своем реестре узлов и базе данных агентов существование узла j и уникального идентификатора робота-агента R_k соответственно. После этого агент-полицейский дает разрешение на миграцию R_k . Время, необходимое для выполнения этих процедур, равно

$$T_1^3 = T_{RV} + T_{RAP} + T_{APU} + T_{BD}.$$

Второй этап – движение агента по маршруту. Во время физического перемещения робота-агента из узла в узел на полицейских участках узла убытия PO_i и узла PO_j прибытия робота-агента выполняются следующие процедуры.

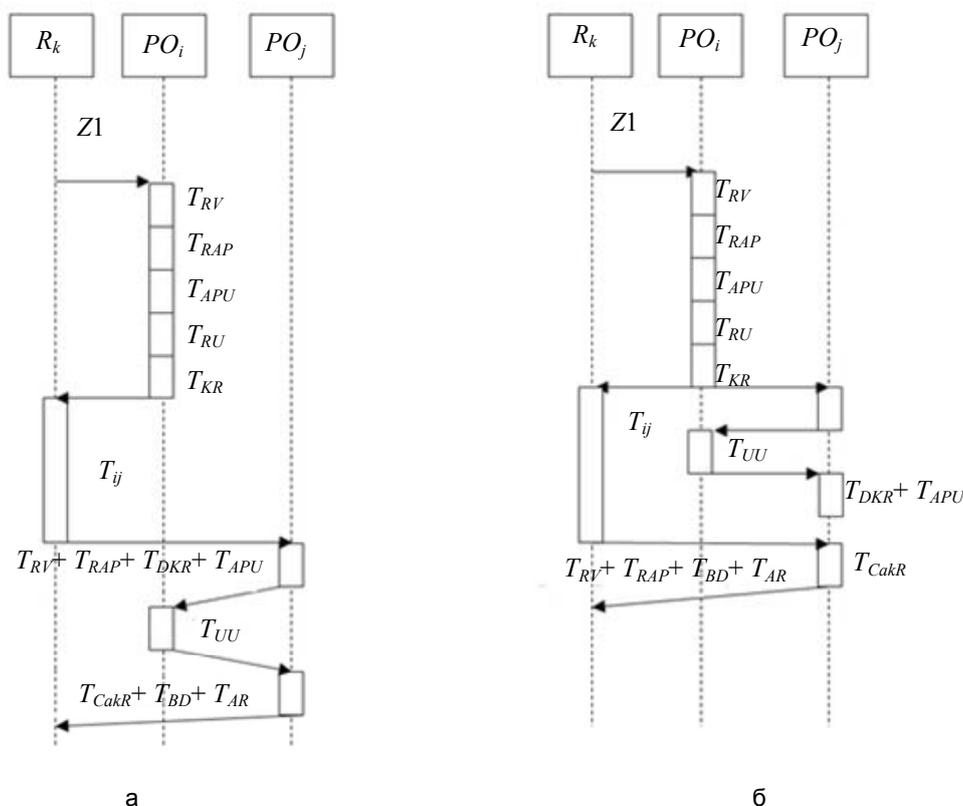


Рис. 4. Диаграмма взаимодействия основных компонентов POM: организация взаимодействия при способе 1 (а); организация взаимодействия при способе 2 (б)

Агент-полицейский AP_i составляет сертификат убывшего робота-агента R_k , в котором содержится информация об остатке энергоресурса, о точке отправления, выбранном маршруте и времени убытия R_k . После этого осуществляется шифрование сертификата и передача его по каналам связи на полицейский участок узла назначения робота R_k . Время, потребное для выполнения этих действий, составляет

$$T_{RU} + T_{KR} + T_{UU} \ll T_{ij}.$$

Агент-полицейский AP_j участка PO_j получает сертификат, дешифрует его и на основе информации, содержащейся в сертификате, осуществляет проверку в своем реестре узлов существования в системе узла i . При положительном результате агент-полицейский AP_j осуществляет расчеты временных и энергетических параметров движения робота-агента из узла i в узел j с целью прогнозирования времени прихода агента и остатка энергоресурса, которые могут быть вычислены исходя из времени выхода из узла убытия и расстояния до него. Полученные результаты расчетов, а также идентификатор и сертификат агента R_k агент-полицейский AP_j вносит в базу данных агентов участка PO_i .

Время работы PO_j по обработке данных ожидаемого агента R_k составляет

$$T_{UU} + T_{DKR} + T_{APU} + T_{CalcR} \ll T_{ij}.$$

Таким образом, на данном этапе работное время защищенной системы равно времени перемещения робота агента из узла i в узел j :

$$T_2^3 = T_{ij}.$$

Третий этап – прибытие в вершину графа. На третьем этапе, прибыв в узел j , робот-агент R_k предъявляет агенту-полицейскому AP_j этого узла свой идентификатор. Агент-полицейский AP_j проверяет наличие идентификатора в базе данных агентов своего участка PO_j . Если таковой имеется, то AP_j сверяет фактическое время прибытия робота, его маршрут следования и остаток энергоресурса с данны-

ми, имеющимися в базе данных участка PO_j в отношении агента R_k . При отсутствии противоречивых сведений AP_j предоставляет агенту R_k доступ к ресурсам узла, необходимым для решения стоящей перед роем задачи. При несоответствии сведений агент-полицейский AP_j осуществляет в отношении робота R_k процедуры блокировки, описанные ранее.

Таким образом, рабочее время на третьем этапе составляет (рис. 4)

$$T_3^3 = T_{RV} + T_{RAP} + T_{AR} + T_{BD},$$

что на величину $T_{DKR} + T_{APU} + T_{UU} + T_{CalcR}$ меньше, чем при реализации защищенного метода обеспечения ИБ роевой системы по способу 1 (формула (3)).

Следует учесть, что при выполнении роем задач на пересеченной местности, а также в условиях воздействия дестабилизирующих факторов естественного и искусственного происхождения, существенную роль приобретают вопросы помехоустойчивости группировки.

Анализ помехоустойчивости защищенного роевого алгоритма

Рассмотрим различия в работе систем, организованных по описанным выше способам при воздействии помехи на каналы связи между РО.

Пусть путь от i -го к j -му участку состоит из n элементарных дистанций (рис. 5). Предположим, что при перемещении агента по этому пути в произвольный момент времени на каналы связи между полицейскими участками PO_i и PO_j воздействовала помеха. Предположим также, что за время воздействия помехи робот-агент преодолел m элементарных дистанций, которые на рисунке выделены желтым цветом. На рис. 5 проиллюстрированы возможные ситуации воздействия помехи на каналы связи между полицейскими участками во время преодоления агентом элементарных дистанций, которые отличаются временем начала постановки помех t_1, t_2, \dots, t_n . Например, при варианте № 3 (рис. 5) воздействие помехи наблюдалось во время передвижения роботом-агентом по дистанциям 2 и 3. Из рисунка видно, что при использовании способа 1 неблагоприятными событиями являются варианты, когда помеха воздействует на каналы связи во время прибытия робота-агента в n -й участок, в котором расположен полицейский участок PO_j (варианты $n+m-2$ и $n+m-1$ рис. 5). В соответствии с диаграммой (рис. 4, а) для выполнения процедур безопасности, доступа R_k к ресурсам системы и выдачи квитанции агенту на продолжение движения PO_j должен установить связь с PO_i , что будет возможно только после прекращения воздействия помехи на каналы связи.

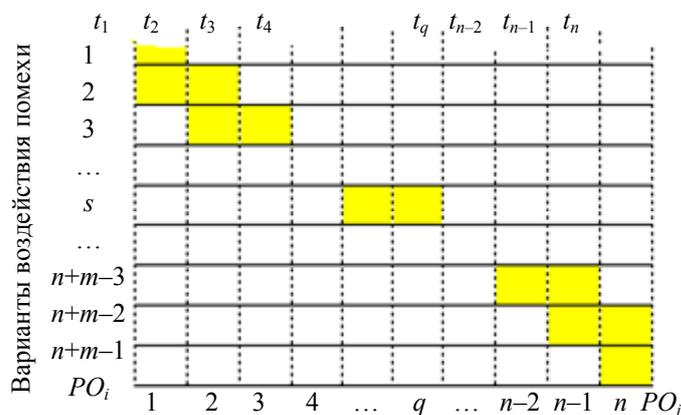


Рис. 5. Варианты воздействия помехи на каналы связи между полицейскими участками во время передвижения робота-агента между i -м и j -м РО при $m=2$

Исходя из этих рассуждений, вероятность отказа РО в обслуживании агента при организации полицейских участков по способу 1 будет равна

$$P_1 = \frac{m}{n + m - 1}. \tag{4}$$

Очевидно, что при использовании способа 2 ситуации, показанные на рис. 5, не представляют сложности для работы системы, так как PO_i начинает устанавливать связь между полицейскими участками сразу после выхода R_k к пункту назначения PO_j . Как видно из рисунка, при таком способе организации взаимодействия неблагоприятное событие (неустановление связи между полицейскими участками)

возможно только при длительности помехи, превышающей время агента в пути ($m > n$), и вероятность его наступления будет равна

$$P_2 = \frac{m-n+1}{n+m-1}. \quad (5)$$

Несложно показать преимущества взаимодействия между РО, организованного по способу 2, по сравнению с организацией по способу 1. Обозначим $k = \frac{m}{n}$. Тогда формулы (4) и (5) примут следующий вид:

$$P_1 = \frac{m}{n+m-1} = \frac{kn}{n(k+1)-1},$$

$$P_2 = \frac{m-n+1}{n+m-1} = \frac{n(k-1)+1}{n(k+1)-1}.$$

Задавшись константой n , построим графики зависимости вероятности отказа РО в зависимости от отношения k (рис. 6).

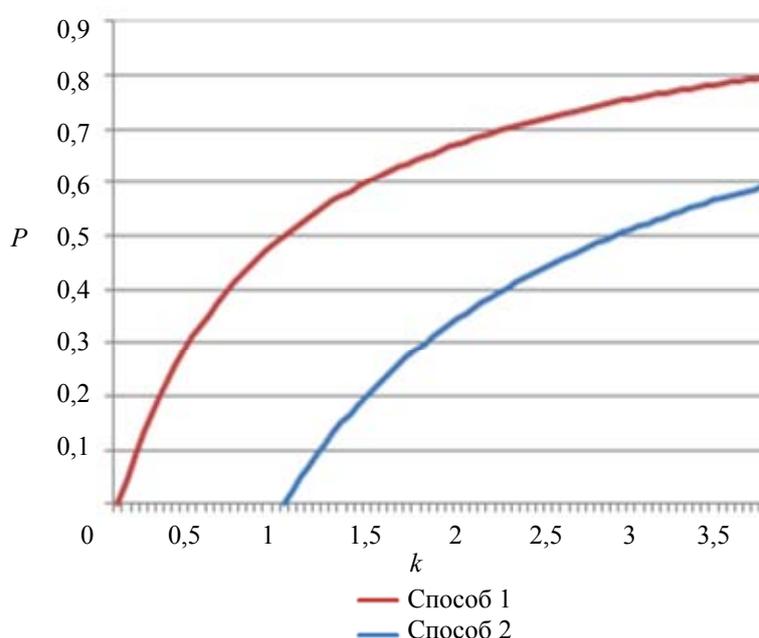


Рис. 6. Зависимость вероятности отказа РО в обслуживании агента от длительности воздействия помехи на канал связи между полицейскими участками

Из рис. 6 видно, что при реализации РОМ способом 2 помеха не влияет на работу защищенного алгоритма при выполнении условия $m < n$. При увеличении времени воздействия помехи, когда $m \gg n$, кривые асимптотически сближаются и приближаются к 1, что свидетельствует о необходимости использования средств радиоэлектронной борьбы в сложной помеховой обстановке.

При моделировании на полигоне Kilobot в среде V-REP предложенная модель безопасности позволила обнаружить и нейтрализовать скрытые атаки на рой со стороны роботов-диверсантов, в том числе в условиях воздействия на каналы связи активных помех. Моделирование скрытых атак на муравьиный алгоритм при наличии системы РОМ показало, что значения вероятности P выбора роем из N агентов рационального пути при действиях диверсантов на этапе поиска кратчайшего пути при разной концентрации диверсантов стремятся к 1 (рис. 2) при различных концентрациях роботов-диверсантов.

Недостатком предлагаемого метода, как указывалось выше, является уязвимость МРТС при выводе из строя полицейских участков. Очевидно, что при практической реализации защищенных роевых алгоритмов будет необходимо использование дополнительных систем контроля их работоспособности.

Заключение

В работе проанализированы виды скрытых атак на робототехнические системы с роевым интеллектом. Рассмотрены современные подходы, ориентированные на решение задач, связанных с обеспечением информационной безопасности агентов робототехнических мультиагентных систем.

В ходе проведенных исследований были получены следующие основные результаты.

1. Выполнено численное моделирование возможных деструктивных информационных воздействий на муравьиный алгоритм поиска кратчайшего пути. Продемонстрированы атаки на муравьиный алгоритм при различной концентрации в рое роботов-диверсантов. Показана зависимость вероятности выбора кратчайшего пути от различной концентрации диверсантов в рое.
2. Обоснованы требования к мягким механизмам обеспечения информационной безопасности роевых робототехнических систем. Рассмотрены достоинства и недостатки подходов к обеспечению информационной безопасности в роевых робототехнических системах, основанных на реализации х централизованного и децентрализованного управления безопасностью мобильных агентов.
3. Разработан метод формирования самоорганизующейся системы управления информационной безопасностью роботов-агентов в роевых коллективах, реализующий модель полицейских участков, которая применяется для обеспечения информационной безопасности мультиагентных систем. Метод заключается в интеграции в коллектив роботов программно-аппаратных компонентов, выполняющих функции выявления и пресечения скрытых атак на роевые алгоритмы.
4. Проведен анализ логики функционирования системы управления информационной безопасностью роботов-агентов в роевых коллективах. Представлен сравнительный анализ реализаций защищенных роевых систем в зависимости от логики функционирования полицейских участков, интегрированных в роевую систему.
5. Проведен численный эксперимент, имитирующий работу защищенного алгоритма в помеховой обстановке.

Литература

1. Губанов Д.А. Обзор онлайн-систем репутации/доверия [Электронный ресурс]. 2009. Режим доступа: http://mtas.ru/bitrix/components/bitrix/forum.interface/show_file.php?fid=1671, свободный. Яз. рус. (дата обращения 24.03.2014).
2. Зикратов А.А., Зикратова Т.В., Лебедев И.С. Доверительная модель информационной безопасности мультиагентных робототехнических систем с децентрализованным управлением // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 2 (90). С. 47–52.
3. Neeran K.M., Tripathi A.R. Security in the Ajanta MobileAgent system. Technical Report. Department of Computer Science, University of Minnesota, 1999. 28 p.
4. Xudong G., Yiling Ya., Yinyuan Y. POM-a mobile agent security model against malicious hosts // Proc. 4th International Conference on High Performance Computing in the Asia-Pacific Region. 2000. V. 2. P. 1165–1166.
5. Page J., Zaslavsky A., Indrawan M. A buddy model of security for mobile agent communities operating in pervasive scenarios // Proc. 2nd Australasian Information Security Workshop (AISW2004). ACS, Dunedin, New Zealand, 2004. V. 32. P. 17–25.
6. Page J., Zaslavsky A., Indrawan M. Countering security vulnerabilities using a shared security buddy model schema in mobile agent communities // Proc. of the First International Workshop on Safety and Security in Multi-Agent Systems (SASEMAS 2004), 2004. P. 85–101.
7. Dorigo M., Maniezzo V., Colomi A. Ant system: optimization by a colony of cooperating agents // IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. 1996. V. 26. N 1. P. 29–41.
8. Wooldridge M. Introduction to MultiAgent Systems. John Wiley & Sons Ltd, 2002. 368 p.
9. Зикратов И.А., Козлова Е.В., Зикратова Т.В. Анализ уязвимостей робототехнических комплексов с роевым интеллектом // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 5 (87). С. 149–154.
10. Комаров И.И., Дранник А.Л., Юрьева Р.А. Моделирование проблем информационной безопасности мультиагентных систем // В мире научных открытий. 2014. № 4 (52). С. 61–70.
11. Дранник А.Л. Использование программ-симуляторов поведения роевых робототехнических систем для исследования вопросов безопасности // Материалы VIII Санкт-Петербургской межрегиональной конференции «Информационная безопасность регионов России (ИБРР-2013)». Санкт-Петербург, 2013. С. 240–251.
12. Schillo M., Funk P., Rovatsos M. Using trust for detecting deceitful agents in artificial societies // Applied Artificial Intelligence. 2000. V. 14. N 8. P. 825–848.
13. Golbeck J., Parsia B., Hendler J. Trust networks on the semantic web // Lecture Notes in Artificial Intelligence. 2003. V. 2782. P. 238–249.
14. Garcia-Morchon O., Kuptsov D., Gurtov A., Wehrle K. Cooperative security in distributed networks // Computer Communications. 2013. V. 36. N 12. P. 1284–1297.
15. Бешта А.А., Кирпо М.А. Построение модели доверия к объектам автоматизированной информационной системы для предотвращения деструктивных воздействий на систему // Известия Томского политехнического университета. 2013. Т. 322. № 5. С. 104–108.

16. Ramchurn S.D., Huynh D., Jennings N.R. Trust in multi-agent systems // Knowledge Engineering Review. 2004. V. 19. N 1. P. 1–25.
17. Gorodetski V., Kotenko I., Karsaev O. Multi-agent technologies for computer network security: Attack simulation, intrusion detection and intrusion detection learning // Computer systems science and engineering. 2003. N 4. P. 191–200.
18. Маслобоев А.В., Путилов В.А. Разработка и реализация механизмов управления информационной безопасностью мобильных агентов в распределенных мультиагентных информационных системах // Вестник МГТУ. 2010. Т. 13. № 4–2. С. 1015–1032.

- | | |
|---------------------------------------|---|
| <i>Зикратов Игорь Алексеевич</i> | – доктор технических наук, профессор, заведующий кафедрой, Университет ИТМО, 197101, Санкт-Петербург, Россия, zikratov@cit.itmo.ru |
| <i>Гуртов Андрей Валерьевич</i> | – PhD, главный научный сотрудник, Хельсинкский институт Информационных Технологий University of Helsinki P.O. Box 33 (Yliopistonkatu 4) 00014, University of Helsinki Finland; адъюнкт-профессор, Аалто Университет, P.O. Box 11000, FI-00076 Aalto, Finland, mailto:gurtov@hiit.fi |
| <i>Зикратова Татьяна Викторовна</i> | – преподаватель, Военный институт (военно-морской политехнический) ВУНЦ ВМФ «Военно-морская академия», 196602, г. Пушкин, Россия, ztv64@mail.ru |
| <i>Козлова Екатерина Владимировна</i> | – студент, Университет ИТМО, 197101, Санкт-Петербург, Россия, kekvlad@gmail.com |
| <i>Igor A. Zikratov</i> | – D.Sc., Professor, Department head, ITMO University, 197101, Saint Petersburg, Russia, zikratov@cit.itmo.ru |
| <i>Andrei V. Gurtov</i> | – PhD, Principal Scientist, Helsinki Institute for Information Technology (HIIT), 00014 Helsinki, Finland; Adjunct Professor, Aalto University, FI-00076 Aalto, Finland, gurtov@hiit.fi |
| <i>Tatyana V. Zikratova</i> | – tutor, Military Institute (Naval Polytechnic) Military Educational and Scientific Center of the Navy "Naval Academy", 196602, Pushkin, Saint Petersburg, Russia, ztv64@mail.ru |
| <i>Ekaterina V. Kozlova</i> | – student, ITMO University, 197101, Saint Petersburg, Russia, kekvlad@gmail.com |

Принято к печати 08.05.14
Accepted 08.05.14

УДК 004.312.2, 004.021, 510.649

ИСПОЛЬЗОВАНИЕ РЕЛЯЦИОННОЙ ТЕОРИИ ПРИ ОПТИМАЛЬНОМ ПРОЕКТИРОВАНИИ ИНТЕГРАЛЬНЫХ СХЕМ

Д.В. Демидов^{a, b}^a Университет ИТМО, 197101, Санкт-Петербург, Россия^b ЗАО «Интел А/О», 196247, Санкт-Петербург, Россия, Daniil.demidov@gmail.com

Аннотация. Предложен подход к адаптации реляционной теории для решения задач систем автоматизированного проектирования интегральных схем. Разработан алгоритм оптимального поиска неявных Don't Care (безразличных) значений. Алгоритм описан в терминах адаптированной теории, что позволило получить простое описание алгоритма как для неформального понимания, так и для формального анализа. Предложенный подход позволяет использовать позитивный опыт реляционных баз данных по эффективной (в том числе распределенной) реализации операций реляционной алгебры.

Проведен сравнительный анализ предложенного алгоритма и классического алгоритма оптимального поиска неявных Don't Care значений. В ходе сравнительного анализа формально доказана корректность предложенного алгоритма. Показано, что предложенный алгоритм, по сравнению с классическим, имеет значительно меньшую асимптотическую сложность в худшем случае.

Поиск неявных Don't Care значений в процессе проектирования интегральных схем позволяет повысить такие качества схем, как занимаемая площадь кристалла, энергопотребление, верифицируемость и надежность. Однако классический алгоритм оптимального поиска неявных Don't Care значений не применяется на практике ввиду его чрезмерно высокой вычислительной сложности. Применение алгоритмов субоптимального поиска не позволяет в полной мере реализовать возможности оптимизации интегральных схем. Реализация предложенного в работе алгоритма в системах автоматизированного проектирования интегральных схем целесообразна в силу значительно меньшей вычислительной сложности, потенциально позволяет улучшить соотношение скорости процесса разработки и качества получаемых интегральных схем в аспектах занимаемой площади кристалла, энергопотребления, верифицируемости и надежности. Предложенный подход делает возможным создание распределенной системы автоматизированного проектирования интегральных схем, что позволит повысить доступную системе вычислительную мощность и автоматизировать проектирование более сложных интегральных схем соответственно.

Ключевые слова: САПР, интегральные схемы, оптимизация комбинационных схем, логические сети, частично определенные булевы функции, неявные Don't Care значения, ODC, реляционная теория, реляционная алгебра, естественное соединение.

RELATIONAL THEORY APPLICATION FOR OPTIMAL DESIGN OF INTEGRATED CIRCUITS

D.V. Demidov^{a, b}^a ITMO University, 197101, Saint Petersburg, Russia^b "Intel A/O", Ltd., 196247, Saint Petersburg, Russia, Daniil.demidov@gmail.com

Abstract. This paper deals with a method of relational theory adaptation for integrated circuits CAD systems. A new algorithm is worked out for optimal search of implicit Don't Care values for combinational multiple-level digital circuits. The algorithm is described in terms of the adapted relational theory that gives the possibility for a very simple algorithm description for both intuitive understanding and formal analysis. The proposed method makes it possible to apply progressive experience of relational databases in efficient implementation of relational algebra operations (including distributed ones).

Comparative analysis of the proposed algorithm and a classic one for optimal search of implicit Don't Cares is carried out. The analysis has proved formal correctness of the proposed algorithm and its considerably less worst-case complexity.

The search of implicit Don't Care values in the integrated circuits design makes it easier to optimize such characteristics of IC as chip area, power, verifiability and reliability. However, the classic algorithm for optimal search of implicit Don't Care values is not used in practice due to its very high computational complexity. Application of algorithms for sub-optimal search doesn't give the possibility to realize the potential of IC optimization to the full. Implementation of the proposed algorithm in IC CAD (a.k.a., EDA) systems is adequate due to much lower computational complexity, and potentially makes it possible to improve the quality-development time ratio of IC (chip area, power, verifiability and reliability). Developed method gives the possibility for creation of distributed EDA system with higher computational power and, consequently, for design automation of more complex IC.

Keywords: CAD, EDA, VLSI, optimization of combinational schemes, Boolean networks, partially defined Boolean functions, implicit Don't Cares, ODC, relational theory, relational algebra, natural join.

Введение

В силу возрастающей сложности и миниатюризации, а также ускоряющихся темпов производства электронных устройств возрастают требования к занимаемой площади кристалла, энергопотреблению, верифицируемости, надежности и скорости разработки интегральных схем (ИС). Неотъемлемой частью процесса проектирования ИС является решение задачи оптимизации логических схем, причем противоречивые ограничения требуют создания многоуровневых логических схем. Однако поиск оптимальной многоуровневой логической схемы неприемлемо сложен, поэтому на практике решается задача поиска субоптимальной схемы, а многие исследования направлены на разработку методов и эвристик, позволяющих снизить сложность решаемой задачи и приблизить результат к оптимальному [1–5].

Одним из ключевых моментов при оптимизации многоуровневых логических схем является поиск так называемых неявных Don't Care (безразличных или DC) значений, который, с одной стороны, позволяет существенно приблизиться к оптимальному решению, но, с другой стороны, заключает в себе значительную часть сложности исходной задачи [6].

Автором было замечено, что используемые при решении этой задачи структуры данных по своей семантике весьма близки к отношениям, используемым в теории реляционных баз данных. Алгоритмы работы с такими структурами хорошо проработаны как теоретически, так и практически, что потенциально позволяет сократить сложность поиска DC-значений, а также разработать распределенную версию алгоритма.

В работе предложен алгоритм оптимального поиска неявных DC-значений, описанный в терминах реляционной теории, адаптированной к задачам систем автоматизированного проектирования интегральных схем (САПР ИС), и доказана корректность последнего. Показано, что для предложенного алгоритма вычислительная сложность в худшем случае значительно ниже, чем для классического (экспоненциальная сложность против более чем факториальной).

Возможно внедрение результатов предлагаемой работы в существующие САПР ИС, что позволит ускорить (соответственно удешевить) и повысить качество разработки ИС (в том числе сократить занимаемую площадь кристалла и энергопотребление, повысить верифицируемость и надежность ИС [7–12]). На основе предложенного подхода к адаптации реляционной теории возможно создание распределенной САПР ИС, что позволит автоматизировать проектирование более сложных ИС, требующее высокой вычислительной мощности.

Объект исследования

В работе исследуется алгоритм оптимального поиска неявных DC-значений, применяемый на одном из этапов оптимизации многоуровневых комбинационных схем – неотъемлемой части проектирования ИС. В силу противоречивых требований к ИС используются многоуровневые комбинационные схемы, позволяющие производить минимизацию как по площади (занимаемой схемой), так и по задержке (при которой схема работает корректно). Математически многоуровневые комбинационные схемы представляют в виде логических сетей [13, 14] – ациклических ориентированных графов, которые подробно описаны в следующем разделе.

Процесс оптимизации многоуровневой комбинационной схемы является итерационным [1–5, 14, 15]. Основные этапы данного процесса показаны на рис. 1.

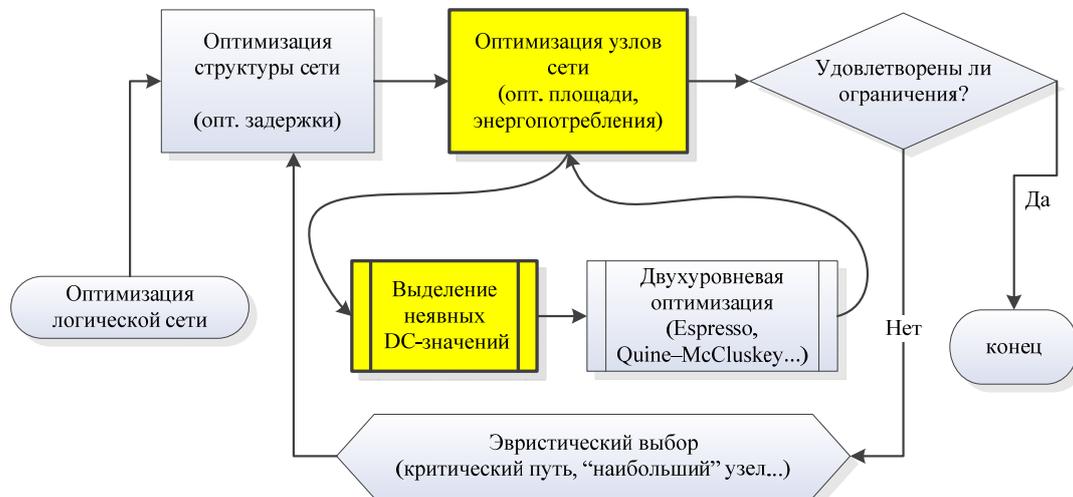


Рис. 1. Основные этапы процесса оптимизации многоуровневых комбинационных схем. Этапы, на которых используются исследуемые алгоритмы, выделены цветом

На этапе оптимизации структуры сети изменяется структура графа, представляющего логическую сеть (объединяются или разделяются узлы), и процесс, в основном, направлен на минимизацию задержки схемы.

На этапе оптимизации узлов сети структура графа остается неизменной, а изменяется внутреннее представление узлов, и процесс направлен на минимизацию площади (занимаемой схемой) и косвенно энергопотребления. Данный этап разделяется на два подэтапа – выделение неявных DC-значений и двухуровневую оптимизацию. На подэтапе двухуровневой оптимизации происходит собственно оптимизация внутреннего представления узла с применением точных (например, метод Квайна–Мак-Класки) или эвристических (например, используемых в семействе программ Espresso) алгоритмов [13–16]. Однако, если не выполнять предшествующий подэтап выделения неявных DC-значений, данный подэтап оперирует

только информацией о внутреннем представлении узла, не учитывая структуру сети, и ограничен в возможностях оптимизации. Кроме того, выполнение подэтапа выделения неявных DC-значений позволяет упростить генерацию тестов единичного отказа, а соответственно, повышает верифицируемость и надежность конечной ИС [7–12].

Представление многоуровневых комбинационных схем

Как упоминалось выше, многоуровневую комбинационную схему представляют в виде логической сети – ориентированного ациклического графа, в котором:

- выделена специальная вершина «Primary Inputs», имеющая только исходящие дуги, помеченные литералами, обозначающими входы схемы;
- выделена специальная вершина «Primary Outputs», имеющая только входящие дуги, помеченные литералами, обозначающими выходы схемы;
- прочие вершины помечены булевыми функциями общего вида, представленными в некоторой канонической форме;
- дуги между прочими вершинами помечены литералами, обозначающими внутренние переменные – выходные значения и аргументы для булевых функций, помечающих начальные вершины и конечные вершины дуги соответственно.

Способы представления булевой функции $f: B^n \rightarrow B, B = \{0,1\}$ в узле можно разделить на два больших класса: бинарные диаграммы решений (БДР или BDD) и структуры, задающие дизъюнктивную или конъюнктивную нормальную форму (ДНФ или КНФ) функции (например, Positional Cube Notation). В настоящей работе рассматриваются способы, задающие ДНФ функции. Семантически такие способы задают множество множеств литералов (аргументов функции), взятых в положительной или отрицательной форме:

$$S \in 2^{L \times B}, \langle S \rangle = \bigvee_{s \in S(l,n) \in s} \langle l \rangle \oplus n,$$

где $L \subset \{a, b, c, \dots\}$ – множество литералов; s – терм ДНФ; $\langle l \rangle$ – значение аргумента обозначаемого литералом l .

В действительности для задаваемой функций обычно существует такое множество DC_f значений аргументов, что значение выхода функции неважно. Примерами причин существования подобных входных значений являются следующие ситуации:

- некоторые сочетания входных значений могут быть запрещенными и, при корректной работе окружения схемы, никогда не будут поданы на ее входы (например, незаданные коды операций);
- в соответствии с некоторым протоколом при определенных условиях (например, схема находится в состоянии сброса или определенном режиме) некоторые выходы схемы могут быть проигнорированы окружением схемы.

Если множество DC_f не пусто, говорят, что функция f определена частично. Математически частично определенная булева функция f задается парой (f_{ON}, f_{OFF}) обычных (полностью определенных) булевых функций при выполнении условия $\forall \mathbf{x}. \overline{f_{ON}(\mathbf{x}) \wedge f_{OFF}(\mathbf{x})}$, где f_{ON} – ON-set, определяющий значения аргументов \mathbf{x} , при которых $f(\mathbf{x})$ должна равняться логической единице, а f_{OFF} – OFF-set, определяющий значения аргументов \mathbf{x} , при которых $f(\mathbf{x})$ должна равняться логическому нулю. Тогда можно определить функцию $f_{DC}(\mathbf{x}) = \overline{f_{ON}(\mathbf{x}) \wedge f_{OFF}(\mathbf{x})}$, такую, что $DC_f = \{\mathbf{x}: f_{DC}(\mathbf{x})\}$. Также можно сказать, что частично определенная функция задает множество функций $F = \left\{ f: (f_{ON}(\mathbf{x}) \rightarrow f(\mathbf{x})) \wedge (f_{OFF}(\mathbf{x}) \rightarrow \overline{f(\mathbf{x})}) \right\}$, среди которых производится выбор наилучшей альтернативы на подэтапе двухуровневой оптимизации.

Неявные Don't Care значения. Классический алгоритм поиска

Для каждого узла f сети множество DC-значений определяется как объединение пяти подмножеств: $DC_f = XCDC_f \cup CDC_f \cup SDC_f \cup ODC_f \cup XODC_f$ (рис. 2). Здесь $XCDC_f$ и $XODC_f$ – так называемые внешние (external) DC-значения, которые задаются инженером для всей схемы (в явном виде или выделяются из описания схемы на уровне регистровых передач (register transfer layer, RTL)). Примеры причин, по которым возникают $XCDC_f$ и $XODC_f$ соответственно, были приведены в предыдущем разделе. Оставшиеся три множества – SDC_f , CDC_f и ODC_f – называют внутренними или неявными

DC-значениями и вычисляют для каждого узла, исходя из структуры сети. Рассмотрим подробнее, по каким причинам возникает и как вычисляется каждое из этих множеств.

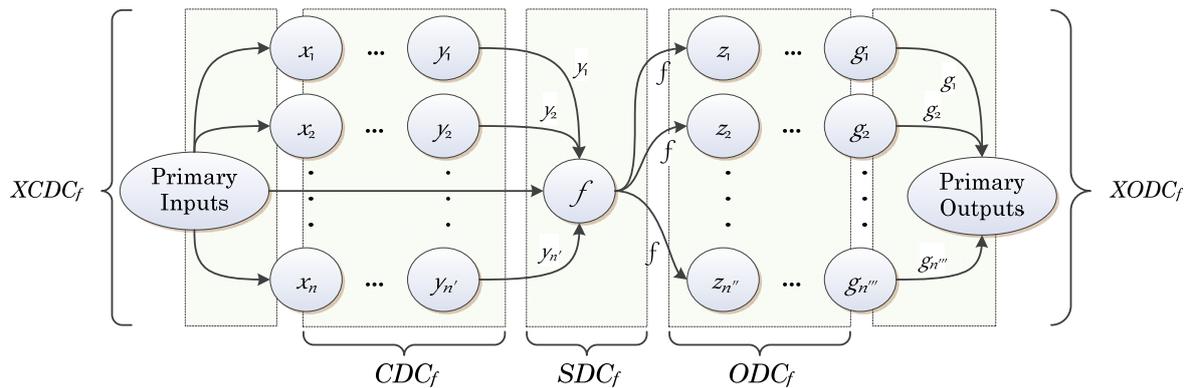


Рис. 2. Подмножества множества DC-значений для узла f логической сети

SDC_f – Satisfiability Don't Cares, определяют для узла f комбинации значений его входных и выходных переменных (аргументов и собственно значений функции), которые являются запрещенными (никогда не наблюдаются при корректной работе узла) в силу зависимости выходного значения от входного. Данное подмножество DC-значений не использует информации о структуре сети, но используется для вычисления прочих DC-значений. Вычисляются $SDC_f = \{(f, \mathbf{x}) : f_{SDC}(f, \mathbf{x})\}$ следующим образом:

$$f_{SDC}(f, \mathbf{x}) = f(\mathbf{x}) \oplus f. \tag{1}$$

CDC_f – Controlability Don't Cares, определяют для узла f комбинации значений его входных переменных, которые являются запрещенными (никогда не наблюдаются при корректной работе данного узла, а также узлов, от которых он зависит) в силу имеющихся зависимостей между ними. Вычисляются $CDC_f = \{\mathbf{x} : f_{CDC}(\mathbf{x})\}$ на основе определенных выше SDC_f :

$$f_{CDC}(\mathbf{x}) = \forall z \notin \text{inputs}(f). \bigvee_{g \in \text{inputs}^*(f)} g_{SDC}, \tag{2}$$

где $\text{inputs}(f)$ – множество входных переменных узла (аргументов функции) f , $\text{inputs}^*(f) = \text{inputs}(f) \cup \bigcup_{g \in \text{inputs}(f)} \text{inputs}^*(g)$ – транзитивное замыкание последних, а выражение вида

$$\forall x. y(x, \mathbf{z}) \text{ вычисляется как } y|_x \wedge y|_{\bar{x}} = y(1, \mathbf{z}) \wedge y(0, \mathbf{z}).$$

ODC_f – Observability Don't Cares, определяют для узла f комбинации значений его входных переменных (аргументов функции), которые делают выходное значение данного узла (функции) ненаблюдаемым всеми узлами, зависящими от него, т.е. при которых изменение выходного значения узла не повлечет изменения выходного значения какого-либо из узлов, зависящих от данного. Вычисляются $ODC_f = \{\mathbf{x} : f_{ODC}(\mathbf{x})\}$ следующим образом:

$$f_{ODC}(\mathbf{x}) = \forall z \notin \text{inputs}(f). \bigwedge_{h \in \text{outputs}^*(f)} \overline{\left(\frac{\partial h}{\partial f}\right)}, \tag{3}$$

где $\text{outputs}(f)$ — множество конечных узлов, выходных дуг узла f , $\text{outputs}^*(f) = \text{outputs}(f) \cup \bigcup_{g \in \text{outputs}(f)} \text{outputs}^*(g)$ – транзитивное замыкание последних, а $\partial h / \partial f$ обозна-

чает так называемую булеву производную функции h по переменной f .

Булева производная $\partial h / \partial f$ является функцией от $\text{inputs}^*(h) \setminus f$ и возвращает единицу, если изменение значения f приводит к изменению h , и ноль – в противном случае. Если f входит в h в явном виде ($f \in \text{inputs}(h)$), то $\partial h / \partial f$ вычисляется как булева разность: $h|_f \oplus h|_{\bar{f}}$. Если же h зависит от f косвенно (через другие функции), то необходимо применять цепное правило (chaining rule):

$$\frac{\partial h}{\partial f} = \left(\frac{\partial h}{\partial g_1} \cdot \frac{\partial g_1}{\partial f} \oplus \dots \oplus \frac{\partial h}{\partial g_n} \cdot \frac{\partial g_n}{\partial f} \right) \oplus \left(\frac{\partial^2 h}{\partial g_1 \partial g_2} \cdot \frac{\partial g_1}{\partial f} \cdot \frac{\partial g_2}{\partial f} \oplus \frac{\partial^2 h}{\partial g_1 \partial g_3} \cdot \frac{\partial g_1}{\partial f} \cdot \frac{\partial g_3}{\partial f} \oplus \dots \right) \oplus \dots =$$

$$= \bigoplus_{G \subset \text{inputs}(h)} \left(\frac{\partial^{|G|} h}{\prod_{g \in G} \partial g} \cdot \bigwedge_{g \in G} \frac{\partial g}{\partial f} \right).$$

Адаптированная реляционная теория

В данном разделе описан используемый в работе частный случай реляционной теории, адаптированной к задачам САПР ИС, и семантика отношений и некоторых операторов реляционной алгебры в контексте решаемой задачи. Будем называть отношением R пару (h, T) , где $h \in 2^L$, $L = \{a, b, c, \dots\}$ – подмножество литералов, называемое заголовком, и $T = \{t \in B^{|h|}\}$ – множество кортежей размера $|h|$, заданных на расширенном булевом множестве $B' = \{0, 1, *\}$, в которое включено специальное значение $*$, называемое телом. Отношение R можно интерпретировать как некоторую булеву функцию $\langle R \rangle = f : B^{|h|} \rightarrow B$, заданную следующей ДНФ:

$$f(\mathbf{x}) = \bigvee_{t \in R, x \in h} \bigwedge \begin{cases} x, & \text{если } t_x = 1 \\ \bar{x}, & \text{если } t_x = 0 \\ 1, & \text{если } t_x = * \end{cases}.$$

Над отношениями можно производить ряд операций; в данной работе используются проекция $\pi_{h'} R$, селекция $\sigma_{p(h)} R$, переименование $\rho_{x=f(h)} R$, объединение $A \cup B$ и две модификации естественного соединения – $A \bowtie B$ и $A \bowtie^{h'} B$. Первые три операции взяты из реляционной алгебры без каких-либо изменений; остальные операции имеют небольшие изменения, позволяющие естественным образом отобразить их на операции булевой алгебры над соответствующими функциями и, в то же время, сохранить все основные свойства операций.

В классической реляционной алгебре операция объединения разрешена только для отношений с одинаковыми заголовками и возвращает отношение с тем же заголовком и объединением тел, а операция естественного соединения объединяет кортежи двух отношений, если они строго равны по всем общим атрибутам [17–19]. В адаптированной реляционной алгебре разрешается объединять отношения с разными заголовками, при этом значения «недостающих» атрибутов картежей принимаются равными $*$, а операция естественного соединения игнорирует значения атрибутов, равные $*$:

$$A \bowtie B = C \Rightarrow \forall t \in T_A \forall t' \in T_B (\forall x \in h_C (t_x \sim t'_x) \Rightarrow \exists s \in T_C (s_x = \text{select}(t_x, t'_x))),$$

где $x \sim y \Leftrightarrow (x = y) \vee (x = *) \vee (y = *)$, $\text{select}(x, *) = x$ и $\text{select}(*, y) = y$.

Вторая модификация естественного соединения $A \bowtie^{h'} B$, названная ограниченным естественным соединением, вводится аналогичным способом с заменой $x \sim y$ на

$$x \sim^{h'} y = \begin{cases} x \sim y, & \text{если } (x \notin h') \wedge (y \notin h') \\ x = y, & \text{иначе} \end{cases}.$$

В таком случае перечисленные операции можно интерпретировать следующим образом (таблица).

Реляционная алгебра	Булева алгебра
Объединение $A \cup B$	Дизъюнкция $\langle A \rangle \vee \langle B \rangle$
Естественное соединение $A \bowtie B$	Конъюнкция $\langle A \rangle \wedge \langle B \rangle$
Проекция $\pi_{h'} R$	Квантор существования $\exists x. \langle R \rangle = \langle R \rangle _x \vee \langle R \rangle _{\bar{x}}$
Селекция $R_x = \pi_{h' \setminus x-1}(\sigma R)$; $R_{\bar{x}} = \pi_{h' \setminus x-0}(\sigma R)$	Кофактор Шеннона $\langle R \rangle _x; \langle R \rangle _{\bar{x}}$

Таблица. Интерпретация основных операций адаптированной реляционной алгебры

Определим понятие «функционального» заголовка отношения R как $fh_R = \{y \in h_R : R_y \triangleright R_y = [\]\}$, где $[\]$ – пустое отношение (с пустым телом). Тогда можно также интерпретировать отношение R относительно атрибута $y \in fh_R$ как частично определенную функцию $\langle R \rangle_y = (\langle R_y \rangle, \langle R_y^- \rangle)$. Также легко показать, что будут верны следующие тождества:

$$\langle R \rangle = \langle R \rangle_y \oplus y; \quad \overline{\langle R \rangle}_y = \left\langle \rho_{y=y}^- R \right\rangle_y; \quad (4)$$

$$\langle F \triangleright G \rangle_g = \langle G \rangle_g \circ \langle F \rangle_f, \text{ если } f \in fh_F, g \in fh_G, f \in h_G, g \notin h_F. \quad (5)$$

Описание предлагаемого алгоритма поиска неявных DC-значений

Особенность предлагаемого алгоритма по сравнению с классическим заключается в том, что множество DC_f вычисляется неявно. При классическом подходе для каждого узла f явным образом конструируется пара $(f_{ON} = f \wedge \overline{f_{DC}}, f_{DC})$, являющаяся входом для следующего подэтапа в процессе оптимизации сети. В предлагаемом алгоритме для каждого узла f вначале предполагается пара $(f_{ON} = f, f_{OFF} = \overline{f})$ (т.е. предполагается, что DC_f пусто), а затем f_{ON} и f_{OFF} уточняются, и на вход следующего подэтапа передается пара $(f'_{ON} = f \wedge \overline{f_{DC}}, f'_{OFF} = \overline{f} \wedge \overline{f_{DC}})$ (для алгоритмов из семейства Espresso в действительности необходимо передать любые две функции из тройки $(f_{ON}, f_{OFF}, f_{DC})$ [13–15]). Таким образом, DC_f не конструируется в явном виде, что во многом определяет более низкую вычислительную сложность.

Описание предлагаемого алгоритма в терминах адаптированной реляционной теории состоит из трех шагов: первый шаг является подготовительным, второй неявно рассчитывает CDC_f для каждого узла f , а третий – ODC_f :

1. для каждого узла $f = \langle R \rangle$ составить отношение $R' = \rho_{f=1} R \cup \rho_{f=0} \overline{R}$,

$$\text{где } \overline{R} = \bigtriangleright \bigcup_{t \in T_R} \rho_{x=t} [\], \quad \overline{\langle R \rangle} = \langle \overline{R} \rangle;$$

2. рекурсивно, начиная с вершины «Primary Inputs» (от входов схемы к выходам) для всех смежных вершин произвести естественное соединение соответствующих отношений и спроецировать результат на конечную вершину;
3. рекурсивно, начиная с вершины «Primary Outputs» (от выходов схемы к входам) для всех смежных вершин произвести естественное соединение соответствующих отношений, ограниченное по помечаемому дугу литералу, и спроецировать результат на начальную вершину.

Первый шаг в общем случае обладает высокой вычислительной сложностью, но он выполняется только на первой итерации процесса оптимизации логической сети. Кроме того, так как логическая сеть на первой итерации получена из RTL-описания схемы и обычно ее узлы помечены булевыми функциями из ограниченного набора, вычисления первого шага можно табулировать. Таким образом, первый этап не влияет на асимптотическую сложность всего алгоритма в целом, что является важным преимуществом предлагаемого алгоритма.

Доказательство корректности предлагаемого алгоритма

В силу ограничений на объем настоящей работы доказательства приведены в сжатом виде.

Под корректностью предлагаемого алгоритма подразумевается то, что для заданной логической сети он рассчитывает в точности то же множество DC_f , что и классический алгоритм (с той лишь разницей, что предлагаемый алгоритм рассчитывает DC_f неявно, см. предыдущий раздел).

Докажем корректность расчета CDC_f , т.е. докажем, что, согласно (2), для

$$f' = \left\langle \pi_{h_f} \left(F \triangleright \left(\bigtriangleright_{g \in \text{inputs}^*(f)} G \right) \right) \right\rangle_f \text{ верно } f'_{DC} = \forall z \notin \text{inputs}(f). \bigvee_{g \in \text{inputs}^*(f)} g_{SDC}. \quad (6)$$

Рассмотрим отрицание правой части тождества (6) и перепишем его в терминах реляционной алгебры (см. (1), (4) и таблицу):

$$\overline{\forall z \notin \text{inputs}(f). \bigvee_{g \in \text{inputs}^*(f)} g_{SDC}} = \overline{\forall z \notin \text{inputs}(f). \left\langle \bigcup_{g \in \text{inputs}^*(f)} \rho_{\bar{g}} G \right\rangle} = \exists z : z \notin \text{inputs}(f). \left\langle \bigcup_{g \in \text{inputs}^*(f)} \rho_{\bar{g}} G \right\rangle =$$

$$= \left\langle \pi_{h_f} \left(\overline{\bigcup_{g \in \text{inputs}^*(f)} \rho_{\bar{g}} G} \right) \right\rangle = \left\langle \pi_{h_f} \left(\bigotimes_{g \in \text{inputs}^*(f)} \overline{\rho_{\bar{g}} G} \right) \right\rangle = \left\langle \pi_{h_f} \left(\bigotimes_{g \in \text{inputs}^*(f)} G \right) \right\rangle.$$

Теперь рассмотрим

$$(f'_{ON}, f'_{OFF}) = (\langle F \wedge \overline{f'_{DC}}, \overline{\langle F \rangle} \wedge \overline{f'_{DC}} \rangle = \left\langle \rho_{f=1} (F_f \bowtie \overline{DC}) \cup \rho_{f=0} (F_{\bar{f}} \bowtie \overline{DC}) \right\rangle = \langle F \bowtie \overline{DC} \rangle, \text{ где } \langle DC \rangle = f'_{DC},$$

т.е. выполняется тождество $f' = \left\langle F \bowtie \pi_{h_f} \left(\bigotimes_{g \in \text{inputs}^*(f)} G \right) \right\rangle$, что и требовалось доказать.

Докажем корректность расчета ODC_f . Заметим, что если $f \in fh_f$, то в результирующее отношение входят только те картежи из отношения F , для которых атрибут f в отношении G имеет значение, отличное от *, т.е. $\partial g / \partial f = 1$.

Аналогично, если $f \in fh_f$ и $g \in fh_g$, то

$$\left\langle \pi_{h_f} \left(F \bowtie \{f\} G \bowtie \{g\} H \right) \right\rangle_f = \left(\exists z : z \notin \text{inputs}(f). f_{ON} \cdot \frac{\partial g}{\partial f} \cdot \frac{\partial h}{\partial g}, \exists z : z \notin \text{inputs}(f). f_{OFF} \cdot \frac{\partial g}{\partial f} \cdot \frac{\partial h}{\partial g} \right).$$

Учитывая (5), получим, что $F \bowtie \{f\} G_1 \bowtie \{f\} G_2 = F \bowtie \{f\} G$, где

$$\langle G \rangle_g = \langle G_1 \bowtie G_2 \rangle_g = \begin{cases} g_1 \vee g_2, & \text{если } g = \{ \} \\ g_1 \circ g_2, & \text{если } g \subset fh_{G_2}, g = (h_{G_1} \cap h_{G_2}); \\ g_2 \circ g_1, & \text{если } g \subset fh_{G_1} \end{cases}$$

тогда

$$\bigvee_{g \in \text{outputs}^*(f)} \frac{\partial g}{\partial f} = F \bowtie \{f\} \left(G_1 \bowtie \{g_1\} \left(H_{1,1} \bowtie \{h_{1,1}\} \dots \right) \bowtie \{g_1\} \dots \right) \bowtie \{f\} \left(G_2 \bowtie \{g_2\} \left(H_{2,1} \bowtie \{h_{2,1}\} \dots \right) \bowtie \{g_2\} \dots \right) \dots$$

Преобразуем (3) следующим образом:

$$\overline{f_{ODC}} = \overline{\forall z \notin \text{inputs}(f). \bigwedge_{g \in \text{outputs}^*(f)} \left(\frac{\partial g}{\partial f} \right)} =$$

$$= \exists z : z \notin \text{inputs}(f). \bigwedge_{g \in \text{outputs}^*(f)} \left(\frac{\partial g}{\partial f} \right) = \exists z : z \notin \text{inputs}(f). \bigvee_{g \in \text{outputs}^*(f)} \frac{\partial g}{\partial f},$$

тогда для f'' , вычисляемого согласно третьему шагу предлагаемого алгоритма,

$$(f''_{ON}, f''_{OFF}) = (f'_{ON} \cdot \overline{f'_{ODC}}, f'_{OFF} \cdot \overline{f'_{ODC}}), \text{ что доказывает корректность расчета } ODC_f.$$

Сравнение вычислительной сложности предложенного и классического алгоритмов

Произведем оценку асимптотической вычислительной сложности в худшем случае для классического и предложенного алгоритмов. Пусть логическая сеть – граф $G=(V, E)$ – имеет $|V|=m$ вершин, а каждая вершина $v_i, i \in \overline{0, m-1}$ имеет $|\text{inputs}(v_i)|=n_i$ входных дуг (равно аргументов соответствующей булевой функции, равно мощность заголовка соответствующего отношения) и $t_i, t_i \leq 2^{n_i-1}$ термов ДНФ, задающей булеву функцию (равно количество кортежей в соответствующем отношении). Будем обозначать $n = \max_i n_i$.

Рассмотрим классический алгоритм. В сложности $f_{SDC_{v_i}}(n_i)$ расчета SDC_{v_i} (см. (1)) доминирует сложность расчета отрицания $v_i : f_{SDC_{v_i}} \in O(n_i^{t_i}) \subseteq O(n_i^{2^{n_i-1}})$, причем результирующая ДНФ будет иметь

$n_i + 1$ аргументов и до $2 \cdot t_i$ термов. Сложность расчета CDC_{v_i} (см. (2)) в худшем случае складывается из сложности расчета SDC_{v_i} для m вершин и сложности квантора всеобщности по $(m-1) \cdot n - 1 \in O(m \cdot n)$ переменным (что в худшем случае сводится к конъюнкции $2^{m \cdot n - 2}$ ДНФ по два терма): $f_{CDC_{v_i}} \in O\left(m \cdot \binom{n}{2}\right) + O\left(2^{m \cdot n}\right)$. Наихудший случай для расчета ODC_{v_i} достигается при такой конфигурации графа, что $n_i = i$ и максимальная длина пути в графе – m . Тогда сложность задается рекуррентным соотношением $f_{ODC}(i) = g(i) + \sum_{k=1}^i f_{ODC}(i-k) \cdot C_i^k$, $g \in O\left(i^{2^{i+1}+1}\right)$. Точно решить данное соотношение не представляется возможным, однако с уверенностью можно сказать, что сложность расчета ODC_{v_i} и всего классического алгоритма не лучше чем $O\left(m! \cdot m^{2^{m^2+1}+1}\right)$.

Анализ сложности предложенного алгоритма значительно проще, причем асимптотическая сложность расчета CDC_{v_i} и ODC_{v_i} совпадают. Достаточно заметить, что в худшем случае необходимо произвести естественное соединение отношений для всех пар (v_i, v_j) вершин графа, причем сложность естественного соединения в худшем случае будет составлять $O(t_i \cdot t_j)$. Таким образом, асимптотическая сложность предложенного алгоритма в худшем случае составляет $O\left(\sum_{i,j} t_i \cdot t_j\right) \subseteq O(|E| \cdot 2^n \cdot 2^n) = O(m^2 \cdot 2^{2n}) \subseteq O(m^2 \cdot 2^m)$.

Заключение

Описан подход к адаптации реляционной теории к решению задач систем автоматизированного проектирования интегральных схем. На основе предложенного подхода разработан алгоритм оптимального поиска неясных Don't Care значений. Доказана корректность предложенного алгоритма.

Применение предложенного алгоритма в системах автоматизированного проектирования потенциально позволит ускорить процесс разработки интегральных схем, а также сократить занимаемую площадь кристалла, энергопотребление, повысить верифицируемость и надежность получаемых схем.

В перспективе создание распределенной версии исследуемых алгоритмов позволит решать задачи систем автоматизированного проектирования для более сложных схем, требующих высокой вычислительной мощности.

Литература

1. De Micheli G. Synchronous logic synthesis: algorithms for cycle-time minimization // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 1991. V. 10. N 1. P. 63–73.
2. Hachtel G.D., Rho J.K., Somenzi F., Jacoby R. Exact and heuristic algorithms for the minimization of incompletely specified state machines // Proc. of European Conference on Design Automation. Amsterdam, Netherlands, 1992. P. 184–191.
3. Bogatyrev V.A., Bogatyrev S.V., Golubev I.Yu. Optimization and the process of task distribution between computer system clusters // Automatic Control and Computer Sciences. 2012. V. 46. N 3. P. 103–111.
4. Sentovich E.M., Singh K.J., Lavagno L., Moon C., Murgai R., Saldanha A., Savoj H., Stephan P.R., Brayton R.K., Sangiovanni-Vincentelli A.L. SIS: A System for Sequential Circuit Synthesis. Technical Report UCB/ERL M92/41, Electronics Research Lab, Univ. of California, Berkeley, 1992.
5. Pederson D.O. The VIS Group. VIS: Verification Interacting with Synthesis, 1995 [Электронный ресурс]. Режим доступа: <http://embedded.eecs.berkeley.edu/Respep/Research/vis>, свободный. Яз. англ. (дата обращения 04.12.2013).
6. Bartlett K.A., Brayton R.K., Hachtel G.D., Jacoby R.M., Morrison C.R., Rudell R.L., Sangiovanni-Vincentelli A., Wang A. Multi-level logic minimization using implicit don't cares // IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 1988. V. 7. N 6. P. 723–740.
7. Fujiwara H. Logic Testing and Design for Testability. Cambridge: MIT Press, 1985. 304 p.
8. Chang A.C.L., Reed I.S., Beans A.V. Path sensitization, partial boolean difference and automated fault diagnosis // IEEE Transactions on Computers. 1972. V. C-21. N 2. P. 189–195.
9. Bogatyrev V.A. Exchange of duplicated computing complexes in fault-tolerant systems // Automatic Control and Computer Sciences. 2011. V. 45. N 5. P. 268–276.
10. Bogatyrev V.A. Fault tolerance of clusters configurations with direct connection of storage devices // Automatic Control and Computer Sciences. 2011. V. 45. N 6. P. 330–337.

11. Богатырев В.А., Богатырев С.В. Критерии оптимальности многоуровневых отказоустойчивых компьютерных систем // Научно-технический вестник СПбГУ ИТМО. 2009. № 5 (63). С. 92–97.
12. Богатырев В.А., Богатырев А.В. Функциональная надежность систем реального времени // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 4 (86). С. 150–151.
13. Muroga S., Kambayashi Y., Lai H.C., Culliney J.N. Transduction method – design of logic networks based on permissible functions // IEEE Transactions of Computers. 1989. V. 38. N 10. P. 1404–1424.
14. Lin B., Touati H.J., Newton A.R. Don't care minimization of multi-level sequential logic networks // Proc. IEEE International Conference on Computer-Aided Design. 1990. P. 414–417.
15. Brayton R.K., Hachtel G., McMullen C., Sangiovanni-Vincentelli A. Logic Minimization Algorithms for VLSI Synthesis. Kluwer Academic Publishers, 1985. 206 p.
16. Theobald M., Nowick S.M., Wu T. Espresso-HF: a heuristic hazard-free minimizer for two-level logic // Proceedings - Design Automation Conference. Las Vegas, USA, 1996. P. 71–76.
17. Codd E.F. A relational model of data for large shared data banks // M.D. Computing. 1998. V. 15. N 3. P. 162–166.
18. Abiteboul S., Hull R., Vianu V. Foundations of Databases. Addison-Wesley, 1995. 685 p.
19. Date C.J., Darwen H. Foundation for Future Database Systems: The Third Manifesto. 2nd ed. Addison-Wesley, 2000. 608 p.

Демидов Даниил Валентинович – аспирант, Университет ИТМО, 197101, Санкт-Петербург, Россия; инженер, ЗАО «Интел А/О», 196247, Санкт-Петербург, Россия, Daniil.demidov@gmail.com

Daniil V. Demidov – postgraduate, ITMO University, 197101, Saint Petersburg, Russia; software engineer, “Intel A/O”, Ltd., 196247, Saint Petersburg, Russia, Daniil.demidov@gmail.com

*Принято к печати 02.07.14
Accepted 02.07.14*

УДК 681.324

ОПТИМИЗАЦИЯ ИНТЕРВАЛОВ ПРОВЕРКИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ СИСТЕМ

В.А. Богатырев^а, А.В. Богатырев^а, С.В. Богатырев^б

^а Университет ИТМО, 197101, Санкт-Петербург, Россия, Vladimir.bogatyrev@gmail.com

^б Санкт-Петербургский государственный университет аэрокосмического приборостроения, 190000, Санкт-Петербург, Россия

Аннотация. Предложена марковская модель защищенных информационных систем, функционирующих в условиях деструктивных воздействий, последствия которых обнаруживаются оперативным и тестовым контролем. Предполагается, что оперативный контроль, в отличие от тестового, характеризуется ограниченной полнотой контроля, но не требует остановки вычислительного процесса. Целью исследований является построение моделей, позволяющих оптимизировать интервалы инициализации тестового контроля по критерию максимизации вероятности нахождения системы в состоянии готовности к безопасному выполнению функциональных запросов и минимизации опасных состояний системы с учетом неопределенности и вариативности интенсивности деструктивных воздействий. Рассмотрены варианты задачи оптимизации интервалов тестирования в зависимости от интенсивности деструктивных воздействий по критерию достижения максимума готовности системы к безопасному выполнению запросов. Оптимизация проведена без адаптации и с адаптацией к изменениям реальной интенсивности деструктивных воздействий.

Показана эффективность адаптивного изменения периодов тестирования в зависимости от наблюдаемой активности деструктивных воздействий. Решение задачи оптимизации проведено с использованием встроенных средств системы компьютерной математики Mathcad 15, включая средства символьной математики решения систем алгебраических уравнений. Предложенные модели и методы определения оптимальных интервалов тестирования могут найти применение при системотехническом проектировании компьютерных систем и сетей критического применения, работающих в условиях дестабилизирующих воздействий при повышенных требованиях к их безопасности.

Ключевые слова: марковская модель, контроль, опасные состояния, деструктивные воздействия, оптимизация.

Благодарности. Работа выполнена в рамках НИР «Методы и модели обеспечения интегрированной безопасности и устойчивости функционирования компьютерных систем».

INTERVALS OPTIMIZATION OF SYSTEMS INFORMATION SECURITY INSPECTION

V.A. Bogatyrev^а, A.V. Bogatyrev^а, S.V. Bogatyrev^б

^а ITMO University, 197101, Saint Petersburg, Russia, Vladimir.bogatyrev@gmail.com

^б Saint Petersburg State University of Aerospace Instrumentation, 190000, Saint Petersburg, Russia, Vladimir.bogatyrev@gmail.com

Abstract. A Markov model is suggested for secure information systems, functioning under conditions of destructive impacts, which aftereffects are found by on-line and test control. It is assumed that on-line control, in contrast to the test one, is characterized by the limited control completeness, but does not require the stopping of computational process. The aim of research is to create models that optimize intervals of test control initialization by the criterion of probability maximization for system stay in the ready state to secure fulfillment of the functional requests and minimization of the dangerous system states in view of the uncertainty and intensity variance of the destructive impacts. Variants of testing intervals optimization are considered depending on the intensity of destructive impacts by the criterion of the maximum system availability for the safe execution of queries. Optimization is carried out with and without adaptation to the actual intensity change of destructive impacts.

The efficiency of adaptive change for testing periods is shown depending on the observed activity of destructive impacts. The solution of optimization problem is obtained by built-in tools of computer mathematics Mathcad 15, including symbolic mathematics for solution of systems of algebraic equations. The proposed models and methods of determining the optimal testing intervals can find their application in the system design of computer systems and networks of critical applications, working under conditions of destabilizing actions with the increased requirements for their safety.

Keywords: Markov model, control, dangerous states, destructive impacts, optimization.

Acknowledgements. The work is done within the framework of S&R “Methods and Models for Integrated Security and Operation Stability of Computer Systems”.

Введение

В настоящее время большое внимание уделяется развитию методов системотехнического проектирования компьютерных систем и сетей с высокой надежностью, отказоустойчивостью и производительностью при минимизации затрат на их реализацию и эксплуатацию [1–7].

Достижение высокой надежности [8–25], функциональной и информационной безопасности систем хранения и обработки данных требует использования комплекса средств оперативного и тестового контроля, направленных на обнаружение и минимизацию последствий деструктивных воздействий как злонамеренного, так и случайного характера. Оперативный контроль [1] позволяет быстро обнаружить последствия деструктивных воздействий, но замедляет вычислительный процесс, и достижимая им полнота контроля, как правило, ограничена, в результате чего возможно нарушение безопасности системы (переход в опасные состояния необнаружения последствий деструктивных воздействий). Для обнаружения опасных состояний в системе может дополнительно проводиться периодический тестовый контроль.

При организации тестового контроля возникает задача определения оптимальных интервалов между тестированием, так как уменьшение этого интервала позволяет снизить вероятности опасных состояний, но приводит к потере реальной производительности и к возрастанию среднего времени пребывания запросов в системе, что отрицательно сказывается на ее эффективности, особенно при работе в реальном времени.

В теории надежности известны марковские модели, учитывающие влияние оперативного и тестового контроля на готовность системы и позволяющие оптимизировать периодичность тестового контроля [1, 7–9, 20] с целью максимизации коэффициента готовности системы. Использование известных надежных моделей для исследования защищенных информационных систем, подверженных злонамеренным деструктивным воздействиям, затруднено тем, что их интенсивность (в отличие от отказов), как правило, характеризуется вариантностью, переменностью и неопределенностью последовательности и частоты смены вариантов.

Таким образом, для защищенных информационных систем с целью повышения их готовности к безопасному обслуживанию запросов возникает потребность определения оптимальных интервалов тестирования в зависимости от вариантности и изменяемой интенсивности деструктивных воздействий. Адаптация периодов тестирования к изменениям интенсивности воздействий, с одной стороны, позволяет увеличить готовность системы при снижении вероятности ее опасных состояний, но, с другой, увеличивает простои системы, вызываемые необходимостью обнаружения изменений интенсивности деструктивных воздействий, что обуславливает целесообразность постановки и решения задачи оптимизации процесса тестирования.

Модель защищенной системы

Для построения марковской модели защищенной информационной системы интервалы между деструктивными воздействиями, инициализацией тестирования, а также время тестирования будем считать распределенными по показательному закону. Граф состояний и переходов моделируемого процесса на рис. 1 соответствует идеальному случаю обнаружения тестовым контролем всех опасных состояний системы в предположении постоянной интенсивности нарушений.

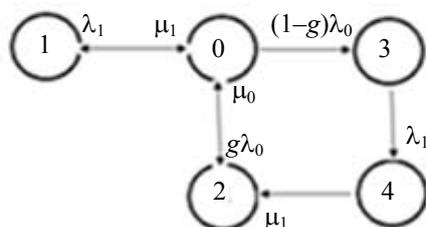


Рис. 1. Граф состояний и переходов системы с контролем злонамеренных воздействий: 0 – готовность системы к безопасному выполнению функциональных запросов; 1 – тестирование при отсутствии нарушений (последствий деструктивных воздействий); 2 – обнаружение последствий деструктивных воздействий и восстановление безопасного состояния; 3 – опасное состояние функционирования при необнаруженных последствиях деструктивных воздействий; 4 – тестирование при наличии нарушений

На рис. 1 введены обозначения: λ_0 – интенсивность потока деструктивных воздействий (величина, обратная среднему времени между деструктивными воздействиями); λ_1 – интенсивность инициализации тестового контроля (величина, обратная среднему времени между тестированием); g – полнота оперативного контроля (вероятность обнаружения нарушений оперативным контролем); μ_1 – интенсивность тестирования (величина, обратная среднему времени тестирования); μ_0 – интенсивность восстановления безопасного состояния.

Представленный на рис. 1 граф состояний и переходов позволяет составить системы алгебраических или дифференциальных уравнений, из которых находятся вероятности всех состояний $P_0 - P_4$ в стационарном или нестационарном режимах [1].

Система алгебраических уравнений Колмогорова для графа на рис. 1 имеет следующий вид:

$$\lambda_1 P_0 - \mu_1 P_1 = 0,$$

$$\lambda_1 P_3 - \mu_1 P_4 = 0,$$

$$\mu_1 P_4 + g\lambda_0 P_0 - \mu_0 P_2 = 0,$$

$$(1 - g)\lambda_0 P_0 - \lambda_1 P_3 = 0,$$

$$\sum_{i=0}^4 P_i = 1.$$

Решение получаем с использованием встроенных средств символьной математики Mathcad 15:

$$P_0 = \lambda_1 \mu_0 \mu_1 W, \quad P_1 = \lambda_1^2 \mu_0 W, \quad P_2 = \mu_1 \lambda_0 \lambda_1 W, \quad P_3 = \lambda_0 \mu_0 \mu_1 (1-g)W, \quad P_4 = \lambda_0 \lambda_1 \mu_0 (1-g)W,$$

где $W = (\lambda_1^2 \mu_0 + \mu_1 \lambda_0 \lambda_1 + \mu_0 \lambda_0 \lambda_1 + \lambda_1 \mu_0 \mu_1 + \lambda_0 \mu_0 \mu_1 - g \lambda_0 \mu_0 (\lambda_1 + \mu_1))^{-1}$.

Результаты расчета вероятностей состояний системы в зависимости от интенсивности инициализации тестового контроля λ_1 при $\lambda_0 = 0,1$ 1/ч, $\mu_1 = 10$ 1/ч, $\mu_0 = 2$ 1/ч и $g = 0,5$ представлены на рис. 2. Кривая 1 соответствует вероятности готовности системы к безопасному выполнению запросов, кривая 2 – вероятности простоя системы при ее тестировании или восстановлении, кривая 3 – вероятности простоя системы при ее тестировании в безопасном состоянии, кривая 4 – вероятности нахождения системы в опасном состоянии. Кривая 5 соответствует вероятности опасных состояний по уточненной шкале правой оси ординат. Представленные зависимости показывают возможность существования оптимального интервала между инициализацией процедуры тестирования, при котором достигается максимум вероятности нахождения системы в состоянии готовности к безопасному выполнению функциональных запросов, что подтверждает целесообразность постановки и решения задачи оптимизации интервалов тестирования.

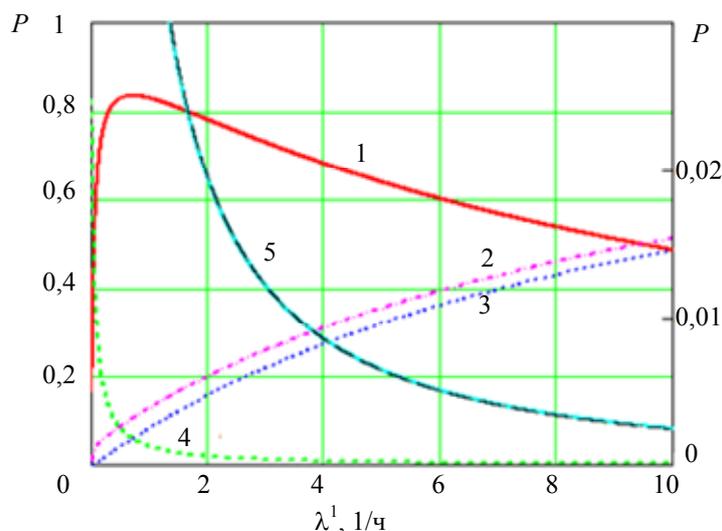


Рис. 2. Вероятности состояний системы в зависимости от интенсивности инициализации контроля: кривые 1–4 соответствуют вероятностям состояний $P_0, P_1+P_2+P_4, P_1, P_3$ соответственно; кривая 5 представляет вероятность P_3 по уточненной шкале правой оси ординат

Оптимизация интервалов тестирования по критерию максимизации готовности системы к безопасному выполнению запросов

Особенность модели контроля для исследования систем, подверженных злонамеренным деструктивным воздействиям, заключается в необходимости учета неопределенности интенсивности этих воздействий и их изменений во времени. При оптимизации интервалов тестирования будем предполагать считать заданными вектор вариантов возможных интенсивностей (q_i) и вектор вероятностей этих вариантов (r_i), имеющие размерность $n \times 1$.

При неопределенности потока деструктивных воздействий рассмотрим различные варианты постановки задачи оптимизации интервалов тестирования, обеспечивающих максимум вероятности нахождения системы в состоянии готовности к безопасному выполнению функций, обозначаемые как В1–В3.

При варианте В1 определяется значение интервалов тестирования $1/\lambda_1$ обеспечивающее максимум вероятности состояния готовности к безопасному выполнению запросов P_0 по критерию

$$\max_{\lambda_1} (\lambda_1 \mu_0 \mu_1 / (\lambda_1^2 \mu_0 + \mu_1 \lambda_0 \lambda_1 + \mu_0 \lambda_0 \lambda_1 + \lambda_1 \mu_0 \mu_1 + \lambda_0 \mu_0 \mu_1 - g \lambda_0 \mu_0 (\lambda_1 + \mu_1)))$$

при средней интенсивности воздействий $\lambda_0 = \sum_{i=0}^{n-1} q_i r_i$.

При варианте В2 определяется значение интервалов тестирования $1/\lambda_1$ (интенсивности инициализации тестирования λ_1), обеспечивающее максимум математического ожидания вероятности состояния готовности к безопасному выполнению запросов с учетом всех возможных интенсивностей воздействий q_i ($i=0, 1, \dots, n-1$). В этом случае критерий оптимальности имеет вид

$$\max_{\lambda_1} \sum_{i=0}^{n-1} r_i (\lambda_1 \mu_0 \mu_1 / (\lambda_1^2 \mu_0 + \mu_1 q_i \lambda_1 + \mu_0 q_i \lambda_1 + \lambda_1 \mu_0 \mu_1 + q_i \mu_0 \mu_1 - g q_i \mu_0 (\lambda_1 + \mu_1))),$$

где q_i – i -й вариант возможной интенсивности деструктивных воздействий, вероятность которой r_i .

При варианте В3 для каждой возможной интенсивности деструктивных воздействий определяется вектор значений интенсивности инициализации тестирования (a_i), обеспечивающий максимум математического ожидания готовности системы к безопасному выполнению запросов. При этом критерий оптимальности имеет вид

$$\max_a \sum_{i=0}^{n-1} r_i \left(a_i \mu_0 \mu_1 / (a_i^2 \mu_0 + \mu_1 q_i a_i + \mu_0 q_i a_i + a_i \mu_0 \mu_1 + q_i \mu_0 \mu_1 - g q_i \mu_0 (a_i + \mu_1)) \right).$$

Вариант В3 предполагает предварительное формирование (на основе оптимизации) вектора интервалов тестирования в зависимости от варианта интенсивности воздействий. Адаптивный переход к выбираемому значению интервала тестирования происходит в результате измерений текущей активности (интенсивности) деструктивных воздействий и идентификации соответствующего варианта градации активности q_i .

Возможна модификация адаптивного варианта назначения интервалов тестирования, когда градации активности не вводятся, а осуществляются измерения активности воздействий в реальном времени, и при обнаружении их изменений решается расчетная задача оптимизации интервалов тестирования. Следует заметить, что реализация рассмотренных вариантов адаптивного задания интервалов тестирования требует дополнительных временных издержек, а поэтому его применение требует обоснования.

Оптимизация интервалов тестирования по критерию минимизации опасных состояний при безопасном выполнении запросов

Определим интервалы тестирования, обеспечивающие минимум вероятности нахождения системы в опасном состоянии необнаруженных последствий деструктивных воздействий в условиях неопределенности и вариантности их интенсивностей. При оптимизации выделим случаи, соответствующие ранее рассмотренным вариантам В1–В3 критериев оптимизации:

$$\min_{\lambda_1} \left(\lambda_0 \mu_0 \mu_1 (1-g) / (\lambda_1^2 \mu_0 + \mu_1 \lambda_0 \lambda_1 + \mu_0 \lambda_0 \lambda_1 + \lambda_1 \mu_0 \mu_1 + \lambda_0 \mu_0 \mu_1 - g \lambda_0 \mu_0 (\lambda_1 + \mu_1)) \right),$$

$$\lambda_0 = \sum_{i=0}^{n-1} q_i r_i,$$

$$\min_{\lambda_1} \sum_{i=0}^{n-1} r_i \left(q_i \mu_0 \mu_1 (1-g) / (\lambda_1^2 \mu_0 + \mu_1 q_i \lambda_1 + \mu_0 q_i \lambda_1 + \lambda_1 \mu_0 \mu_1 + q_i \mu_0 \mu_1 - g q_i \mu_0 (\lambda_1 + \mu_1)) \right),$$

$$\min_a \sum_{i=0}^{n-1} r_i \left(q_i \mu_0 \mu_1 (1-g) / (a_i^2 \mu_0 + \mu_1 q_i a_i + \mu_0 q_i a_i + a_i \mu_0 \mu_1 + q_i \mu_0 \mu_1 - g q_i \mu_0 (a_i + \mu_1)) \right),$$

где (a_i) – вектор значений интенсивностей инициализации тестирования, обеспечивающий минимум вероятности опасного состояния P_3 для $i = 0, 1 \dots n-1$ возможных вариантов интенсивностей деструктивных воздействий (q_i).

При многокритериальной оптимизации максимизации готовности системы к безопасному выполнению запросов и минимизации опасных состояний воспользуемся аддитивным скалярным критерием, имеющим для рассматриваемых вариантов В1–В3 следующий вид:

$$\max_{\lambda_1} \left([\alpha \lambda_1 \mu_0 \mu_1 - (1-\alpha) \lambda_0 \mu_0 \mu_1 (1-g)] / (\lambda_1^2 \mu_0 + \mu_1 \lambda_0 \lambda_1 + \mu_0 \lambda_0 \lambda_1 + \lambda_1 \mu_0 \mu_1 + \lambda_0 \mu_0 \mu_1 - g \lambda_0 \mu_0 (\lambda_1 + \mu_1)) \right),$$

$$\lambda_0 = \sum_{i=0}^{n-1} q_i r_i,$$

$$\max_{\lambda_1} \sum_{i=0}^{n-1} r_i \left([\alpha \lambda_1 \mu_0 \mu_1 - (1-\alpha) q_i \mu_0 \mu_1 (1-g)] / (\lambda_1^2 \mu_0 + \mu_1 q_i \lambda_1 + \mu_0 q_i \lambda_1 + \lambda_1 \mu_0 \mu_1 + q_i \mu_0 \mu_1 - g q_i \mu_0 (\lambda_1 + \mu_1)) \right),$$

$$\max_a \sum_{i=0}^{n-1} r_i \left([\alpha a_i \mu_0 \mu_1 - (1-\alpha) q_i \mu_0 \mu_1 (1-g)] / (a_i^2 \mu_0 + \mu_1 q_i a_i + \mu_0 q_i a_i + a_i \mu_0 \mu_1 + q_i \mu_0 \mu_1 - g q_i \mu_0 (a_i + \mu_1)) \right).$$

Пример оптимизации интервалов тестирования

Проведем оптимизацию интервалов тестирования защищенной информационной системы по критерию максимизации готовности системы к безопасному выполнению запросов. Предположим, что интенсивность тестирования и восстановления безопасного состояния $\mu_1 = 1$ 1/ч, $\mu_0 = 0,1$ 1/ч при полноте оперативного контроля $g = 0,6$. Пусть возможны варианты деструктивных воздействий и их вероятности, представленные векторами (q_i) и (r_i) соответственно:

$$q_i = \begin{bmatrix} 0,001 \\ 0,002 \\ 0,003 \\ 0,004 \\ 0,005 \\ 0,006 \end{bmatrix}, r_i = \begin{bmatrix} 0,1 \\ 0,15 \\ 0,3 \\ 0,2 \\ 0,15 \\ 0,1 \end{bmatrix}.$$

Тогда оптимальным интервалам тестирования, определяемым по критерию максимизации готовности системы к безопасному выполнению запросов, для вариантов В1, В2 соответствуют значения $\lambda_1 = 0,03672$ 1/ч и $\lambda_1 = 0,03714$ 1/ч. Для варианта В3 вектор оптимальных интенсивностей инициализации тестирования (a_i), обеспечивающий минимум вероятности опасного состояния P_3 для $i = 0, 1, \dots, n-1$ возможных вариантов интенсивностей деструктивных воздействий (q_i), определен в результате оптимизации как

$$a_i = \begin{bmatrix} 0,020 \\ 0,028 \\ 0,035 \\ 0,040 \\ 0,045 \\ 0,049 \end{bmatrix}.$$

Результаты расчета вероятности готовности системы к безопасному выполнению запросов в зависимости от варианта интенсивности деструктивных воздействий представлены на рис. 3. Кривая 1 соответствует варианту В3 адаптивного изменения периодов тестирования в зависимости от наблюдаемой интенсивности деструктивных воздействий. Кривая 2 соответствует вариантам В1, В2, при которых задается постоянный период тестирования независимо от реальной интенсивности деструктивных воздействий. Кривые 3, 4 показывают увеличение вероятности готовности системы к безопасному выполнению запросов в результате адаптации периода тестирования к изменениям интенсивности деструктивных воздействий по варианту В3 относительно вариантов В1 и В2 соответственно.

Представленные графики позволяют сделать вывод об эффективности адаптивного изменения периодов тестирования в зависимости от наблюдаемой интенсивности деструктивных воздействий.

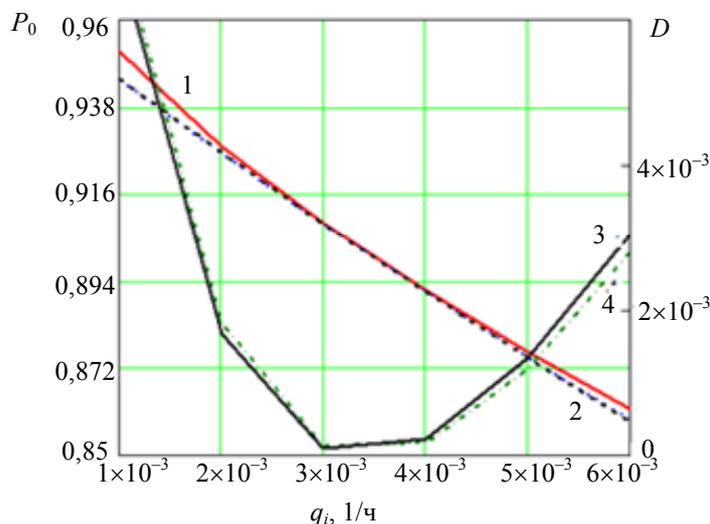


Рис. 3. Вероятности готовности системы к безопасному выполнению запросов: кривая 1 соответствует варианту В3, кривая 2 – вариантам В1, В2, кривые 3, 4 представляют разницу D готовности системы к безопасному выполнению запросов между вариантами В3–В1 и В3–В2

Предложенные модели и методы определения оптимальных интервалов тестирования могут найти применение при системотехническом проектировании компьютерных систем и сетей критического применения, работающих в условиях дестабилизирующих воздействий при повышенных требованиях к их безопасности [14–25].

Таким образом, предложена марковская модель защищенных информационных систем, функционирующих в условиях деструктивных злонамеренных и случайных воздействий, последствия которых обнаруживаются оперативным и тестовым контролем.

Поставлена и решена задача оптимизации интервалов инициализации тестового контроля по критерию максимизации вероятности нахождения системы в состоянии готовности к безопасному выполнению функциональных запросов и минимизации опасных состояний системы с учетом неопределенности и вариантности интенсивности деструктивных воздействий.

Рассмотрены варианты задачи оптимизации интервалов тестирования без их адаптации и с адаптацией к изменениям реальной интенсивности деструктивных воздействий.

Показана эффективность адаптивного изменения периодов тестирования в зависимости от наблюдаемой активности деструктивных воздействий. Так, из рис. 3 видно, что, например, при интенсивности деструктивных воздействий $6 \cdot 10^{-3}$ 1/ч готовность к безопасному выполнению запросов в результате адаптации (вариант В3) увеличивается на $3 \cdot 10^{-3}$ относительно вариантов без адаптации (В1, В2).

Решение задачи оптимизации проведено с использованием встроенных средств системы компьютерной математики Mathcad 15, включая средства символьной математики решения систем алгебраических уравнений.

Литература

1. Черкесов Г.Н. Надежность аппаратно-программных комплексов. СПб: Питер, 2005. 479 с.
2. Kopetz H. Real-Time Systems: Design Principles for Distributed Embedded Applications. Springer, 2011. 396 p.
3. Wang S.-C., Yan K.-Q., Ho C.-L., Wang S.-S. The optimal generalized Byzantine agreement in cluster-based wireless sensor networks // Computer Standards and Interfaces. 2014. V. 34. N 5. P. 821–830.
4. Abd-El-Barr M., Gebali F. Reliability analysis and fault tolerance for hypercube multi-computer networks // Information Sciences. 2014. V. 276. P. 295–318.
5. Dolev D., Függer M., Posch M., Schmid U., Steininger A., Lenzen C. Rigorously modeling self-stabilizing fault-tolerant circuits: an ultra-robust clocking scheme for systems-on-chip // Journal of Computer and System Sciences. 2014. V. 80. N 4. P. 860–900.
6. Li H., Liu H., Gao H., Shi P. Reliable fuzzy control for active suspension systems with actuator delay and fault // IEEE Transactions on Fuzzy Systems. 2012. V. 20. N 2. P. 342–357.
7. Shooman M.L. Reliability of Computer Systems and Networks: Fault Tolerance, Analysis, and Design. John Wiley & Sons Inc., 2002. 527 p.
8. Sorin D.J. Fault Tolerant Computer Architecture. Morgan & Claypool, 2009. 103 p.
9. Koren I., Krishna C.M. Fault Tolerant Systems. San Francisco: Morgan Kaufmann Publishers, 2009. 378 p.
10. Gómez A., Carril L.M., Valin R., Mouriño J.C., Cotelo C. Fault-tolerant virtual cluster experiments on federated sites using BonFIRE // Future Generation Computer Systems. 2014. V. 34. P. 17–25.
11. Bogatyrev V.A., Bogatyrev S.V., Golubev I.Yu. Optimization and the process of task distribution between computer system clusters // Automatic Control and Computer Sciences. 2012. V. 84. N 3. P. 103–111.
12. Bogatyrev V.A. Fault tolerance of clusters configurations with direct connection of storage devices // Automatic Control and Computer Sciences. 2011. V. 45. N 6. P. 330–337.
13. Bogatyrev V.A. Exchange of duplicated computing complexes in fault tolerant systems // Automatic Control and Computer Sciences. 2011. V. 46. N 5. P. 268–276.
14. Алиев Т.И. Проектирование систем с приоритетами // Изв. вузов. Приборостроение. 2014. Т. 57. № 4. С. 30–35.
15. Богатырев В.А., Богатырев С.В., Богатырев А.В. Функциональная надежность вычислительных систем с перераспределением запросов // Изв. вузов. Приборостроение. 2012. Т. 55. № 10. С. 53–56.
16. Колбанев М.О., Татарникова Т.М., Воробьев А.И. Модель обработки клиентских запросов // Телекоммуникации. 2013. № 9. С. 42–47.
17. Богатырев В.А. Отказоустойчивость и сохранение эффективности функционирования многомагистральных распределенных вычислительных систем // Информационные технологии. 1999. № 9. С. 44–48.
18. Богатырев В.А. К повышению надежности вычислительных систем на основе динамического распределения функций // Изв. вузов СССР. Приборостроение. 1981. Т. 23. № 8. С. 62–65.
19. Богатырев В.А., Богатырев С.В. Критерии оптимальности многоуровневых отказоустойчивых компьютерных систем // Научно-технический вестник СПбГУ ИТМО. 2009. № 5 (63). С. 92–97.
20. Перегуда А.И., Тимашов Д.А. Вероятностный анализ показателей надежности подсистем СУЗ с учетом периодического контроля исправности // Изв. вузов. Ядерная энергетика. 2009. № 4. С. 45–53.
21. Богатырев В.А. Мультипроцессорные системы с динамическим перераспределением запросов через общую магистраль // Изв. вузов СССР. Приборостроение. 1985. № 3. С. 33–38.
22. Богатырев В.А. Оценка вероятности безотказной работы функционально-распределенных вычислительных систем при иерархической структуре узлов // Изв. вузов. Приборостроение. 2000. Т. 43. № 3. С. 67–70.

23. Богатырев В.А., Богатырев С.В. Надежность системы управления агрегатами и машинами коммунального хозяйства // Техничко-технологические проблемы сервиса. 2008. № 4 (6). С. 23–27.
24. Богатырев В.А., Богатырев С.В., Парантаев Г.В. Балансировки нагрузки в системах управления машинами и агрегатами коммунально-бытовой сферы // Техничко-технологические проблемы сервиса. 2008. № 3 (5). С. 54–58.
25. Гатчин Ю.А., Жаринов И.О., Коробейников А.Г. Математические модели оценки инфраструктуры системы защиты информации на предприятии // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 2 (78). С. 92–95.

- Богатырев Владимир Анатольевич*** – доктор технических наук, профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия, Vladimir.bogatyrev@gmail.com
- Богатырев Анатолий Владимирович*** – аспирант, Университет ИТМО, 197101, Санкт-Петербург, Россия, Vladimir.bogatyrev@gmail.com
- Богатырев Станислав Владимирович*** – младший научный сотрудник, Санкт-Петербургский государственный университет аэрокосмического приборостроения, 190000, Санкт-Петербург, Россия, Vladimir.bogatyrev@gmail.com
- Vladimir A. Bogatyrev*** – D.Sc., Professor, ITMO University, 197101, Saint Petersburg, Russia, Vladimir.bogatyrev@gmail.com
- Anatoly V. Bogatyrev*** – postgraduate, ITMO University, 197101, Saint Petersburg, Russia, Vladimir.bogatyrev@gmail.com
- Stanislav V. Bogatyrev*** – junior scientific researcher, Saint Petersburg State University of Aerospace Instrumentation, 190000, Saint Petersburg, Russia, Vladimir.bogatyrev@gmail.com

*Принято к печати 13.01.14
Accepted 13.01.14*

УДК 81'322.2

АЛГОРИТМ СЕМАНТИЧЕСКОГО АНАЛИЗА ТЕКСТА, ОСНОВАННЫЙ НА БАЗОВЫХ СЕМАНТИЧЕСКИХ ШАБЛОНАХ С УДАЛЕНИЕМ

А.В. Мочалова^а^а Петрозаводский государственный университет, 185910, г. Петрозаводск, Россия, stark345@gmail.com

Аннотация. В связи с ростом объема текстовой информации все более актуальными становятся системы автоматической обработки текста. Одной из основных задач таких систем является задача семантического анализа. В работе предлагается алгоритм поиска семантических зависимостей между частями предложений анализируемого текста, основанный на сопоставлении текста с базовыми семантическими шаблонами. Каждое предложение, поступающее на вход программы, постепенно сокращается: некоторые части предложения в соответствии с правилами, описанными в семантических шаблонах, добавляются в очередь с приоритетом, после чего на каждой итерации алгоритма из анализируемого предложения изымается та его часть, которая имеет в очереди наибольший приоритет. Для определения приоритета в такой очереди используются два значения: значение приоритета группы, к которой принадлежит семантическая зависимость, описанная в шаблоне, и позиция слова (или последнего слова из набора, если элемент, хранимый в очереди, состоит из нескольких слов) в анализируемом предложении. В ходе работы составлено 2160 базовых семантических шаблонов, а также на языке программирования Java реализован предлагаемый в статье алгоритм. Применение в процессе реализации алгоритма экспертной системы Drools, использующей алгоритм быстрого сопоставления с шаблонами PHREAK, обеспечило высокую скорость работы семантического анализатора. По результатам тестирования сделан вывод, что предложенный алгоритм семантического анализа без использования экспертной системы Drools работает медленнее в среднем в 6–8 раз. Программная реализация алгоритма показала, что результаты работы программы быть использованы в системах автоматической обработки текстов. Разработанный семантический анализатор используется в качестве составного модуля интеллектуальной вопросно-ответной системы.

Ключевые слова: семантические зависимости, семантический анализатор, семантические шаблоны.

ALGORITHM FOR SEMANTIC TEXT ANALYSIS BY MEANS OF BASIC SEMANTIC TEMPLATES WITH DELETION

A.V. Mochalova^а^а Petrozavodsk State University, 185910, Petrozavodsk, Russia, stark345@gmail.com

Abstract. The systems of automatic text processing have become more and more important due to the constant growth of textual data. One of the main issues arising in such systems is a problem of semantic analysis. The paper deals with an algorithm for finding semantic dependencies by means of basic semantic templates with deletion. While working with the Drools expert system (and PHREAK algorithm for fast pattern matching) we have developed and implemented a semantic analyzer for construction of semantic dependencies between parts of a sentence. During the semantic analysis we add some text parts to the priority queue according to the rules described in the semantic templates, and then at each iteration of the sentence being analyzed we drop some segment of the analyzed text which has the highest priority in the queue. To determine the priority in this queue two values are used: the priority of semantic relationship group and word position. The proposed algorithm is implemented in Java. We have prepared 2160 rules using Drools expert system. The software implementation of the proposed algorithm has shown its applicability for the systems of automatic text processing. Testing results have proved that suggested algorithm of semantic analysis without Drools expert system operates 6-8 times slower, on the average. We use proposed semantic analyzer as a composite module to intellectual question-answering system.

Keywords: semantic dependencies, semantic analyzer, semantic templates.

Введение

В наши дни все больше времени люди тратят на анализ текстов, предположительно содержащих интересующие их факты. Для сокращения этого времени создаются системы автоматической обработки текста (АОТ), призванные упростить задачу нахождения нужной информации в большом объеме текста. Одной из задач автоматической обработки текстов является семантический анализ, способ реализации которого предлагается в настоящей работе.

Одним из самых распространенных способов АОТ является его сопоставление различным шаблонам. Например, в работе [1] автором описывается метод автоматического построения онтологий на основе лексико-синтаксических шаблонов. Метод синтаксических шаблонов, основанный на концепции падежной грамматики Ч. Филлмора, описанный в работах [2, 3], используется для автоматического преобразования структур знаний, хранимых в базе данных (БД), в тексты естественного языка.

Шаблоны также используются для формализации текстовой информации, что описано автором работы [4]. В этой же работе предлагается метод автоматического формирования шаблонов для идентификации сущностей и событий, а также алгоритмы формирования графа синтаксико-семантических отношений с помощью синтаксико-семантических шаблонов, создание которых предлагается автоматизировать. Синтаксические шаблоны сборки именных групп применяются для извлечения терминов-словосочетаний [5, 6].

Одним из эффективных методов извлечения семантических отношений является метод лексических шаблонов [7, 8]. Marti Hearst [8] показал, что данный метод показывает «достаточно адекватный» результат для идентификации родо-видовых отношений.

Шаблоны являются неотъемлемой частью машинного поиска в коллекции документов, а также во многих других областях автоматического анализа текста.

Все вышеупомянутые шаблоны, хотя и отличаются друг от друга, как отличаются и цели их составления, однако имеют одну общую характеристику: они сопоставляются с естественно-языковым текстом, который на протяжении всей операции сопоставления остается неизменным. В данной работе предлагается метод сопоставления текста с базовыми семантическими шаблонами, в результате чего формируются семантические зависимости, связывающие части анализируемого предложения. Особенность этого алгоритма и отличие от вышеупомянутых способов сопоставления текста с шаблонами заключается в том, что анализируемое предложение, поступающее на вход семантического анализатора, в ходе анализа постепенно сокращается: некоторые части предложения удаляются из последующего анализа при выполнении определенных условий, описанных в базовых семантических шаблонах.

Предлагаемый способ семантического анализа естественно-языкового текста предполагает формирование базовых семантических шаблонов вручную, однако количество подобных шаблонов значительно меньше количества шаблонов, сопоставление с которыми происходит по классическим алгоритмам, не подразумевающим последовательное сокращение анализируемого текста. Вследствие небольшого количества базовых семантических шаблонов работа семантического анализатора значительно ускоряется.

Базовые семантические шаблоны

Семантическое отношение. Семантическое отношение – это некая универсальная связь, усматриваемая носителем языка в тексте. Эта связь бинарна, т.е. она идет от одного семантического узла к другому узлу [9]. В качестве семантических узлов удобно рассматривать неделимые смысловые единицы языка, которые могут быть представлены, например, именованными сущностями. Будем говорить, что два различных семантических узла α и β из одного предложения связывает семантическая зависимость с именем R (обозначим $R(\alpha, \beta)$), если между α и β существует некая универсальная бинарная связь. Для конкретных семантических узлов α , β и зависимости R направление выбирается таким образом, чтобы формула была эквивалентна утверждению, что « β является R для α ». В идеале множество семантических зависимостей, используемое при машинном анализе текста, должно покрывать все возможные связи между частями текста [1].

Главным аргументом в семантической связи назовем тот узел, от которого можно задать вопрос ко второму семантическому узлу из рассматриваемой семантической связи. Например, для связи ПРИЗНАК (хижина, ветхая) главным аргументом является слово «хижина», так как можно задать вопрос: «хижина какая?» – «ветхая». Для определенности главный атрибут в семантической связи всегда будем располагать первым.

Базовые семантические шаблоны. Базовым семантическим шаблоном назовем правило, по которому в анализируемом тексте находится семантическая зависимость. Базовый семантический шаблон состоит из 4 основных частей:

1. последовательность слов или неделимых смысловых единиц, для которых указаны их морфологические признаки, а в некоторых случаях, когда это особенно важно для семантического анализа, приведены названия этих слов и смысловых единиц;
2. название семантического отношения, которое должно быть сформировано в случае обнаружения в тексте последовательности, описанной в предыдущем пункте;
3. последовательность чисел, определяющая позиции в последовательности из п. 1, элементы которой должны быть добавлены в очередь с приоритетом, в соответствии с которой впоследствии будут удаляться слова из анализируемого предложения, подаваемого на вход семантическому анализатору;
4. число, обозначающее значение приоритета, группы семантических зависимостей, к которой относится данное семантическое отношение.

Очередь с приоритетом. В классическом определении очередь с приоритетом определяется как абстрактный тип данных, позволяющий хранить пары (ключ, значение) и поддерживающий следующие операции [10]:

- `init` – инициализация новой пустой очереди;
- `insert` – добавление нового элемента в очередь;
- `remove` – удаление и возвращение элемента с наивысшим приоритетом из очереди;
- `isEmpty` – проверка очереди на наличие в ней элементов.

В рамках настоящей работы «очередь с приоритетом» используется для хранения слов или набора слов, являющихся правым аргументом некоторой семантической связи, найденной в анализируемом предложении. Для определения приоритета элемента в такой очереди используются два значения:

- значение приоритета группы, к которой принадлежит данная семантическая связь;
- позиция слова (или последнего слова из набора, если элемент, хранимый в очереди, состоит из нескольких слов) в анализируемом предложении.

Будем говорить, что элемент из описываемой очереди обладает наивысшим приоритетом, если значение приоритета семантической группы минимально, а значение позиции последнего слова из набора, образующего элемент, максимально. Таким образом, элементы очереди с приоритетом сортируются по возрастанию приоритетов групп семантических зависимостей. Если в очереди нашлось несколько элементов с одинаковыми значениями приоритетов семантических групп, то такие элементы сортируются по убыванию позиции последнего слова, относящегося к рассматриваемому элементу, в анализируемом предложении. Приведем код метода `compareTo`, реализованный на языке программирования Java и осуществляющий сравнение двух объектов класса `Unit`, каждый из которых имеет свой приоритет (`prioritet`) и позицию в анализируемом предложении (`positionInSentence`):

```
public int compareTo(Unit o)
{
    int ret = Integer.compare(prioritet, o.prioritet);

    if (ret == 0)
        ret = Integer.compare(o.positionInSentence, positionInSentence);

    return ret;
}
```

Ниже описаны правила добавления в очередь Q с приоритетом элемента (α', sp', pos') , где α' – значение элемента; sp' – приоритет семантической группы; pos' – позиция элемента в анализируемом предложении для случая, когда Q содержит элемент (α, sp, pos) такой, что $(\alpha' = \alpha)$ и $(pos' = pos)$:

1. $(sp' > sp)$, следовательно, $Q = Q \setminus (\alpha, sp, pos) \cup (\alpha', sp', pos')$;
2. $(sp' \leq sp)$, следовательно, Q не изменяется.

На рис. 1 представлен с пояснениями пример базового семантического шаблона. Здесь последовательность слов представлена тремя составляющими: глаголом (Г) мужского рода (мр), в единственном числе (ед), инфинитивом (ИНФ) и существительным (С) в именительном падеже (им), единственном числе (ед), мужского рода (мр). В случае обнаружения в тексте указанной последовательности, не разделенной другими словами и знаками препинания, формируется семантическая связь ДЕЙСТВИЕ, аргументы в которой располагаются в той последовательности, в которой они указаны в круглых скобках, учитывая, что нумеруются элементы последовательности с нуля (нумерация указана в квадратных скобках над описанием элементов последовательности). Левый и правый аргументы семантической связи в шаблоне разделены знаком «|». После того, как семантическая связь ДЕЙСТВИЕ сформирована, слово из последовательности с номерами 2 ставится в очередь на удаление, организованную в виде очереди с приоритетом, как это было описано выше.

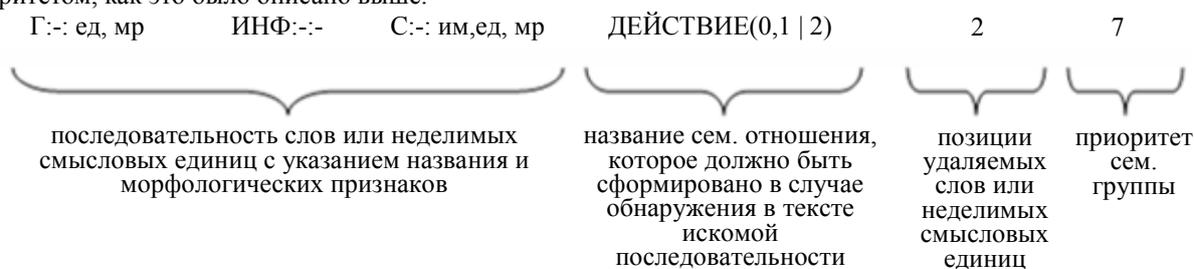


Рис. 1. Пример базового семантического шаблона

Очередь с приоритетом, хранящая части предложения, которые будут удаляться из анализируемого текста T , организована таким образом, что каждый раз после того, как все базовые семантические шаблоны в T найдены, из T удаляются по одному слову или набору слов, образующему правый аргумент некоторой семантической связи. При этом удаляемый из T элемент принадлежит к паре из Q с наивысшим приоритетом.

В данной работе предлагается способ поиска семантических отношений с помощью сопоставления текста анализируемого предложения с набором базовых семантических шаблонов.

Алгоритм нахождения семантических зависимостей с помощью базовых семантических шаблонов

Обозначения, используемые при описании алгоритма:

- T – анализируемое предложение;
- t_i – i -ая неделимая смысловая единица анализируемого предложения T ;
- S – множество всех базовых семантических шаблонов;

- s_i - i -й шаблон множества S ;
- sp_i - приоритет шаблона s_i ;
- $R_i(t_{i_1}, t_{i_2})$ - семантическая зависимость R_i , определяемая шаблоном s_i и связывающая две неделимые смысловые единицы t_{i_1} и t_{i_2} ;
- $pos(t_i)$ - позиция в анализируемом предложении последнего слова из t_i ;
- Q - очередь с приоритетами;
- (t_{i_2}, sp_i, pos_i) - элемент очереди Q , образованный посредством шаблона s_i .

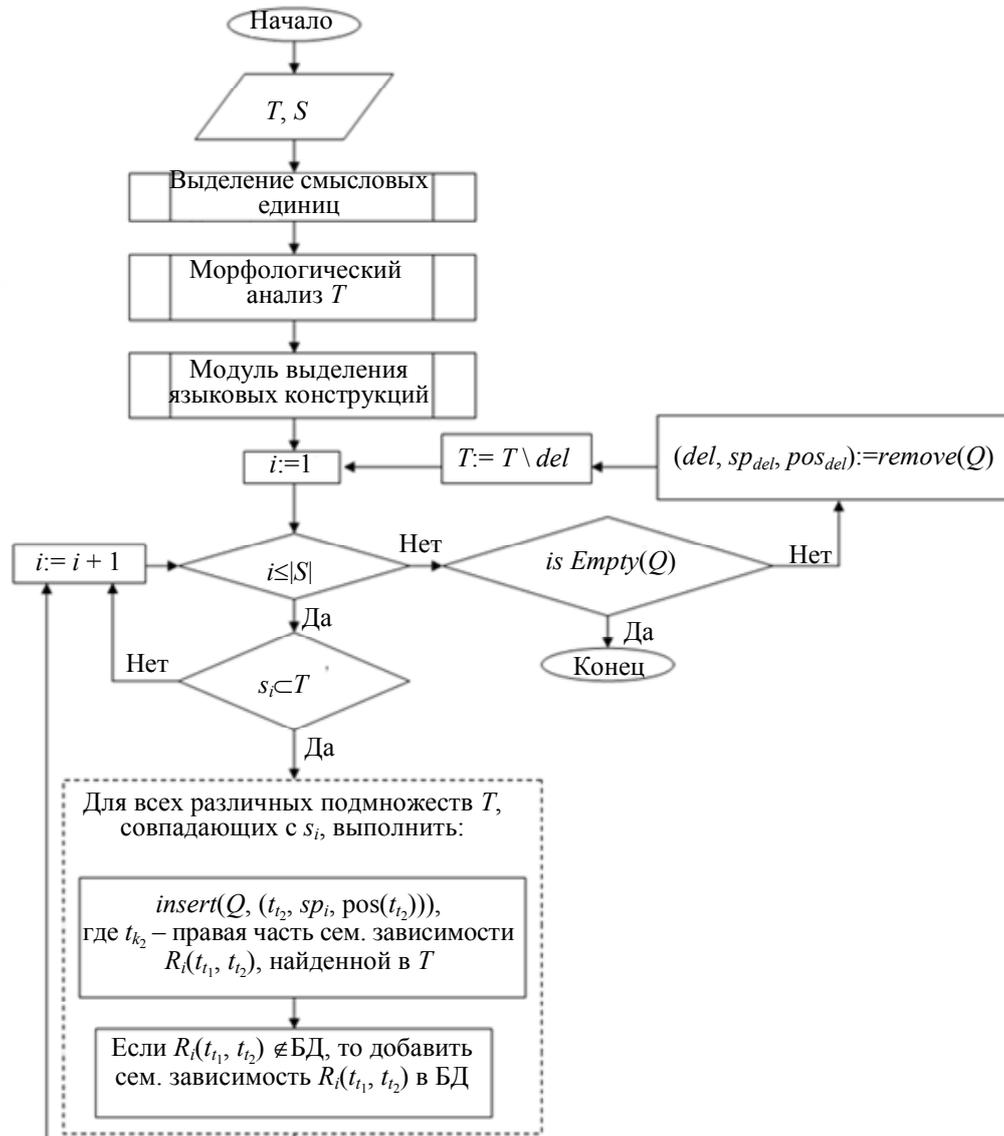


Рис. 2. Блок-схема алгоритма поиска семантических зависимостей

Алгоритм нахождения семантических зависимостей. На вход семантическому анализатору подается предложение T на естественном языке и множество $S = \{s_1, s_2, s_3, \dots\}$ базовых семантических шаблонов, где s_i - отдельный семантический шаблон. Предложение T разделяется на неделимые смысловые единицы, обозначаемые $t_1, t_2, t_3, \dots, t_n$, состоящие либо из одного слова, либо из набора слов, который может являться именованной сущностью (например, название государства, название мероприятия, титул человека и т. п.). Таким образом, получается представление T в виде упорядоченного набора t_i , где $i = 1..n$. Набор $t_1, t_2, t_3, \dots, t_n$ упорядочен в том смысле, что при последовательном написании всех t_i , получится анализируемое предложение T , т.е. T можно представить так: $T = t_1 \cup t_2 \cup t_3 \cup \dots \cup t_n$, сохраняя при этом порядок следования t_i .

После формирования набора неделимых смысловых единиц производится морфологический анализ каждой такой единицы. Затем в модуле выделения языковых конструкций происходит поиск таких сложных языковых конструкций, как вводные, причастные и деепричастные обороты, придаточные предложения и т. д.

Далее последовательно происходит поиск совпадений каждого базового семантического шаблона из множества S в множестве T , при этом как для базовых семантических шаблонов, так и для всех t_i учитываются морфологические характеристики.

В случае обнаружения совпадения семантического шаблона s_i , с некоторыми подмножествами множества $T = t_1 \cup t_2 \cup t_3 \cup \dots \cup t_n$ (на блок-схеме алгоритма обозначено как $s_i \subset T$), для всех различных подмножеств T , совпадающих с s_i , формируются семантические зависимости $R_i(t_{i_1}, t_{i_2})$, которые записываются в БД, если они обнаружены в анализируемом тексте впервые. При этом в очередь Q с приоритетом добавляется новый элемент, представленный тройкой $(t_{i_2}, sp_i, pos(t_{i_2}))$, где sp_i – приоритет группы, к которой относится семантическая зависимость R_i . Поиск базовых семантических шаблонов в T происходит до тех пор, пока не будут проверены на совпадение все шаблоны. После окончания поиска в T шаблонов происходит проверка очереди Q с приоритетом на пустоту с помощью функции isEmpty: если она пуста, это означает, что на очередном этапе сопоставления шаблонов с T новых семантических зависимостей не найдено, и программа завершает свою работу. В противном случае посредством функции remove, описанной выше, получаем элемент $(del, sp_{del}, pos_{del})$ из очереди Q с наивысшим приоритетом, после чего значение del удаляется из текущего множества T , представляющего оставшиеся для дальнейшего анализа слова из анализируемого предложения. По определению функции remove после ее вызова происходит удаление элемента $(del, sp_{del}, pos_{del})$ из Q .

Далее повторяется поиск базовых семантических шаблонов $S = \{s_1, s_2, s_3, \dots\}$ среди оставшихся неделимых семантических единиц множества T . Так продолжается до тех пор, пока множество Q не станет пустым (это означает, что в анализируемом предложении найдены все возможные семантические зависимости, описанные базовыми семантическими шаблонами S). Блок-схема описанного алгоритма представлена на рис. 2.

Программная реализация семантического анализатора

В ходе работы в соответствии с описанным алгоритмом поиска семантических зависимостей с помощью базовых семантических шаблонов с удалением на языке программирования Java был реализован семантический анализатор русского языка. Для ускорения работы программы была использована экспертная система Drools 6.0 [11], использующая алгоритм быстрого сопоставления с шаблоном PHREAK. На вход программе поступают предложения на русском языке, а на выходе программа предоставляет набор семантических отношений, сформированных по анализируемому тексту.

Для реализации семантического анализатора было построено 2160 базовых семантических шаблонов, определяющих на данный момент 539 семантических зависимостей. Все семантические зависимости разбиты на 17 групп – по значениям приоритетов удаления, значения которых используются для формирования очереди с приоритетом на удаление. Морфологический анализ предложения, предшествующий семантическому, реализован на базе грамматического словаря А.А. Зализняка [12], а морфология неизвестных слов, не найденных в словаре, определяется по алгоритму, предложенному авторами [13] и основанному на статистическом анализе последовательностей последних букв предложения. Морфологическая омонимия в текущей программной реализации частично снимается посредством использования программы *mystem* [14] и специальных правил, учитывающих возможность морфологического согласования различных частей речи, обладающими определенными морфологическими признаками. В таблице приведено несколько примеров таких правил.

В будущем для снятия морфологической омонимии предполагается использовать вероятностную модель, предложенную авторами работы [15], которая основывается на нормализующих подстановках.

Снятие омонимии со слов, имеющих одинаковые морфологические характеристики, но различающиеся по смыслу (например, «детский лагерь» и «лагерь демократов») в текущей программной реализации не реализовано. В будущем данную проблему планируется разрешать посредством интеграции семантического анализатора с онтологией, а также путем анализа контекста слова, которому могут соответствовать несколько смыслов.

Описанный в этой работе алгоритм позволяет устанавливать связь между частями предложения, разделенными такими сложными языковыми конструкциями, как вводные, причастные и деепричастные обороты, придаточные предложения и т. д. Достигается это путем использования соответствующих семантических шаблонов и синтаксического анализатора в модуле выявления языковых конструкций. При

этом аргументами семантических отношений могут быть как сами обороты или придаточные предложения целиком, так слова или словосочетания внутри них.

Правило	Пример морфологической омонимии	Снятие морфологической омонимии
Ближайшее справа от предлога существительное не может находиться в именительном падеже	На (ПРЕДЛ) стол (С:им,ед,мр С:вн,мн,мр)	На (ПРЕДЛ) стол (С:вн,мн,мр)
Ближайшее справа от предлога «для» существительное может находиться только в родительном падеже	Для (ПРЕДЛ) школы (С:им, мн, жр С:рд, ед, жр С:вн,мн,жр)	Для (ПРЕДЛ) школы (С:рд,ед,жр)
Если перед словом X, определенным морфологическим анализатором как существительное или глагол стоит качественное наречие, то слово X является глаголом	Мыла (С:ср,рд,ед С:ср,им,мн С:ср,вн,мн Г:прш,жр,ед) быстро (Н:кач)	Мыла (Г:прш,жр,ед) быстро (Н:кач)

Таблица. «Правила снятия морфологической омонимии»

Ниже приведен пример работы семантического анализатора, на вход которому подается предложение «Отличники школы яхтенного спорта, завоевав переходящий кубок, выехали в лагерь.»:

- ДЕЕПРИЧ_ОБОРОТ (выехали | завоевав переходящий кубок)
- ПРИЗНАК (кубок | переходящий)
- ЧТО (завоевав | кубок)
- ЧЕГО (отличники | школы)
- ЧЕГО (школы | спорта)
- ПРИЗНАК (спорта | яхтенного)
- МЕСТО (выехали | в лагерь)
- ДЕЙСТВИЕ (выехали | отличники).

Заключение

В ходе работы был разработан и программно реализован алгоритм работы семантического анализатора русскоязычного текста, основанный на базовых семантических шаблонах с удалением.

Программная реализация данного алгоритма показала, что при достаточном количестве базовых семантических шаблонов, используемых анализатором, работа программы может быть сопоставима с результатами работы такого известного семантического анализатора русских текстов, как продукт, разработанный группой Aot.ru [16].

Использование в предлагаемом алгоритме базовых семантических шаблонов, подразумевающее постепенное сокращение анализируемого текста, а также применение быстрого алгоритма сопоставления с шаблонами PHREAK обеспечивает высокую скорость работы семантического анализатора. Так, например, описанный в работе алгоритм семантического анализа текста, использующий экспертную систему Drools и 2160 базовых семантических шаблонов, в тексте сказки Э.Т.А. Гофмана «Золотой горшок» определил 8213 семантических отношений за 6930 мс. Предложенный алгоритм семантического анализа без использования экспертной системы Drools работает медленнее в среднем в 6–8 раз. Тестирование проводилось на процессоре Intel Core i3 M 350 CPU 2.27 ГГц в операционной системе Ubuntu 12.04.

Разработанный семантический анализатор используется в качестве составного модуля интеллектуальной вопросно-ответной системы, описанной в работе [17]. В дальнейшем планируется интегрировать данный семантический анализатор с онтологией.

Литература

1. Рабчевский Е.А. Автоматическое построение онтологий на основе лексико-синтаксических шаблонов для информационного поиска // Труды XI Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». Петрозаводск, 2009. С. 69–77.
2. Филлмор Ч. Дело о падеже // Новое в зарубежной лингвистике. Вып. X. М.: Прогресс, 1981. С. 369–495.
3. Филлмор Ч. Дело о падеже открывается вновь // Новое в зарубежной лингвистике. Вып. X. М.: Прогресс, 1981. С. 496–530.
4. Чубинидзе К.А. Метод синтактико-семантических шаблонов и его применение в информационной технологии интерпретации текстов: дис... канд. техн. наук. М., 2006. 156 с.
5. Большаков И.А. Какие словосочетания следует хранить в словарях? // Труды международного семинара Диалог'2002 по компьютерной лингвистике и ее приложениям. Протвино: 2002. Т. 2. С. 61–69.

6. Загоруйко Ю.А., Сидорова Е.А. Система извлечения предметной терминологии из текста на основе лексико-синтаксических шаблонов // Труды XIII Международной конференции «Проблемы управления и моделирования в сложных системах». Самара, 2011. С. 506–511.
7. Hearst M.A. Automatic acquisition of hyponyms from large text corpora // Proc. 14th International Conference on Computational Linguistics, 1992. P. 539–545.
8. Лайонз Дж. Введение в теоретическую лингвистику. М.: Прогресс, 1978. 544 с.
9. Сокирко А.В. Семантические словари в автоматической обработке текста (по материалам системы ДИАЛИНГ): дис. ... канд. техн. наук. М., 2001. 120 с.
10. Downey A.V. Think Python. O'Reilly Media, 2012. 300 p.
11. Drools Documentation [Электронный ресурс]. Режим доступа: http://docs.jboss.org/drools/release/6.0.1.Final/drools-docs/html_single, свободный. Яз. англ. (дата обращения 25.05.2014).
12. Зализняк А.А. Грамматический словарь русского языка. Словоизменение. М.: Русский язык, 1980. 880 с.
13. Белоногов Г.Г., Зеленков Ю.Г. Алгоритм морфологического анализа русских слов // Вопросы информационной теории и практики. 1985. № 53. С. 62–93.
14. О программе mystem [Электронный ресурс]. Режим доступа: <http://api.yandex.ru/mystem>, свободный. Яз. рус. (дата обращения 17.03.2014)
15. Зеленков Ю.Г., Сегалович И.В., Титов В.А. Вероятностная модель снятия морфологической омонимии на основе нормализующих подстановок и позиций соседних слов // Компьютерная лингвистика и интеллектуальные технологии. 2005. С. 188–197.
16. Автоматическая обработка текста [Электронный ресурс]. Режим доступа: <http://www.aot.ru>, свободный. Яз. рус. (дата обращения 12.05.2014).
17. Мочалова А.В., Мочалов В.А. Интеллектуальная вопросно-ответная система // Информационные технологии. 2011. № 5. С. 6–12.

Мочалова Анастасия Викторовна – соискатель, Петрозаводский государственный университет, 185910, г. Петрозаводск, Россия, stark345@gmail.com

Anastasia V. Mochalova – applicant, Petrozavodsk State University, 185910, Petrozavodsk, Russia, stark345@gmail.com

*Принято к печати 03.06.14
Accepted 03.06.14*

УДК 004.415.28

АРХИТЕКТУРА ВЕБ-ОРИЕНТИРОВАННОЙ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ТЕХНОЛОГИЧЕСКОЙ ПОДГОТОВКИ ПРОИЗВОДСТВА

Н.Е. Филюков^а^а Университет ИТМО, 197101, Санкт-Петербург, Россия, badfilin@gmail.com

Аннотация. Рассматривается построение веб-ориентированной автоматизированной системы технологической подготовки производства. Предложено в качестве основы такой системы использовать удаленные приложения и базы данных, находящиеся в «частном облаке». В основе подхода предлагается сервис-ориентированная архитектура, использующая в качестве модулей веб-приложения и веб-сервисы, мультиагентные технологии для реализации функции обмена информацией между составными частями системы и использование PDM-системы для управления технологическими проектами. Предложенная архитектура подразумевает преобразование автоматизированной системы технологической подготовки производства в корпоративную информационную систему, что позволит обеспечить согласованное функционирование подсистем на основе единого информационного пространства, распараллеливание коллективной работы над технологическими проектами и эффективный контроль технологической подготовки производства. В рамках данной архитектуры была разработана система, которая предоставляет возможность достаточно простого подключения технологических подсистем к системе и их взаимодействия между собой. Предоставляется возможность производить конфигурирование автоматизированной системы технологической подготовки производства применительно к конкретному предприятию на множестве разработанных подсистем и баз данных с заданием соответствующих прав доступа для сотрудников этого предприятия. Предложенный подход упрощает сопровождение программного и информационного обеспечения подсистем автоматизированной системы технологической подготовки производства за счет их централизованного нахождения в дата-центре. Результаты работы могут быть применены как основа для построения и апробации автоматизированной системы технологической подготовки производства в рамках учебного процесса для отработки и модернизации алгоритмов работы системы, после чего возможна апробация в рамках расширенного предприятия.

Ключевые слова: технологическая подготовка, мультиагентные технологии, PDM-система, веб-сервисы, облачные технологии, веб-ориентированная система, архитектура АСТПП, веб-ориентированная АСТПП.

ARCHITECTURE OF WEB BASED COMPUTER-AIDED MANUFACTURING SYSTEM

N.E. Filyukov^а^а ITMO University, 197101, Saint Petersburg, Russia, badfilin@gmail.com

Abstract. The paper deals with design of a web-based system for Computer-Aided Manufacturing (CAM). Remote applications and databases located in the "private cloud" are proposed to be the basis of such system. The suggested approach contains: service - oriented architecture, using web applications and web services as modules, multi-agent technologies for implementation of information exchange functions between the components of the system and the usage of PDM - system for managing technology projects within the CAM. The proposed architecture involves CAM conversion into the corporate information system that will provide coordinated functioning of subsystems based on a common information space, as well as parallelize collective work on technology projects and be able to provide effective control of production planning. A system has been developed within this architecture which gives the possibility for a rather simple technological subsystems connect to the system and implementation of their interaction. The system makes it possible to produce CAM configuration for a particular company on the set of developed subsystems and databases specifying appropriate access rights for employees of the company. The proposed approach simplifies maintenance of software and information support for CAM subsystems due to their central location in the data center. The results can be used as a basis for CAM design and testing within the learning process for development and modernization of the system algorithms, and then can be tested in the extended enterprise.

Keywords: production planning, multi-agent technologies, PDM - system, web - services, cloud computing, web - based system, CAM architecture, web based CAM.

Введение

Одним из важнейших этапов жизненного цикла является технологическая подготовка производства (ТПП), уровень которой во многом определяет качество производимого изделия, сроки его выхода на рынок и, в конечном счете, конкурентоспособность предприятия в целом. Одним из главных направлений совершенствования ТПП изделий является создание АСТПП.

На данный момент в Российской Федерации начинают формироваться расширенные предприятия как одна из форм кооперации множества предприятий и организаций, работающих под управлением организатора кооперационной сети, позволяющая уменьшить стоимость и ускорить выпуск промышленной продукции [1, 2]. Предприятия, участвующие в кооперации, используют различные автоматизированные системы для решения различных технологических задач, однако возникают сложности взаимного функционирования систем и их интеграции между собой, связанные с:

- трудностью информационной интеграции;
- необходимостью коллективной работы над технологическими проектами;
- необходимостью сопровождения программного обеспечения АСТПП;
- необходимостью сопровождения большого комплекса баз данных и знаний.

Преодоление указанных трудностей возможно на базе использования современных информационных технологий, что позволит обеспечить эффективное решение технологических задач всех участников кооперации и в то же время удовлетворить все требования расширенного предприятия.

АСТПП как корпоративная система

В настоящее время начинает развиваться подход к АСТПП как к корпоративной информационной системе. Такой подход особенно важен для расширенного предприятия, для которого характерен интенсивный информационный обмен между отдельными подразделениями. Исходя из этого, в настоящей работе предлагается использовать концепцию облачных технологий [3–5] для создания АСТПП как «веб-ориентированной» системы, в которой подсистемы функционируют на основе удаленных кросс-платформенных приложений (веб-сервисов) с использованием удаленных баз данных и знаний.

На наш взгляд, такой подход позволит:

1. обеспечить согласованное функционирование подсистем на основе единого информационного пространства;
2. упростить сопровождение подсистем АСТПП и баз данных (БД) в едином дата-центре;
3. распараллелить коллективную работу над технологическими проектами;
4. использовать PDM-систему для управления технологическими проектами и обеспечить эффективный контроль ТПП.

«Облако» для АСТПП создается как частное, предназначенное для использования группой родственных предприятий, что позволяет для территориально разделенных подразделений предприятия организовать коллективную работу над технологическим проектом, включая проектирование технологических процессов и технологического оснащения, разработку управляющих программ, изготовление и внедрение технологической оснастки и т.д. [1].

Отсюда следует, что в организационном плане целесообразно использовать подход «программное обеспечение в качестве услуги» (Software as a Service, SaaS) [6], при котором программные компоненты АСТПП сосредотачиваются на центральном сервере, а обращение к ним выполняется через браузеры, установленные на компьютере пользователя. Применение SaaS позволяет снизить затраты на закупку, сопровождение и использование компонент АСТПП, при этом новые версии этих компонент сразу становятся доступными всем пользователям. Кроме того, появляется возможность коллективного сопровождения удаленных баз данных и знаний, особенностями которых являются большой объем и динамический характер проблемной среды. Каждое предприятие может как иметь свои персональные базы данных и знаний, так и обращаться к общим для группы предприятий базам данных и знаний, содержащих нормативно-справочную информацию [7].

Из существующих информационных технологий большой интерес представляют мультиагентные технологии [8, 9]. Их использование позволяет повысить уровень интеллектуальности АСТПП, при этом агенты, хотя и не зависят друг от друга, но имеют возможность взаимодействовать между собой и предлагать технологю варианты решения технологических задач [10].

В настоящее время на кафедре технологии приборостроения (ТПС) Университета ИТМО ведется разработка технологической интегрированной среды (ТИС), которая представляет собой комплекс, предназначенный для решения основных задач ТПП на основе вышеуказанных подходов. Важнейшей компонентой системы является «ТИС-Админ», рассмотрению архитектуры которой, как и архитектуре АСТПП в целом, и посвящена настоящая работа.

Архитектура «ТИС-Админ»

Согласно рекомендациям [11], АСТПП разделяется по функциональному значению на два типа: общего назначения и специального назначения. Документ [12] предлагает следующую укрупненную схему АСТПП (рис. 1).

«ТИС-Админ» представляет собой подсистему общего назначения и состоит из комплекса веб-сервисов, реализующих функции взаимодействия между компонентами ТИС. Неотъемлемыми компонентами подсистемы являются также внешние подключаемые программы и системы (рис. 2):

- подсистема управления функциями администрирования, имеющая развитый интерфейс для работы с пользователями;
- агент «Словарь» – словарная система, представленная в виде подсистемы, которая предоставляет онтологический словарь для взаимодействия с подсистемами ТИС;
- PDM-система – система, через которую на предприятии или в кластере предприятий производится документооборот и управление технологическим проектом.

Для расширения функциональных возможностей ТИС к ней с помощью «ТИС-Админ» могут подключаться сторонние подсистемы, функционирующие в АСТПП.

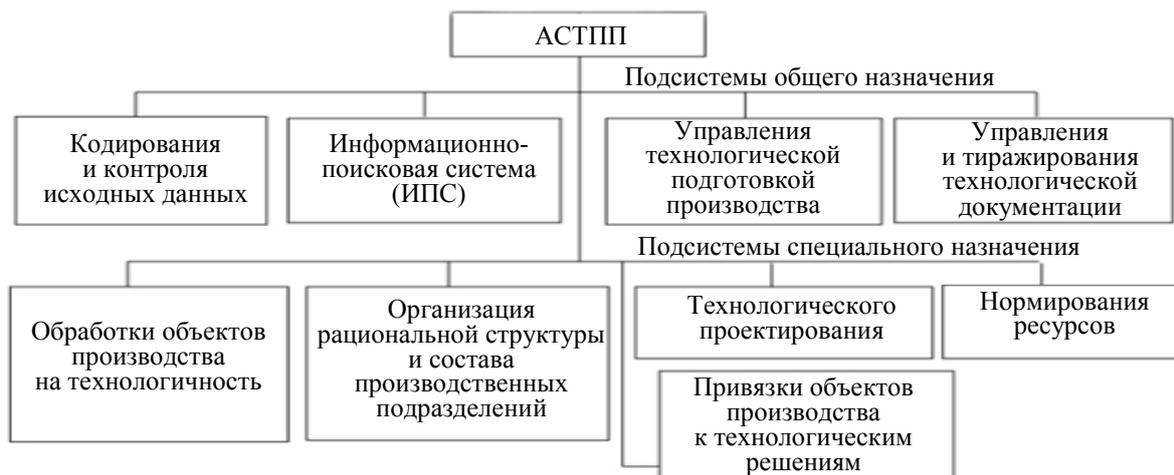


Рис. 1. Укрупненная схема АСТПП

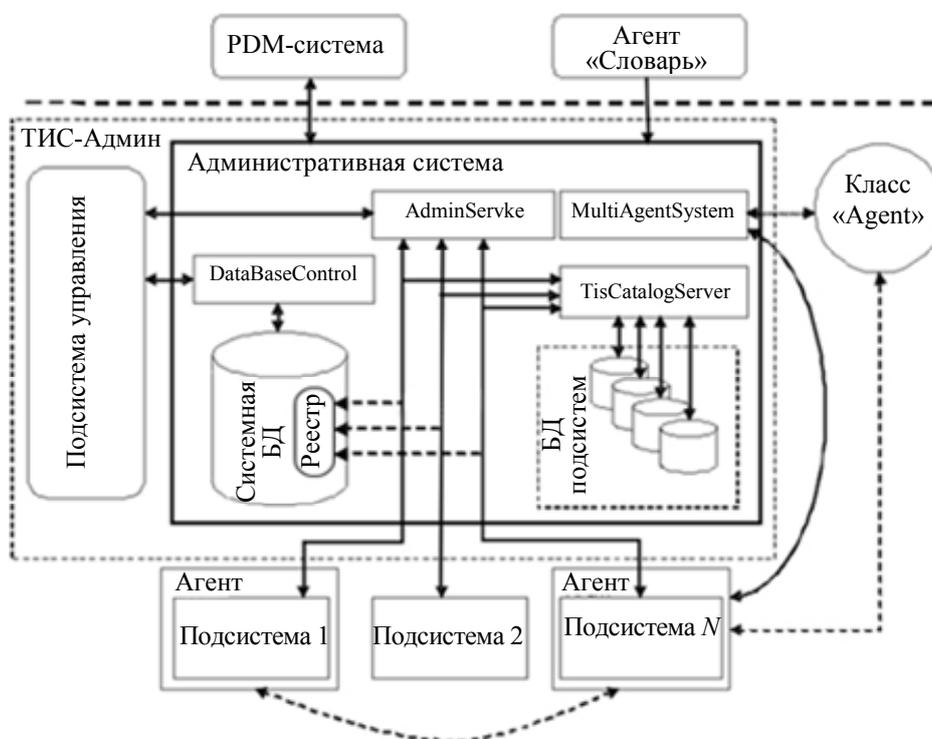


Рис. 2. Архитектура «ТИС-Админ»

Архитектура «ТИС-Админ» включает следующие системы.

1. Административная система (АС) – центральный узел системы, представляющий собой набор веб-сервисов, баз данных и знаний [13]. Через АС производится аутентификация, авторизация пользователей, идентификация подключенных подсистем. В рамках АС производится защищенное хранение данных для подключенных подсистем. АС обеспечивает взаимодействие подсистем между собой и содержит следующие компоненты.
 1. DataBaseControl – веб-сервис, взаимодействующий только с подсистемой управления, предоставляет доступ к системной БД.
 2. AdminService – веб-сервис, через который осуществляется аутентификация и авторизация пользователей. Этот веб-сервис производит идентификацию подсистем при их запросах к системе. Предоставляет подсистемам список прав доступа пользователя для их обработки.
 3. TisCatalogServer – веб-сервис, который является функциональным аналогом веб-сервиса DataBaseControl, но работает не с подсистемой управления и системной БД, а с подсистемами, подключенными к ТИС, и соответствующими им БД. С его помощью подсистемы имеют возможность создания наборов данных в рамках своих БД для последующей работы с ними.

4. MultiagentSystem совместно с абстрактным классом «Агент» позволяет подсистемам взаимодействовать между собой и обмениваться информацией [14]. Для обмена информацией используется база знаний, основанная на онтологическом словаре «Словарь».
 5. Системная БД – база данных АС, в которой хранится вся системная информация о ТИС, а именно, информация о предприятиях, использующих систему, пользователей, об их ролях и правах доступа, также хранится информация о конфигурировании системы применительно к каждому предприятию.
 6. Одной из важнейших частей системной БД является Реестр. В Реестре хранится информация о подключенных подсистемах, а также информация о данных, предоставляемых подсистемами для последующего взаимодействия между собой и другая информация.
 7. БД подсистем – в рамках ТИС имеется возможность при помощи вышеописанного веб-сервиса TisCatalogServer создавать БД с таблицами, содержащими информацию подключенных подсистем. К каждой БД имеет доступ соответствующая ей подсистема.
2. Подсистема управления – веб-приложение, предоставляющее визуальный интерфейс пользователя. Позволяет:
1. проходить аутентификацию в системе, как в рамках подсистемы управления, так и для дальнейшей переадресации к подключенным подсистемам;
 2. при помощи представленного интерфейса по управлению ТИС, производить администрирование, как ТИС, так и подключенных подсистем, управление БД подсистем и т.д.;
 3. производить конфигурирование системы применительно к каждому предприятию, входящему в расширенное предприятие на основе множества технологических подсистем;
 4. возможен обмен информацией между участниками расширенного предприятия.
3. Приложение «Агент «Словарь» является внешней программой и представляет собой словарную систему ТИС. Приложение описывает онтологию ТПП и предоставляет доступ к онтологической базе технологических терминов [15, 16]. Под онтологией ТПП понимается система, содержащая спецификацию задач и понятий, применяемых в ТПП. В ней формально описываются сферы деятельности, выполняемые приложениями, а также термины, применяемые при описании моделей объектов, циркулирующих в АСТПП. Онтология основана на применении единого «языка», содержащего лингвистический словарь и правила манипулирования единицами словаря. Словарь, который имеется в этой базе, используется при разработке новых агентов технологической системы. Словарь пополняется пользователями – разработчиками и технологами. Такой подход позволяет агентам понимать друг друга, что является необходимым условием для их общения при приеме и передачи моделей объектов и управляющих воздействий. Сложность разработки онтологии обусловлена различной природой объектов, информация о которых циркулирует в АСТПП: это 3D- и 2D-модели деталей и сборочных единиц, модели технологических процессов, разнообразные модели средств технологического оснащения и т.д. Каждая подсистема АСТПП имеет свои форматы данных, и приведение их, например, в нейтральный формат STEP требует разработки нетривиального комплекса преобразователей.

В рамках ТИС допускается использование нескольких подобных приложений, работающих параллельно и описывающих свои онтологические словари. Такие приложения не привязаны к ТИС, они развернуты параллельно и пополняются инженерно-технологическими сотрудниками. В дальнейшем, при формировании технологического агента на базе подсистемы, разработчик использует необходимый «Агент «Словарь» для описания предоставляемой функциональности.

4. PDM-система – внешняя система, используемая для создания единого информационного пространства расширенного предприятия, позволяющая осуществлять документооборот между предприятиями и подразделениями расширенного предприятия. ТИС должна иметь возможность интегрироваться и хранить выходные данные в PDM-системе. На кафедре ТПС Университета ИТМО разработана методика интеграции ТИС с PDM-системой Eovia Smarteam, что позволяет хранить данные в базах этой PDM-системы и совместно с ТИС выполнять технологические проекты.

Для выполнения технологических задач в рамках архитектуры ТИС к системе подключаются подсистемы специального назначения, которые делятся на два типа.

1. Подсистемы – веб-приложения, решающие технологические задачи и входящие в состав ТИС. Подсистемы запускаются из «ТИС-Админ» с учетом прав пользователя, вызывающего приложение, и прав доступа к базам данных (знаний), которые будет использовать приложение.
2. Агенты – подсистемы, способные взаимодействовать между собой в рамках ТИС. Подсистемы имеют возможность расширения до агентов с помощью абстрактного класса «Агент», который запрашивается у администратора системы, что дает возможность взаимодействовать и обмениваться информацией между собой, используя веб-сервис MultiAgentSystem.

Класс «Агент» – абстрактный класс, имеющий начальный набор методов и структур для осуществления обмена информацией. Используя класс «Агент» в качестве родительского, подсистемы получают возможность взаимодействия с другими подсистемами, тем самым получая свойства агента.

Экспериментальное исследование предложенного подхода

Для проверки правильности предложенного подхода были проведены эксперименты, в которых, кроме системы ТИС-Админ, участвовали следующие системы: ТИС-Процесс для проектирования технологических процессов, ТИС-СТО для сопровождения и поиска технологического оснащения, ТИС-ТАП для формирования базы знаний технологического оснащения и ТИС-Словарь для сопровождения и поиска в словаре. Указанные системы оформлены в виде веб-сервисов. Подсистемы в рамках эксперимента были распределены по различным серверам. Аутентификация, авторизация пользователей и взаимодействие подсистем осуществлялось согласно архитектуре через обращение к ТИС-Админ. Подсистемы запускались на различных веб-браузерах (Opera, Chrome, Internet Explorer), а также на мобильных аналогах браузеров в системах Android и Windows Phone. Эксперименты показали платформенезависимость при работе с системой. На данном этапе производилась функциональная отработка, из которой были сделаны выводы в правильности предложенного подхода и продолжении исследований.

Заключение

1. Предложенная архитектура веб-ориентированной автоматизированной системы технологической подготовки производства позволяет преобразовать такую систему в корпоративную информационную систему и использовать все достоинства последней.
2. Возможность последовательного расширения создаваемой автоматизированной системы технологической подготовки производства позволяет достаточно просто подключать новые подсистемы и интегрировать их на основе использования словарной системы. Конфигурирование системы применительно к конкретному предприятию на множестве разработанных подсистем и баз данных достигается путем задания соответствующих прав доступа для сотрудников этого предприятия.
3. Многоагентные технологии позволили организовать эффективное взаимодействие между подсистемами и повысить их интеллектуальный уровень при решении технологических задач.
4. Предложенный подход упрощает сопровождение программного и информационного обеспечения подсистем автоматизированной системы технологической подготовки производства за счет централизованного их нахождения в дата-центре.
5. Предложенная архитектура построения автоматизированной системы технологической подготовки производства, в основе которой лежит преобразование системы в корпоративную информационную систему, отвечает, на наш взгляд, стратегическому направлению развития автоматизации технологической подготовки производства.

Литература

1. Саломатина А.А. Методы и алгоритмы функционирования технологической подготовки производства в информационной среде виртуального предприятия: Дис... канд. техн. наук. СПб: НИУ ИТМО, 2011. 149 с.
2. Yang N., Zheng H., Yang T., Han X., Xiao T. Modeling of a cooperation environment in a virtual enterprise // *Tsinghua Science and Technology*. 2002. V. 7. N 3. P. 294–298.
3. Jaeger P.T., Lin J., Grimes J.M., Simmons S.N. Where is the cloud? Geography, economics, environment, and jurisdiction in cloud computing // *First Monday*. 2009. V. 14. N 5 [Электронный ресурс]. URL: <http://firstmonday.org/ojs/index.php/fm/article/view/2456/2171>, свободный. Яз. англ. (дата обращения 03.08.2014).
4. Sadiku M.N.O., Musa S.M., Momoh O.D. Cloud computing: opportunities and challenges // *IEEE Potentials*. 2014. V. 33. N 1. P. 34–36.
5. Kaufman L.M. Data security in the world of cloud computing // *IEEE Security and Privacy*. 2009. V. 7. N 4. P. 61–64.
6. Laplante P.A., Zhang J., Voas J. What's in a Name? Distinguishing between SaaS and SOA // *IT Professional*. 2008. V. 10. N 3. P. 46–50.
7. Gagnon S., Nabelsi V., Passerini K., Cakici K. The next web apps architecture: challenges for SaaS vendors // *IT Professional*. 2011. V. 13. N 5. P. 44–50.
8. Евгеньев Г.Б. Технология создания многоагентных прикладных систем // Труды XI национальной конференции по искусственному интеллекту с международным участием. Москва, Дубна, 2008. Т. 2. С. 306–312.
9. Hermans B. Intelligent software agents on the Internet: an inventory of currently offered functionality in the information society and a prediction of (near) future developments // *First Monday*. 1997. V. 2. N 3. [Электронный ресурс]. URL: <http://firstmonday.org/ojs/index.php/fm/article/view/514/435>, свободный. Яз. англ. (дата обращения 03.08.2014).
10. Филюков Н.Е. Технологическая подготовка производства как многоагентная система // Сборник тезисов докладов конференции молодых ученых. Вып. 2. СПб: СПбГУ ИТМО, 2011. С. 276–277.

11. Р 50-54-86-88. Автоматизированная система технологической подготовки производства. Состав и порядок разработки. Введ. 04.07.1988. М.: ВНИИНМАШ, 1988. 20 с.
12. РТМ 1.4.908-81 Автоматизированная система управления технологической подготовкой производства (изготовления и сборки) агрегатов. Организационные принципы построения. Введ. 01.01.82. НИАТ, 1982. 39 с.
13. Филюков Н.Е. Модуль администрирования веб-центричной АСТПП // Современное машиностроение. Наука и образование. 2013. № 3. С. 610–616.
14. Филюков Н.Е. Взаимодействие технологических интеллектуальных агентов в технологической интегрированной среде // IX Всероссийская межвузовская конференция молодых ученых. V сессия научной школы «Проблемы механики и точности в приборостроении». Сборник докладов. СПб: НИУ ИТМО, 2012. С. 137–141.
15. Филюков Н.Е. Онтология мультиагентных агентов в веб-центричной АСТПП // Труды XII международной конференции «Системы проектирования, технологической подготовки производства и управления этапами жизненного цикла промышленного продукта». Москва, 2012. С. 313–314.
16. Sundresh T.S. Semantic reliability of multi-agent intelligent systems // Bell Labs Technical Journal. 2006. V. 11. N 3. P. 225–236.

Филюков Николай Евгеньевич – аспирант, Университет ИТМО, 197101, Санкт-Петербург, Россия, badfilin@gmail.com

Nikolay E. Filyukov – postgraduate, ITMO University, 197101, Saint Petersburg, Russia, badfilin@gmail.com

*Принято к печати 03.04.14
Accepted 03.04.14*

УДК 004.942, 533.9

МОДЕЛИРОВАНИЕ ИМПУЛЬСНОГО ПРОБОЯ В ГЕЛИИ
С ИСПОЛЬЗОВАНИЕМ АДАПТИВНЫХ МЕТОДОВС.И. Елисеев^a, В.И. Демидов^{b, c}, А.С. Чирцов^{a, c}, В.И. Колобов^d, А.А. Кудрявцев^{a, c}, Е.А. Богданов^b^a СПбГУ, 198504, Санкт-Петербург, Россия, step.eliseev@yandex.ru^b Университет Западной Вирджинии, WV 26506, Morgantown, West Virginia, USA^c Университет ИТМО, 197101, Санкт-Петербург, Россия^d CFD Research Corporation, AL 35806, Huntsville, Alabama, USA

Аннотация. Исследуется возможность использования адаптивных методов для моделирования процессов, происходящих при электрических пробоях в газах. Численное моделирование выполнялось для разряда в гелии при атмосферном давлении между игольчатыми электродами. Физическая модель сопровождающих пробой процессов строилась на базе самосогласованной системы уравнений непрерывности для потоков заряженных частиц (электронов и положительных ионов) и уравнения Пуассона для электрического потенциала. Резкая неоднородность плазмы в области стримеров потребовала использования при моделировании адаптивных алгоритмов построения вычислительных сеток. Дано описание метода адаптивного построения сетки и обоснование его эффективности для решения резко нестационарных задач пробоя газа при атмосферном давлении. Численное моделирование процессов осуществлялось в расширенном варианте свободно распространяемого пакета Gerris. Первоначально ориентированный на решение нелинейных задач динамики жидкостей пакет оказался пригодным для моделирования процессов в нестационарной плазме, описание которых строится на базе уравнений непрерывности. Использование адаптивных сеток позволило получить адекватную численную модель развития пробоя в системе игольчатых электродов. Динамика пробоя проиллюстрирована контурными графиками распределения концентраций электронов и напряженности электрического поля, полученными в ходе решения. Показан и проанализирован механизм пробоя с образованием положительного и аноднаправленного отрицательного стримеров. Показано соответствие между адаптивным построением вычислительной сетки и образующимися в ходе решения градиентами параметров плазмы. Результаты работы могут быть взяты за основу при проведении полномасштабных численных экспериментов по пробую газового промежутка.

Ключевые слова: моделирование плазмы, пробой газа, адаптивные методы, стримеры, импульсный разряд.

SIMULATION OF PULSED BREAKDOWN IN HELIUM
BY ADAPTIVE METHODSS.I. Eliseev^a, V.I. Demidov^{b, c}, A.S. Chirtsov^{a, c}, A.A. Kudryavtsev^{a, c}, V.I. Kolobov^d, E.A. Bogdanov^b^a Saint Petersburg State University, 198504, Saint Petersburg, Russia, step.eliseev@yandex.ru^b West Virginia University, WV 26506, Morgantown, West Virginia, USA^c ITMO University, 197101, Saint Petersburg, Russia^d CFD Research Corporation, AL 35806, Huntsville, Alabama, USA

Abstract. The paper deals with the processes occurring during electrical breakdown in gases as well as numerical simulation of these processes using adaptive mesh refinement methods. Discharge between needle electrodes in helium at atmospheric pressure is selected for the test simulation. Physical model of the accompanying breakdown processes is based on self-consistent system of continuity equations for streams of charged particles (electrons and positive ions) and Poisson equation for electric potential. Sharp plasma heterogeneity in the area of streamers requires the usage of adaptive algorithms for constructing of computational grids for modeling. The method for grid adaptive construction together with justification of its effectiveness for significantly unsteady gas breakdown simulation at atmospheric pressure is described. Upgraded version of Gerris package is used for numerical simulation of electrical gas breakdown. Software package, originally focused on solution of nonlinear problems in fluid dynamics, appears to be suitable for processes modeling in non-stationary plasma described by continuity equations. The usage of adaptive grids makes it possible to get an adequate numerical model for the breakdown development in the system of needle electrodes. Breakdown dynamics is illustrated by contour plots of electron densities and electric field intensity obtained in the course of solving. Breakdown mechanism of positive and negative (orientated to anode) streamers formation is demonstrated and analyzed. Correspondence between adaptive building of computational grid and generated plasma gradients is shown. Obtained results can be used as a basis for full-scale numerical experiments on electric breakdown in gases.

Keywords: plasma modeling, gas breakdown, adaptive methods, streamers, pulsed discharge.

Введение

Методы численного моделирования играют важную роль при описании низкотемпературной плазмы и явлений переноса в газах, так как сложность и разнообразие происходящих в них процессов не позволяют получить их всестороннего описания обычными экспериментальными и теоретическими методами [1–4]. Так, например, в случае тлеющего разряда в воздушной смеси при пониженных давлениях

различия между результатами численных расчетов простейших электронных характеристик и данными эксперимента могут различаться более чем в 2–3 раза [5]. При подходе к решению задач физики плазмы зачастую критическим является вопрос выбора способов расчета. Так, в классическом тлеющем разряде низкого давления изменение физических величин происходит на различных пространственных масштабах. Характеристики большей части плазмы меняются на расстояниях порядка L длины разрядного промежутка. Исключения составляют слои пространственного заряда, примерное местоположение которых в рассматриваемой системе нетрудно указать заранее. При численном моделировании тлеющих разрядов зачастую используется метод конечных элементов. Шаг разрешения вычислительной сетки должен быть малым в областях основных неоднородностей (слоев и т.д.) и большим в области плазмы. Наличие предварительных данных о структуре разряда позволяет сильно сэкономить время расчета.

Разряды атмосферного давления характеризуются стримерами, нитями, искрами и другими динамически развивающимися пространственными структурами, характеризующимися резкими пространственными градиентами и быстро протекающими плазмохимическими процессами. Обычно стримеры и искры развиваются в сильно неоднородных электрических полях около заостренных электродов. Они распространяются в виде узких каналов с плотностью плазмы в них, варьирующейся в диапазоне концентраций частиц 10^{19} – 10^{21} м⁻³. Электрическое поле из-за высокой проводимости плазмы принимает большие значения в области головки стримера, где сильная ионизация электронными столкновениями приводит к его дальнейшему развитию. По своей природе стример является примером сильно неоднородной, неравновесной, нестабильной плазмы, требующей многоуровневого анализа.

Основой описания явления пробоя газового промежутка является классическая теория Таунсенда, берущая за основу лавинный механизм образования электронов в результате прямой ионизации. Когда напряженность электрического поля такова, что позволяет электрону на длине свободного пробега набрать энергию, достаточную для последующей ионизации, происходит пробой, выражающийся в резком экспоненциальном росте концентрации электронов и плотности тока. Следствием этого является закон Пашена, выражающий зависимость пробойного напряжения от произведения давления на длину межэлектродного промежутка и согласующийся с соответствующими экспериментально полученными кривыми Пашена. В случае, когда источником вторичных электронов является электронная эмиссия ионами с поверхности катода, реализуется таунсендовский или диффузионный пробой, критерием которого является выражение

$$\gamma(e^{\alpha d} - 1) = 1,$$

где γ – коэффициент вторичной электронной эмиссии; α – ионизационный коэффициент Таунсенда; d – межэлектродное расстояние.

Когда число электронов в лавине достаточно велико, в ней происходит разделение заряда, экранирующее внешнее поле, и образуется стример, движущийся со скоростью, превышающей скорость дрейфа электронов. Источником вторичных электронов являются фотоионизация, выталкивание электронов на фронте головной части лавины и убегание быстрых электронов. Критерий лавинно-стримерного перехода весьма условен и основывается на экранировке внешнего поля. Ситуация еще больше усложняется в случае импульсных разрядов ввиду их разнообразия. Возможность применения законов подобия для исследования импульсных разрядов обсуждается в [6]. Особое значение при этом приобретает пространственное распределение заряженных частиц и потенциала в различные моменты времени. Стримерный пробой реализуется при высоких давлениях, когда концентрации заряженных частиц велики. Однако в случае малых межэлектродных промежутков лавинно-стримерный переход может не осуществиться. Как известно, напряжение пробоя межэлектродного промежутка описывается кривыми Пашена, дающими зависимость его значения от параметра pd – произведения давления на расстояние между электродами (рис. 1).

В районе минимума кривой Пашена осуществляется стабильное горение тлеющего разряда. При атмосферных давлениях в гелии это будет соответствовать расстояниям между электродами порядка 400 мкм. А так как механизм пробоя тлеющего разряда является лавинным по своей природе, то в этих условиях представляется возможным выявить и исследовать лавинно-стримерный переход. Из-за малых межэлектродных расстояний экспериментальное исследование динамики пробоя в таких системах затруднено и может дать лишь ограниченную информацию.

Целью настоящей работы является исследование динамики пробоя в гелии при атмосферном давлении с использованием численных методов. Пробой рассматривался в системе игольчатых электродов с малым межэлектродным расстоянием. В связи с тем, что геометрия исследуемого разрядного промежутка и сама газоразрядная среда в области существования движущегося в пространстве стримера оказываются сильно неоднородными, корректное численное моделирование изучаемых плазменных процессов требует развития специальных методов, обеспечивающих необходимую точность расчетов, с одной стороны, и разумную экономию вычислительных ресурсов – с другой.

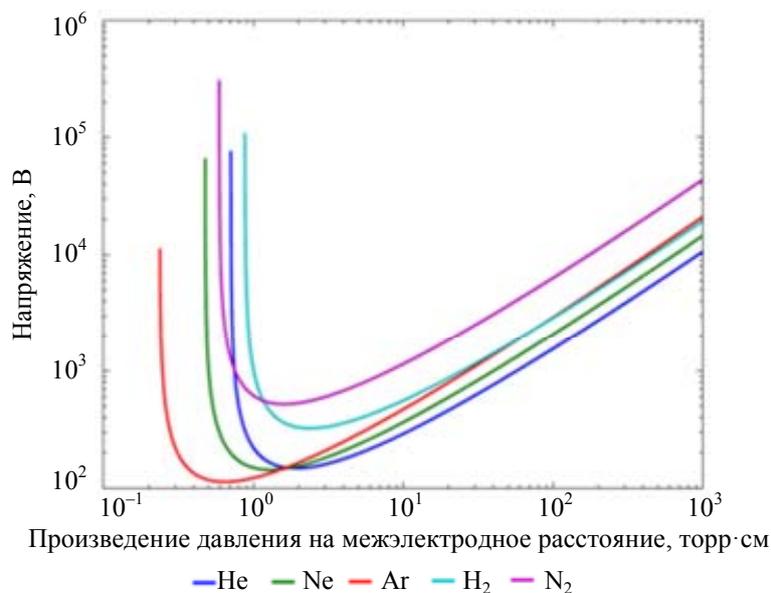


Рис. 1. Вид кривых Пашена для различных газов

Описание задачи моделирования пробоя с использованием гидродинамической модели плазмы

В ходе работы было проведено численное моделирование импульсного пробоя в гелии при атмосферном давлении в системе электродов «игла-игла» с использованием гидродинамической модели, широко используемой при описании процессов в плазме стационарных разрядов [7–13]. Она состоит из уравнений баланса электронов и ионов и уравнения Пуассона на электростатический потенциал:

$$\begin{cases} \Delta\varphi = 4\pi e(n_e - n_i) \\ \frac{\partial n_e}{\partial t} - \nabla(\mu_e n_e \mathbf{E} + D_e \nabla n_e) = I - R, \\ \frac{\partial n_i}{\partial t} + \nabla(\mu_i n_i \mathbf{E} - D_i \nabla n_i) = I - R \end{cases}$$

где I – источник заряженных частиц за счет прямой ионизации; R – гибель заряженных частиц в результате объемной рекомбинации; n_e – концентрация электронов; n_i – концентрация ионов; φ – потенциал; \mathbf{E} , – напряженность электрического поля, равная градиенту потенциала $-\nabla\varphi$; μ_e – подвижность электронов, определенная как $\frac{q_e}{m_e v_m}$ (q_e – заряд электрона, m_e – масса электрона, v_m – частота столкновений электронов с атомами газа, равная $6,1 \cdot 10^9 (\text{с} \cdot \text{торр})^{-1} \cdot P$, P – давление газа в торрах); D_e – коэффициент диффузии электронов, определяемый через соотношение Эйнштейна как $\mu_e k_B T_e$, где постоянная Больцмана k_B равна $8,62 \cdot 10^{-5} \text{эВ} \cdot \text{К}^{-1}$, T_e – температура электронов, взятая равной 3 эВ. Аналогично, подвижность ионов μ_i и коэффициент диффузии ионов D_i определялись как $\frac{q_i}{M_i v_m}$ (M_i – масса иона гелия) и $\mu_i k_B T_i$ соответственно ($T_i = 0,025 \text{эВ}$ – температура ионов, равная температуре газа).

Задача решалась в осесимметричной постановке. Геометрия задачи представлена на рис. 2: рассматривается система игольчатых электродов, находящихся друг от друга на расстоянии 600 мкм. Пробой в системе осуществлялся при подаче короткого (около 40 нс) импульса напряжения, вид которого представлен на рис. 3. Ввиду этого главную роль в динамике пробоя играли быстропротекающие процессы прямой ионизации гелия, скорость которой задавалась при помощи коэффициента ионизации Таунсенда α [11] в зависимости от локального значения напряженности электрического поля:

$$I = \alpha n_e = n_e A \exp\left(-\frac{BP}{E}\right),$$

где $A = 3 (\text{см} \cdot \text{торр})^{-1}$, $B = 34 \text{В}/(\text{см} \cdot \text{торр})$; P – давление, равное атмосферному; E – абсолютная величина напряженности электрического поля [12]. Рекомбинационные процессы в этих условиях вступают в силу только при больших концентрациях заряженных частиц. Скорость этих процессов описывалась в пороговом приближении:

$$R = \begin{cases} 0, n_e < 10^{14} \text{см}^{-3} \\ 10^{10} n_e, n_e > 10^{14} \text{см}^{-3} \end{cases}$$

В качестве граничных условий на открытых границах использовались условия Неймана на концентрации частиц и потенциал. Правый электрод заземлен, на левый подавался импульс напряжения длительностью 40 нс (рис. 3). Условия на потоки ионов и электронов задавались в соответствии с [13]. Данная модель реализована в плазменном модуле программного пакета Gerris, анализ возможностей которого и сопоставление с другими пакетами моделирования плазмы проведено в [14].

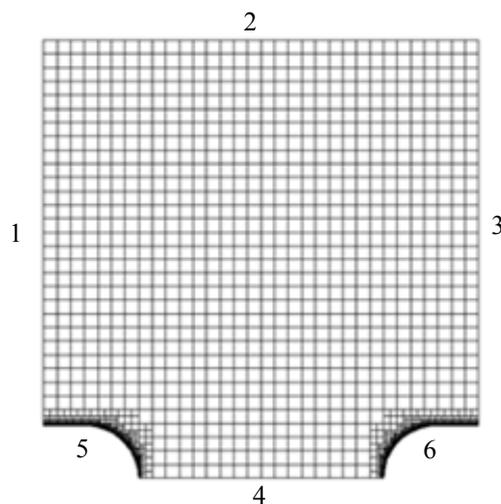


Рис. 2. Геометрия и изначальная сетка: 1, 2, 3 – открытые границы; 4 – ось симметрии; 5, 6 – электроды

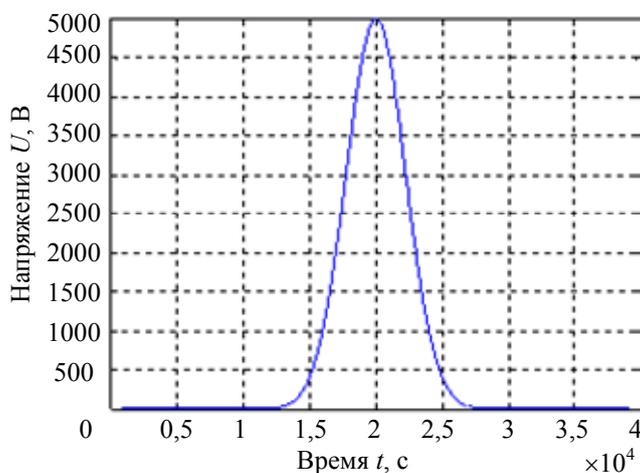


Рис. 3. Профиль импульса напряжения

Особенности численного моделирования неоднородной плазмы в условиях электрического пробоя

Использование метода конечных элементов с фиксированной вычислительной сеткой является нецелесообразным, так как простые оценки пространственных и временных масштабов изменения физических величин в стримерах [11, 15] показывают, что при рассмотрении развития стримера в атмосферном воздухе на расстоянии 1 см потребовало бы вычислительной сетки, состоящей из 10^9 ячеек.

При моделировании явлений газового пробоя представляется целесообразным применение метода конечных элементов с использованием адаптивных алгоритмов построения вычислительной сетки [16]. Такой подход позволяет автоматически генерировать вычислительную сетку с минимальным вмешательством со стороны пользователя и динамически подстраивать ее под электростатический потенциал и потоки заряженных частиц. Первоначально адаптивные методы построения вычислительных сеток разрабатывались для решения резко нестационарных задач гидродинамики и аэродинамики с решением уравнений Эйлера или Навье–Стокса. Гидродинамическое описание потоков заряженных частиц в плазме позволило распространить эти методы на задачи низкотемпературной плазмы и газовых разрядов.

Суть метода заключается в следующем. Вся вычислительная область представляет собой ячейку нулевого уровня (квадрат), которая затем может улучшаться разбиением на 4 ячейки первого уровня и т.д., образуя древовидную структуру с корнями и листьями. Разбиение ячеек производится в зависимости от получающихся в ходе решения градиентов физических величин, которые определяются как разность их значений в соседних ячейках. Затем локальные значения этих градиентов используются для подсчета коэффициента адаптации α . Затем, если его значение превышает 1, то обе соседние ячейки разбиваются на 4. В случае, когда значение α лежит в диапазоне от $\frac{1}{4}$ до 1, то ячейки остаются неизменными. Если же значение α меньше $\frac{1}{4}$ для всех ячеек, имеющих общий корень, то ячейки объединяются в одну. Размеры соседних ячеек могут отличаться максимум в два раза, что обеспечивает плавные градиенты, значительно упрощает процедуру интерполяции и ускоряет расчеты. Отметим, что выбор вида зависимости коэффи-

циента адаптации от градиентов искомых величин зависит от пользователя и зачастую является ключевым фактором, влияющим на точность и скорость решения.

На рис. 4 представлен пример построения адаптивной сетки и направление расчета градиентов. Вычисление неизвестных величин происходит исключительно в листьях древовидной структуры уровней.

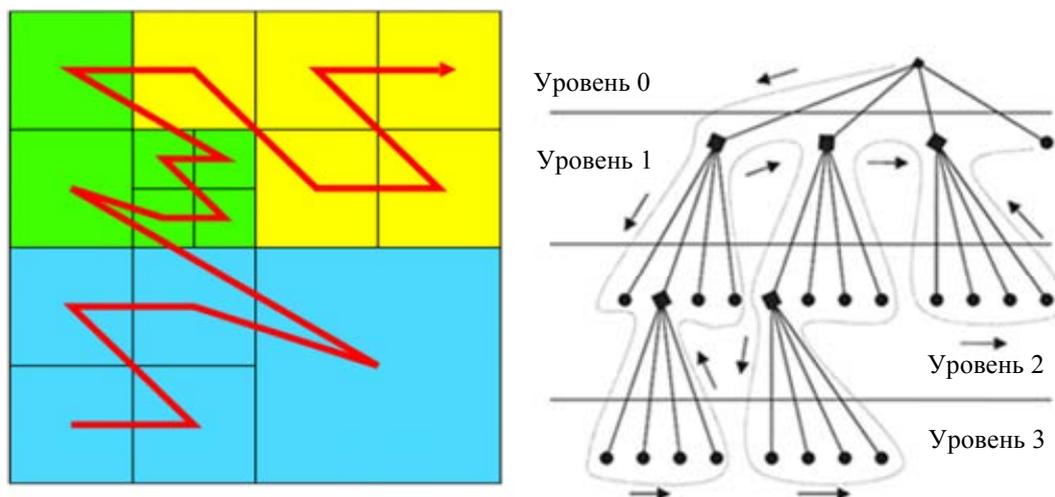


Рис. 4. Пример адаптивного построения вычислительной сетки: направление расчета градиентов величин при вычислении коэффициента адаптации (а); древовидная структура уровней и направление движения по сетке при проведении вычислений (б)

Адаптация вычислительной сетки во время решения происходила в зависимости от параметра α [12], заданного как

$$\alpha = 20 \left(\frac{n_e}{\max(n_e)} + \frac{n_e}{\max(n_e)} \right) + 3 \left(\log_{10}(n_e) + \log_{10}(n_i) + \log_{10}(v_i) + \log_{10}(|E|) \right),$$

Каждая отдельная ячейка проходила процедуру улучшения тогда, когда градиент α превосходил фиксированное значение $\nabla \alpha > 1$ в этой ячейке. Загрубление сетки происходило при значениях $\nabla \alpha < 1/4$ в этой ячейке. Отметим, что данный вид зависимости коэффициента адаптации от параметров плазмы был получен авторами [12] в ходе многочисленных пробных расчетов.

Результаты и анализ

В ходе моделирования была получена детальная динамика пробоя, начавшемся при напряжениях порядка 1,5 кВ.

На рис. 5 представлены напряженности электрического поля и вычислительные сетки в разные моменты пробоя. Конфигурация сетки изменялась в соответствии с вышеописанным алгоритмом и обеспечивала требуемую точность вычислений.

На рис. 6 представлено пространственное распределение концентрации электронов в разные моменты пробоя в разрядном объеме и на оси симметрии. Как видно, пробой имеет стримерный характер, динамика которого может быть описана следующим образом. В предпробойной стадии процессы ионизации преобладают в областях у электродов с наибольшим значением напряженности поля. Дрейф электронов от катода к аноду объясняет образование локального максимума концентрации электронов и малые значения у анода. В начале пробоя у анода образуется слой положительного заряда, который постепенно смещается к катоду за счет дрейфа ионов и диффузии электронов, образуя положительный стример, экранирующий внешнее поле. Разбиение в области около катода становится меньше для лучшего расчета образующегося градиента напряженности электрического поля. Экранирование поля стримером, в свою очередь, приводит к увеличению напряженности поля у катода, интенсификации дрейфа электронов и образованию анодонаправленного стримера. Слияние двух стримеров (рис. 5, г) приводит к образованию локального максимума концентрации электронов. Адаптация сетки в левой части модели происходит в соответствии с образующимися градиентами концентраций в слое у катода и на границе анодонаправленного стримера.

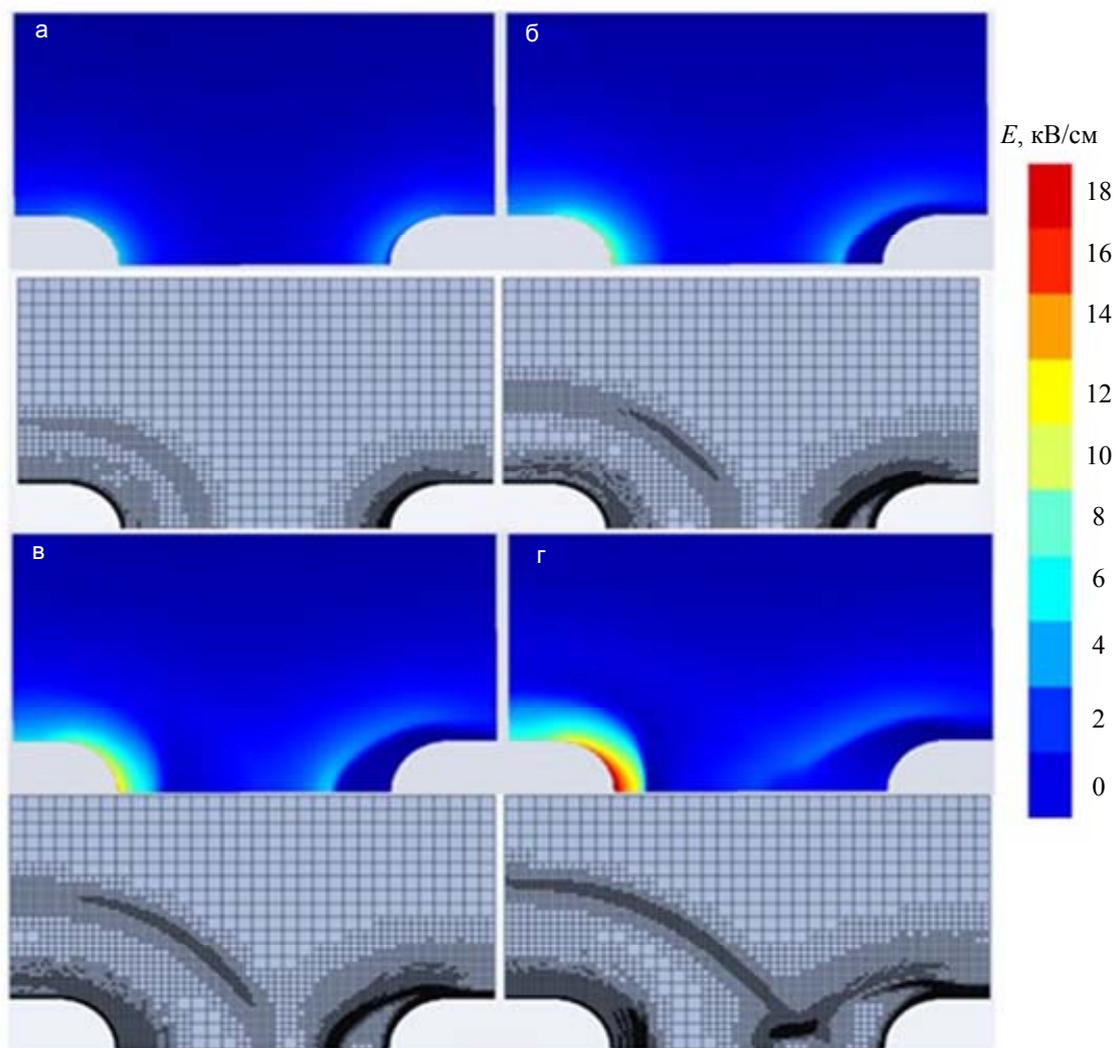


Рис. 5. Контурные графики распределения напряженности электрического поля и изменение вычислительной сетки в процессе пробоя в моменты времени: 7,8 нс (а); 9 нс (б); 9,4 нс (в); 9,8 нс (г)

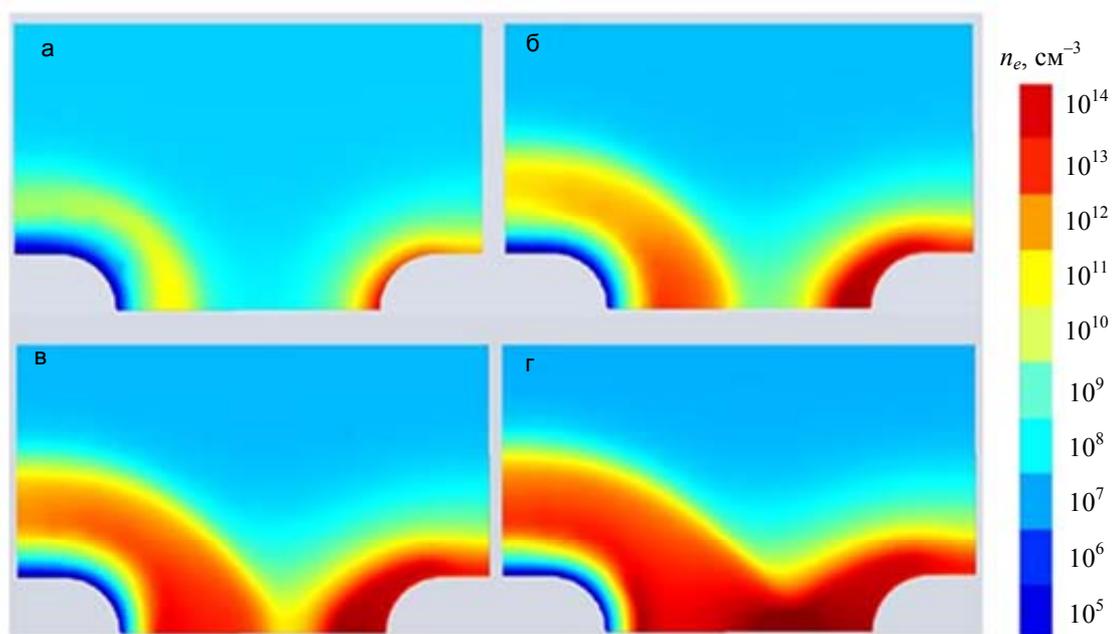


Рис. 6. Распределение концентрации электронов в моменты пробоя: 7,4 нс (а); 9 нс (б); 9,4 нс (в); 9,8 нс (г)

Заключение

Проведены расчеты динамики пробоя в гелии при атмосферном давлении с использованием адаптивных алгоритмов. Продемонстрирован в действии процесс адаптации сетки под градиенты получающихся в ходе решения параметров плазмы – напряженности электрического поля и концентрации заряженных частиц. Показано формирование и распространение стримеров в процессе пробоя. Относительно простая система уравнений позволила описать процесс пробоя и формирование стримера. Динамика пробоя представляла собой образование катодо- и анодо-направленных стримеров, движущихся навстречу друг другу, с характерным экранированием поля пространственным зарядом, образующимся в головках стримеров. Следует отметить, что полученная динамика пробоя в системе не учитывала процессы фотоионизации, которые зачастую считаются основными в формировании стримеров.

Результаты моделирования могут являться основой для дальнейшего исследования явлений электрического пробоя путем учета сложных плазмохимических процессов в многокомпонентных газах.

Литература

1. Bogdanov E.A., Kolobov V.I., Kudryavtsev A.A., Tsendin L.D. Scaling laws for oxygen discharge plasmas // *Technical Physics*. 2002. V. 47. N 8. P. 946–954.
2. Bogdanov E.A., Kudryavtsev A.A., Tsendin L.D., Arslanbekov R.R., Kolobov V.I., Kudryavtsev V.V. Substantiation of the two-temperature kinetic model by comparing calculations within the kinetic and fluid models of the positive column plasma of a DC oxygen discharge // *Technical Physics*. 2003. V. 48. N 8. P. 983–994.
3. Bogdanov E.A., Kudryavtsev A.A., Tsendin L.D., Arslanbekov R.R., Kolobov V.I., Kudryavtsev V.V. Scaling laws for the spatial distributions of the plasma parameters in the positive column of a DC oxygen discharge // *Technical Physics*. 2003. V. 48. N 9. P. 1151–1158.
4. Gutsev S.A., Kudryavtsev A.A., Zamchiy R.Yu., Demidov V.I., Kolobov V.I. Diagnostics and modeling of a short (without positive column) glow discharge in helium with nonlocal plasma // *Proc. 40th European Physical Society Conference on Plasma Physics*. Finland, 2013. N 06.502.
5. Чернышева М.В., Марек В.П., Чирцов А.С., Швагер Д.А. Компьютерное моделирование при изучении физических процессов в тлеющем разряде в воздушных смесях при низких давлениях // *Научно-технический вестник информационных технологий, механики и оптики*. 2014. № 3 (91). С. 140–146.
6. Mesyats G.A. Similarity laws for pulsed gas discharges // *Physics–Uspekhi*. 2006. V. 49. N 10. P. 1045–1065.
7. Bogdanov E.A., Chirtsov A.S., Kudryavtsev A.A. Fluxes of charged particles in two-chamber ICP discharge in oxygen // *IEEE Transactions on Plasma Science*. 2011. V. 39. N 11 part 1. P. 2562–2563.
8. Chirtsov A.S., Kapustin K.D., Kudryavtsev A.A., Bogdanov E.A. Nonlocal behavior of electron fluxes and excitation rates for «local» EEDF in moderate and high pressures DC positive column plasmas // *IEEE Transactions on Plasma Science*. 2011. V. 39. N 11 part 1. P. 2580–2581.
9. Rafatov I. Bogdanov E.A., Kudryavtsev A.A. On the accuracy and reliability of different fluid models of the direct current glow discharge // *Physics of Plasmas*. 2012. V. 19. N 3. Art. 033502.
10. Bogdanov E.A., Chirtsov A.S., Kudryavtsev A.A. Fundamental nonambipolarity of electron fluxes in 2D plasmas // *Physical Reviews Letters*. 2011. V. 106. N 19. Art. 195001.
11. Райзер Ю.П. Физика газового разряда. 3-е изд. Долгопрудный: Интеллект, 2009. 736 с.
12. Kolobov V.I., Arslanbekov R.R. Simulations of low-temperature plasmas with adaptive cartesian mesh // *ASP Conference Series*. 2012. V. 459. P. 328–333.
13. Hagelaar G.J.M., Kroesen G.M.W., Van Slooten U., Schreuders H. Modeling of the microdischarges in plasma addressed liquid crystal displays // *Journal of Applied Physics*. 2000. V. 88. N 5. P. 2252–2262.
14. Kolobov V.I., Arslanbekov R.R., Bogdanov E.A., Eliseev S., Kudryavtsev A.A. Comparison of computational tools for simulations of glow and corona discharges // *Proc. XXXI International Conference on Phenomena in Ionized Gases (ICPIG)*. Granada, Spain, 2013. PS2-024.
15. Базелян Э.П., Райзер Ю.П. Искровой разряд. М.: МФТИ, 1997. 320 с.
16. Kolobov V.I., Arslanbekov R.R. Towards adaptive kinetic-fluid simulations of weakly ionized plasmas // *Journal of Computational Physics*. 2012. V. 231. N 3. P. 839–869.

<i>Елисеев Степан Иванович</i>	– инженер-исследователь, СПбГУ, 198504, Санкт-Петербург, Россия, step.eliseev@yandex.ru
<i>Демидов Владимир Иванович</i>	– кандидат физико-математических наук, доцент, Research Professor, Университет Западной Вирджинии, WV 26506, Morgantown, West Virginia
<i>Чирцов Александр Сергеевич</i>	– кандидат физико-математических наук, доцент, профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия; доцент, СПбГУ, 198504, Санкт-Петербург, Россия, Alex_chirtsov@mail.ru

- Кудрявцев Анатолий Анатольевич* – кандидат физико-математических наук, доцент, старший научный сотрудник, Университет ИТМО, 197101, Санкт-Петербург, Россия; доцент, СПбГУ, 198504, Санкт-Петербург, Россия, akud53@gmail.ru
- Колобов Владимир Иванович* – кандидат физико-математических наук, технический сотрудник, CFD Research Corporation, AL 35806, Хантсвилль, Алабама, США, vik@cfdr.com
- Богданов Евгений Анатольевич* – кандидат физико-математических наук, доцент, доцент, СПбГУ, 198504, Санкт-Петербург, Россия, eugene72@mail.ru
- Vladimir I. Demidov* – PhD, Associate professor, Research Professor, West Virginia University, WV26506, Morgantown, West Virginia, USA
- Alexander S. Chirtsov* – PhD, Associate professor, ITMO University, 197101, Saint Petersburg, Russia; Associate professor, Saint Petersburg State University, 198504, Saint Petersburg, Russia, Alex_chirtsov@mail.ru
- Anatoly A. Kudryavtsev* – PhD, Associate professor, ITMO University, 197101, Saint Petersburg, Russia; Associate professor, Saint Petersburg State University, 198504, Saint Petersburg, Russia, akud53@gmail.ru
- Vladimir I. Kolobov* – PhD, technical fellow, CFD Research Corporation, AL 35806, Huntsville, Alabama, USA
- Evgeny A. Bogdanov* – PhD, Associate professor, Associate professor, Saint Petersburg State University, 198504, Saint Petersburg, Russia, eugene72@mail.ru

Принято к печати 02.07.14
Accepted 02.07.14

УДК 531.383-1:537.2

О ПОСТРОЕНИИ АЛГОРИТМОВ РАБОТЫ БЕСКАРДАННОГО ГИРОГОРИЗОНТКОМПАСА НА ЭЛЕКТРОСТАТИЧЕСКОМ ГИРОСКОПЕ

Г.И. Емельянцеv^{а, b}, А.А. Медведков^{а, b}, Цай Тицзин^с

^а ОАО «Концерн «ЦНИИ «Электроприбор», 197046, Санкт-Петербург, Россия;

^б Университет ИТМО, 197101, Санкт-Петербург, Россия, medvedcov@yandex.ru

^с Юго-Восточный университет, 210096, г. Нанкин, Китай, caitij@seu.edu.cn

Аннотация. Разработаны алгоритмы работы возможной схемы построения бескарданного гироскопа на электростатическом гироскопе для подвижного объекта. Для реализации режима начальной выставки и калибровки коэффициентов модели дрейфа электростатического гироскопа в условиях подвижного объекта необходимо привлечение эталонных данных о параметрах ориентации (по курсу и углам качки) и координат места. Требуемые эталонные значения параметров ориентации могут вырабатываться при совместной обработке данных измерительного блока на микромеханических датчиках (гироскопах и акселерометрах) и GPS-компыа. В зависимости от уровня динамических условий на объекте и требуемой точности выработки курса для построения вертикали места в системе может использоваться вместо микромеханических датчиков измерительный блок на волоконно-оптических гироскопах и акселерометрах.

Рассмотрены особенности алгоритмов выработки курса для бескарданного гироскопа. Описываются калибровочный и рабочий (корректируемый) режимы работы системы. Особенность алгоритма работы бескарданного гироскопа заключается в использовании двух электростатических гироскопов с ортогонально расположенными векторами кинетических моментов, при этом один гироскоп является опорным (орт его кинетического момента направляется по оси Мира), а второй является «виртуальным» – погрешности его положения относительно инерциальной системы координат и коэффициенты модели ухода являются нулевыми. Совместная обработка данных бескарданного гироскопа и внешней информации о координатах места осуществляется с использованием алгоритма обобщенного фильтра Калмана с обратной связью по всему вектору состояния системы.

Приведены результаты имитационного моделирования алгоритмов работы системы, подтверждающие наличие компасного эффекта у системы и характеризующие необходимое время для калибровки электростатического гироскопа со сплошным ротором. Результаты внедрены в разработки ОАО «Концерн «ЦНИИ «Электроприбор».

Ключевые слова: электростатический гироскоп, бескарданый гироскоп, волоконно-оптический гироскоп, микромеханические датчики.

ON ALGORITHMS CREATION FOR STRAPDOWN STABILIZED GYROCOMPASS OPERATION BASED ON ELECTRICALLY SUSPENDED GYROSCOPE

G.I. Emelyantsev^{а, b}, A.A. Medvedkov^{а, b}, C. Tijing^с

^а State Research Center of the Russian Federation Concern CSRI Elektropribor, JSC, 197046, Saint Petersburg, Russia;

^б ITMO University, 197101, Saint Petersburg, Russia, medvedcov@yandex.ru

^с Southeast University, 210096, Nanjing, P.R.China

Abstract. The paper presents operation algorithms of ESG-based strapdown stabilized gyrocompass (SSGC) located onboard a mobile vehicle. Initial alignment mode and calibration of drift model coefficients onboard a vehicle is aided by reference attitude (heading, pitch and roll angles) and position data. The required reference attitude parameters can be generated by joint processing of data from MEMS IMU with gyros and accelerometers and GPS compass. Depending on the vehicle dynamics and required accuracy of generated heading, the system may use IMU based on the fiber-optic gyros and accelerometers instead of MEMS to construct the place vertical.

Specific features of SSGC algorithms in heading generation are considered. Calibration and corrected operation modes of the system are described. The SSGC uses two ESGs with orthogonal angular momentum vectors, where one gyro is the reference (unit vector of its angular momentum is aligned with the celestial axis) and the other one is virtual (with zero misalignments with respect to the inertial frame, and zero drift model coefficients). Joint processing of SSGC data and external position aiding is realized by extended Kalman filter with full-state feedback control.

Simulation modeling results of the system operation algorithms are presented. Simulation modeling has confirmed the system compass effect and determined the time required for calibration of ESG with solid-rotor. The results have been applied at «Concern CSRI “Elektropribor”», JSC.

Keywords: electrically suspended gyroscope (ESG), strapdown stabilized gyrocompass, fiber-optic gyroscope, micromechanical sensors (MEMS).

Введение

Использование позиционных гироскопов, к числу которых относится электростатический гироскоп (ЭСГ) [1], представляет интерес для построения бескарданных инерциальных модулей [2–4] при решении задачи ориентации для подвижных объектов типа автоматических подводных аппаратов (ПА), внутритрубных инспектирующих снарядов [5], используемых для мониторинга нефтяных и газовых скважин, а также трубопроводов.

В работе [6] рассматривались схема построения и алгоритмы работы бескарданного гироскопа (БГТК) на электростатическом гироскопе и микромеханических датчиках. Было показано, что для реализации режима начальной выставки и калибровки коэффициентов модели ухода (КМУ) ЭСГ в условиях подвижного объекта необходимо привлечение эталонных данных о параметрах ориента-

ции (по курсу и углам качки) и координат места. Требуемые эталонные значения параметров ориентации в надводном положении ПА могут вырабатываться при совместной обработке данных измерительного блока на микромеханических датчиках (гироскопах и акселерометрах) и GPS-компаса [7, 8].

В зависимости от уровня динамических условий на объекте и требуемой точности выработки курса для построения вертикали места в системе может использоваться вместо микромеханических датчиков измерительный блок на волоконно-оптических гироскопах [9] и акселерометрах. При этом остается актуальным вопрос повышения точности курсоуказания БГГК в рабочем режиме работы.

Рассмотрим построение исследуемого БГГК на ЭСГ. Измерительный модуль (оси x_b, y_b, z_b) БГГК состоит из одного ЭСГ с полярной ориентацией, малогабаритного блока гироскопов (датчиков угловой скорости – ДУС) и акселерометров, установленных на основании прибора в связанных с объектом осях и предназначенных для выработки углов качки. При начальной выставке системы в данном случае, в отличие от [6], корпус ЭСГ (оси $x_{kn1}, y_{kn1}, z_{kn1}$) разворачивается относительно основания прибора (оси x_c, y_c, z_c , связанные с объектом) и устанавливается приближенно по оси Мира. Ось y_b измерительного модуля направлена к Северному полюсу, а соответствующая ей ось z_{kn1} корпуса ЭСГ – к Южному. После этого ротор гироскопа разгоняется с направлением вектора кинетического момента по оси z_{kn1} корпуса гироскопа.

Рассматривается один из возможных алгоритмов определения курса, при котором вводится понятие дополнительного идеального «виртуального» ЭСГ, ориентируемого в плоскости земного экватора по одной из инерциальных осей. Для определения точного начального положения орта кинетического момента рабочего ЭСГ₁, оценки КМУ и погрешностей привязки его измерительных осей к осям объекта сразу после выставки корпуса ЭСГ и запуска системы осуществляется работа БГГК в режиме калибровки. Для этого привлекается внешняя эталонная информация о курсе K_{et} , координатах места объекта $\lambda_{et}, \varphi_{et}$ и звездном времени $S_{гр}$ на гринвичском меридиане (рис. 1). С помощью блока ДУС и акселерометров осуществляется выработка углов качки, т.е. углов тангажа ψ_{pr} и крена θ_{pr} объекта. В итоге формируются эталонные значения матрицы ориентации $C_{c,in}$, характеризующей положение связанных с объектом осей x_c, y_c, z_c (c) относительно инерциальных осей in_1, in_2, in_3 (in).

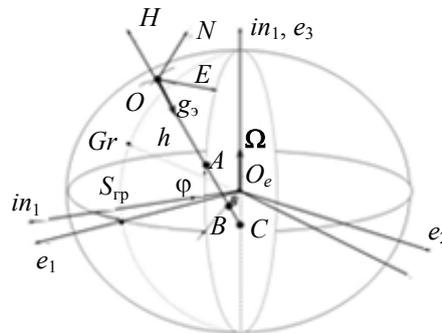


Рис. 1. Ориентация географического сопровождающего трехгранника относительно ИСК

В настоящее время принята детерминированная модель ухода ЭСГ со сплошным ротором, которая представляется в виде аналитических функций, связывающих геометрические параметры несферического и несбалансированного ротора с параметрами физических полей – источников уводящих моментов [1],

$$\omega_k = f(k_0, k_{1N}, k_2, k_3, k_4, k_5, \mu, \nu),$$

где $k_0, k_{1N}, k_2, k_3, k_4, k_5$ – КМУ ЭСГ₁, обусловленные действием моментов от взаимодействия соответствующих гармоник формы ротора с полем подвеса; μ – КМУ, характеризующие консервативную часть момента от взаимодействия неравножесткого подвеса с радиально несбалансированным ротором, а коэффициенты ν – диссипативную часть данного момента; ω_k – корпусной дрейф ЭСГ. Для обеспечения наблюдаемости оценок КМУ и погрешностей привязки измерительных осей ЭСГ, а также снижения уровня дрейфа ЭСГ используется модуляционное вращение корпуса ЭСГ вокруг направления его кинетического момента. С завершением режима калибровки происходит переход БГГК в рабочий режим (режим коррекции), в котором используется внешняя информация только о координатах места объекта.

Особенности математического обеспечения системы

Основные обозначения систем координат и кинематических параметров, используемые в статье:

- ИСК (in) – инерциальная система координат (ИСК) (in_1, in_2, in_3), правый ортогональный трехгранник с началом в центре масс (точка O_e) Земли (ось in_3 направлена по оси суточного вращения Земли, ось

- in_1 – в точку весеннего равноденствия (рис. 1)); $e_1e_2e_3$ – гринвичский навигационный трехгранник, вращается вокруг оси Мира относительно ИСК с угловой скоростью Ω ;
- $ENH(h)$ – географический сопровождающий трехгранник, правый ортогональный трехгранник с началом в центре масс (точка O) объекта (ось H направлена по нормали к эллипсоиду Земли, ось N – лежит в плоскости меридиана места (рис. 1));
- $x_c y_c z_c(c)$ – связанная с основанием (объектом) система координат (y_c – продольная ось, ось x_c направлена в правый борт);
- $x_b y_b z_b(b)$ – оси измерительного блока БГК и $x_{kn1} y_{kn1} z_{kn1}(kn)$, $x_{kp1} y_{kp1} z_{kp1}(kp)$ – оси, связанные соответственно с измерительными осями ЭСГ₁ и его корпусом. Их взаимная ориентация характеризуется следующими матрицами:

$$\begin{aligned}
 C_{c,bn} &= \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, C_{bn,bi} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sin \varphi_0 & \cos \varphi_0 \\ 0 & -\cos \varphi_0 & \sin \varphi_0 \end{bmatrix}, C_{bi,b} = \begin{bmatrix} \cos \rho & 0 & -\sin \rho \\ 0 & 1 & 0 \\ \sin \rho & 0 & \cos \rho \end{bmatrix}, \\
 C_{c,bi} &= C_{bn,bi} \cdot C_{c,bn} = \text{const}, C_{c,b} = C_{bi,b} \cdot C_{c,bi}; \\
 C_{kp,b} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, C_{kn1,b} = C_{kp,b} \cdot C_{kn,kp}.
 \end{aligned} \tag{1}$$

где φ_0 – широта места; ρ – угол модуляционного вращения корпуса ЭСГ; $C_{kn,kp}$ – матрица привязки измерительных осей ЭСГ к его корпусным осям, подлежащая оценке при калибровке системы.

Особенности алгоритмов работы БГК на ЭСГ заключаются в следующем:

- используются два ЭСГ_{*i*} ($i = 1, 2$) с ортогональными векторами кинетических моментов;
- рабочий (опорный) ЭСГ₁ формируется таким образом, что орт его кинетического момента ориентируется по оси Мира;
- второй, «виртуальный» ЭСГ (ЭСГ₂) формируется идеальным: погрешности его положения относительно ИСК $in_1 in_2 in_3$ и КМУ равны нулю. Приведение его данных к связанным $h_{c_2}^{et}$ с основанием осям $x_c y_c z_c$ осуществляется с точностью до погрешностей матрицы ориентации $C_{c,in}$, значения которой в режиме калибровки вычисляются по эталонным данным о координатах места, курсе объекта (основания) и углам качки (ψ_{pr} и θ_{pr});
- на основе выходных данных ЭСГ_{*i*} ($h_{kn_1}^S$ и $h_{c_2}^{et}$ – векторов значений направляющих косинусов ортов кинетических моментов ЭСГ_{*i*} соответственно в корпусных $x_k y_k z_k$ и связанных $x_c y_c z_c$ осях) моделируется в пространстве ортогональный гироскопический трехгранник $q_1 q_2 q_3$, вычисляя текущие значения матрицы $C_{q,c}$, характеризующей угловое положение трехгранника $q_1 q_2 q_3$ относительно связанных с основанием БГК осей $x_c y_c z_c$;
- прогнозирование ухода калибруемого ЭСГ₁ осуществляется в ИСК, однако расчетная модель погрешностей описана в квазиинерциальной системе координат (квази-ИСК) $int_1 int_2 int_3$, дискретно (в моменты коррекции положения ЭСГ₁) учитывающей прецессию гироскопического трехгранника $q_1 q_2 q_3$;
- введение квази-ИСК [10] позволяет осуществить линеаризацию матрицы динамики погрешностей ЭСГ₁ и измерений в точках пространства состояния, дискретно движущихся вместе с вектором кинетического момента ЭСГ₁; переход от ИСК к квази-ИСК характеризуется матрицей $C_{in,int}$;
- для обработки соответствующих измерений как в режиме калибровки БГК, так и в режиме коррекции, используется алгоритм обобщенного фильтра Калмана (ФК) с обратной связью по всему вектору состояния системы.

Алгоритм выработки курса

В рассматриваемом БГК исходными данными являются направляющие косинусы орта h_1 вектора кинетического момента ЭСГ₁ относительно правой ортогональной системы координат $x_{kn} y_{kn} z_{kn}(kn)$, связанной с корпусом гироскопа,

$$\mathbf{h}_{kn_1}^S = [h_{11}^k \quad h_{21}^k \quad h_{31}^k]^T, \quad (2)$$

и направляющие косинусы орта \mathbf{h}_2 «виртуального» ЭСГ₂ в ИСК (*in*)

$$\begin{aligned} \mathbf{h}_{in_2}^{et}(t_0) &= [1, 0, 0]^T; \\ \mathbf{h}_{in_2}^R(t) &= \mathbf{h}_{in_2}^{et}(t_0) = \text{const}. \end{aligned} \quad (3)$$

Направляющие косинусы этих же ортов в связанной с объектом (основанием БГК) системе координат (x_c, y_c, z_c) могут быть найдены в соответствии с исходными положениями (2)–(3) как

$$\begin{aligned} \mathbf{h}_{c_2}^R &= (\mathbf{C}_{c,in})^T \mathbf{h}_{in_2}^R(t_0), \\ \mathbf{h}_{c_1}^S &= \mathbf{C}_{kn_1,c} \mathbf{h}_{kn_1}^S, \end{aligned} \quad (4)$$

где $\mathbf{C}_{c,in}$ – расчетные значения матрицы ориентации, формируемой как $\mathbf{C}_{c,in}^{et}$ (с использованием эталонных значений курса) в режиме калибровки и как $\mathbf{C}_{c,in}^{et-pr}$ (с использованием приборных значений курса) в режиме коррекции. Учитывая (1), можно записать матрицу перехода от корпусных осей ЭСГ к осям, связанным с основанием прибора:

$$\mathbf{C}_{kn_1,c} = \mathbf{C}_{bi,c} \mathbf{C}_{b,bi}(\rho) \mathbf{C}_{kp,b} \mathbf{C}_{kn,kp}. \quad (5)$$

Для построения алгоритма выработки курса введем правый ортогональный трехгранник $q_1q_2q_3$ (рис. 2), орты которого $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ построены на ортах $\mathbf{h}_1, \mathbf{h}_2$ векторов кинетических моментов ЭСГ_{*i*} (т.е. необходимо решение задачи ортогонализации):

$$\mathbf{q}_1 = \frac{1}{\sin \Theta} \cdot (\mathbf{h}_1 \times \mathbf{h}_2), \quad \mathbf{q}_2 = \mathbf{h}_1, \quad \mathbf{q}_3 = \mathbf{q}_1 \times \mathbf{h}_1 = \frac{1}{\sin \Theta} \cdot (\mathbf{h}_2 - \cos \Theta \cdot \mathbf{h}_1), \quad (6)$$

где Θ – угол между векторами $\mathbf{h}_1, \mathbf{h}_2$, причем $\cos \Theta = \mathbf{h}_1 \cdot \mathbf{h}_2$.

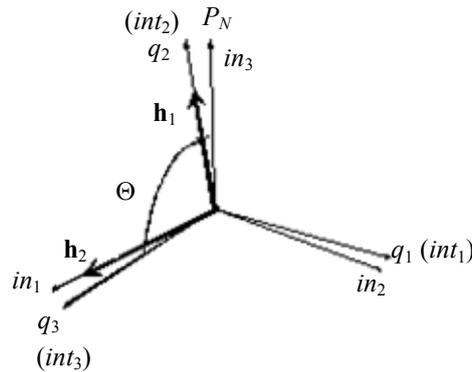


Рис. 2. Система координат $q_1q_2q_3$, связанная с ортами $\mathbf{h}_1, \mathbf{h}_2$ кинетических моментов ЭСГ_{*i*}

Ориентация трехгранника $q_1q_2q_3$ относительно связанной с объектом (основанием БГК) системы координат x_c, y_c, z_c определяется в этом случае матрицей направляющих косинусов $\mathbf{C}_{q,c}$, аналогично [11]

$$\mathbf{C}_{q,c} = \begin{bmatrix} \frac{1}{\sin \Theta} \cdot (h_{21}^c \cdot h_{32}^c - h_{31}^c \cdot h_{22}^c) & h_{11}^c & \frac{1}{\sin \Theta} \cdot (h_{12}^c - \cos \Theta \cdot h_{11}^c) \\ \frac{1}{\sin \Theta} \cdot (h_{31}^c \cdot h_{12}^c - h_{11}^c \cdot h_{32}^c) & h_{21}^c & \frac{1}{\sin \Theta} \cdot (h_{22}^c - \cos \Theta \cdot h_{21}^c) \\ \frac{1}{\sin \Theta} \cdot (h_{11}^c \cdot h_{22}^c - h_{21}^c \cdot h_{12}^c) & h_{31}^c & \frac{1}{\sin \Theta} \cdot (h_{32}^c - \cos \Theta \cdot h_{31}^c) \end{bmatrix}, \quad (7)$$

где h_{ij}^c – элементы векторов $\mathbf{h}_{c_1}^S, \mathbf{h}_{c_2}^{et}$.

Прогнозируемое положение орта $\mathbf{h}_{in_1}^R$ кинетического момента рабочего гироскопа ЭСГ₁ в ИСК вычисляется (с учетом (2), (3), (5)) на рабочей частоте следующим образом:

$$d\mathbf{h}_{in_1}^R / dt = \boldsymbol{\omega}_{in_1} \times \mathbf{h}_{in_1}^R; \quad \mathbf{h}_{in_1}^R(t_0) = \mathbf{C}_{c,in}^{et}(t_0) \mathbf{C}_{kn_1,c} \mathbf{h}_{kn_1}^S(t_0), \quad (8)$$

где $\boldsymbol{\omega}_{in_1}$ – расчетные значения систематических дрейфов ЭСГ₁ в ИСК:

$$\omega_{in_1} = C_{c,in} C_{kn_1,c} \cdot \omega_{kn_1} + (C_{in,h}^{et})^T \cdot C_h ; \quad (9)$$

здесь ω_{kn_1} – систематические дрейфы ЭСГ₁ в корпусных осях, КМУ которых, согласно [12–15], вычисляются при стендовых испытаниях и подлежат уточнению при новом запуске прибора; $C_h = [C_E \ C_N \ C_H]^T$ – дополнительно введенные систематические дрейфы ЭСГ в географических осях (необходимость их введения была выявлена в ходе стендовых испытаний бескарданного ЭСГ с полярной ориентацией); $C_{in,h}^{et}$ – матрица направляющих косинусов, определяющая взаимную ориентацию географического сопровождающего трехгранника ENH относительно ИСК, вычисляемая по эталонным значениям координат места объекта $\lambda_{et}, \varphi_{et}$ и звездному времени $S_{гр}$ на гринвичском меридиане.

Положение (построение) ИСК относительно трехгранника $q_1q_2q_3$ характеризуется матрицей $C_{q,in}$, орты-столбцы которой вычисляются по данным $\mathbf{h}_{in_1}^R, \mathbf{h}_{in_2}^R(t) = \mathbf{h}_{in_2}^{et}(t_0) = \text{const}$ согласно принятому условию ортогонализации (6).

Элементы искомой матрицы $C_{c,h}$ направляющих косинусов, определяющей взаимную ориентацию связанной $x_c y_c z_c$ системы координат и географического сопровождающего трехгранника ENH , могут быть вычислены, учитывая (7), в соответствии с матричным соотношением

$$C_{c,h} = C_{in,h}^{et} \cdot C_{q,in} \cdot (C_{q,c})^T, \quad (10)$$

откуда текущее значение курса объекта вычисляется как

$$K = \arctg(d_{12} / d_{22}), \text{ при } \arctg(d_{12} / d_{22}) < 0, \\ K = \arctg(d_{12} / d_{22}) + 2\pi, \quad (11)$$

где d_{ij} – соответствующие элементы матрицы $C_{c,h}$.

Обработка информации в режимах калибровки и коррекции системы

Измерения выполняются в соответствии с выражением

$$z1 = (\mathbf{h}_{in_1}^R \times \mathbf{h}_{in_2}^R) - (\mathbf{h}_{c_1}^S \times \mathbf{h}_{c_2}^R) = \cos \Theta_R - \cos \Theta_S, \\ z2 = h_{int1_1}^R - h_{int1_1}^{et}, \quad (12)$$

где $z1$ – скалярное измерение, представляющее собой разность косинусов расчетного Θ_R и измеренного Θ_S угла между ортами \mathbf{h}_i векторов \mathbf{H}_i кинетических моментов калибруемого «опорного» ($i = 1$) и виртуального ($i = 2$) гироскопов (разность скалярных произведений соответствующих ортов \mathbf{h}_i); $z2$ – измерение как первый элемент вектора

$$z2 = \mathbf{h}_{int_1}^R - \mathbf{h}_{int_1}^{et} = (C_{int_pr}^{int_et} - \mathbf{E}) \mathbf{h}_{int_1}^{et}. \quad (13)$$

Здесь $\mathbf{h}_{in_1}^R, \mathbf{h}_{in_2}^R, \mathbf{h}_{int_1}^R, \mathbf{h}_{int_1}^{et}$ – расчетные \mathbf{h}^R (прогнозируемые) и эталонные \mathbf{h}^{et} значения ортов \mathbf{h}_i ЭСГ в проекциях на оси соответственно ИСК $in_1 in_2 in_3$ и квази-ИСК $int_1 int_2 int_3$; $C_{int_pr}^{int_et}$ – матрица перехода от истинных осей int к их приборной реализации, характеризующая прецессию ЭСГ; \mathbf{E} – единичная матрица. Необходимые для формирования измерений (12) расчетные значения ортов ЭСГ_{*i*} равны

$$\mathbf{h}_{int_i}^R = C_{in,int} \cdot \mathbf{h}_{in_i}^R. \quad (14)$$

Эталонные значения орта $\mathbf{h}_{int_1}^{et}$ для ЭСГ₁ формируются следующим образом:

$$\mathbf{h}_{int_1}^{et} = C_{in,int} C_{c,in} C_{kn_1,c} \mathbf{h}_{kn_1}^S. \quad (15)$$

где матрица $C_{in,int}$ характеризует положение квази-ИСК относительно ИСК и считается равной значению матрицы $(C_{q,in})^T$ в моменты коррекции положения ЭСГ₁. При вычислении значений ортов $\mathbf{h}_{int_1}^{et}, \mathbf{h}_{c_2}^R$ для выполнения измерений (12) необходимо знание матрицы ориентации $C_{c,in}$, согласно (15), (4). В режиме калибровки БГК ее значения вычисляются с использованием эталонных значений курса и координат места, а в режиме коррекции – с использованием приборных значений курса (рассчитанных по (10)–(11)) от БГК и эталонных значений координат места.

Расчетная модель погрешностей

Модель погрешностей $\Delta \mathbf{h}_{int_1}^R$ прогнозирования текущих значений орта $\mathbf{h}_{int_1}^R$ ЭСГ₁ в проекциях на оси квази-ИСК $int_1 int_2 int_3$, требуемая для обработки в ФК-измерений (12), была получена варьированием уравнений (8), (9). Линеаризация измерений (12) и матрицы динамики модели погрешностей системы осуществлялась согласно условию ортогонализации (6) относительно значений орта ЭСГ₁

$$\mathbf{h}_{int_lo}^R = [0 \ 1 \ 0]^T$$

и оценок КМУ ЭСГ₁ на предыдущем шаге решения задачи фильтрации.

Погрешности построения на ЭСГ квази-ИСК могут быть представлены вектором малого поворота $\Lambda_{int} = [\Lambda_{int1} \ \Lambda_{int2} \ \Lambda_{int3}]^T$, характеризующим текущие погрешности построения ИСК в проекциях на оси квази-ИСК.

Анализируя из соотношения (13) матрицу

$$\mathbf{C}_{int_pr}^{int_et} = \mathbf{E} - \delta \mathbf{C}_{int_et}^{int_pr},$$

где

$$\delta \mathbf{C}_{int_et}^{int_pr} = \begin{bmatrix} 0 & -\Lambda_{int3} & \Lambda_{int2} \\ \Lambda_{int3} & 0 & -\Lambda_{int1} \\ -\Lambda_{int2} & \Lambda_{int1} & 0 \end{bmatrix}, \quad (16)$$

можно показать [12], что вблизи точки линеаризации имеют место следующие приближенные соотношения:

$$\begin{aligned} \Lambda_{int1} &= -\Delta h_{int3_1}^R, \\ \Lambda_{int2} &= -\frac{1}{\sin \Theta} \Delta h_{int1_2}^R + \text{ctg} \Theta \cdot \Delta h_{int1_1}^R, \\ \Lambda_{int3} &= \Delta h_{int1_1}^R, \end{aligned} \quad (17)$$

где $\Delta h_{int_j_i}^R$ – составляющие ($j=1,2,3$) векторов погрешностей прогнозирования уходов ЭСГ_i ($i=1,2$) в проекциях на оси квази-ИСК.

Очевидно, что в рассматриваемой задаче при формировании идеального «виртуального» ЭСГ₂ можно положить, что $\Delta h_{int1_2}^R \cong 0$.

Составляющие $\Delta h_{int1_1}^R, \Delta h_{int3_1}^R$ погрешностей прогнозирования ухода ЭСГ₁ в проекциях на оси квази-ИСК были описаны линеаризованной моделью типа [12] с учетом дополнительно введенных дрейфов ЭСГ₁ в географических осях в следующем виде:

$$\begin{aligned} \Delta \dot{h}_{int1}^R &= -\omega_{q1} \frac{\cos \Theta}{\sin \Theta} \Delta h_{int1}^R + \omega_{q3} \frac{\cos \Theta}{\sin \Theta} \Delta h_{int3}^R - (ck_{31}f_{x-ko} + ck_{32}f_{y-ko} + ck_{33}f_{z-ko}) \Delta k0 - \\ &- (ck_{31}f_{x-\mu12} + ck_{32}f_{y-\mu12}) \Delta \mu_{12} - (ck_{32}f_{y-\mu23} + ck_{33}f_{z-\mu23}) \Delta \mu_{23} - \\ &- (ck_{31}f_{x-\mu31} + ck_{33}f_{z-\mu31}) \Delta \mu_{31} - ck_{33}f_{z-v12} \cdot \Delta v_{12} - ck_{31}f_{x-v23} \cdot \Delta v_{23} - ck_{32}f_{y-v31} \cdot \Delta v_{31} - \\ &- (ck_{32} \cdot \omega_{k3} - ck_{33} \cdot \omega_{k2}) \Lambda_1 - (ck_{33} \cdot \omega_{k1} - ck_{31} \cdot \omega_{k3}) \Lambda_2 - (ck_{31} \cdot \omega_{k2} - ck_{32} \cdot \omega_{k1}) \Lambda_3 - \\ &- (ck_{31}f_{x-k1n} + ck_{32}f_{y-k1n} + ck_{33}f_{z-k1n}) \Delta k1n - (ck_{31}f_{x-k2} + ck_{32}f_{y-k2} + ck_{33}f_{z-k2}) \Delta k2 - \\ &- (ck_{31}f_{x-k3n} + ck_{32}f_{y-k3n} + ck_{33}f_{z-k3n}) \Delta k3n - ch_{31} \cdot \Delta C_E - ch_{32} \cdot \Delta C_N - ch_{33} \cdot \Delta C_H; \\ \Delta \dot{h}_{int3}^R &= -\omega_{q3} \frac{\cos \Theta}{\sin \Theta} \Delta h_{int1}^R - \omega_{q1} \frac{\cos \Theta}{\sin \Theta} \Delta h_{int3}^R + (ck_{11}f_{x-ko} + ck_{12}f_{y-ko} + ck_{13}f_{z-ko}) \Delta k0 + \\ &+ (ck_{11}f_{x-\mu12} + ck_{12}f_{y-\mu12}) \rightarrow \Delta \mu_{12} - (ck_{12}f_{y-\mu23} + ck_{13}f_{z-\mu23}) \Delta \mu_{23} + \\ &+ (ck_{11}f_{x-\mu31} + ck_{13}f_{z-\mu31}) \Delta \mu_{31} + ck_{13}f_{z-v12} \cdot \Delta v_{12} + ck_{11}f_{x-v23} \cdot \Delta v_{23} + ck_{12}f_{y-v31} \cdot \Delta v_{31} + \\ &+ (ck_{12} \cdot \omega_{k3} - ck_{13} \cdot \omega_{k2}) \Lambda_1 + (ck_{13} \cdot \omega_{k1} - ck_{11} \cdot \omega_{k3}) \Lambda_2 + (ck_{11} \cdot \omega_{k2} - ck_{12} \cdot \omega_{k1}) \Lambda_3 + \\ &+ (ck_{11}f_{x-k1n} + ck_{12}f_{y-k1n} + ck_{13}f_{z-k1n}) \Delta k1n + (ck_{11}f_{x-k2} + ck_{12}f_{y-k2} + ck_{13}f_{z-k2}) \Delta k2 + \\ &+ (ck_{11}f_{x-k3n} + ck_{12}f_{y-k3n} + ck_{13}f_{z-k3n}) \Delta k3n + ch_{11} \cdot \Delta C_E + ch_{12} \cdot \Delta C_N + ch_{13} \cdot \Delta C_H, \end{aligned} \quad (18)$$

где $\Delta k0, \Delta \mu_{ij}, \Delta v_{ij}, \Delta k1n, \Delta k2, \Delta k3n$ – погрешности априорных значений КМУ ЭСГ₁ в корпусных осях [12], описываемые в расчетной модели винеровскими процессами; f_{i-j} ($i=x,y,z$), ($j=ko, k1n, \dots, \mu, v$)

– функции связи, соответствующие модели корпусных дрейфов ЭСГ [12]; $\Lambda_j (j=1,2,3)$ – погрешности привязки измерительных осей ЭСГ₁ к его корпусным осям; $\Delta C_j (j=E, N, H)$ – погрешности априорных оценок дополнительно введенных в расчетную модель систематических дрейфов ЭСГ₁ в географических осях, описываемые винеровскими процессами или случайными константами; $\omega_{qj}, \omega_{kj} (j=1,2,3)$ – значения прогнозируемых систематических дрейфов ЭСГ₁ в проекциях на оси гироскопического трехгранника $q_1q_2q_3$ и корпусные оси $x_{kn}y_{kn}z_{kn}$ соответственно; ck_{ij}, ch_{ij} – элементы матриц соответственно $C_{kp,int} = C_{in,int} C_{c,in} (C_{c,b})^T C_{kp,b}$ и $C_{h,int} = C_{in,int} (C_{in,h}^{et})^T$ соответственно.

Учитывая (16) и соотношения (17), получим из измерений (12) вблизи точки линеаризации следующие приближенные выражения:

$$\begin{aligned} z1 &= \sin \Theta \cdot \Delta h_{int3_1}^R - \sin \Theta \cdot \Delta h_{int3_1}^S - \Delta h_{int2_2}^R; \\ z2 &= \Delta h_{int1_1}^R - \Delta h_{int1_1}^{et}; \end{aligned} \quad (19)$$

где $\Delta h_{int3_1}^S$ – составляющая вектора погрешностей $\Delta h_{c_1}^S$ ЭСГ₁ в проекциях на оси квази-ИСК; $\Delta h_{int2_2}^R$ – составляющая вектора погрешностей $\Delta h_{c_2}^R$ ЭСГ₂ в проекциях на оси квази-ИСК; $\Delta h_{int1_1}^{et}$ – составляющая погрешностей формирования орта $h_{int1_1}^{et}$ согласно (15).

Проанализируем погрешности, входящие в измерения (19).

Согласно (4) и (15) и учитывая, что $\Delta C_{kn_i,kp} = -\delta C_{kn_i,kp} C_{kn_i,kp}$ [11], получим:

$$\begin{aligned} \Delta h_{int1_1}^S &= C_{c,int} \cdot \Delta h_{c_1}^S, \\ \Delta h_{c_1}^S &= C_{kp,c} (-\delta C_{kn_i,kp} \cdot h_{kp} + C_{kn_1,kp} \cdot \Delta h_{kn_1}^S), \\ \Delta h_{int1_1}^{et} &= C_{h,int} \cdot \delta C_{in,h}^{et} \cdot h_{h_1}^S - C_{h,int} \cdot \delta C_{c,h} \cdot h_{h_1}^S - C_{kp,int} \cdot \delta C_{kn_i,kp} \cdot h_{kp} + C_{kn_1,int} \cdot \Delta h_{kn_1}^S, \end{aligned} \quad (20)$$

где $\Delta h_{kn_1}^S$ – вектор погрешностей измеренных значений направляющих косинусов орта кинетического момента ЭСГ₁ в корпусных осях (погрешности оптико-электронной системы съема данных);

$$\delta C_{kn_1,kp} = \begin{bmatrix} 0 & -\Lambda_3 & \Lambda_2 \\ \Lambda_3 & 0 & -\Lambda_1 \\ -\Lambda_2 & \Lambda_1 & 0 \end{bmatrix}, \quad \Lambda_{-1} = [\Lambda_1 \quad \Lambda_2 \quad \Lambda_3]^T$$

– матрица и вектор малого поворота, характеризующие погрешности формирования матрицы $C_{kn_1,kp}$, обусловленные погрешностями привязки измерительных осей ЭСГ₁ к осям основания $x_c y_c z_c$;

$$\delta C_{c,h} = \begin{bmatrix} 0 & -\Lambda_{h3}^c & \Lambda_{h2}^c \\ \Lambda_{h3}^c & 0 & -\Lambda_{h1}^c \\ -\Lambda_{h2}^c & \Lambda_{h1}^c & 0 \end{bmatrix}, \quad \Lambda_{h,c} = \begin{bmatrix} \Lambda_{h1}^c \\ \Lambda_{h2}^c \\ \Lambda_{h3}^c \end{bmatrix} = \begin{bmatrix} \beta \\ \gamma \\ \Delta K \end{bmatrix}$$

– матрица и вектор погрешностей в построении географического сопровождающего трехгранника (ΔK – погрешность по курсу, β, γ – погрешности построения в системе вертикали места) [11]; $\delta C_{in,h}^{et}$ – матрица, обусловленная погрешностями эталонных координат места, которые вследствие малости формируют шумы измерений; h_{h_1} – значения элементов вектора $h_{kn_1}^S$ в географических осях.

Проанализируем погрешность $\Delta h_{int2_2}^R$, входящую в измерение $z1$. Согласно (4) и (14), имеем:

$$\begin{aligned} h_{c_2}^R &= C_{in,c} \cdot h_{in_2}^R = C_{h,c} C_{in,h}^{et} \cdot h_{in_2}^R, \\ \Delta h_{int2_2}^R &= C_{c,int} \Delta h_{c_2}^R. \end{aligned}$$

Учитывая, что для «виртуального» ЭСГ₂ $\Delta h_{in_2}^R = 0$, получим соотношение

$$\Delta h_{int2_2}^R = C_{h,int}^{et} \delta C_{c,h} \cdot h_{h_2}^R - C_{h,int}^{et} \delta C_{in,h}^{et} \cdot h_{h_2}^R. \quad (21)$$

Анализ полученных выражений показывает, что в режиме выставки и калибровки погрешности $\Delta h_{int2_2}^R$ и $\Delta h_{int1_1}^{et}$ обусловлены, в основном, погрешностями эталонных значений координат места и курса объекта (основания прибора), а также погрешностями измерения углов качки. Погрешности $\Delta h_{int1_1}^{et}$ и $\Delta h_{int3_1}^S$ дополнительно содержат также погрешности привязки измерительных осей ЭСГ₁ к осям основания системы и погрешности списывающих устройств ЭСГ.

В режиме коррекции погрешности $\Delta h_{int2_2}^R$ и $\Delta h_{int1_1}^{et}$ будут дополнительно содержать с весами, соответствующими (21) и (20), погрешность ΔK приборного курса БГТК, которую представим, согласно [11], следующим соотношением:

$$\Delta K = \frac{1}{\cos \varphi} (\cos \lambda_* \cdot \Delta h_{int1_1}^R - \sin \lambda_* \cdot \Delta h_{int3_1}^R) + \operatorname{tg} \varphi \cdot \gamma, \quad (22)$$

где $\lambda_* = \lambda + S_{гр}$ – инерциальная долгота.

Таким образом, линеаризованная расчетная модель погрешностей системы и измерения могут быть представлены в следующем виде:

$$\mathbf{x}_{k+1} = \Phi_{k/k+1} \cdot \mathbf{x}_k + \Gamma_{k+1} \mathbf{w}_k, \quad k = 0, 1, 2, \dots$$

$$\mathbf{z}_{k+1} = \mathbf{H}_{k+1} \mathbf{x}_{k+1} + \mathbf{v}_{k+1},$$

где

$$\mathbf{x} = [\Delta h_1 \quad \Delta h_3 \quad \Delta k_0 \quad \Delta \mu_{12} \quad \Delta \mu_{23} \quad \Delta \mu_{31} \quad \Delta v_{12} \quad \Delta v_{23} \quad \Delta v_{31}]^T$$

$$[\Lambda_1 \quad \Lambda_2 \quad \Lambda_3 \quad \Delta k_{1n} \quad \Delta k_2 \quad \Delta k_{3n} \quad \bar{\beta} \quad \bar{\gamma} \quad \Delta C_E \quad \Delta C_N \quad \Delta C_H]^T$$

– вектор состояния системы, здесь индексы «int», «R» и «1» при $\Delta h_{intj_1}^R$ опущены,

$\Phi_{j/j+1} \cong \mathbf{E}_{n \times n} + \mathbf{F}(t_j) \cdot dT + \frac{1}{2} [\mathbf{F}(t_j) \cdot dT]^2$ – значение переходной матрицы $\Phi_{j/j+1}$ состояния системы на рабочей частоте (шаг dT). Здесь $\mathbf{F}(t_j)$ (20×20) – матрица динамики системы, соответствующая уравнениям (18) и учитывающая винеровский характер изменения КМУ ЭСГ₁;

$$\Phi_{k/j+1} = \Phi_{j/j+1} \cdot \Phi_{k/j},$$

$$\Phi_{k/j} = \mathbf{E} \quad \text{при } j = 0,$$

$$\Phi_{k/j+1} = \Phi_{k/k+1} \quad \text{при } j = k$$

– значение переходной матрицы Φ_{k+1} на шаге T_z поступления измерений; $\Gamma_{k+1} \cong \Phi_{k+1} \cdot dT$ – матрица, определяющая влияние вектора входных шумов \mathbf{w}_k с ковариациями \mathbf{Q}_k ; \mathbf{H}_{k+1} – матрица измерений, соответствующая уравнениям (19)–(22), значения элементов которой различны для режимов калибровки и коррекции; \mathbf{v}_{k+1} – шумы измерений с матрицей ковариаций \mathbf{R}_{k+1} .

Результаты имитационного моделирования

Моделирование проводилось в пакете MATLAB (Simulink) на основе имитационных данных направляющих косинусов и напряжений гироскопа ЭСГ₁, координат места и углов качки. Период модуляционного вращения составлял 10 мин.

Истинные (модельные) значения КМУ в осях корпуса ЭСГ были приняты на уровне: $k_0 = 3,7$ %/ч; $k_{1N} = 0,1$ %/ч; $k_2 = -1$ %/ч; $k_{3N} = 0,2$ %/ч; $k_4 = 3,72$ %/ч; $k_{5N} = 0,3$ %/ч; $m_u = 0,2$ %/ч; $n_u = 0,1$ %/ч. При этом начальные значения погрешностей априорных оценок КМУ находились на уровне 0,05–0,06 %/ч, а погрешностей привязки измерительных осей к корпусным осям ЭСГ были заданы на уровне 10'.

Погрешности выработки углов качки и погрешности списывающих устройств ЭСГ аппроксимировались дискретными белыми шумами величиной 30" и 15" на уровне (1σ) соответственно. Погрешности эталонной информации по курсу и координатам места принимались как дискретные белые шумы величиной 10' и 10 м на уровне (1σ) соответственно.

На рис. 3, а–в, представлены ошибки оценки начальных значений погрешностей КМУ и погрешностей привязок измерительных осей ЭСГ к его корпусным осям. Из рисунков видно, что ошибки КМУ k_{1N} и привязок Λ_i (рис. 3, б, и рис. 3, в, соответственно) измерительных осей к корпусным осям ЭСГ приходят в установившиеся значения примерно за 20 ч после начала калибровки, а ошибки коэффициентов k_0 , μ_{ij} – более чем за 30 ч, вследствие менее эффективной их наблюдаемости.

На рис. 3, г, приведена кривая погрешности по курсу при работе БГК в режиме коррекции с предварительно проведенной калибровкой погрешностей ЭСГ, а на рис. 3, д, – фрагмент кривой на интервале 40–90 ч. При этом начальная погрешность по курсу была задана на уровне 100'.

Характер погрешности ΔK и анализ ковариационного канала работы ФК указывают на наличие компасного эффекта у системы. Время переходного процесса не превышает 25 ч. В установившемся режиме (рис. 3, г) погрешность имеет колебательный характер с периодами прецессионного движения гироскопа и модуляционного вращения корпуса гироскопа.

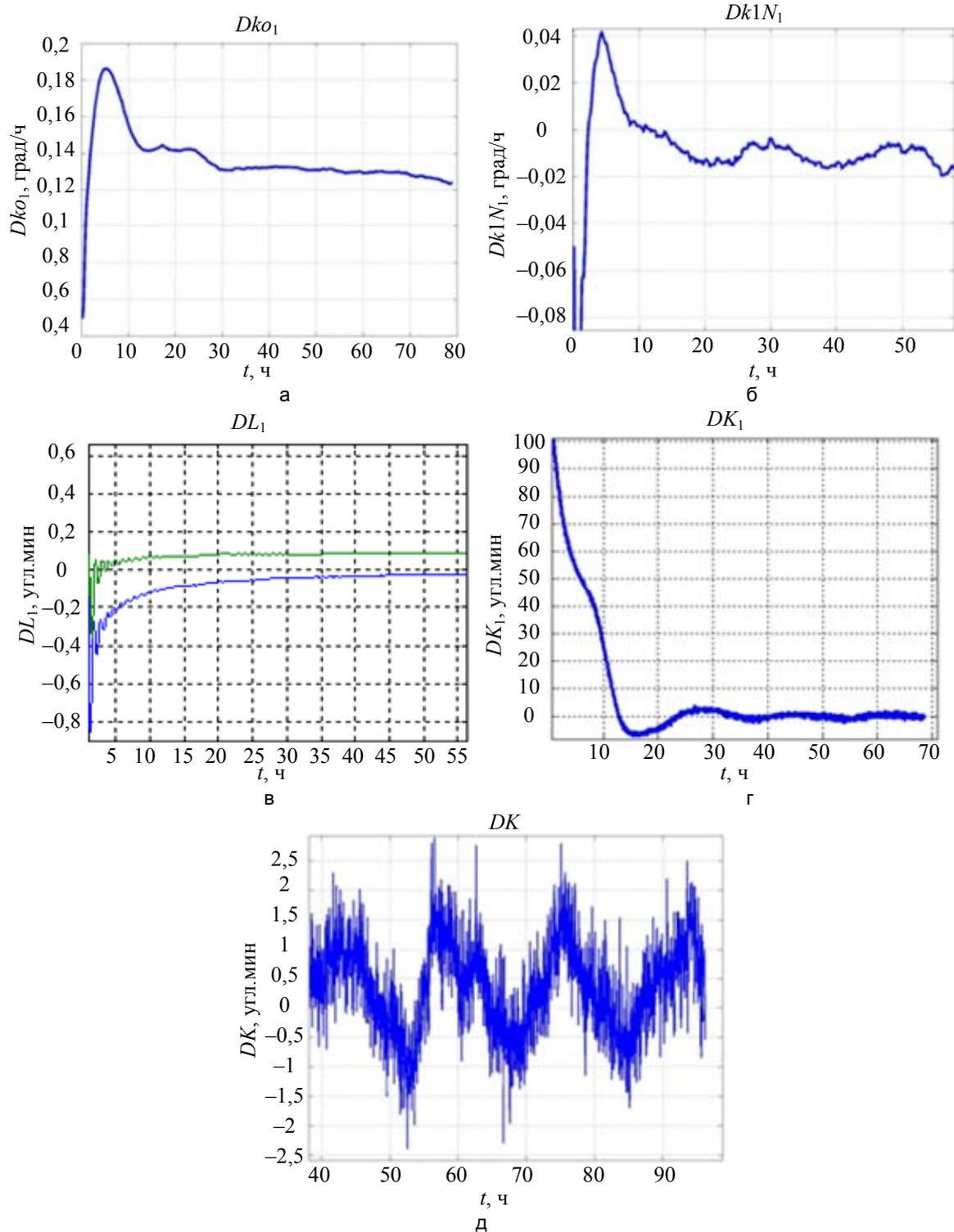


Рис. 3. Результаты моделирования: ошибки оценки начального значения погрешности во время калибровки на протяжении 50 ч: коэффициента k_0 (Dk_0) ($^{\circ}/ч$) (а); коэффициента k_{1N} (Dk_{1N}) ($^{\circ}/ч$) (б); ошибка оценки начальных значений погрешностей привязок измерительных осей ЭСГ к его корпусным осям (DL_1) (') во время калибровки на протяжении 50 ч (в); погрешность (') по курсу (DK_1) в режиме коррекции (г); погрешность определения курса (DK) (') в корректируемом режиме на протяжении 50 ч (д)

Заключение

В заключение сформулируем основные выводы и положения проведенного исследования.

1. Разработаны алгоритмы работы бескарданного гироскопического компаса на электростатическом гироскопе для режимов калибровки и коррекции. Существенное влияние на точность калибровки электроста-

тического гироскопа оказывают погрешности построения вертикали места и погрешности его оптико-электронной системы съема данных.

2. Привлечение внешней информации только о координатах места при работе системы в режиме коррекции обеспечивает ей компасный эффект, т.е. стационарный характер погрешности по курсу в установленном режиме. При этом обеспечивается также уточнение некоторых наблюдаемых коэффициентов модели корпусных дрейфов электростатического гироскопа.
3. Для повышения точности выработки курса необходимо модуляционное вращение корпуса гироскопа, обеспечивающее наблюдаемость коэффициентов модели ухода и погрешностей привязки измерительных осей гироскопа, а также снижение уровня непрогнозируемых составляющих его дрейфа.

Литература

1. Буравлев А.П., Кузин В.М., Ландау Б.Е., Сумароков В.В. Бескарданный электростатический гироскоп с подвесом на двойных электродах // XXVI научно-техническая конференция памяти Н.Н.Острякова: доклады. СПб, 2008. С. 17–18.
2. Emel'yantsev G.I., Landau B.E., Levin S.L., Gurevich S.S., Romanenko S.G. Integrated attitude reference and navigation system for orbital spacecraft // Gyroscopy and Navigation. V. 2. N 3. 2011. P. 146–151.
3. Ландау Б.Е., Белаш А.А., Гуревич С.С., Емельянец Г.И., Левин С.Л., Романенко С.Г. Бескарданная инерциальная система ориентации на электростатических гироскопах для орбитального космического аппарата // Изв. вузов. Приборостроение. 2011. Т. 54. № 6. С. 66–74.
4. Емельянец Г.И., Ландау Б.Е., Левин С.Л., Гуревич С.С., Романенко С.Г. Особенности построения интегрированной системы ориентации и навигации для орбитального космического аппарата // Гироскопия и навигация. 2011. № 1. С. 17–25.
5. Никишин В.Б., Синев А.И., Плотников П.К., Наумов С.Г. Повышение точности подземной навигации на основе интеграции БИНС, одометров и приемников GPS/ГЛОНАСС // Сб. материалов XVII международной конференции по интегрированным навигационным системам. Санкт-Петербург, 2010. С. 169–174.
6. Емельянец Г.И., Лочехин А.В. О погрешностях бескарданного гирогоризонтокомпаса на электростатическом гироскопе и микромеханических датчиках // Изв. вузов. Приборостроение. 2010. Т. 53. № 10. С. 42–48.
7. Блажнов Б.А., Волынский Д.В., Емельянец Г.И., Несенюк Л.П., Степанов А.П. Интегрированная инерциально-спутниковая система ориентации и навигации с микромеханическим инерциальным модулем. Результаты испытаний на автомобиле // Гироскопия и навигация. 2008. № 4 (63). P. 77.
8. Blazhnov B.A., Emeliantsev G.I., Koshayev D.A., Semenov I.V., Stepanov A.P., Zhilinskii V.M., Korotkov A.N., Timofeev E.A., Tsekhanovich G.S. Integrated tightly coupled inertial satellite orientation and navigation system // Gyroscopy and Navigation. 2010. V. 1. N 1. P. 10–18.
9. Volynskii D.V., Odintsov A.A., Dranitsyna E.V., Untilov A.A. Calibration of fiber-optic gyros within strapdown inertial measurement units // Gyroscopy and Navigation. 2012. V. 3. N 3. P. 194–200.
10. Gusinsky V.Z., Lesyuchevsky V.M., Litmanovich Yu.A. Calibration and alignment of inertial navigation systems with multivariate error state vector // Proc. 4th St. Petersburg International Conference on Integrated Systems. St. Petersburg, 1997. P. 371–378.
11. Анучин О.Н., Емельянец Г.И. Интегрированные системы ориентации и навигации для морских подвижных объектов. СПб: ЦНИИ «Электроприбор», 2003. 390 с.
12. Ландау Б.Е., Гуревич С.С., Емельянец Г.И., Левин С.Л., Романенко С.Г., Одинцов Б.В. Результаты калибровки электростатических гироскопов в бескарданной инерциальной системе ориентации // Материалы XV международной конференции по интегрированным навигационным системам. Санкт-Петербург, 2008. С. 122–129.
13. Landau B.E., Gurevich S.S., Emeliantsev G.I., Levin S.L., Romanenko S.G. Calibrating the error of a strapdown ESG-based attitude reference system under conditions of orbital flight // Gyroscopy and Navigation. 2010. V. 1. N 3. P. 176–182.
14. Ландау Б.Е., Левин С.Л., Гуревич С.Г., Емельянец Г.И., Завгородний В.И., Романенко С.Г., Одинцов Б.В. Наземная отработка методики полетной калибровки БИСО на ЭСГ для орбитальных космических аппаратов с произвольной ориентацией // Материалы XIX Санкт-Петербургской международной конференции по интегрированным навигационным системам. Санкт-Петербург, 2012. С. 127–135.
15. Emeliantsev G.I., Landau B.E., Levin S.L., Romanenko S.G. Refining the drift model of a gimballess inertial attitude control system based on electrostatic gyros: methods of calibration on a ground-based test bench and on board an orbiting space vehicle // Gyroscopy and Navigation. 2010. V. 1. N 2. P. 134–140.

Емельянец Геннадий Иванович – доктор технических наук, профессор, главный научный сотрудник, ОАО «Концерн «ЦНИИ «Электроприбор», 197046, Санкт-Петербург, Россия; профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия, emeliantsev_gi@mail.ru

Медведков Андрей Александрович – младший научный сотрудник, аспирант, ОАО «Концерн «ЦНИИ

- Цай Тицзин** – «Электроприбор», 197046, Санкт-Петербург, Россия; ассистент кафедры, Университет ИТМО, 197101, Санкт-Петербург, Россия, medvedcov@yandex.ru
- Gennady I. Emelyantsev** – профессор, профессор кафедры, Юго-Восточный университет, 210096, г. Нанкин, Китай, caitij@seu.edu.cn
- Andrei A. Medvedkov** – D.Sc., Professor, chief scientific researcher, State Research Center of the Russian Federation Concern CSRI Elektropribor, JSC, 197046, Saint Petersburg, Russia; Professor, ИТМО University, 197101, Saint Petersburg, Russia, emeliantsev_gi@mail.ru
- Cai Tijing** – junior scientific researcher, State Research Center of the Russian Federation “Concern CSRI Elektropribor”, JSC, 197046, Saint Petersburg, Russia; postgraduate, ИТМО University, 197101, Saint Petersburg, Russia, medvedcov@yandex.ru
- Professor, Southeast University, Nanjing, 210096, P.R.China, caitij@seu.edu.cn

*Принято к печати 01.07.14
Accepted 01.07.14*

УДК 681.2.087

МОДЕЛИРОВАНИЕ И ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ЧУВСТВИТЕЛЬНОГО ЭЛЕМЕНТА ВОЛОКОННО-ОПТИЧЕСКОГО ГИДРОФОНА

М.Е. Ефимов^а, М.Ю. Плотников^а, А.В. Куликов^а^а Университет ИТМО, 197101, Санкт-Петербург, Россия, spbeok@hotmail.com

Аннотация. Предложена модель волоконно-оптического чувствительного элемента гидрофона. Конструкция гидрофона содержит податливый сердечник из полимерного материала с регламентированными упругими свойствами, на который намотано оптическое волокно. При моделировании использован встроенный модуль среды Comsol Multiphysics – Acoustic Solid Interaction, позволяющий оценить воздействие акустического поля различной частоты и амплитуды на величину деформации поверхности чувствительного элемента. Предложенная модель позволяет имитировать работу гидрофона в различных средах, материалы и размеры чувствительного элемента выбираются на этапе проектирования с целью обеспечения требуемых эксплуатационных характеристик – частотной характеристики и чувствительности волоконно-оптического гидрофона.

Правильность построения модели проверена сопоставлением результатов компьютерного моделирования и экспериментального исследования макета в акустическом бассейне. Макет представлял собой волоконный фазовый интерферометрический гидрофон на брэгговских решетках. Чувствительный элемент выполнен в виде цилиндрического сердечника, на который намотано оптическое волокно. Характеристики сердечника: показатель затухания материала (damping) – 0,1, модуль Юнга сердечника – 6 МПа, коэффициент Пуассона – 0,49. Макет испытан в экспериментальном бассейне, конструкция которого позволяет проводить измерения на частотах выше 3000 Гц при отсутствии переотражений акустического сигнала. Оценка воздействия акустического поля проведена с помощью аттестованного пьезоэлектрического гидрофона: амплитуда акустического поля плоской волны 0,5 и 1 Па, частоты акустического воздействия 3000–8000 Гц. Согласно полученным результатам, чувствительность изготовленного макета составила 0,1 рад/Па на частоте 3000 Гц. Исследования показали, что чувствительность моделируемого волоконно-оптического гидрофона будет уменьшаться с ростом частоты гидроакустического воздействия. На частоте 8000 Гц чувствительность снижается до 0,01 рад/Па. Результаты макетных испытаний подтвердили адекватность компьютерной модели, что дает возможность рекомендовать предложенную модель при разработке и исследовании волоконно-оптических гидрофонов.

Ключевые слова: волоконно-оптический гидрофон, моделирование, Comsol Multiphysics, чувствительность.

Благодарности. Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (проект №02.G25.31.0044).

MODELING AND EXPERIMENTAL STUDY OF A FIBER OPTIC HYDROPHONE SENSING ELEMENT

M.E. Efimov^a, M.Yu. Plotnikov^a, A.V. Kulikov^a^a ITMO University, 197101, Saint Petersburg, Russia, spbeok@hotmail.com

Abstract. A model of the fiber-optic hydrophone sensor is suggested. Hydrophone construction comprises a malleable core made of a polymeric material with regulated elastic properties to which the optical fiber is wound. The built-in module of Comsol Multiphysics - Acoustic Solid Interaction is used in the simulation; it evaluates the impact of the acoustic field of different frequencies and amplitudes on the value of the sensor surface deformation. The proposed model gives the possibility for simulating the hydrophone in various environments; materials and dimensions of sensor are selected at the design stage to ensure the required performance: frequency response and sensitivity of fiber optic hydrophone. Correctness of the model construction was verified by results comparison of the computer simulation and experimental study in the acoustic pool. The prototype was represented as the phase interferometric fiber-optic hydrophone on the Bragg gratings. The sensing element is formed as a cylindrical core round which the optical fiber is wound. Core characteristics are: the material attenuation (damping) – 0.1, Young's modulus of the core - 6 MPa, Poisson's ratio - 0.49. The prototype was tested in the experimental pool, which design makes it possible to carry out measurements at frequencies above 3 kHz in the absence of reflections of the acoustic signal. The impact assessment of the acoustic field is carried out by means of an approved piezoelectric hydrophone: the amplitude of the acoustic field of a plane wave is 0.5 and 1 Pa, frequencies of the acoustic impact are 3000 - 8000 Hz. According to the findings fabricated prototype sensitivity was equal to 0.1 rad / Pa at the frequency of 3 kHz. Studies have shown that the sensitivity of the simulated fiber optic hydrophone will decrease with increasing frequency of hydroacoustic exposure. At 8 kHz frequency the sensitivity is decreased to 0.01 rad / Pa. Prototype testing results have confirmed the adequacy of the computer model that makes it possible to recommend the proposed model for the development and study of fiber optic hydrophones.

Keywords: fiber-optic hydrophone, modeling, Comsol Multiphysics, sensitivity.

Acknowledgements. The work was financially supported by the Ministry of Education and Science of the Russian Federation (project №02.G25.31.0044).

Введение

В настоящее время хорошо изучены и получили широкое распространение в гидроакустике пьезо-керамические чувствительные элементы [1]. Их главными недостатками являются большие массогабаритные параметры, сложность объединения их в гидроакустические комплексы (сложность их мультиплексирования) [2], высокая чувствительность к электромагнитному воздействию и небольшая акустическая чувствительность по сравнению с волоконно-оптическими гидрофонами.

Сегодня все более широкое распространение получают оптические датчики различных типов [3] благодаря высокой чувствительности измерений, большому динамическому диапазону, коррозионной стойкости, взрыво- и пожаробезопасности, а также устойчивости к электромагнитному излучению [4]. Эта тенденция коснулась и гидроакустики. Датчики, фиксирующие звуковое давление в водной среде, называются гидрофонами.

Целью настоящей работы является создание модели чувствительного элемента волоконно-оптического гидрофона (ВОГ) и сопоставление результатов моделирования в среде Comsol Multiphysics с результатами, полученными в ходе экспериментального исследования макета чувствительного элемента.

Волоконно-оптические датчики обладают рядом преимуществ по сравнению с пьезокерамическими чувствительными элементами. ВОГ обладают лучшими метрологическими параметрами, имеют более выгодные весогабаритные характеристики, малое энергопотребление, независимость от влияния электромагнитного поля. Все эти преимущества делают волоконно-оптические датчики более перспективными для использования в газонефтяной промышленности, а также в геолокационных и гидролокационных комплексах [5]. Моделирование в среде Comsol Multiphysics позволяет провести предварительную оценку частотной характеристики чувствительного элемента ВОГ без непосредственного изготовления опытного образца.

Фазовые интерферометрические датчики

В любых волоконно-оптических датчиках оптическое волокно может служить либо средой для передачи информации, либо чувствительным элементом. В последнем случае датчик обычно строится на основе волоконно-оптического интерферометра, одно волокно в котором служит сигнальным плечом, а другое – опорным. На рис. 1 представлена схема работы фазовых интерферометрических датчиков на брэгговских решетках с интерферометром Маха–Цендера.

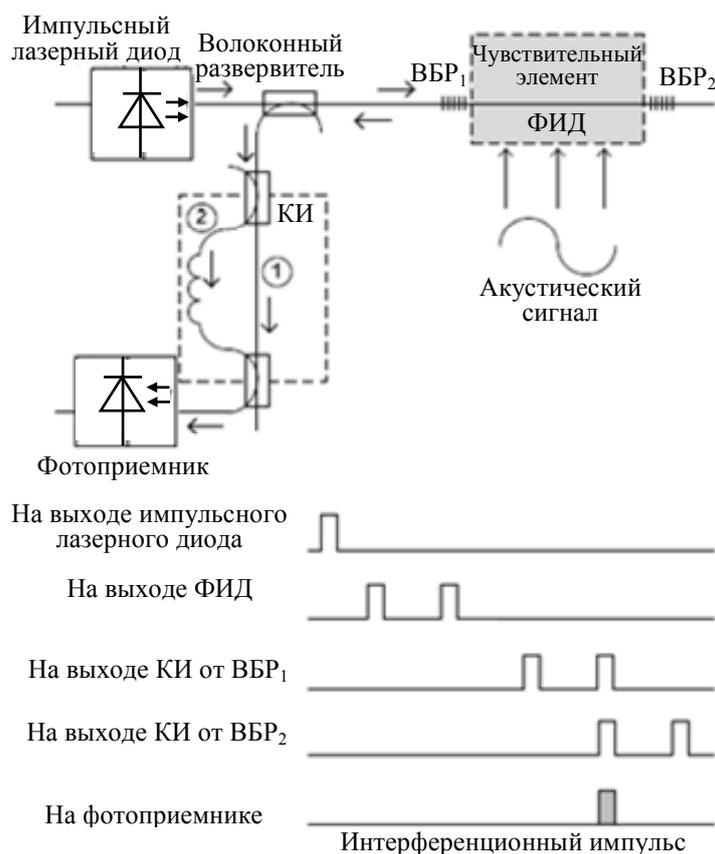


Рис. 1. Принцип действия фазовых интерферометрических датчиков на брэгговских решетках: КИ – компенсационный интерферометр; ВБР – волоконная брэгговская решетка; ФИД – фазовый интерферометрический датчик

Измерение параметров среды в этом случае осуществляется за счет изменения оптической длины пути волокна или фазы излучения в нем. Модуляция оптической длины или фазы при этом может происходить либо через вспомогательный элемент, связаный с волокном, либо через изменение параметров самого волокна. Например, в волоконно-оптических гидрофонах обычно используется интерферометр Маха–Цендера и когерентный источник излучения, свет из которого распространяется по обоим плечам

интерферометра. Намотанное на катушку оптическое волокно, выступающее в роли чувствительного элемента, подвергается акустическому воздействию, которое модулирует фазу проходящего по нему света. Изменение фазы света меняет итоговую интерференционную картину, что регистрируется фотоприемником (ФП) [6].

Акустическое воздействие изменяет диаметр сердцевинки, длину и показатели преломления оболочки и сердцевинки волокна, тем самым осуществляя фазовую модуляцию распространяющегося по волокну светового излучения. Изменение длины волокна в этом случае является преобладающим эффектом [7].

Изменение во времени деформации чувствительного элемента датчика и показателя преломления оптического волокна в результате внешнего воздействия вызывают изменение разности фаз между интерферирующими импульсами. Интерференционный сигнал преобразуется ФП в изменение величины тока, которое после демодуляции несет информацию об акустическом сигнале [8].

Моделирование чувствительного элемента ВОГ

Чувствительность волоконно-оптических датчиков давления зависит от многих факторов, таких как длина намотанного оптического волокна, физические свойства материала, из которого он изготовлен, конструкция чувствительного элемента и т.д. [9]. Под физическими свойствами материала подразумеваются такие его параметры, как модуль Юнга и коэффициент Пуассона, полностью характеризующие упругие свойства изотропного материала. Чувствительность гидрофона тем выше, чем больше модуль Юнга, коэффициент Пуассона должен быть примерно равен 0,5.

Так как чувствительность ВОГ зависит от упругих свойств материала изготовления, было проведено исследование деформации модели от частоты воздействия акустического поля. С этой целью была реализована модель в среде Comsol Multiphysics.

Моделируемый чувствительный элемент ВОГ представляет собой податливый цилиндрический сердечник, размеры которого задаются параметрически, покрытый более жестким полимерным материалом, на который наматывается оптическое волокно (рис. 2). Полученный чувствительный элемент с намотанным оптическим волокном для защиты покрывается тонким слоем более жесткого полимерного материала, толщина этого слоя также задается параметрически.

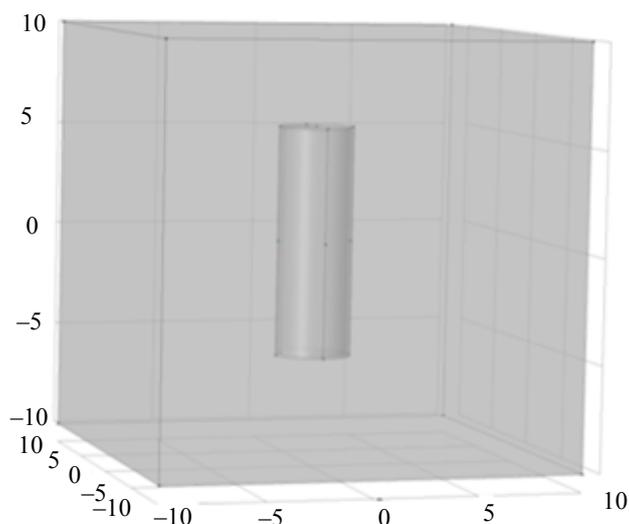


Рис. 2. Геометрическая модель ВОГ (масштаб: 5см/клетка)

Для построения модели был использован встроенный модуль среды Comsol Multiphysics – Acoustic Solid Interaction, Frequency Domain, позволяющий моделировать акустическое поле плоской волны заданной амплитуды в диапазоне частот и оценивать ее влияние на чувствительный элемент [10]. В модели использовались следующие параметры: амплитуда акустического поля плоской волны – 0,5 и 1 Па; частота акустического воздействия – 3000–8000 Гц; показатель затухания материала (damping) – 0,1; модуль Юнга сердечника – 6 МПа; коэффициент Пуассона сердечника – 0,49; модуль Юнга полимерного покрытия – 3,2 ГПа; коэффициент Пуассона полимерного покрытия – 0,35. При изменении этих параметров результат будет меняться соответственно. При уменьшении коэффициента Пуассона и (или) модуля Юнга величина деформации будет уменьшаться. При изменении геометрической формы будет меняться характер частотной характеристики. Частотная характеристика становится менее плавной при уменьшении показателя затухания материала и более плавной при его увеличении [11].

Параметры свойств сердечника и полимерного покрытия были заданы таким образом, чтобы они были максимально близки к имеющимся в лаборатории материалам, из которых позже планировалось изготовить экспериментальный образец. Диапазон исследуемых частот акустического воздействия был

выбран в соответствии с параметрами бассейна, в котором проводились испытания экспериментального образца. Показатель затухания был выбран в соответствии с используемыми параметрами (как у резины – 0,1). Амплитуды акустического воздействия соответствовали возможностям генератора, используемого для проведения экспериментального исследования.

На рис. 3, 4 представлены результаты моделирования. Рис. 3 отображает деформацию чувствительного элемента на частотах 3000 Гц (рис. 3, а) и 8000 Гц (рис. 3, б) при амплитуде акустического воздействия 1 Па. Из рис. 3 можно сделать вывод о большей деформации на 3000 Гц и о зонах максимальной деформации поверхности гидрофона. Рис. 4 отображает среднеквадратичную величину смещения центральной точки на боковой поверхности цилиндрической формы чувствительного элемента волоконного гидрофона при амплитуде акустического воздействия 0,5 и 1 Па.

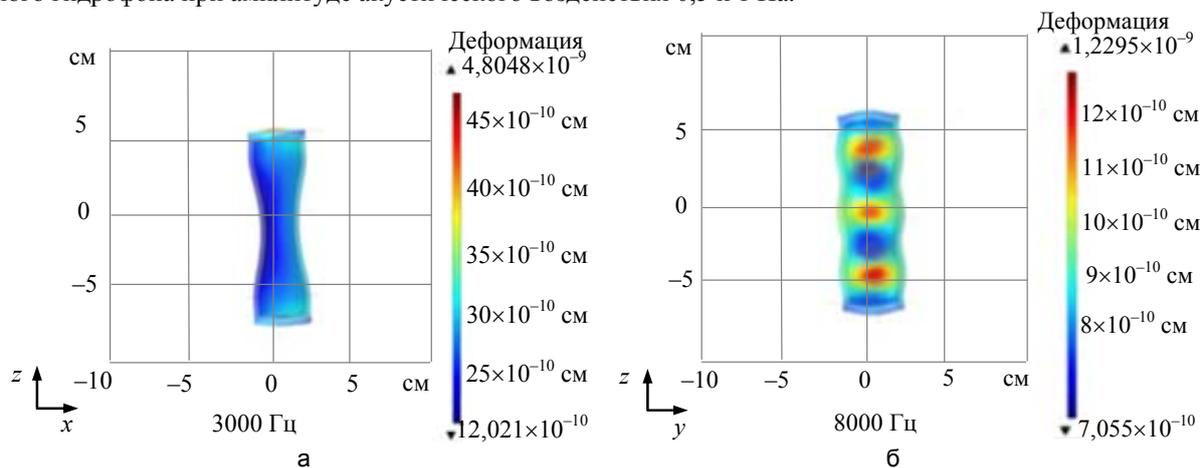


Рис. 3. Деформация поверхности ВОГ на частотах: 3000 Гц (а); 8000 Гц (б)

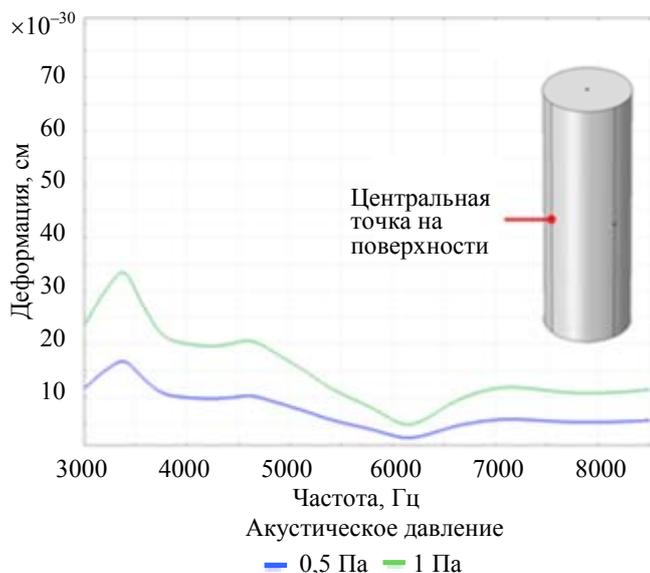


Рис. 4. Среднеквадратичная величина деформации в центральной точке на боковой поверхности цилиндрической формы чувствительного элемента

Для более детального анализа полученных результатов моделирования рассмотрим график зависимости среднеквадратичной величины смещения в центральной точке на боковой поверхности цилиндрической формы чувствительного элемента от частоты и амплитуды акустического поля (рис. 4). Представленная на рис. 4 частотная характеристика среднеквадратичной величины деформации в полосе частот 3000–8000 Гц показывает, что чувствительность ВОГ снижается с увеличением частоты воздействия акустического поля при амплитудах воздействия акустического поля 0,5 Па и 1 Па, а деформация поверхности модели гидрофона при амплитуде акустического воздействия 0,5 Па в два раза меньше, чем при амплитуде 1 Па.

Полученная частотная характеристика деформации модели ВОГ должна согласоваться с частотной характеристикой чувствительности экспериментального образца, полученной в ходе экспериментального исследования, что докажет достоверность предложенного метода разработки и исследования ВОГ.

Создание и экспериментальное исследование чувствительного элемента ВОГ

Для проверки полученных результатов моделирования был изготовлен макет волоконного фазового интерферометрического гидрофона [12] на брэгговских решетках (рис. 5). При изготовлении экспериментального образца чувствительного элемента ВОГ был залит цилиндрический сердечник, на который было намотано оптическое волокно [13]. Полученный чувствительный элемент с намотанным оптическим волокном для защиты покрывается тонким слоем более жесткого полимерного материала [14].

Изготовление чувствительного элемента ВОГ проходило в несколько этапов:

- заливка цилиндрического сердечника;
- намотка оптического волокна на цилиндрический сердечник;
- покрытие тонким слоем более жесткого полимера.



Рис. 5. Макет волоконно-оптического гидрофона

Изготовленный макет ВОГ был испытан в бассейне, конструкция которого позволяет проводить измерения на частотах выше 3000 Гц (из-за большого количества переотражений на более низких частотах). Для оценки воздействия акустического поля был использован аттестованный пьезоэлектрический гидрофон, располагающийся в одной плоскости с макетом ВОГ на расстоянии 0,2 м. Излучатель был установлен на расстоянии 6 м от испытуемого волоконного гидрофона. Результаты испытаний представлены на рис. 6 [15].

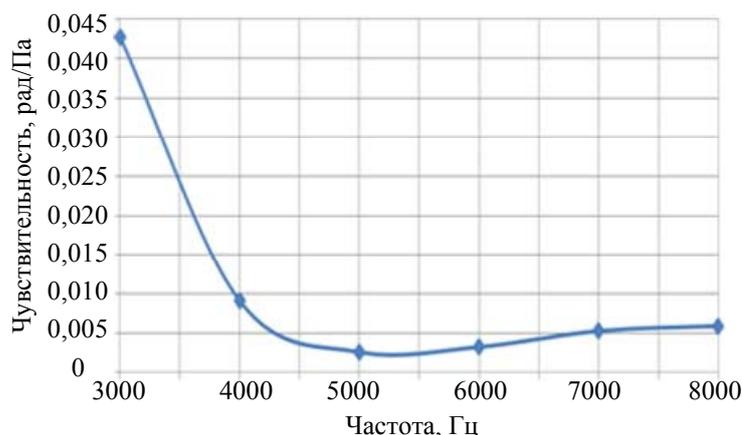


Рис. 6. График зависимости чувствительности гидрофона от частоты

Согласно полученной экспериментальной частотной характеристике макета ВОГ, чувствительность составила 0,1 рад/Па на частоте 3000 Гц и снизилась до 0,01 рад/Па – на 8000 Гц. Приведенные значения чувствительности экспериментального образца в рад/Па соответствуют отношению величины фазового сдвига (рад) в чувствительном плече интерферометрического датчика к величине акустического давления (Па).

При сравнении полученных экспериментальных данных с результатами моделирования было установлено их хорошее совпадение. Некоторая рассогласованность полученных результатов может быть обусловлена небольшим различием свойств материалов макета-аналога и модели. Это позволяет утверждать о достоверности полученных результатов моделирования в среде Comsol Multiphysics и корректности модели. Полученные результаты представляют практический интерес с точки зрения разработки новых чувствительных элементов и могут быть использованы для создания чувствительных элементов волоконно-оптических гидрофонов.

Заключение

Предложена компьютерная модель ВОГ с заданными параметрами, позволяющая качественно оценить частотную характеристику чувствительности ВОГ без его изготовления и испытаний.

Для подтверждения результатов моделирования был изготовлен и экспериментально исследован

макет чувствительного элемента в диапазоне частот 3000–8000 Гц. Чувствительность макета волоконного гидрофона составила 0,1 рад/Па на частоте 3000 Гц и снизилась до 0,01 рад/Па на частоте 8000 Гц.

Результаты экспериментальных исследований показали, что характеристика чувствительности макета ВОГ качественно совпадает с результатами моделирования, что позволяет судить о правильности построенной модели и корректности полученных в результате моделирования данных.

Актуальность проведенных исследований определяется необходимостью построения адекватных моделей волоконно-оптических чувствительных элементов, востребованной для проектирования гидрофонов с заданными параметрами.

Литература

1. Рыбнянец А.Н., Сахненко В.П. Современное состояние и перспективы развития пьезоэлектрической керамики за рубежом // Микросистемная техника. 2002. № 3. С. 16–22.
2. Варжель С.В., Куликов А.В., Брунов В.С., Асеев В.А. Метод понижения коэффициента отражения волоконных брэгговских решеток с помощью эффекта фотохромизма // Научно-технический вестник информационных технологий, механики и оптики. 2012. № 1 (77). С. 151–152.
3. Fiber Optic Hydrophones. Stockbridge: 2011. 5 p.
4. Wurster C., Staudenraus J., Eisenmenger W. Fiber optic probe hydrophone // Proc. of the IEEE Ultrasonics Symposium. 1994. V. 2. P. 941–944.
5. Cole J.H., Kirkendall C., Dandridge A., Cogdell G., Giallorenzi T.G. Twenty-five years of interferometric fiber optic acoustic sensors at the naval research laboratory // Washington Academy of Sciences. 2004. P. 40–57.
6. Куликов А.В. Волоконно-оптические акустические сенсоры на брэгговских решетках: автореф. дис. ... канд. техн. наук. СПб: НИУ ИТМО, 2012. 20 с.
7. Bucaro J.A., Lagakos N., Cole J.H., Giallorenzi T.G. Fiber optic acoustic transduction // Physical Acoustics. 1982. V. 16. P. 385–457.
8. Giallorenzi T.G., Bucaro J.A., Dandridge A., Sigel G.H. Jr., Cole J.H., Rashleigh S.C., Priest R.G. Optical fiber sensor technology // IEEE Journal of Quantum Electronics. 1982. V. QE-18. N 4. P. 626–665.
9. Guo K., Zhang M., Liao Y., Lai S., Wang Z., Tang J. Fiber-optic hydrophone with increased sensitivity // Proceedings of SPIE – The International Society for Optical Engineering. 2006. V. 6293. Art. 629312.
10. Acoustic-Structure Interaction [Электронный ресурс]. 2012. Режим доступа: https://www.comsol.com/model/download/121005/models.aco.acoustic_structure.pdf, свободный. Яз. англ. (дата обращения 10.05.2014).
11. Wang Y., Wang C. Simulation of high-sensitivity hydrophone based on ANSYS // Proc. of International Conference on Mechanical Engineering and Material Science (MEMS 2012). 2012. P. 697–699.
12. Jameson P., Jameson P., Burton T., Ordubadi A., Africk S. Design of rubber mandrel fiber optic hydrophones // Journal of Acoustical Society of America. 1981. V. 70. P. 100.
13. Аксарин С.М., Архипов С.В., Варжель С.В., Куликов А.В., Стригалева В.Е. Исследование зависимости параметров анизотропных одномодовых волоконных световодов от диаметра намотки // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 6 (88). С. 22–26.
14. Балицкий В.А., Городецкий В.С., Лямшев Л.М. и др. // Акустический журнал. 1985. Т. 31. № 5. С. 47.
15. Бутусов М.М., Латинский В.С., Тарасюк Ю.Ф., Галкин С.Л. Волоконная оптика в судовом приборостроении. Л.: Судостроение, 1990. 82 с.

Ефимов Михаил Евгеньевич – студент, Университет ИТМО, 197101, Санкт-Петербург, Россия, spbeok@hotmail.com

Плотников Михаил Юрьевич – аспирант, Университет ИТМО, 197101, Санкт-Петербург, Россия, plotnikov-michael@yandex.ru

Куликов Андрей Владимирович – соискатель, Университет ИТМО, 197101, Санкт-Петербург, Россия, a.kulikov86@gmail.com

Mikhail E. Efimov – student, ITMO University, 197101, Saint Petersburg, Russia, spbeok@hotmail.com

Mikhail Yu. Plotnikov – postgraduate, ITMO University, 197101, Saint Petersburg, Russia, plotnikov-michael@yandex.ru

Andrey V. Kulikov – applicant, ITMO University, 197101, Saint Petersburg, Russia, a.kulikov86@gmail.com

Принято к печати 30.06.14

Accepted 30.06.14

УДК 37:004

К ВОПРОСУ ПОВЫШЕНИЯ РЕЗУЛЬТАТИВНОСТИ
МАССОВОГО ОНЛАЙН-КУРСАЛ.С. Лисицына^а, А.А. Першин^а, В.Л. Усков^б^а Университет ИТМО, 197101, Санкт-Петербург, Россия, lisizina@mail.ifmo.ru^б Бредли университет, 61625, Пеория, штат Иллинойс, США

Аннотация. Поставлена проблема повышения результативности массового онлайн-курса для обучения в области программирования и IT-технологий, определены некоторые пути ее решения. Детально проанализирована проблема организации мгновенной обратной связи в онлайн-курсе. Предложено решение для разработки обучающих интерфейсов курса на основе механики «сделал – увидел результат – понял, что сделал правильно или неправильно», которое ускоряет не только визуализацию результата, но и проверку правильности кода. Рассмотрены и проанализированы три варианта сценариев запуска визуализации и проверки правильности кода. Сделан вывод о том, что для повышения интерактивности онлайн-курса необходимо реализовать мгновенную обратную связь автоматически, без дополнительных действий со стороны обучаемого. При этом сценарий запуска визуализации и проверки кода может содержать этап анализа кода на корректность, позволяющий обучаемому понять правильность хода выполнения данного задания. Предложен подход для поэтапного решения больших по объему заданий, который в сочетании с автоматической мгновенной связью повышает наглядность и скорость обучения. Приведены результаты реализации игровой механики перфекционизма в массовом онлайн-курсе Университета ИТМО по подготовке (повышению квалификации) разработчиков веб-интерфейсов на основе технологий HTML5 и CSS3. Результаты экспериментального исследования на 5588 человек подтвердили, что использование игровой механики перфекционизма не только повышает уровень вовлечения в обучение, но и способствует более глубокому усвоению материала. Приведены результаты экспериментального исследования, которые свидетельствуют о существенном преимуществе синхронного обучения по сравнению с асинхронным: количество успешно завершивших онлайн-курс выросло до 27,9%. Материалы статьи, будут полезны исследователям, занимающимся разработкой проблем электронного обучения при разработке новых массовых онлайн-курсов.

Ключевые слова: массовый онлайн-курс, результативность электронного обучения, мгновенная обратная связь, обучающие интерфейсы, сценарий запуска визуализации и проверки правильности кода, игровые механики, перфекционизм, синхронное обучение.

NEW APPROACHES TO EFFICIENCY OF MASSIVE ONLINE COURSE

L.S. Lisitsyna^а, A.A. Pershin^а, V.L. Uskov^б^а ITMO University, 197101, Saint Petersburg, Russia, lisizina@mail.ifmo.ru^б Bradley University, 61625, Peoria, IL, USA

Abstract. This paper is focused on efficiency of e-learning, in general, and massive online course in programming and information technology, in particular. Several innovative approaches and scenarios have been proposed, developed, implemented and verified by the authors, including 1) a new approach to organize and use automatic immediate feedback that significantly helps a learner to verify developed code and increases an efficiency of learning, 2) a new approach to construct learning interfaces – it is based on “develop a code – get a result – validate a code” technique, 3) three scenarios of visualization and verification of developed code, 4) a new multi-stage approach to solve complex programming assignments, 5) a new implementation of “perfectionism” game mechanics in a massive online course. Overall, due to implementation of proposed and developed approaches, the efficiency of massive online course has been considerably increased, particularly 1) the additional 27.9 % of students were able to complete successfully “Web design and development using HTML5 and CSS3” massive online course at ITMO University, and 2) based on feedback from 5588 students a “perfectionism” game mechanics noticeably improves students’ involvement into course activities and retention factor.

Keywords: massive online course, efficiency of e-learning, immediate feedback, learning interfaces, scenario of code visualization and verification, game mechanics, perfectionism, synchronous learning.

Введение

С развитием электронного обучения и появлением обширного мирового рынка образовательных услуг на основе технологии MOOC (Massive Open Online Courses) внимание разработчиков онлайн-курсов обращено к вопросам повышения их результативности (количество успешно завершивших такой курс не превышает и 5%).

Как построить результативный онлайн-курс? Каким образом усилить интерес у обучаемых, мотивируя их к успешному завершению такого курса? Во-первых, такой курс должен иметь актуальное проблемно-ориентированное содержание, построенное на планировании результатов обучения, ожидаемых потенциальным обучаемым. С методикой создания такого курса, а также с подходом к моделированию ожидаемых результатов обучения в нем можно ознакомиться в работах авторов [1–4]. Во-вторых, для

того, чтобы обучение было активным, в нем должна быть реализована немедленная обратная связь [5], обеспечивающая необходимую интерактивность процесса обучения. И в-третьих, курс должен быть построен на основе игровых механик, делающих процесс обучения увлекательным и затягивающим – процесс, сопровождаемый мониторингом успеваемости обучаемого на фоне других достижений; процесс, в котором организовано сетевое общение обучаемых по обмену опытом прохождения данного курса. Применение компьютерных игр вызывает много споров вокруг их эффективности, так как до недавних пор считалось, что они имеют негативное влияние на человека, способны вызывать психическую зависимость [6, 7]. Могут ли игровые механики повысить результативность онлайн-курса? И верно ли, что синхронное обучение в массовом онлайн-курсе имеет существенное преимущество перед асинхронным?

Ответам на эти вопросы и посвящена данная статья, в которой обобщен опыт разработки и применения на практике одного из самых популярных массовых онлайн-курсов Университета ИТМО – курса для подготовки (повышения квалификации) разработчиков пользовательских графических веб-интерфейсов с применением технологий HTML5 и CSS3 [8–10]. Доступ к этому курсу открыт для всех желающих по адресу <http://htmlacademy.ru>, а также из системы открытого онлайн-обучения Университета ИТМО <http://courses.ifmo.ru>. На данный момент для прохождения курса уже зарегистрировалось более 38 тысяч желающих, которые в совокупности выполнили более 2 миллионов практических упражнений.

Основные проблемы в организации механизма мгновенной обратной связи

Во многих интерактивных курсах данная механика частично реализована в виде так называемого «лайвкодинга», когда результат работы кода, создаваемого обучаемым, практически мгновенно отображается в визуализаторе. Это делает обучение максимально наглядным, так как обучаемый сразу видит результат работы своего кода. Но при этом он не может сразу понять, соответствует ли его результат условиям задания? Происходит это потому, что в настоящее время в таких курсах проверка запускается лишь тогда, когда обучаемый завершил написание кода и отправил его на проверку. Опишем далее основные проблемы организации обратной связи в онлайн-курсе и пути их решения.

Первая проблема связана с тем, каким образом построить интерфейсы курса с эффективной обратной связью. Механизм обратной связи «сделал – увидел результат» работает очень быстро, а механизм «сделал – понял, что сделал правильно или неправильно» работает очень медленно. На рис. 1 приведена временная диаграмма работы интерфейсов с запуском проверки кода со стороны обучаемого, из которой виден основной недостаток данного механизма – наличие множества итераций, которые должен выполнить обучаемый при написании правильного кода. Исходя из этого, предлагается новое решение на основе комбинации обоих механизмов – «сделал – увидел результат – понял, что сделал правильно или неправильно» (рис. 2), которое ускоряет не только визуализацию результата, но и проверку правильности кода. Суть решения заключается в том, чтобы запускать механизм проверки кода сразу же после визуализации, без дополнительных действий со стороны обучаемого.

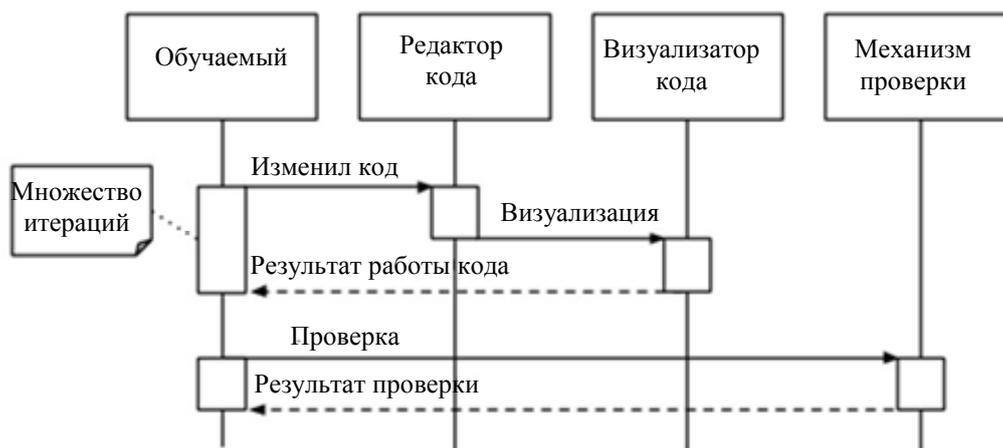


Рис. 1. Механизм работы интерфейсов с запуском проверки кода со стороны обучаемого

Вторая проблема связана с выбором сценария запуска визуализации и механизма проверки результата. Для начала условимся, что обучаемый работает с текстовым редактором, в котором пишет программный код или код разметки, результат работы этого кода отображается в визуализаторе, а механизм проверки кода определяет, выполнил ли обучаемый поставленную задачу или нет? Рассмотрим далее возможные варианты сценариев.

Первый вариант сценария (рис. 3) состоит в том, чтобы запускать визуализацию и механизм проверки мгновенно после каждого нажатия клавиши в редакторе. Несомненно, этот сценарий дает самую быструю обратную связь, но обладает следующими недостатками:

- чрезвычайно большое количество избыточных и некорректных проверок, а также визуализаций с ошибками из-за того, что используется еще не заверченный код, часто оборванный «на середине слова»;
- большая нагрузка на программный механизм из-за большого количества проверок, что может привести к зависаниям и замедлению работы обучающего интерфейса;
- обучаемый слишком часто будет видеть «сломанный» результат в визуализаторе, что может негативно сказаться на его интересе к дальнейшему обучению в таком курсе.

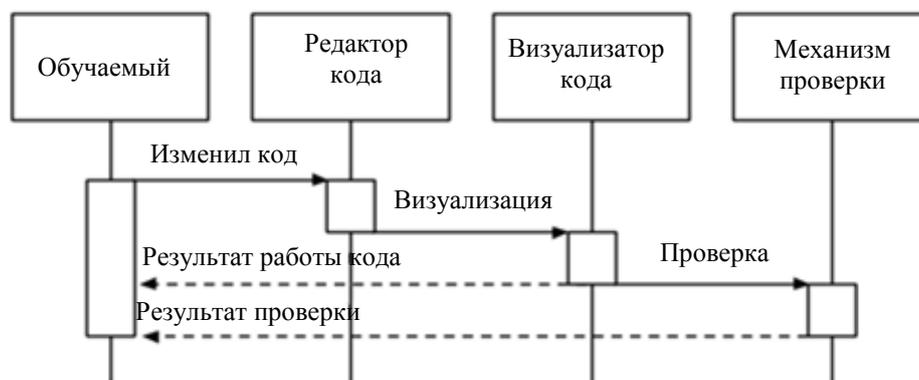


Рис. 2. Механизм работы интерфейсов с автоматической мгновенной проверкой

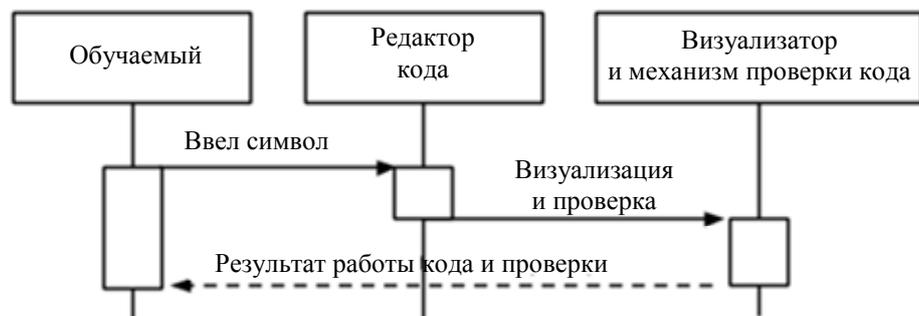


Рис. 3. Сценарий запуска визуализации и проверки кода после каждого нажатия клавиши в редакторе

Второй вариант сценария (рис. 4) состоит в том, чтобы запускать визуализацию и проверку после паузы в наборе кода. Длительность паузы может варьироваться от одной до нескольких секунд. Такой временной интервал соответствует времени переключения фокуса внимания от создаваемого кода к результату его выполнения и проверки. Данный временной интервал также является достаточно продолжительным для того, чтобы успевать набирать в редакторе завершённые конструкции кода. Этот сценарий избавлен от некоторых недостатков первого сценария, а именно:

- снижается количество ненужных циклов визуализаций и проверок;
- снижается количество визуализаций с ошибками, так как в большинстве случаев обучаемый будет успевать набирать значимые части кода.

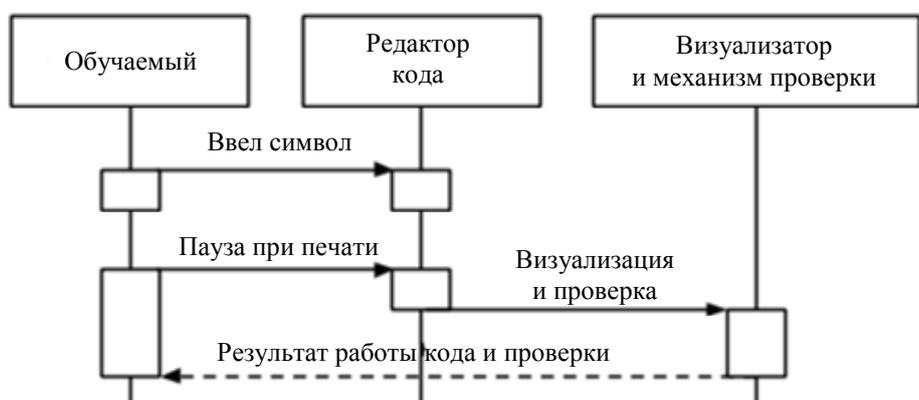


Рис. 4. Сценарий запуска визуализации и проверки кода после паузы в наборе кода

Третий вариант сценария (рис. 4) состоит в том, что в работу механизма добавляется этап анализа кода на корректность, а визуализация и проверка кода запускаются только тогда, когда введенный обучаемым код является работающим. Когда код некорректен, либо отображается предыдущее состояние, либо выдается сообщение о том, что код содержит ошибки. Проверку на корректность кода можно запускать при каждом изменении или после паузы, что снизит количество ненужных проверок. Основная проблема этого сценария заключается в том, что ученик не видит результатов работы «сломанного» кода, а это может негативно сказаться на качестве обучения.

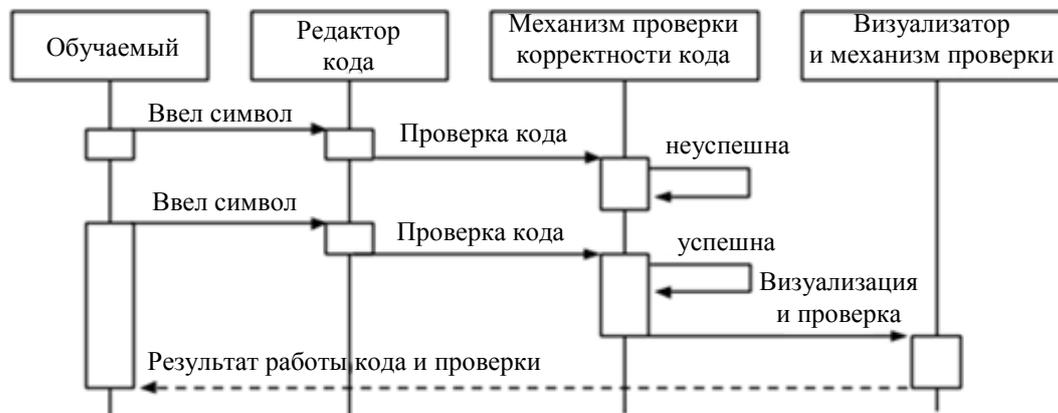


Рис. 5. Сценарий запуска визуализации и проверки кода с анализом кода на корректность

Таким образом, для повышения интерактивности онлайн-курса необходимо реализовать мгновенную обратную связь автоматически, без дополнительных действий со стороны обучаемого. При этом сценарий запуска визуализации и проверки кода может содержать этап анализа кода на корректность, позволяющий обучаемому понять правильность хода выполнения данного задания.

Проблема дозирования объема заданий для практических упражнений

В существующих аналогах [11–15] объем заданий для практических упражнений весьма невелик: время их выполнения варьируется от десятков секунд до нескольких минут. Общей особенностью таких заданий является то, что они неделимы и проверяются системой целиком. В то же время на практике многие задания для обучения разработчиков веб-интерфейсов являются сложными, требуют для выполнения много времени. В связи с этим предлагается подход, основанный на дроблении сложных задач на мелкие подзадачи, которые проверяются по отдельности. Подобная организация обучения делает его максимально похожим на приемы работы профессионалов, которые обычно пишут код очень маленькими порциями и проверяют его буквально после написания каждой новой строчки.

Предположим, что обучаемому предложено решить объемную задачу поэтапно, мелкими шагами с продолжительностью выполнения не более одной минуты. При таком решении обратная связь «выполнение задачи – визуализация – проверка результата» разбивается на несколько более мелких и быстрых решений, связанных с конкретными подзадачами. Это особенно важно, когда задание достаточно сложное и выполнить его сразу не получается. В этом случае обучаемому приходится несколько раз решать всю задачу целиком, перепроверя весь объем кода, необходимый для ее решения, а затем несколько раз ждать полной проверки. Представим, что один цикл выполнения и проверки сложной задачи составляет 2 мин, тогда для решения задачи с 10 проверками потребуется 20 мин. Раздробим сложную задачу на четыре подзадачи так, чтобы цикл выполнения и проверки каждой составлял 30 с. Предположим, что три подзадачи будут простыми и будут выполнены с первого раза, а одна из подзадач будет сложной, и для нее обучаемому понадобится 10 проверок, как и в случае с исходной задачей. Тогда общее время выполнения четырех подзадач составит $30 + 30 + 30 + 30 \times 10 = 6$ мин 30 с.

Если из четырех подзадач две будут простыми, а две – сложными (требующими 10 проверок), то общее время выполнения составит 11 мин. Заметим, что при расчете мы не учитывали дополнительные временные затраты на работу с более медленным интерфейсом, содержащим кнопку проверки.

Таким образом, подход с дополнительным дроблением задач на подзадачи в сочетании с мгновенной автоматической проверкой результата повышает наглядность и скорость обучения.

Перфекционизм как игровая механика для повышения мотивации к достижению лучших результатов обучения

Представленная игровая механика строится на принципе необязательности достижения максимальных результатов обучения, а для ее реализации часто применяются так называемые «прогрессбары» или шкалы. Данная механика в сочетании со шкалами достаточно распространена во многих областях,

например, во многих социальных сетях отображается процент заполненности профиля. Однако в образовании перфекционизм используется незаслуженно редко. В связи с этим предлагается подход, основанный на включении механики перфекционизма непосредственно в обучающий интерфейс, что, как будет показано далее, повышает увлеченность обучаемого, создает мотивацию к достижению лучших результатов обучения.

Чтобы применение данной механики было возможным, необходимо, чтобы задания оценивались не по бинарной шкале «выполнено/не выполнено», а по более широкой шкале, например, по 100-балльной шкале. Суть механики заключается в том, что шкала оценок разбивается минимум на два интервала («плохо» и «хорошо»), причем граница второго интервала «хорошо» располагается ниже максимального уровня. Самое главное заключается в том, чтобы явным образом указать обучаемому, что для успешного выполнения задания достаточно достигнуть нижней границы интервала «хорошо», а достижение максимального результата вовсе и необязательно.

Такой подход позволяет достичь глубокого вовлечения обучаемого в выполнение задания, так как для него работа с заданием после «неидеального» выполнения становится схожа с игрой, что характеризуется следующими условиями:

- отсутствием давления со стороны (задание уже выполнено, дальше можно заниматься с ним в свое удовольствие);
- отсутствием наказания (никто не будет ругать, если сделаешь еще лучше).

В этом случае обучаемым руководит любопытство, у него активизируется встроенная особенность мозга искать самые эффективные паттерны, приводящие к максимальному результату. Многие бросают вызов самому себе, от которого очень сложно отказаться, и получают удовольствие от достижения максимального результата.

Стоит отметить, что в большинстве аналогов [11–15] создание обучающих интерфейсов с использованием перфекционизма невозможно, так как в них отсутствуют задания с интервальными оценками. Очевидно, что в курсах должны существовать разные типы обучающих интерфейсов: с более простыми заданиями, которые оцениваются бинарно, и с более сложными заданиями, которые можно оценивать по 100-балльной шкале.

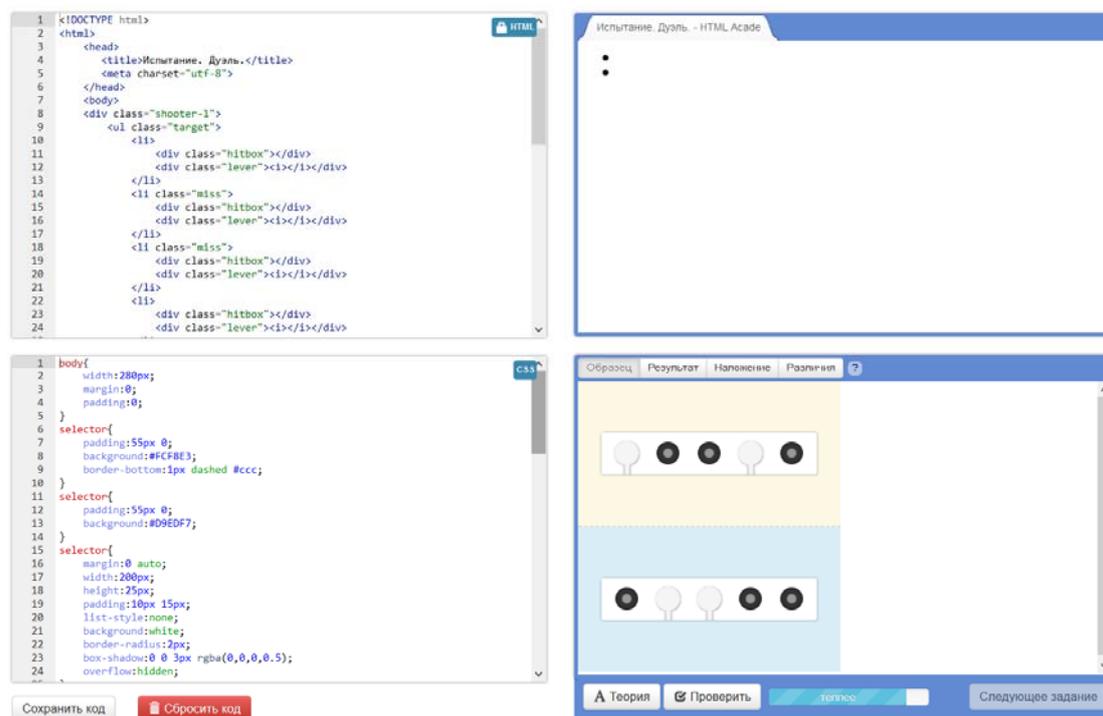


Рис. 6. Пример обучающего интерфейса с применением механики перфекционизма

Пример обучающего интерфейса, использующего эту механику, приведен на рис. 6. Здесь предложено испытание, в котором обучаемый должен сверстать такой же интерфейс, как на картинке-образце в правом нижнем окне. Результат вычисляется как уровень совпадения двух изображений и отображается на шкале. Испытание считается пройденным при уровне совпадения 90%, а оставшиеся 10% отданы для перфекционистов. Результаты от применения перфекционизма в этом примере следующие: на неидеальное выполнение задания в среднем уходит 30 мин, а на идеальное выполнение – несколько часов.

Для изучения влияния механики перфекционизма на результативность обучения в курсе [10] было проведено экспериментальное исследование в объеме 5588 чел., зарегистрировавшихся в системе htmlacademy.ru в период с 01.03.2014 г. по 18.04.2014 г. Так как обучение было асинхронным, то для сбора и обработки статистики был применен когортный анализ (в когорту объединялись обучаемые, зарегистрировавшиеся на одной неделе, всего было создано 7 когорт). Результаты исследования, приведенные в табл. 1, подтверждают положительное влияние данной механики на результативность обучения. Здесь все обучаемые разбиты на две группы: идеалисты обучались с использованием данной механики, а не-идеалисты – без ее использования.

Показатель	Прогресс					
	Все		Идеалисты		Не идеалисты	
	значения	%	значения	%	значения	%
Кол-во регистраций	5588	100,00	971	100,00	4402	100,00
Начали обучение	4361	78,04	971	100,00	3177	72,17
Прошли 1 раздел	2791	49,95	971	100,00	1609	36,55
Прошли 2 раздела	2265	40,53	971	100,00	1084	24,63
Прошли 3 раздела	1719	30,76	967	99,59	552	12,54
Прошли 4 раздела	1471	26,32	867	89,29	409	9,29
Прошли 5 разделов	1043	18,66	623	64,16	245	5,57
Прошли 6 разделов	865	15,48	502	51,70	199	4,52
Прошли 7 разделов	733	13,12	415	42,74	162	3,68
Прошли 8 разделов	606	10,84	348	35,84	119	2,70
Прошли 9 разделов	503	9,00	285	29,35	94	2,14
Прошли 10 разделов	423	7,57	239	24,61	76	1,73
Прошли 11 разделов	349	6,25	193	19,88	60	1,36
Прошли 12 разделов	299	5,35	163	16,79	46	1,04
Прошли 13 разделов	239	4,28	121	12,46	35	0,80
Прошли 14 разделов	201	3,60	99	10,20	25	0,57
Прошли 15 разделов	148	2,65	67	6,90	17	0,39
Прошли 16 разделов	119	2,13	54	5,56	11	0,25

Таблица 1. Результаты экспериментальных исследований влияния перфекционизма на результативность курса

Также проводилось исследование влияния перфекционизма на уровень вовлечения в обучение на основе опросов и отзывов обучаемых. Выяснилось, что на неидеальное выполнение задания в среднем уходит 30 мин, а на идеальное выполнение – в среднем несколько часов, причем максимальное время выполнения доходило до 6 ч. Оказалось, что люди, выполняющие задания идеально, часто повторно проходят предшествующие испытанию задания, а также самостоятельно ищут дополнительную информацию с помощью поисковых систем. Исходя из этого очевидно, что перфекционизм не только повышает уровень вовлечения в обучение, но и способствует более глубокому усвоению материала.

Влияние синхронного обучения на результативность онлайн-курса

С 17 февраля по 27 апреля 2014 г. проводился экспериментальный запуск разработанного курса в системе открытого онлайн-обучения Университета ИТМО. Обучение 258 человек, зарегистрировавшихся в системе, из которых приступило к обучению 245 чел. (95%), было организовано синхронно следующим образом: каждую неделю там открывались последовательно все 16 разделов курса. Результаты, приведенные в табл. 2, свидетельствуют о существенном преимуществе синхронного обучения по сравнению с асинхронным (табл. 1): количество успешно завершивших курс выросло до 27,9%.

Номер раздела	Прогресс															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Человек	222	210	199	189	171	165	160	150	138	131	127	117	104	91	81	72
%	86	81,4	77,1	73,3	66,3	64	62	58,1	53,5	50,8	49,2	45,3	40,3	35,3	31,4	27,9

Таблица 2. Результаты экспериментального исследования влияния синхронного обучения на результативность массового онлайн-курса

Заключение

Представленный подход способствует повышению результативности массового онлайн-курса за счет усиления его интерактивности (мгновенная обратная связь), дозированной подачи заданий при выполнении практических упражнений, погружения процесса обучения в игровую среду (игровая механика перфекционизма). Выявлено существенное влияние синхронного обучения на результативность курса за счет сетевого общения обучаемых. Данный подход имеет большое практическое значение для разработки новых массовых онлайн-курсов для обучения в области программирования и IT-технологий.

Литература

1. Васильев В.Н., Стафеев С.К., Лисицына Л.С., Ольшевская А.В. От традиционного дистанционного обучения к массовым открытым онлайн-курсам // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 1 (89). С. 199–205.
2. Васильев В.Н., Лисицына Л.С. Основные направления развития информационно-образовательной среды вуза в связи с переходом на ФГОС ВПО // Компьютерные инструменты в образовании. 2012. № 4. С. 62–68.
3. Васильев В.Н., Лисицына Л.С. Планирование и оценивание ожидаемых результатов освоения компетенций ФГОС ВПО // Научно-технический вестник информационных технологий, механики и оптики. 2013. № 2 (84). С. 142–148.
4. Lisitsyna L., Lyamin A. Approach to development of effective e-learning courses // Smart Digital Futures 2014. 2014. V. 262. P. 732–738.
5. Boyle E.A., Connolly T.M., Hainey T. The role of psychology in understanding the impact of computer games // Entertainment Computing. 2011. V. 2. N 2. P. 69–74.
6. Sim T., Gentile D.A., Bricolo F., Serpollini G., Gulamoydeen F. A conceptual review of research on the pathological use of computers, video games, and the Internet // International Journal of Mental Health and Addiction. 2012. V. 10. N 5. P. 748–769.
7. Messias, E., Castro J., Saini A., Usman M., Peeples D. Sadness, suicide, and their association with video game and Internet overuse among teens: results from the youth risk behavior survey 2007 and 2009 // Suicide and Threatening Behavior. 2011. V. 41. N 3. P. 307–315.
8. Лисицына Л.С. Першин А.А. К вопросу создания Fab lab для подготовки разработчиков пользовательских веб-интерфейсов на основе HTML и CSS // Дистанционное и виртуальное обучение. 2014. № 1 (79). С. 32–38.
9. Першин А.А. Средства обучения HTML-верстке, построенные на базе методики автоматической оценки качества верстки // Материалы V Международной научно-практической конференции «Информационная среда вуза XXI века». Петрозаводск, 2011. С. 221.
10. Лисицына Л.С. Першин А.А. Электронный курс для обучения технологиям разработки веб-интерфейсов на основе HTML и CSS // Сб. материалов I международной научно-практической конференции «Современные проблемы компьютерных наук». Пенза: ПГУ, 2013. С. 83–84.
11. Codecademy [Электронный ресурс]. Режим доступа: <http://www.codecademy.com>, свободный. Яз. англ. (дата обращения 17.07.2014).
12. Code School [Электронный ресурс]. Режим доступа: <http://www.codeschool.com>, свободный. Яз. англ. (дата обращения 17.07.2014).
13. Advance Your Career Through Project-Based Online Classes – Udacity [Электронный ресурс]. Режим доступа: <https://www.udacity.com>, свободный. Яз. англ. (дата обращения 17.07.2014).
14. Code Avengers: Learn to Code Games, Apps and Websites [Электронный ресурс]. Режим доступа: <http://www.codeavengers.com>, свободный. Яз. англ. (дата обращения 17.07.2014).
15. Learn Web Design, Web Development, and More [Электронный ресурс]. Режим доступа: <http://teamtreehouse.com>, свободный. Яз. англ. (дата обращения 17.07.2014).

- Лисицына Любовь Сергеевна* – доктор технических наук, профессор, заведующая кафедрой, Университет ИТМО, 197101, Санкт-Петербург, Россия, lisizina@mail.ifmo.ru
- Першин Александр Александрович* – ассистент кафедры, Университет ИТМО, 197101, Санкт-Петербург, Россия, pershin_zan@mail.ru
- Усков Владимир Львович* – PhD, профессор кафедры, Бредли университет, 61625, Пеория, Иллинойс, США, uskov@fsmail.bradley.edu
- Liubov S. Lysitsina* – D.Sc., Professor, Head of Department, ITMO University, 197101, Saint Petersburg, Russia, lisizina@mail.ifmo.ru
- Alexander A. Pershin* – Assistant, ITMO University, 197101, Saint Petersburg, Russia, pershin_zan@mail.ru
- Vladimir L. Uskov* – PhD, Professor, Bradley University, 61625, Peoria, IL, USA, uskov@fsmail.bradley.edu

Принято к печати 19.07.14
Accepted 19.07.14

УДК 371.69

ИНФОРМАЦИОННАЯ ИНФРАСТРУКТУРА ОБРАЗОВАТЕЛЬНОЙ СРЕДЫ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ ВИРТУАЛЬНЫХ МАШИН

А.Д. Береснев^{a, b}, Н.Ф. Гусарова^a, Р.В. Иванов^a, Д.Н. Федотов^{c, d}^a Университет ИТМО, 197101, Санкт-Петербург, Россия^b ООО «Ротэк Технолоджиз», 192007, Санкт-Петербург, Россия^c Институт информатики и вычислительной техники (IATI AS), 10416, Таллин, Эстония^d Учебный центр Omis OÜ, Satelles Saturni OÜ, 10117, Таллин, Эстония**Аннотация.**

Предмет исследования. Создана и исследована информационная инфраструктура среды обучения с применением технологии виртуальных машин для малых педагогических систем (отдельные классы, авторские курсы).

Методика исследования. В методологии ARIS построена модель жизненного цикла информационной инфраструктуры для малых педагогических систем с использованием виртуальных машин. Предложена методика формирования информационной инфраструктуры с виртуальными машинами на базе процессного подхода. В качестве базовой модели использована модель событийной цепочки в сочетании с диаграммой окружения. Для каждой функции событийной цепочки определен необходимый набор средств информационной и программной поддержки. Применение методики проиллюстрировано на примере проектирования информационной инфраструктуры образовательной среды, учитывающей специфику малых педагогических систем. К преимуществам спроектированной информационной инфраструктуры относятся: максимальное использование открытых или бесплатных компонентов; использование стандартных протоколов (в основном HTTP и HTTPS); максимальная переносимость (серверы приложений могут быть развернуты на любой из распространенных операционных систем); единый интерфейс к управлению различными платформами виртуализации; возможность инвентаризации содержимого виртуальной машины без ее запуска; гибкое управление инвентаризацией виртуальной машины с помощью настраиваемых цепочек правил.

Апробация. Апробация результатов работы проводилась на базе учебного центра «Институт информатики и вычислительной техники» (Таллин, Эстония). Применение методики в рамках курса «Пользование компьютером и программным обеспечением» позволило более чем в 2 раза сократить число отказов компонентов информационной инфраструктуры, требующих вмешательства технического специалиста, а также время на устранение такого рода неисправностей. Кроме того, учащиеся, получившие более широкий опыт работы с компьютером и программным обеспечением за счет применения информационной инфраструктуры с виртуальными машинами, показали более хорошие результаты в рамках зачетных испытаний.

Практическое применение. Результаты работы могут быть рекомендованы для внедрения в практику работы образовательных учреждений среднего и высшего образования, а также при разработке авторских и специализированных курсов, ориентированных на коллективную работу учащихся.

Ключевые слова: виртуальные машины, платформа виртуализации, процессный подход, малые педагогические системы, информационная инфраструктура, цепочки правил, событийная цепочка.

INFORMATION INFRASTRUCTURE OF THE EDUCATIONAL ENVIRONMENT WITH VIRTUAL MACHINE TECHNOLOGY

A.D. Beresnev^{a, b}, N.F. Gusarova^a, R.V. Ivanov^a, D.N. Fedotov^{c, d}^a ITMO University, 197101, Saint Petersburg, Russia, natfed@list.ru;^b "ROTEK Technologies", Ltd, 192007, Saint Petersburg, Russia, open.look@gmail.com^c Institute of Informatics and Computers (IATI AS), 10416, Tallinn, Estonia^d Education Center OMIS OÜ, Satelles Saturni OÜ, 10117, Tallinn, Estonia

Subject of research. Information infrastructure for the training environment with application of technology of virtual computers for small pedagogical systems (separate classes, author's courses) is created and investigated.

Research technique. The life cycle model of information infrastructure for small pedagogical systems with usage of virtual computers in ARIS methodology is constructed. The technique of information infrastructure formation with virtual computers on the basis of process approach is offered. The model of an event chain in combination with the environment chart is used as the basic model. For each function of the event chain the necessary set of means of information and program support is defined. Technique application is illustrated on the example of information infrastructure design for the educational environment taking into account specific character of small pedagogical systems.

Advantages of the designed information infrastructure are: the maximum usage of open or free components; the usage of standard protocols (mainly, HTTP and HTTPS); the maximum portability (application servers can be started up on any of widespread operating systems); uniform interface to management of various virtualization platforms, possibility of inventory of contents of the virtual computer without its start, flexible inventory management of the virtual computer by means of adjusted chains of rules.

Approbation. Approbation of obtained results was carried out on the basis of training center "Institute of Informatics and Computer Facilities" (Tallinn, Estonia). Technique application within the course "Computer and Software Usage" gave the possibility to get half as much the number of refusals for components of the information infrastructure demanding intervention of the technical specialist, and also the time for elimination of such malfunctions. Besides, the pupils who have got broader experience with computer and software, showed better results within tests due to the usage of information infrastructure with virtual computers.

Practical application. Obtained output can be recommended for introduction into the activity of educational institutions of secondary and higher education, and also for developing the author's and specialized courses focused on collective actions of pupils.

Keywords: virtual computers, virtualization platform, process approach, small pedagogical systems, information infrastructure, chains of rules, event chain.

Введение

Реализация задач, поставленных Законом «Об образовании в Российской Федерации»¹ и государственными программами развития образования^{2,3}, требует новых подходов к формированию информационной инфраструктуры (ИИ) в рамках единой образовательной среды⁴. Среди таких подходов важное место занимает создание сред обучения на базе технологии виртуальных машин (ВМ). ВМ определяется как «эффективный, изолированный дубликат реальной машины» [1] или как «программная реализация машины (т.е. компьютера), который выполняет программы как физическая машина» [2]. Содержательно под ВМ понимается вычислительная среда, реализованная в физическом компьютере и выделяющая его аппаратные ресурсы таким образом, чтобы предоставить возможность запуска нескольких операционных систем на одном компьютере. При этом каждая операционная система (ОС) выполняется в собственной ВМ и имеет выделенные логические экземпляры процессоров, жестких дисков, сетевых карт и других аппаратных ресурсов компьютера.

Виртуализация – одна из наиболее быстро развиваемых информационных технологий, которая активно внедряется и в сферу образования [3, 4]. Например, в [5] сообщается о разработке, основанной на виртуализации системы дистанционного обучения vMoodle, интегрированной с обучающей средой Moodle; vMoodle поддерживает ВМ, размещенные как в частных облаках, так и в общественном облаке (Amazon EC2). Несколько учебных курсов открытого университета Великобритании используют технологию ВМ [6]. Образовательные сети США [7] предоставляют ВМ как сетевой сервис. В практике работы образовательных учреждений России также имеется опыт использования ВМ (например, [8–11]).

Однако при попытке внедрения технологии виртуализации в практику своей работы педагог сталкивается с целым рядом проблем.

- В исследованиях и проектах последних лет широко представлены работы, посвященные оценке эффективности электронных образовательных ресурсов и информационных сервисов (см., например, [12–19]). Однако эти работы ориентированы на педагогические системы макро- и мезоуровней [20], т.е. на государственные и региональные системы образования, а также на учебные заведения в целом. Предлагаемые в них методики не позволяют учесть индивидуальный опыт и подготовленность педагога и обучаемых, существенные на микроуровне, т.е. в малых педагогических системах (МПС), например, в конкретном классе или авторском курсе, и не обеспечивают гибкой адаптации ИИ к конкретной педагогической задаче.
- Существуют различные технологии виртуализации, каждая из которых реализована в наборе коммерческих ВМ. Однако их сравнительные оценки, представленные в литературе, основываются на маркетинговых данных компаний-разработчиков (например, [21]) и не учитывают специфику работы ВМ в конкретном инфраструктурном решении, что затрудняет встраивание ВМ в систему средств обучения, уже используемую педагогом.

Таким образом, очевидные преимущества технологии ВМ в образовательном процессе реализуются только при наличии комплексного инфраструктурного решения, поддерживающего работу с ВМ в условиях конкретной педагогической системы. В то же время как показывает анализ, для МПС такие инфраструктурные решения отсутствуют. Целью настоящей работы является разработка методики формирования ИИ для использования технологии ВМ в условиях перехода к единой информационной образовательной среде. Применение методики иллюстрируется на примере проектирования ИИ образовательной среды, учитывающей специфику МПС.

Методика формирования ИИ МПС с использованием технологии ВМ

ИИ как компонент архитектуры информационных технологий в общем случае является одним из механизмов реализации целей организации. В рассматриваемом случае в качестве основной цели выступает адекватное решение педагогической задачи, возникающей в рамках МПС, и ИИ с применением технологии ВМ правомерно рассматривать как педагогическое программное средство [20], обладающее специфическими особенностями (СО).

СО1. Компоненты ВМ, как правило, должны встраиваться в уже существующую информационную архитектуру МПС, причем потребность такого встраивания может возникнуть на любых этапах жизненного цикла ИИ.

¹ Федеральный закон от 29 декабря 2012 года №273-ФЗ «Об образовании в Российской Федерации»

² Государственная программа Российской Федерации «Развитие образования» на 2013–2020 годы, утверждена распоряжением Правительства РФ от 22.11.2012 № 2148-р

³ Федеральная целевая программа развития образования на 2011–2015 годы, утверждена постановлением Правительства РФ от 7.2.2011 № 61

⁴ Концепция развития единой информационной образовательной среды в Российской Федерации. – http://raec.ru/upload/files/eios_conception.pdf

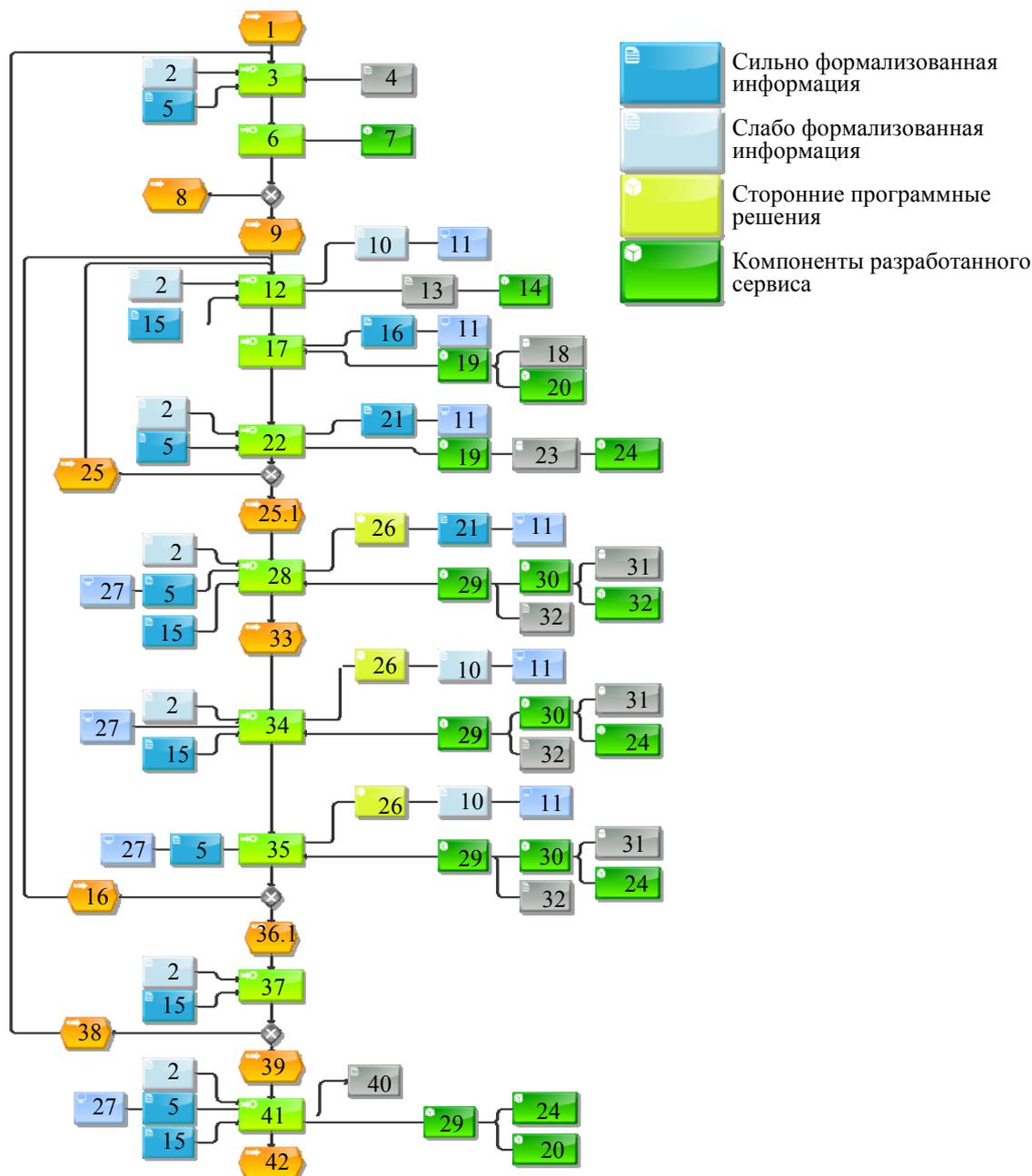


Рис. 1. Модель жизненного цикла ИИ МПС с использованием VM в методологии ARIS: 1 – старт педагогической задачи; 2 – информация о составе МПС; 3 – идентификация параметров педагогической задачи; 4 – шаблон параметров для оценки рисков; 5 – параметры имеющейся ИИ; 6 – оценка рисков применения ИИ; 7 – модуль оценки рисков; 8 – выбор ИИ без VM; 9 – переход к формированию ИИ с VM; 10 – сторонний опыт и примеры описания VM; 11 – среда Интернет; 12 – лингвистическое описание необходимой VM; 13 – структурированная информация о необходимой VM; 14 – тематический форум по VM; 15 – критерии достижения целей педагогической задачи; 16 – информация об имеющихся VM; 17 – отбор первичного набора возможных реализаций VM; 18 – уточнение технических параметров; 19 – система отбора VM; 20 – модуль семантического анализа; 21 – информация о параметрах имеющихся VM и (или) среды виртуализации; 22 – запуск VM; 23 – информация о технических параметрах выбранного прототипа VM; 24 – модуль сбора технических параметров; 25 – возможный вариант VM отсутствует; 25.1 – вариант VM выбран; 26 – среда виртуализации; 27 – характеристики существующей ИИ; 28 – тонкая настройка VM; 29 – модуль визуализации; 30 – варианты сред виртуализации; 31 – репозиторий VM; 32 – информация с форума; 33 – VM готова к эксплуатации в МПС; 34 – использование VM учащимися; 35 – контроль технических параметров ИИ с VM; 36 – технические параметры ИИ с VM не соответствуют педагогической задаче МПС; 36.1 – технические параметры ИИ с VM соответствуют педагогической задаче МПС; 37 – контроль соответствия параметров ИИ с VM критериям достижения цели педагогической задачи; 38 – педагогическая задача МПС не выполнена; 39 – педагогическая задача МПС выполнена; 40 – зафиксированные результаты выполнения педагогической задачи; 41 – завершение выполнения педагогической задачи; 42 – педагогическая задача завершена

СО2. Решение об использовании ВМ, как правило, принимается в рамках самой МПС, а не «спускается сверху» как обязательный компонент ИИ; целесообразно предусмотреть возможность гибкого пересмотра этого решения (вплоть до его отмены) при изменении ситуации в МПС.

СО3. При принятии решения об использовании ВМ необходимо уделять особое внимание человеческому фактору, т.е. индивидуальным особенностям каждого участника МПС.

СО4. ИИ с применением технологии ВМ является программным средством, и при оценке его качества необходимо использовать соответствующие ГОСТы¹.

Как показал сравнительный анализ, наиболее полно учесть эту специфику позволяет процессный подход к моделированию ИИ, реализуемый в методологии ARIS [22]. Эта методология дает возможность комплексно описать ИИ не только как сугубо программную сущность, но и как компонент социально-технической системы. В качестве базовой модели для формирования ИИ МПС выбрана модель событийной цепочки (Extended event driven process chain, eEPC), которая описывает последовательность управляющих событий и функций, выполняемых ИИ в ходе решения педагогической задачи.

Существенным преимуществом методологии ARIS является возможность комбинирования модели eEPC с диаграммой окружения (Function allocation diagram), что позволяет сопоставить каждой функции модели событийной цепочки необходимый набор средств информационной и программной поддержки. На рис. 1 представлена объединенная модель в нотации ARIS, причем, в дополнение к типовым графическим обозначениям элементов нотации ARIS, для средств информационной и программной поддержки применен цветовой код (см. врезку). Модель отражает основные этапы жизненного цикла использования ВМ в ИИ МПС, содержательно соответствующие следующим вопросам:

1. Вводить ли ВМ в состав ИИ для решения педагогической задачи?
2. Какой тип ВМ использовать?
3. Не требуется ли текущая модификация ВМ в составе ИИ?

Первый этап (блоки 1, 3, 6, 8, 9 событийной цепочки) предваряет переход к ИИ с применением ВМ, поэтому здесь производится проактивная оценка рисков такого перехода для решения заданной педагогической задачи в условиях конкретной МПС и существующего состоянии ИИ. Для поддержки этого этапа авторами разработан проблемно-ориентированный инструментарий оценки рисков [23].

На втором этапе (блоки 12, 17, 22, 25, 25.1 событийной цепочки) выполняется выбор типа ВМ для построения ИИ как компонента образовательной среды. Критерии этого выбора определены вышеупомянутыми стандартами, однако, как показывает анализ, для решения большинства реальных педагогических задач многие из этих критериев оказываются взаимосвязанными. В связи с этим здесь необходим инструментарий многокритериального анализа, позволяющий выявить взаимосвязи между конкретными критериями и сократить их количество, представленный в работе авторов [24]. Этап завершается «тонкой настройкой», т.е. кастомизацией выбранного прототипа в соответствии с параметрами уже существующих компонентов ИИ, куда встраивается ВМ, а также спецификой МПС (блоки 28, 33 событийной цепочки).

После введения настроенной ВМ в состав ИИ начинается собственно решение педагогической задачи. На этом этапе осуществляется постоянный мониторинг адекватности текущего состояния структуры и параметров ИИ решаемой педагогической задаче и, при необходимости, модификация структуры и параметров ИИ вплоть до возвращения к структуре ИИ без ВМ (блоки 34–39 событийной цепочки). Здесь требуется программный инструментарий для поддержки мониторинга. Отдельные программные решения для такого мониторинга, разработанные авторами, защищены свидетельствами на интеллектуальную собственность [25–27]. Этап завершается фиксацией результатов деятельности МПС (блоки 41, 42 событийной цепочки). Проведенный анализ событийной цепочки жизненного цикла ИИ в сочетании с его особенностями (СО1–СО4) позволяет для каждой из выделенных функций модели выделить необходимые средства информационной и программной поддержки, которые также представлены на рис. 1 (расшифровка – в подписочной подписи). Заметим, что многие из этих средств для различных функций дублируют друг друга, что упрощает построение ИИ в целом; с другой стороны, особенности СО1–СО4 вынуждают учитывать при выборе программных средств для каждой функции возможный уровень формализации ее информационной поддержки, что отражено в легенде рис. 1.

Таким образом, методика формирования ИИ МПС с использованием технологии ВМ может быть представлена в виде следующей последовательности шагов:

1. идентифицировать педагогическую задачу;
2. определить параметры списка СО1–СО4 применительно к конкретной МПС;
3. представить функции, выполняемые ИИ под управлением педагога в ходе решения педагогической задачи, в виде событийной цепочки;

¹ ГОСТ 25195-89 «Оценка качества программных средств. Общие положения»; ГОСТ Р ИСО/МЭК 9126 «Информационная технология. Оценка программной продукции. Характеристика качества и руководства по их применению»; ГОСТ 28806-90. Качество программных средств. Термины и определения

4. используя результаты п.п. 1 и 2, для каждой функции событийной цепочки определить необходимый набор средств информационной и программной поддержки (объектов окружения);
5. в результатах п. 4 устранить дублирование;
6. используя результаты п.п. 2 и 5, перейти к проектированию ИИ МПС.

Архитектура ИИ МПС с использованием технологии VM

Содержательный анализ и факторизация объектов окружения событийной цепочки (рис. 1) позволяют перейти к проектированию ИИ МПС с использованием технологии VM.

На рис. 2 приведена архитектура ИИ в нотации UML в виде диаграммы развертывания. Для раскрытия основных принципов функционирования ИИ диаграмма дополнена компонентами системы. ИС построена по трехзвенной архитектуре с разнесением слоя сервера приложений на два узла. К основным задачам ИС относятся:

- сбор формальных данных о VM;
- сбор слабо формализуемой информации о VM;
- подготовка сводных отчетов о свойствах VM;
- единообразное управление VM на разных гипервизорах.

В связи с выделенными выше особенностями МПС (CO1 и CO3) в архитектуре предусматривается возможность работы VM на гипервизорах, развернутых на отдельных устройствах, причем с одновременной поддержкой нескольких платформ. В данной реализации предусмотрена поддержка трех платформ: Microsoft Hyper-V, VMware vSphere и Proxmox Virtual Environment, каждая из которых может быть представлена несколькими экземплярами. На каждой платформе может выполняться множество VM. На схеме они сгруппированы в пакет «Платформы виртуализации».

Архитектурным ограничением является использование в VM исключительно эмулируемых дисковых устройств (образов виртуальных жестких дисков), даже в том случае, если гипервизор поддерживает прямой доступ к дисковым устройствам. Хранение образов VM и их файлов конфигурации организовано в единой сетевой системе хранения данных (NAS). Доступ к файлам VM с платформ виртуализации осуществляется по протоколу SMB или NFS. Единая сетевая система хранения данных вместо сети хранения данных выбрана для упрощения совместного доступа к файлам образов VM и конфигураций со стороны гипервизоров и подсистемы инвентаризации VM.

Клиентская часть построена в виде Web-приложения с ограничением браузеров по реализации на актуальных программных ядрах Trident, WebKit и Gecko, поддерживающих HTML5. Клиентская часть не требует установки специальных плагинов и не предполагает выполнение апплетов Java или flash-программ. Web-клиент, выполняя стандартную задачу интерфейсного ввода-вывода, взаимодействует с Web-сервером, на котором развернуты Web-модули сайта front-end (пакет «Web-модули front-end»).

На сервере приложений подсистемы поддержки принятия решения (ППР) функционирует сервер приложений Tomcat, выполняющий сервлеты (пакет «Web-модули back-end»). На этом сервере работает подсистема инвентаризации VM, осуществляющая управляемый сбор формальных данных о конфигурациях VM, о конфигурации операционных систем VM и об установленном в VM программном обеспечении (ПО). Для реализации прозрачного и единообразного управления VM, выполняемыми на разных гипервизорах, в системе предусмотрен промежуточный сервер приложений подсистемы управления VM. На нем развернут управляющий модуль, который через стандартные компоненты обеспечивает управление VM, а также доступ к консолям VM прямо в окне браузера. На сервере баз данных (БД) развернуты экземпляры MS SQL Server и MS SQL Server Analysis Services. Система управления базами данных обеспечивает работу трех баз данных, показанных на рис. 2. Состав пакетов «Web-модули front-end» и «Web-модули back-end» показан на рис. 3.

Модуль пользовательского интерфейса является точкой входа клиента на сайт. Через него подключаются внешние модули. Модуль управления реализует пользовательский интерфейс к управлению VM. Команды пользователя об изменении состояния VM (создание, клонирование, остановка, запуск, редактирование свойств и т.п.) передаются к подсистеме управления VM, которая, в свою очередь, или через Python-скрипты по протоколу SSH или через скрипты PowerShell по протоколу RPC\HTTPS передает их гипервизорам. Также модуль обеспечивает доступ к консолям VM в окне браузера через Proxu-модуль подсистемы управления VM.

Модуль форума jforum обеспечивает коммуникацию пользователей в режиме отложенного времени. Следует отметить, что, в соответствии с выявленным содержанием объектов окружения в составе ИИ (рис. 1), основное назначение форума – сбор слабо формализованной информации о VM. В сообщениях форума каждый экземпляр VM идентифицируется уникальным составным идентификатором, состоящим из имени сервера виртуализации и ID VM.

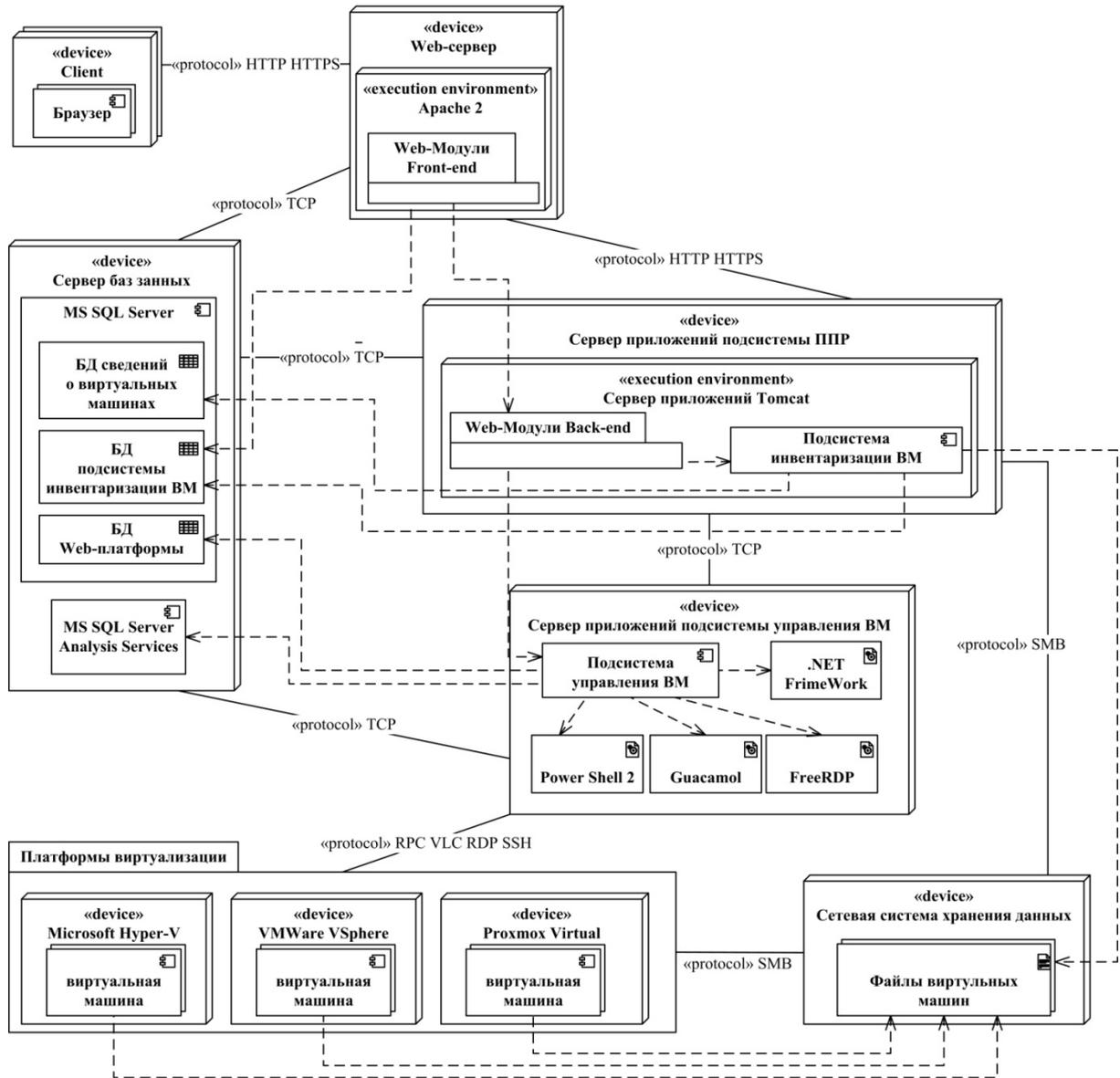


Рис. 2. Диаграмма развертывания ИИ МПС с использованием VM

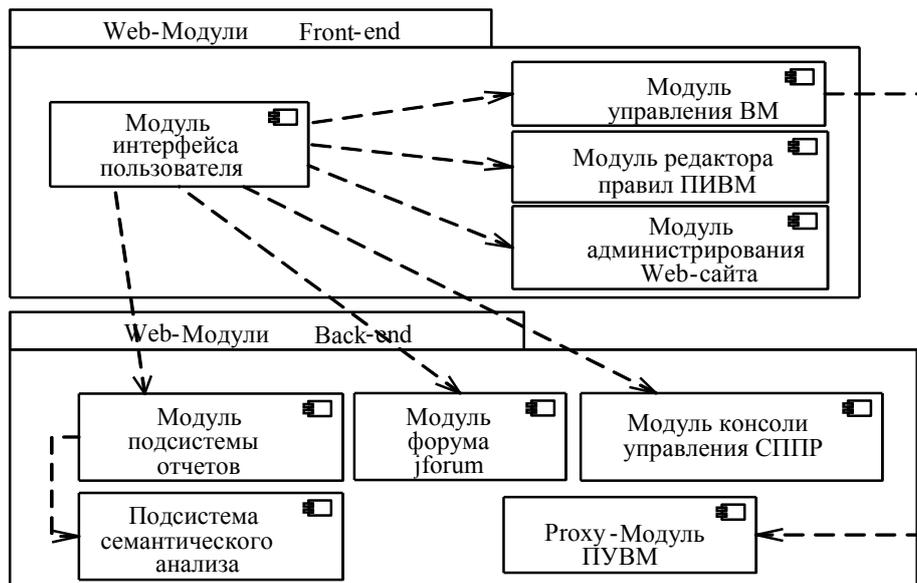


Рис. 3. Диаграмма компонентов пакетов Web-сайтов

В подсистему поддержки принятия решения объединены:

- модуль подсистемы отчетов, отвечающий за передачу запросов и за формирование сводных отчетов;
- подсистема семантического анализа, обеспечивающая формирование запросов и доступ к результатам обработки системой MS SQL Server Analysis Services данных о ВМ;
- модуль консоли управления, предназначенного для управления шаблонами отчетов.

Специфика реализации ИИ МПС с использованием ВМ

Как подчеркивалось выше, в связи с особенностями МПС к архитектуре и программной реализации ИИ с ВМ предъявляются специфические требования:

- в ИИ могут включаться ВМ на гипервизорах, развернутых на отдельных устройствах, причем с одновременной поддержкой нескольких платформ; педагог-администратор должен иметь возможность контролировать текущий состав ИИ без запуска ВМ;
- педагог-администратор ИИ должен иметь возможность контролировать текущее состояние каждой запущенной ВМ и при необходимости удаленно осуществлять необходимые корректирующие действия.

Ключевыми компонентами ИИ для реализации этих требований являются подсистема инвентаризации ВМ и подсистема удаленного управления.

Подсистема инвентаризации ВМ предназначена для сбора формальных данных о виртуальных машинах. К этим данным относятся:

- составное наименование ВМ;
- информация о платформе виртуализации (модель, версия);
- информация о конфигурации виртуального оборудования (количество ядер CPU, объем, количество и объем жестких дисков, интерфейс подключения жестких дисков, количество и скорость работы сетевых адаптеров);
- информация о средствах интеграции (наличие, версия, поддержка управления питанием, поддержка heartbeat);
- версия виртуальной ОС;
- информация об установленном в ВМ программном обеспечении и, при необходимости, параметрах его конфигурации.

Особенностью функционирования подсистемы является отсутствие необходимости запуска ВМ для получения необходимой информации. Исходными объектами для нее служат файлы конфигурации ВМ и файл виртуального диска. Функционирование программы не зависит от типа ОС, установленной в ВМ.

Подсистема инвентаризации ВМ осуществляет анализ ВМ, выполняя набор специальных управляющих директив (правил). Правила представляют собой функции, получающие в качестве аргумента данные об объекте анализа и возвращающие строковые данные и (или) вызывающие следующие правила. Структурно правило представляет собой совокупность обработчика и дополнительных данных для обработчика. Обработчик – это шаблонная функция, имеющая одинаковый формат входных и выходных данных. Это и позволяет составлять из них цепочки, работающие по принципу конвейера – выход одной функции является входом другой. В описании обработчика содержатся типы данных, которые требуются на входе, и тип данных, получаемых на выходе. Это описание используется при проверке цепочки правил при добавлении в базу и выполнении. В случае нарушения осуществляется переход к следующей цепочке и выдается сообщение об ошибке. Кроме данных, полученных из предыдущего правила, обработчик принимает дополнительные данные, содержащиеся в описании правила, или в описании цепочки.

Эксперт заранее готовит упорядоченные наборы (цепочки) правил, определяя объекты для анализа, конкретные параметры ВМ, значения которых следует определить, а также последовательность проверки и условия перехода между цепочками. Для этого служит модуль редактора правил.

Цепочки правил хранятся в специальной базе данных. В программе цепочка правил представлена стеком правил. Обработка цепочки правил реализована в виде рекурсивного вызова обработки правил. В качестве входных данных в следующее правило передается результат работы текущего правила (для первого элемента передается NULL) и стек цепочки правил, из которого текущее правило удаляет свою запись.

Таким образом, подсистема инвентаризации по файлам конфигурации ВМ на сетевом хранилище определяет параметры ВМ и по содержимому виртуальных дисков проводит инвентаризацию установленного в ВМ программного обеспечения и его конфигурации. В этом случае объектом анализа выступают пары «файл конфигурации с регулярной структурой – параметр в этом файле» или «файл бинарных данных с регулярной структурой – указатель на данные». В настоящее время реализован анализ файлов ini, conf (linux) и реестра MS Windows. Запуск проверки осуществляется или по выполнению триггера, или автоматически при фиксации релиза ВМ, или вручную для любого экземпляра ВМ, даже без фиксации релиза. На рис. 4 представлена sequence-диаграмма процесса сбора данных о ПО, развернутом в ВМ.

Дополнением к подсистеме инвентаризации ВМ может служить подсистема семантического анализа, которая обрабатывает тексты обсуждений ВМ на форуме, выделяя дополнительную информацию о ВМ.

Подсистема удаленного управления ВМ предназначена для управления различными виртуальными машинами, работающими на различных платформах виртуализации (вне зависимости от конкретной платформы виртуализации), при помощи Web-браузера. Она представляет собой Web-сервис, позволяющий управлять ВМ в сети, и реализует единую точку управления набором серверов виртуализации, в том числе:

- предоставление интерфейса регистрации серверов виртуализации;
- управление иерархией ВМ;
- управление отдельной ВМ (модификация, запуск, перезагрузка и т.п.);
- создание новой ВМ с управлением параметрами машины;
- работа с ВМ через консоль в окне браузера.

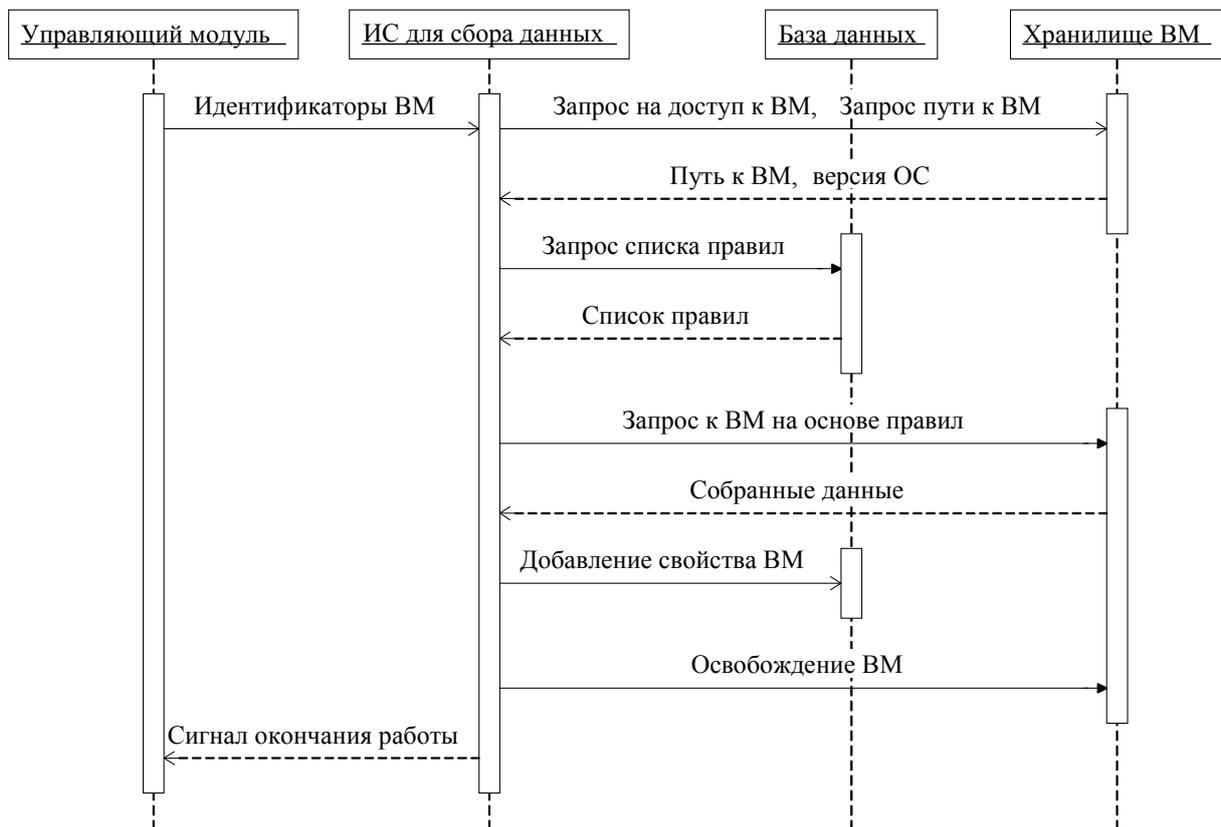


Рис. 4. Диаграмма процесса сбора данных

В подсистеме осуществляется поддержка трех платформ виртуализации: Microsoft Hyper-V, VMware vSphere и Proxmox Virtual Environment. Каждая платформа в системе может быть представлена несколькими экземплярами. Отметим, что номенклатура поддерживаемых платформ может быть расширена за счет незначительной модификации подсистемы.

Подсистема выполнена в виде модуля, реализованного с использованием технологий Microsoft .NET на языке C#. Взаимодействие с гипервизорами осуществляется через стандартные для платформ виртуализации командные интерфейсы. Задача модуля сводится к установлению защищенного сеанса связи с платформой виртуализации и передачи высокоуровневых команд с интерпретацией ответов систем для их представления в едином интерфейсе пользователя.

Для платформы MS Hyper-V связующим элементом выбран Windows PowerShell – расширяемое средство автоматизации, состоящее из оболочки с интерфейсом командной строки и сопутствующего языка сценариев. Защищенный канал формируется с помощью службы WinRM, запущенной на стороне платформы. Трафик инкапсулируется в протокол HTTPS. Для остальных платформ защищенный канал формируется по протоколу SSH с взаимной аутентификацией по сертификатам.

Для доступа к консолям операционных систем ВМ используется схема на основе свободно доступных компонентов. Поток консольного ввода–вывода через стандартные для платформ виртуализации протоколы (VNC для VMware vSphere и Proxmox Virtual Environment, и RDP для Microsoft Hyper-V) поступает через открытые компоненты Guacamole и FreeRDP на Proxy-модуль, реализованный в виде JAVA сервлета; тот, в свою очередь, транслирует их в поток протокола HTML5 и передает их в модуль управления ВМ. Обратная связь реализуется аналогично.

Апробация

Апробация результатов работы проводилась на базе учебного центра «Институт информатики и вычислительной техники» (Таллинн, Эстония). Для апробации была выбрана дисциплина «Пользование компьютером и программным обеспечением», входящая в базовый цикл. Длительность изучения дисциплины – 1 семестр. Занятия проводятся в группах по 12–15 человек. В рамках дисциплины учащиеся знакомятся с физическим и логическим устройством компьютера, обучаются пользованию основным набором прикладных программных продуктов. Использование ВМ в рамках данной дисциплины было направлено на то, чтобы учащиеся познакомились не только с нативным ПО фирмы Apple, которое является базовым для Центра, но и с ПО, разработанным под Windows и Linux. Выбор типа использованной ИИ, а также мониторинг деятельности учащихся выполнялись на основе разработанной методики.

Для сравнения было взято два потока: в одном из них проводились занятия с использованием ВМ, в другом – без использования ВМ. Сравнение проводилось по следующим характеристикам.

1. Количество отказов ПО, с использованием которого проводится занятие – отказов типа 1 (фиксировалось по обращениям учащегося к преподавателю, рассматривались только те отказы, для которых было необходимо вмешательство технического специалиста).
2. Количество отказов аппаратного обеспечения и ОС – отказов типа 2 (фиксировалось по обращениям преподавателя к техническому специалисту).
3. Среднее время восстановления работоспособности (оценивалось время, потраченное на устранение неисправности типов 1 и 2).
4. Личный опыт участников (оценивалось количество и сложность заданий, выполненных на зачетном занятии. Оценка приводится в 10 бальной шкале – средняя по группе).

Результаты апробации приведены в таблице.

Характеристика	Группа с ВМ	Группа без ВМ
Общее количество отказов ПО, ед.	3	7
Количество отказов техники, ед.	2	2
Среднее время восстановления (тип 1/тип 2), ч	0,5/2	1/2
Личный опыт участников, баллов (по 10 бальной шкале)	9	7

Таблица. Результаты апробации

Полученные результаты подтверждают эффективность предложенной методики создания ИИ МПС на базе ВМ: более чем в 2 раза сократилось число отказов компонентов ИИ, требующих вмешательства технического специалиста, а также снизилось время на устранение такого рода неисправностей. Кроме того, учащиеся, получившие более широкий опыт работы с компьютером и ПО за счет применения ИИ с ВМ показали более хорошие результаты в рамках зачетных испытаний.

Заключение

В работе предложена методика формирования информационной инфраструктуры для использования технологии виртуальных машин в условиях перехода к единой информационной образовательной среде. Применение методики проиллюстрировано на примере проектирования информационной инфраструктуры образовательной среды, учитывающей специфику малой педагогической системы. К особенностям спроектированной информационной инфраструктуры относятся:

- максимальное использование открытых или бесплатных компонентов. Так, в качестве сред выполнения используются серверы Apache и Tomcat (лицензия Apache), модуль Guacamole (лицензии AGPL 3.0), модуль FreeRDP (лицензия Apache), PowerShell (бесплатный в рамках EULA);
- использование стандартных протоколов (в основном HTTP и HTTPS);
- максимальная переносимость (серверы приложений могут быть развернуты на любой из распространенных операционных систем);
- единый интерфейс к управлению различными платформами виртуализации;
- возможность инвентаризации содержимого виртуальной машины без ее запуска;
- гибкое управление инвентаризацией виртуальной машины с помощью настраиваемых цепочек правил.

Литература

1. Popek G.J., Goldberg R.P. Formal requirements for virtualizable third generation architectures // Communications of the ACM. 1974. V. 17. N 7. P. 412–421.
2. Smith J.E., Nair R. Virtual Machines: Versatile Platforms for Systems and Processes. Morgan Kaufmann, 2005. 656 p.
3. Гультьев А.К. Виртуальные машины: несколько компьютеров в одном. СПб: Питер, 2006. 224 с.
4. Блинков Ю.В. Моделирование компьютерных систем на виртуальных машинах. Пенза: ПГУАС, 2011. 268 с.

5. vMoodle Virtualization-Based Online Learning System [Электронный ресурс]. Режим доступа: <http://visa.cis.fiu.edu/tiki/vMoodle>, свободный. Яз. англ. (дата обращения 15.07.2014).
6. Cursory Thoughts on Virtual Machines in Distance Education Courses [Электронный ресурс]. Режим доступа: <http://blog.ouseful.info/2013/12/02/packaging-software-for-distance-learners-vms-101>, свободный. Яз. англ. (дата обращения 15.07.2014).
7. Virtual Machines - Kansas Research and Education Network [Электронный ресурс]. Режим доступа: <http://www.kanren.net/index.php/services/net-services/kanrenvm.html>, свободный. Яз. англ. (дата обращения 15.07.2014).
8. Скуидлевский А.А., Лямин А.В. Разработка стенда для экспериментальных исследований типовых схем развертывания системы AcademicNT // Научно-технический вестник СПбГУ ИТМО. 2011. № 3 (73). С. 126–129.
9. Усманов Ш.Н. Виртуальные машины в преподавании информатики // Информатика и образование. 2007. № 6. С. 120–121.
10. Чурилов И.А. Применение виртуальных машин в процессе обучения ИТ-специальностям // Преподавание информационных технологий в Российской Федерации: материалы XI открытой Всероссийской конференции. Воронеж: Воронежский государственный университет, 2013. С. 75–77.
11. Лисовецкий К. Виртуальная Windows в системе Linux [Электронный ресурс]. Режим доступа: http://gimnaz.com/blog/win_to_linux.90.aspx, свободный. Яз. рус. (дата обращения 16.02.2014).
12. Гаврилов С.И. Модели, методы и программные средства оценки качества информационно-образовательных ресурсов: автореф. дис. ... канд. техн. наук. М.: МИФИ, 2011. 22 с.
13. Развитие электронных образовательных интернет-ресурсов нового поколения [Электронный ресурс]. Режим доступа: <http://kkidppo.ru/razvitie-eog-novogo-pokoleniya>, свободный. Яз. рус. (дата обращения 30.08.2014).
14. Ившина Г.В. Разработка электронных образовательных ресурсов: мониторинг качества и внедрение. Казань, КГУ, 2008. 53 с.
15. Структура системы оценки качества программных комплексов для дистанционного обучения [Электронный ресурс]. Режим доступа: <http://www.hr-portal.ru/article/struktura-sistemy-ocenki-kachestva-programmnyh-kompleksov-dlya-distancionnogo-obucheniya>, свободный. Яз. рус. (дата обращения 15.03.2014).
16. Журавлёва Е.В. Проблема качества педагогических программных средств в контексте инноватизации преподавания математики // Вестник РУДН. Серия: Информатизация образования. 2011. № 1. С. 85–90.
17. Бабешко В.Н., Нежурина М.И. Комплексные решения ЦДО МИЭМ в области качества электронного обучения [Электронный ресурс]. Режим доступа: <http://tm.ifmo.ru/tm2005/src/264b.pdf>, свободный. Яз. рус. (дата обращения 15.03.2014).
18. Wu H.-Y., Lin Y.-K., Chang C.-H. Performance evaluation of extension education centers in universities based on the balanced scorecard // Evaluation and Program Planning. 2011. V. 34. N 1. P. 37–50.
19. Tzeng G.H., Chiang C.H., Li C.W. Evaluating intertwined effects in e-learning programs: a novel hybrid MCDM model based on factor analysis and DEMATEL // Expert Systems with Applications. 2007. V. 32. N 4. P. 1028–1044.
20. Безрукова В.С. Основы духовной культуры (энциклопедический словарь педагога). Екатеринбург, 2000. 937 с.
21. Сравнение виртуальных машин [Электронный ресурс]. Режим доступа: http://ru.wikipedia.org/wiki/Сравнение_виртуальных_машин, свободный. Яз. рус. (дата обращения 15.03.2014).
22. Шеер А.В. ARIS – моделирование бизнес-процессов. Вильямс, 2009. 175 с.
23. Артемова Г.О., Гусарова Н.Ф., Иванов Р.В. Оценка рисков применения виртуальных машин в инфраструктуре образовательного учреждения // Дистанционное и виртуальное обучение. 2014. № 11. (принято к печати).
24. Артемова Г.О., Гусарова Н.Ф., Иванов Р.В. Управление информационной инфраструктурой образовательной среды с использованием технологии виртуальных машин // Дистанционное и виртуальное обучение. 2014. № 10. (принято к печати).
25. Гусарова Н.Ф., Добренко Н.В., Иванов Р.В., Решетников В.В. Среда управления виртуальными машинами. Свидетельство о государственной регистрации программы для ЭВМ 2013661663. Заявл. 29.10.2013.
26. Гусарова Н.Ф., Добренко Н.В., Иванов Р.В., Прокофьев Е.С. Утилита извлечения структурированной информации о программном обеспечении. Свидетельство о государственной регистрации программы для ЭВМ 2013661662. Заявл. 29.10.2013.
27. Гусарова Н.Ф., Добренко Н.В., Иванов Р.В., Строганов П.С. Программа сопровождения поиска виртуальных машин в репозитории. Свидетельство о государственной регистрации программы для ЭВМ 2013661664. Заявл. 29.10.2013.

- Береснев Артем Дмитриевич*** – старший преподаватель, Университет ИТМО, 197101, Санкт-Петербург, Россия; руководитель ИТ-отдела, ООО «РОТЭК Технолоджиз», 192007, Россия, Санкт-Петербург, open.look@gmail.com
- Гусарова Наталия Федоровна*** – кандидат технических наук, старший научный сотрудник, доцент, Университет ИТМО, 197101, Санкт-Петербург, Россия, natfed@list.ru
- Иванов Роман Владимирович*** – старший преподаватель, Университет ИТМО, 197101, Санкт-Петербург, Россия, старший преподаватель, rvivanov@mail.ifmo.ru
- Федотов Дмитрий Николаевич*** – ведущий преподаватель, Институт информатики и вычислительной техники (ИАТИ АС), 10416 Таллинн, Эстония; ведущий преподаватель, Учебный центр Omis OÜ, Satelles Saturni OÜ (отдел информационных и методических технологий), 10117, Таллинн, Эстония, dmitry@satellessaturni.eu
- Artem D. Beresnev*** – lecturer, ITMO University, 197101, Saint Petersburg, Russia; Head of IT-department, “ROTEK Technologies”, Ltd, 192007, Saint Petersburg, Russia, open.look@gmail.com
- Natalia F. Gusarova*** – PhD, senior scientific researcher, Associate professor, ITMO University, 197101, Saint Petersburg, Russia, natfed@list.ru
- Roman V. Ivanov*** – senior lecturer, ITMO University, 197101, Saint Petersburg, Russia, rvivanov@mail.ifmo.ru
- Dmitry N. Fedotov*** – – leading lecturer, Institute of Informatics and Computers (IATI AS), 10416, Tallinn, Estonia; Education Center OMIS OÜ, Satelles Saturni OÜ, 10117, Tallinn, Estonia; dmitry@satellessaturni.eu

Принято к печати 30.07.14

Accepted 30.07.14

УДК 535.41

ДВУХСЛОЙНЫЕ ФАЗО-КОМПЕНСИРУЮЩИЕ ИНТЕРФЕРЕНЦИОННЫЕ СИСТЕМЫ

Г.В. Никандров^а, Э.С. Путилин^а, Л.А. Губанова^а, Д. Стародубов^б^аУниверситет ИТМО, 197101, Санкт-Петербург, Россия, nikandrov.spb@mail.ru^бУниверситет Южной Калифорнии, 90089, Лос Анджелес, Калифорния, США,

Аннотация. Разработаны оптические интерференционные покрытия, позволяющие формировать волновой фронт без изменения энергетических характеристик падающего и отраженного излучения. Корректировка достигается за счет того, что в состав покрытия вводится слой, толщина которого является функцией координаты поверхности оптического элемента. Для устранения изменения коэффициента отражения по поверхности оптического элемента, возникающего из-за наличия в составе покрытия слоя непостоянной толщины, предложена методика выбора показателей преломления материалов, формирующих двухслойное интерференционное покрытие, позволяющая создавать покрытия с постоянным коэффициентом по поверхности оптического элемента. В качестве пленкообразующих материалов использованы оксид магния и диоксид циркония. В работе приводится экспериментально полученное распределение толщины слоя, входящего в состав фазо-компенсирующего покрытия. Предложенный в работе новый класс оптических покрытий может быть использован для корректировки формы волнового фронта.

Ключевые слова: волновой фронт, интерференционные покрытия, диэлектрические слои, фаза отраженного излучения.

Благодарности. Работа выполнена при государственной финансовой поддержке Российского научного фонда (Соглашение № 14-23-00136).

TWO-LAYER PHASE COMPENSATING INTERFERENCE SYSTEMS

G.N. Nikandrov^а, E.S. Putilin^а, L.A. Gubanova^а, D. Starodubov^б^аITMO University, 197101, Saint Petersburg, Russia, nikandrov.spb@mail.ru^бUniversity of Southern California, 90089, Los Angeles, CA, USA

Abstract. The paper deals with creation of optical interferential coatings, giving the possibility to form the wave front without the change of energy characteristics of the incident and reflected radiation. Correction is achieved due to the layer, which thickness is a function of coordinate of an optical element surface. Selection technique is suggested for refractive index materials, forming two-layer interference coating that creates a coating with a constant coefficient of reflection on the surface of the optical element. By this procedure the change of coefficient of reflection for the optical element surface, arising because of the variable thickness is eliminated. Magnesium oxide and zirconium dioxide were used as the film-forming materials. The paper presents experimentally obtained thickness distribution of the layer, which is a part of the phase compensating coating. A new class of optical coatings proposed in the paper can find its application for correcting the form of a wave front.

Keywords: wave front, interference coatings, dielectric layers, reflected radiation phase.

Acknowledgements. The work was partially financially supported by the Russian Scientific Foundation (Agreement № 14-23-00136).

Диэлектрические интерференционные покрытия, сформированные на оптических элементах, как правило, выполняют только одну функцию – изменение энергетических характеристик прошедшего или отраженного излучения. Однако им можно придать дополнительные функции, связанные с формированием фронта световой волны [1–4], например, исправление дефектов подложки, исправление аберраций, получение длиннофокусных сферических зеркал, создание асферических поверхностей, использование во внутренних резонаторах для аподизации пучков лазеров большой мощности, фазированное объединение пучков лазеров и другие. При этом, как правило, присоединение второй функции приводит к ухудшению первой, связанной с формированием энергетических характеристик отраженного или прошедшего излучения. Задачей работы является создание покрытий, позволяющих формировать волновой фронт без существенного изменения отражения (пропускания) оптического элемента, на поверхности которого сформировано такое покрытие.

Для решения такой задачи могут быть использованы фазо-компенсирующие интерференционные системы, сформированные на поверхности оптического элемента. У этих систем толщина одного или нескольких слоев является функцией координаты подложки, при этом фаза отраженного и прошедшего излучения меняется по поверхности оптического элемента по заданной зависимости, а энергетические коэффициенты пропускания и отражения не зависят от координаты. Подобием таких систем является киноформы [5]. С их помощью можно управлять волновым фронтом световых волн на структурах с заданным изменением толщины и (или) показателя преломления вещества. Киноформные элементы имеют существенные преимущества по сравнению с традиционными оптическими элементами: малый вес, небольшие габариты. Но также эти элементы обладают рядом существенных недостатков: влияние износа резца при их изготовлении на параметры элемента, недостаточная точность, ограниченные возможности

изготовления неосесимметричных элементов, высокая трудоемкость и, как следствие, высокая стоимость изготовления. Фазо-компенсирующие покрытия лишены подобных недостатков.

Фазо-компенсирующие системы могут содержать разное количество слоев. Энергетические и фазовые характеристики тонкослойной интерференционной системы могут быть найдены с помощью описания таких систем матрицей интерференции [6]

$$\mathbf{M} = \mathbf{M}_1 \cdot \mathbf{M}_2 \dots \cdot \mathbf{M}_j,$$

где j – количество слоев. Рассмотрим двухслойную интерференционную систему, у которой толщина слоя, граничащего с воздухом, является функцией координаты по поверхности подложки, а второй слой имеет постоянную толщину, которая равна четверти рабочей длины волны. Такая система может быть описана следующей матрицей интерференции:

$$\mathbf{M} = \mathbf{M}_1 \cdot \mathbf{M}_2 = \begin{vmatrix} \cos \varphi_1(x) & \frac{i}{n_1} \sin \varphi_1(x) \\ in_1 \sin \varphi_1(x) & \cos \varphi_1(x) \end{vmatrix} \begin{vmatrix} i & 0 \\ n_2 & 0 \end{vmatrix} = \begin{vmatrix} \frac{n_2}{n_1} \sin \varphi_1(x) & \frac{i}{n_2} \cos \varphi_1(x) \\ in_2 \cos \varphi_1(x) & \frac{n_1}{n_2} \sin \varphi_1(x) \end{vmatrix}.$$

В этом случае для двухслойной системы амплитудный коэффициент отражения может быть определен из соотношения

$$r = \frac{(n_o \mathbf{M}_{11} - n_m \mathbf{M}_{22}) + i(n_o n_m \mathbf{M}_{12} - \mathbf{M}_{21})}{(n_o \mathbf{M}_{11} + n_m \mathbf{M}_{22}) + i(n_o n_m \mathbf{M}_{12} + \mathbf{M}_{21})} = \frac{\left(\frac{n_o n_2}{n_1} - \frac{n_m n_1}{n_2}\right) \sin \varphi_1(x) + i\left(\frac{n_o n_m}{n_2} - n_2\right) \cos \varphi_1(x)}{\left(\frac{n_o n_2}{n_1} + \frac{n_m n_1}{n_2}\right) \sin \varphi_1(x) + i\left(\frac{n_o n_m}{n_2} + n_2\right) \cos \varphi_1(x)}, \quad (1)$$

а фазовые параметры – из соотношения [7]

$$r = \frac{A'(\cos \varphi_1(x) + i \sin \varphi_1(x))}{B'(\cos \varphi_1(x) - i \sin \varphi_1(x))} = \frac{A'}{B'} e^{-2i\varphi_1(x)}.$$

Здесь n_1 и n_2 – показатели преломления материалов, из которых формируется первый и второй слой соответственно; x – координата на поверхности оптического элемента, отсчет ведется от центра оптического элемента, $\varphi_1 = 2\pi n_1 d_1(x)/\lambda$ и $d_1(x)$ – фазовая и геометрическая толщина слоя в точке с координатой x соответственно, λ – рабочая длина волны. Зависимость (1) позволяет определить соотношение между показателями преломления материалов слоев, входящих в состав покрытия, которые бы обеспечивали постоянство коэффициента отражения при изменяющейся толщине слоя, граничащего с воздухом [8]. Это соотношение имеет вид $n_1 = n_2^2/n_m$.

На основе полученных результатов была рассчитана конструкция фазо-компенсирующего покрытия, состоящего из слоев MgO и ZrO₂, и сформирована на оптическом элементе, изготовленном из стекла марки Borofloat 33.

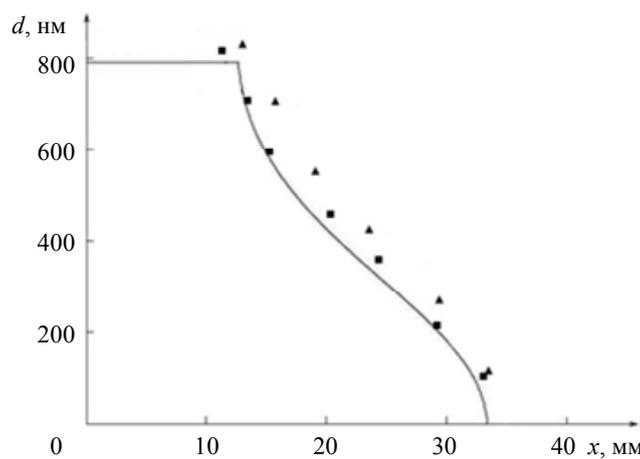


Рисунок. Распределение толщины градиентного слоя (d)

— — рассчитанное распределение, ■ — экспериментально полученное распределение для $\lambda=465$ нм, ▲ — экспериментально полученное распределение для $\lambda=525$ нм

Покровтия такого типа целесообразно изготавливать методом термического испарения пленкообразующих веществ в вакууме. Для получения слоя с заданным распределением толщины слоя по поверхности оптического элемента можно воспользоваться схемой формирования слоев с заданным распределением толщины по поверхности оптического элемента, представленной в [9]. Для получения заданного рас-

пределения слоя была изготовлена оснастка, параметры которой были выбраны с учетом соотношений, представленных в [10]. Перед формированием слоев была проведена аттестация пленкообразующих материалов с использованием метода, изложенного в [11, 12]. Данный метод заключается в том, что были изготовлены четвертьволновые слои из оксида магния и диоксида циркония, используемых в качестве пленкообразующих материалов. Далее были измерены спектральные характеристики полученных слоев и по величине экстремумов определены их показатели преломления.

Полученное покрытие было аттестовано с целью определения распределения толщины слоя по поверхности оптического элемента. Как видно из рисунка, фазо-компенсирующее покрытие имеет распределение толщины по поверхности оптического элемента, близкое к заданному распределению толщины слоя по поверхности оптического элемента.

Рассмотренный в работе новый класс оптических покрытий может быть использован для корректировки формы волнового фронта.

Литература

1. Жиглинский А.Г., Путилин Э.С. Формирование волнового фронта с помощью интерференционных покрытий // Оптика и спектроскопия. 1972. Т. 32. № 6. С. 27–31.
2. Yasui K., Tanaka M., Yagi S. Unstable resonator with phase-unifying coupler for high-power laser // Applied Physics Letters. 1988. V. 52. N 7. P. 530–531.
3. Paré C., Bélanger P.-A. Optical resonators with graded-phase mirrors // Proceedings of SPIE – The International Society for Optical Engineering. 1998. V. 3267. P. 226–233.
4. Kotlikov E.N., Prokashev V.N., Khonineva E.V. Synthesis of unstable resonator output mirrors with phase front compensation // Proceedings of SPIE – The International Society for Optical Engineering. 2001. V. 4353. P. 69–74.
5. Фишман А.И. Фазовые оптические элементы - киноформы // Соросовский образовательный журнал. 1999. № 12. С. 76–83.
6. Бернинг П.Х. Теория и методы расчета оптических свойств тонких пленок // Физика тонких пленок. Т. 1. / Под ред. Г.Хасса, Э. Туна: Пер. с англ. М.: Мир, 1967. С. 91–151.
7. Белашенков Н.Р., Губанова Л.А., Путилин Э.С., Фимин П.Н. Формирование волнового фронта отраженного излучения градиентными зеркалами // Тезисы докладов Международной конференции «Прикладная оптика-96». СПб, 1996. С. 197.
8. Никандров Г.В. Фазо-компенсирующие покрытия // Научно-технический вестник СПбГУ ИТМО. 2008. № 4 (49). С. 3–6.
9. Губанова Л.А. Критерий выбора и синтез условий осаждения градиентных систем // Научно-технический вестник СПбГУ ИТМО. 2004. № 4 (15). С. 145–149.
10. Никандров Г.В. Разработка технологии создания фазо-компенсирующих покрытий // Сборник тезисов VIII всероссийской межвузовской конференции молодых ученых. Вып. 2. Оптотехника и оптические материалы. СПб: СПбГУ ИТМО, 2011. С. 53–54.
11. Эльгарт З.Э. Балансно-двухволновой метод контроля оптических толщин слоев: дис. ... д-р. техн. наук. СПб: СПбГУ ИТМО, 1998. 217 с.
12. Канцельсон Л.Б. Методы контроля оптической толщины интерференционных пленок, наносимых в вакууме. Обзор // Оптико-механическая промышленность. 1969. № 4. С. 50–58.

<i>Никандров Георгий Васильевич</i>	– аспирант, Университет ИТМО, 197101, Санкт-Петербург, Россия, nikandrov.spb@mail.ru
<i>Путилин Эдуард Степанович</i>	– доктор технических наук, профессор, профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия
<i>Губанова Людмила Александровна</i>	– доктор технических наук, доцент, профессор, Университет ИТМО, 197101, Санкт-Петербург, Россия, la7777@mail.ru
<i>Стародубов Дмитрий</i>	– кандидат физико-математических наук, кандидат физико-математических наук, Университет Южной Калифорнии, 90089, Лос Анджелес, Калифорния, США,
<i>Georgiy V. Nikandrov</i>	– postgraduate, ITMO University, 197101, Saint Petersburg, Russia, nikandrov.spb@mail.ru
<i>Edward S. Putilin</i>	– D.Sc., Professor, Professor, ITMO University, 197101, Saint Petersburg, Russia
<i>Lujdmila A. Gubanova</i>	– D.Sc., Associate professor, Professor, ITMO University, 197101, Saint Petersburg, Russia, la7777@mail.ru
<i>Dmitry Starodubov</i>	– PhD, scientist, University of Southern California, 90089, Los Angeles, CA, USA, dstarodubov@gmail.com

Принято к печати 16.07.14
Accepted 16.07.14

УДК 004.42, 004.89

**ПРОТОТИП ПРОФЕССИОНАЛЬНОЙ СОЦИАЛЬНОЙ СЕТИ BARENTSNET
ДЛЯ ИНФОРМАЦИОННОЙ ПОДДЕРЖКИ УПРАВЛЕНИЯ РАЗВИТИЕМ БАРЕНЦЕВА
ЕВРО-АРКТИЧЕСКОГО РЕГИОНА**

А.В. Маслобоев^{a, b}, Р.М. Странд^c

^a Институт информатики и математического моделирования технологических процессов Кольского научного центра РАН, 184209, г. Апатиты Мурманской обл., Россия;

^b Кольский филиал Петрозаводского государственного университета, 184209, г. Апатиты Мурманской обл., Россия, masloboev@iimm.ru

^c ОАО «Arctic Development AS» (г. Вардѳ, Норвегия), Nordere Langgate 20, 9950 Vardѳ, Norway

Аннотация. Разработан прототип профессиональной социальной сети BarentsNet для информационной поддержки деятельности субъектов, участвующих в процессах управления развитием и освоением ресурсного потенциала Баренцева Евро-Арктического региона. Система BarentsNet реализована как мультипредметный веб-ресурс и обеспечивает интеграцию формализованных на базе онтологий экспертных знаний об особенностях развития арктических и субарктических территорий, а также автоматизацию установления профессиональных контактов для сотрудничества и реализации совместных проектов в этой сфере. Разработаны модель функциональной организации, исполнительное ядро и программные компоненты системы BarentsNet.

Ключевые слова: социальная сеть, информационная поддержка, управление, региональное развитие.

Благодарности. Работа выполнена при поддержке РФФИ (проект № 12-07-00138-а «Разработка когнитивных моделей и методов формирования интегрированной информационной среды поддержки управления безопасностью Арктических регионов России»). Авторы выражают благодарность проф. В.А. Путилову и проф. А.И. Смирнову за участие в обсуждении результатов работы.

**A PROTOTYPE OF BARENTSNET PROFESSIONAL SOCIAL NETWORK FOR INFORMATION
SUPPORT OF DEVELOPMENT MANAGEMENT FOR BARENTS EURO-ARCTIC REGION**

A.V. Masloboev^{a, b}, R.M. Strand^c

^aInstitute for Informatics and Mathematical Modeling of Technological Processes Kola Science Center of the Russian Academy of Sciences, 184209, Apatity, Russia;

^b Department head, Kola Branch of Petrozavodsk State University, 184209, Apatity, Russia, masloboev@iimm.ru

^c Partner/CEO (Chief Executive Officer), JSC «Arctic Development AS», 9950 Vardѳ, Norway

Abstract. A prototype of professional social network BarentsNet has been developed for management activities information support of entities participating in the management process of developing and resource potential settling of the Barents Euro-Arctic region. BarentsNet system is implemented as a multi-domain web-service and provides formalized ontology-based expert knowledge integration of the Arctic and sub-Arctic territories development features, and professional contacts linking automation within the system for cooperation and joint project realization in this sphere. The functionality organization model, executive core and software components of the BarentsNet system have been developed.

Keywords: social network, information support, management, regional development.

Acknowledgements. The research is supported by the Russian Foundation for Basic Research (grant 12-07-00138-a “Cognitive models and methods development of integrated information environment formation for security management support of the Arctic regions of the Russian Federation”). Grateful thanks are given to Prof. V.A. Putilov and Prof. A.I. Smirnov for assistance with the paper preparing and participation within the discussion of the findings.

В настоящее время наблюдается эскалация геополитической ситуации в Баренцевом Евро-Арктическом регионе (БЕАР), обусловленная возникновением ряда проблем внешнего и внутреннего характера (военно-политических, экономических, демографических, экологических и т.д.), касающихся различных аспектов социально-экономического развития и обеспечения комплексной безопасности арктических и субарктических территорий. Усиливается борьба за ресурсы (природные, кадровые, информационные), коммуникации (транспортные, цифровые, деловые), пространства (подводное, надводное, воздушное, космическое), что формирует вектор угроз национальным интересам России в БЕАР – экономическим, оборонным, демографическим и экологическим. От инновационного и безопасного развития БЕАР зависят как национальная безопасность нашей страны, так и ее позиционирование в мировой экономике. В связи с этим создание зоны безопасного международного сотрудничества и обеспечение глобальной и региональной безопасности в БЕАР являются задачами государственной важности [1].

Решение этих задач, согласно результатам исследования [2], во многом затрудняется необходимостью интеграции, обработки и анализа больших объемов семантически и организационно разнородной информации для информационного обеспечения деятельности хозяйствующих субъектов в БЕАР, а также удовлетворения их информационных потребностей.

Настоящая работа направлена на развитие исследований, проводимых в области создания мультипредметных веб-ориентированных информационных систем, основанных на знаниях и направленных на решение задач информационно-аналитической поддержки управления региональным развитием в БЕАР. Под мультипредметной информационной системой [3] понимается информационная система, предназначенная для эксплуатации пользователями разных категорий. Под категорией пользователя, в свою очередь, в данном контексте понимается некоторое множество субъектов использования информационной

системы, характеризующихся близкими по структуре и содержанию ментальными моделями одной или более предметных областей. Как правило, такие системы строятся на базе когнитивных информационных технологий (мультиагентные технологии [4], технологии семантического веба [5] и др.).

Для повышения эффективности взаимодействия и удовлетворения информационных потребностей субъектов, участвующих в процессах управления развитием и освоением арктических регионов, разработан прототип виртуальной интеграционной площадки VarentsNet, представляющей собой мультипредметный веб-ресурс, реализующий функционал профессиональной социальной сети, объединяющей экспертов, заинтересованные бизнес-сообщества и государственные структуры для сотрудничества в области обеспечения безопасности и инновационного развития БЕАР.

В рамках системы VarentsNet реализованы средства интеграции формализованных на базе онтологий экспертных знаний о процессах и особенностях развития БЕАР и средства автоматизации установления профессиональных контактов, что обеспечивает сокращение рутинных операций пользователей в процессе информационного поиска и подбора потенциальных партнеров для взаимодействия.

Социальная сеть VarentsNet обеспечивает базовый набор функций по предоставлению доступа к различным источникам проблемно-ориентированной информации и веб-сервисам, заложенный в известные системы-аналоги, такие как Facebook, MySpace, Google+ или ВКонтакте. К базовым функциям в данном случае относятся:

- возможность создания индивидуальных профилей;
- возможности взаимодействия пользователей посредством внутренней почты, комментариев (в режиме блогов или микроблогов) и т.п., а также обмена информацией;
- возможность создания тематических виртуальных сообществ – групп пользователей по интересам (открытого или закрытого типа);
- разграничение прав пользователей на доступ к информации и оперирование ею, инвайты;
- проблемно-ориентированный информационный поиск (люди, группы, идеи, контакты, профили, проекты и т.д.).

Вместе с тем, в системе предусмотрена реализация специфических функций. Речь идет о возможности использования средств оперативной аналитической обработки данных на базе компьютерного моделирования (системно-динамических и агентных моделей, синтезируемых из типовых модельных шаблонов), средств интеграции разнородных информационных ресурсов на базе онтологий, а также средств распределенного семантического поиска информации и средств автоматизированного синтеза проблемно-ориентированных виртуальных организационных структур.

Разработанный прототип системы не имеет аналогов и является эффективным инструментом получения, интеграции, анализа и обработки информации, а также коллективных экспертных знаний, необходимых для решения конкретных задач управления и принятия решений в сфере информационного обеспечения инновационного развития и региональной безопасности в БЕАР. Однако стоит отметить, что близким по технологической реализации решением без учета реализованного набора специальных функций и предметной ориентации (арктической специфики) является современная виртуальная сеть профессиональных контактов LinkedIn (<http://www.linkedin.com>), имеющая сервисно-ориентированную архитектуру и созданная на базе передовых веб-технологий с поддержкой облачных и веб-сервисов.

Построение единой электронной социальной сети VarentsNet с целью создания виртуальной интеграционной площадки по сотрудничеству в сфере управления развитием арктических территорий является одним из главных пунктов новой Киркенесской декларации, а также важным этапом на пути создания единого информационного пространства БЕАР при реализации «Стратегии развития Арктической зоны Российской Федерации и обеспечения национальной безопасности на период до 2020 года».

Мировой опыт создания и использования социальных сетей как инструмента информационного влияния по принципу «мягкой силы 2.0», согласно работам [1, 6, 7], показывает их высокую эффективность.

В качестве средства коммуникации пользователей VarentsNet с компонентами единой информационной среды и друг с другом предложено использовать их онлайн-автоматизированные рабочие места («личные кабинеты») и их виртуальных представителей в этой среде – программных агентов. Структура онлайн-автоматизированных рабочих мест («личных кабинетов») пользователей системы VarentsNet включает в себя средства, реализующие как базовые функциональные блоки современных социальных сетей, так и специализированные программные компоненты.

При создании мультипредметного веб-ресурса VarentsNet использовались язык PHP, СУБД MySQL и технология Ajax. В качестве программного обеспечения, позволяющего управлять содержимым и структурой веб-сайта, использована CMS-система LiveStreet. Программные агенты разработаны на языке Java с помощью инструментальных средств AgentBuilder и Cougaar. Прикладные OWL-онтологии, используемые в системе, созданы в среде Protege. Типовые модельные шаблоны разработаны средствами Anylogic и PowerSim. В качестве средств автоматизированного синтеза имитационных моделей из шаб-

лонов и интеграции разнородных информационных ресурсов на базе онтологий использованы собственные разработки.

В ходе исследований получены следующие основные результаты. Разработаны модель функциональной организации, исполнительное ядро («движок») и активные программные компоненты профессиональной социальной сети BarentsNet на базе инструментария онтологического и имитационного моделирования, а также агентных и веб-технологий. Предложены функциональная структура и базовые модули онлайн-автоматизированных рабочих мест («личных кабинетов») хозяйствующих субъектов в БЕАР в составе системы BarentsNet

Созданная система BarentsNet (<http://www.barents-online.net>) интегрирована в инфраструктуру виртуального когнитивного центра управления региональной безопасностью [8], разработанного в рамках предыдущих исследований.

Литература

1. Смирнов А.И., Григорьев В.Р., Кохтюлина И.Н., Куроедов Б.В., Сандаров О.В. Глобальная безопасность в цифровую эпоху: стратагемы для России. М.: ВНИИГеосистем, 2014. 394 с.
2. Маслобоев А.В. Реализация трансграничных ИТ-проектов в сфере информационного обеспечения комплексной безопасности развития арктических регионов: состояние и перспективы // Информационные ресурсы России. 2014. № 3(139). С. 13–20.
3. Шишаев М.Г., Ломов П.А., Диковицкий В.В. Использование концепции «User as an expert» в разработке мультимедийных веб-ресурсов, основанных на онтологиях // Труды ИСА РАН. 2012. Т. 62. № 3. С. 40–47.
4. Russell S., Norvig P. Artificial Intelligence. A Modern Approach. 3rd ed. New Jersey: Prentice Hall, 2010. 1132 p.
5. Breitman K.K., Casanova M.A., Truszkowski W. Semantic Web: Concepts, Technologies and Applications. Springer, 2007. 344 p.
6. Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства. М.: Физматлит, 2010. 228 с.
7. Jackson M.O. Social and Economic Networks. Princeton University Press, 2010. 520 p.
8. Маслобоев А.В. Виртуальные когнитивные центры как интеллектуальные системы для информационной поддержки управления региональной безопасностью // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 2 (90). С. 167–170.

- Маслобоев Андрей Владимирович** – кандидат технических наук, доцент, старший научный сотрудник, Институт информатики и математического моделирования технологических процессов Кольского научного центра РАН, 184209, г. Апатиты Мурманской обл., Россия; заведующий кафедрой, Кольский филиал Петрозаводского государственного университета, 184209, г. Апатиты Мурманской обл., Россия, masloboev@iimm.ru
- Странд Реми Мортен** – генеральный директор, ОАО «Arctic Development AS» (г. Вардё, Норвегия), Nordere Langgate 20, 9950 Vardø, Norway, rs@arcticdevelopment.no
- Andrey V. Masloboev** – PhD, Associate professor, senior research fellow, Institute for Informatics and Mathematical Modeling of Technological Processes Kola Science Center of the Russian Academy of Sciences, 184209, Apatity, Russia; Department head, Kola Branch of Petrozavodsk State University, 184209, Apatity, Russia, masloboev@iimm.ru
- Remi Strand** – Partner/CEO (Chief Executive Officer), JSC «Arctic Development AS», 9950 Vardø, Norway, rs@arcticdevelopment.no

Принято к печати 11.07.14
Accepted 11.07.14